2018

High Dimensional Multivariate Inference Under General Conditions

Xiaoli Kong
University of Kentucky, kongxl.math@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-6072-0817
Digital Object Identifier: https://doi.org/10.13023/etd.2018.341

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Xiaoli Kong, Student

Dr. Solomon W. Harrar, Major Professor

Dr. Constance L. Wood, Director of Graduate Studies

High Dimensional Multivariate Inference Under General Conditions

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Arts and Sciences
at the University of Kentucky

By

Xiaoli Kong
Lexington, Kentucky

Director: Dr. Solomon W. Harrar, Professor of Statistics
Lexington, Kentucky
2018

ABSTRACT OF DISSERTATION

High Dimensional Multivariate Inference Under General Conditions

In this dissertation, we investigate four distinct and interrelated problems for high-dimensional inference of mean vectors in multi-groups.

The first problem concerned is the profile analysis of high dimensional repeated measures. We introduce new test statistics and derive its asymptotic distribution under normality for equal as well as unequal covariance cases. Our derivations of the asymptotic distributions mimic that of Central Limit Theorem with some important peculiarities addressed with sufficient rigor. We also derive consistent and unbiased estimators of the asymptotic variances for equal and unequal covariance cases respectively.

The second problem considered is the accurate inference for high-dimensional repeated measures in factorial designs as well as any comparisons among the cell means. We derive asymptotic expansion for the null distributions and the quantiles of a suitable test statistic under normality. We also derive the estimator of parameters contained in the approximate distribution with second-order consistency. The most important contribution is high accuracy of the methods, in the sense that p-values are accurate up to the second order in sample size as well as in dimension.

The third problem pertains to the high-dimensional inference under non-normality. We relax the commonly imposed dependence conditions which has become a standard assumption in high dimensional inference. With the relaxed conditions, the scope of applicability of the results broadens.

The fourth problem investigated pertains to a fully nonparametric rank-based comparison of high-dimensional populations. To develop the theory in this context, we prove a novel result for studying the asymptotic behavior of quadratic forms in ranks.

The simulation studies provide evidence that our methods perform reasonably well in the high-dimensional situation. Real data from Electroencephalograph (EEG) study of alcoholic and control subjects is analyzed to illustrate the application of the results.

KEYWORDS: Profile analysis, MANOVA, High-dimension, Repeated measure, Non-parametric, Rank transforms.

Author's signature: Xiaoli Kong

Date: August 2, 2018

High Dimensional Multivariate Inference Under General Conditions

By

Xiaoli Kong

Director of Dissertation: Solomon W. Harrar

Director of Graduate Studies: Constance L. Wood

Date: August 2, 2018

ACKNOWLEDGMENTS

I would like to thank all of the people who have taught, helped and supported me throughout my graduate work at University of Kentucky.

In particular, I would like to express my deepest appreciation to my advisor Dr. Solomon W. Harrar for the continuous support of my Ph.D study and research. He patiently provided the encouragement and advice necessary for me to complete the program and the dissertation. Without his help, I could not have accomplished what I have been done.

Special thanks to all my committee, Dr. Arnold J. Stromberg, Dr. Xiangrong Yin, Dr. David W. Fardo, Dr. Chi Wang, and Dr. Licong Cui, for their support, guidance and helpful suggestions, and to Dr. Mai Zhou, for his support and help during my qualify exam and Ph.D study. I owe them my deepest gratitude.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Nowadays, more and more big data arise in various research areas due to the invention of high-throughput data collection technologies. To cope with the growth of data volume, there is an increasing demand for efficiently (computationally as well as statistically) analyzing the high-dimensional data. Throughout the dissertation, by high dimension is meant that both the sample size and dimension are large but one could be substantially larger relative to the other.

Existing high-dimensional multivariate methods for comparing groups (treatments or populations) formulate hypothesis in terms of mean or location vectors. Some of these results assume multivariate normality (Dempster, 1958, 1960; Fujikoshi et al., 2004; Schott, 2007a; Srivastava and Du, 2008; Yamada and Srivastava, 2012; Dong et al., 2017), while others assume existence of higher moments and pseudo-independence in the sense that higher-order mixed moments can be factored into the product of the corresponding univariate moments (Bai and Saranadasa, 1996; Chen and Qin, 2010; Srivastava and Kubokawa, 2013; Hu et al., 2017). A few others require a different form of weaker dependence but they are still parametric methods (Cai et al., 2014; Cai and Xia, 2014; Feng et al., 2015; Gregory et al., 2015). The nonparametric methods (Wang et al., 2015; Ghosh and Biswas, 2016) are also essentially mean based and assume (generalized) elliptically symmetric populations.

This dissertation aims to solve four distinct but interrelated problems. Two of them pertain to high-dimensional inference about mean profiles, namely parallelism, flatness and coincidence of the mean vectors; under high dimensional asymptotic framework but assume multivariate normality. The other two problems consider high-dimensional group comparisons, but do not need normality assumption. One of them is designed for metric type data and the other one is rank-based, and hence, can be used for non-metric data such as ordered categorical data.

The dissertation is organized in six chapters. In Chapter 2, test statistics for high-dimensional profile analysis in multi-group are introduced and the asymptotic null distributions are derived. Here, multivariate normality is assumed but the covariance matrices can be unequal and unstructured.

The subject of Chapter 3 is high-dimensional asymptotic expansions for the test statistics derived in Chapter 2. Here, our approach treats factorial designs in a unified and succinct manner, especially allowing multiple between-subject and within-subject factors, which may be crossed or nested. The most important contribution is the high accuracy of the methods, in the sense that second-order accuracy in sample size as well as in dimension is achieved by obtaining asymptotic expansion of the distribution of the test statistics, and the estimation of the parameters of the approximate distribution with second-order consistency.

Chapter 4 is concerned with high-dimensional inference about equality of mean vectors under non-normality. As mentioned above, recent results for comparison of the high-dimensional mean vectors under non-normality make strong assumptions that require the dependence between the variables to be rather too weak (see Bai and Saranadasa, 1996; Chen and Qin, 2010; Srivastava and Kubokawa, 2013; Hu et al., 2017). We relax these commonly imposed dependence conditions and broaden the scope of applicability of the results. The theory is worked out in detail for the two-group case and, later, extended to the multi-group situation. The extension of the results for testing hypotheses in profile analysis and factorial mean structures are formally illustrated.

A nontrivial application of the theory developed in Chapter 4 is provided in Chapter 5. More precisely, we investigate rank-based method for comparing groups (treatments or populations) in the high-dimensional asymptotic setting. As pointed out above, existing high-dimensional nonparametric methods are essentially mean-based and they assume (generalized) elliptically symmetric populations (see Wang et al., 2015; Ghosh and Biswas, 2016). The rank-based test we construct is a fully nonparametric method. No assumption is made on the distribution except that the dependences between the variables are required to satisfy some mild conditions. The method is applicable for ordered categorical, skewed and heavy tailed variables or a mixture of them. To develop the theory, we prove a novel result for studying the asymptotic behavior of quadratic forms in ranks.

Appendices containing the proofs and other technical details are included at the end of each Chapters 2 to 5. Also included in these chapters are simulation studies to evaluate the numerical performance of the methods; analyses of data from an Electroencephalogram (EEG) experiment to illustrate the application of the methods; and possible directions for future research. The findings of the dissertation are summarized in Chapter 6.

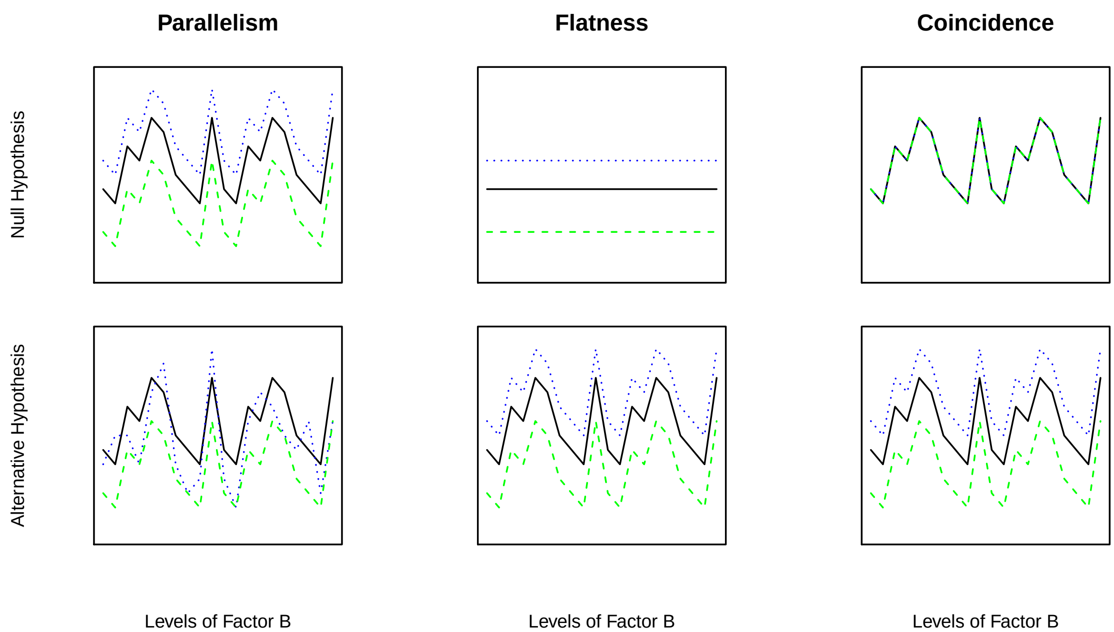
Chapter 2 Multivariate Analysis for Repeated Measures in High-Dimensions with Unequal Covariance Matrices

2.1 Introduction

Consider b measurements taken from n subjects which are classified into a groups. The a groups may represent naturally existing groups such as gender, geographical regions or ethnicity. They may also represent between-subject treatment groups as commonly done in clinical trials. The b repeated measurements could be measurements from b within-subject treatment conditions as in crossover design or from b different tissues of the body or may simply be repeated measurements over time as typically arises in time course studies. For the sake of brevity, in the remainder of this Chapter we will refer to the a groups as the levels of a between-subject factor (A) and to the b repeated measurements as arising from b levels of a within-subject factor (B). Research questions (hypotheses) that are typically tested with this type of data are (i) whether there is interaction effect between the between-subject and within-subject factors (ii) whether there is a between-subject factor effect and (iii) whether there is a within-subject factor effect.

Analysis addressing these research questions are also referred to as *Profile Analyses* in multivariate statistics. Consider a independent b-dimensional normal populations with mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_a$ and covariance matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_a$, respectively. Graphically, the profile of the mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ib})^\top$ of population $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ can be plotted as a line graph connecting the points $(1, \mu_{i1})$, …, (b, μ_{ib}). Profile analysis is the study of the relationship between these lines. In Figure 2.1 below, the three hypotheses of interest are shown graphically. In the terminology of profile analyses the hypotheses (i), (ii) and (iii) are refereed to as parallelism, level and flatness (see, for example, Rencher and Christensen, 2012; Johnson and Wichern, 2007). The level hypothesis is, alternatively, referred to as coincidence hypothesis. The level and flatness hypothesis are typically tested if the parallelism hypothesis holds. This scenario is clearly illustrated in the alternative hypotheses in Figure 2.1.

Figure 2.1: Graphical display of null and alternative hypotheses in profile. Each line plot corresponds to mean vector of one group.



For $i = 1, \ldots, a$, consider n_i independent b-dimensional observations are available from population $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ denoted by $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$ and assume that the a samples are mutually independent. The total sample size is $n = \sum_{i=1}^{a} n_i$. The aim of this Chapter is to derive tests for the three hypothesis in the repeated measures analysis (profile analysis) when both the groups sample sizes n_i and number of repeated measurements b tend to infinity. The approach followed in this Chapter is multivariate in the sense that no structure on the covariance matrices are made other than requiring them to be symmetric positive definite.

Although first analysis of such data dates back to as early as several decades ago, the methods developed so far assume either fixed and bounded number of repeated measures or specialized covariance matrices. From mathematical stand point, tests in profile analysis were first tackled from likelihood ratio point view by Srivastava (1987). Asymptotic expansions for null distributions of the test statistics in profile analyses were derived by Okamoto et al. (2006) under elliptical populations and by Maruyama (2007) under more general populations but both these works focused on the two-group case. Harrar and Xu (2014) considered asymptotic expansion for the null distributions of the likelihood-ratio tests in Srivastava (1987) for several sample situation. On the other hand, Harrar (2009) and Bathke et al. (2010) derived tests for repeated measures analysis for the case when a is large but n_i and b are bounded. Recall that the hypotheses regarding the within-subject and between-subject treatments are considered under the parameter space constrained by the no-interaction (parallelism) hypothesis. Without this constraint, the problem of testing for between-subject factor level effects is the same as in one-way MANOVA. Harrar and Xu (2014) derived likelihood ratio tests for the hypothesis of no within-subject factor level effects under the full parameter space.

In the high-dimensional framework with $b/n \to c \in (0, 1)$, likelihood ratio test statistics together with null distributions derived for MANOVA, e.g., Tonda and Fujikoshi (2004), can be used to get valid tests for the interaction hypothesis. Since the exact distribution of the likelihood ratio test for within subject and between subject factor level effects are known, the same distribution will hold under high-dimensional case as long as the degrees of freedom for the within-covariance estimator is larger than the dimension. For the high-dimensional situation where $b \geq n - a$, the likelihood-ratio tests are not well defined because they involve the determinants or inverses of the estimate of the within covariance matrix which will be singular. This problem has been tackled by many authors in the MANOVA context. Among others, Schott (2007a) and Yamada and Srivastava (2012) developed tests under normality whereas Bai and Saranadasa (1996); Chen and Qin (2010), and Srivastava and Kubokawa (2013) derived tests under non-normality. In repeated measures or profile analysis context, Pauly et al. (2015) consider high-dimensional repeated measures analysis for one sample situation but with the possibility of several within subject factors. The two-sample situation was considered by Takahashi and Shutoh (2016) assuming equal covariance matrices for the two populations. Wang and Akritas (2010a) and Wang and Akritas (2010b) are also high-dimensional asymptotic results applicable for repeated measures but assume that the repeated measurements are inherently ordered and the dependence between the measurements decays as the separation between them increases. The present manuscript provides a complete solution to the analysis of high-dimensional repeated measures design by allowing for several samples as well as unequal and unstructured covariance matrices. Furthermore, no assumption is made about ordering of the observations. It bears some similarity with Pauly et al. (2015) and Takahashi and Shutoh (2016) in the way the tests are constructed.

This Chapter is organized as follows. Section 2.2 introduces the statistical model, hypotheses and notations used in the remainder of the Chapter. Tests for interaction and main effects under equal covariance matrices assumption are the subject of Section 2.3. These tests are again studied in Section 2.4 without assuming equal covariance matrices. Numerical accuracy of the asymptotic results in Section 2.3 and 2.4 is investigated in Section 2.5 for various choices for the parameters of the model. Also in Section 2.5, the power of the tests proposed in this Chapter will be compared against an existing method. The application of the results will be illustrated in Section 2.6 with data from an electroencephalograph (EEG) experiment. Section 2.7 contains discussions and conclusions. All proofs and preliminary results are placed in the Appendix.

2.2 Model and Hypotheses

Let

$$\boldsymbol{X} = (\boldsymbol{X}_{11}^\top, \ldots, \boldsymbol{X}_{1n_1}^\top, \boldsymbol{X}_{21}^\top, \ldots, \boldsymbol{X}_{2n_2}^\top, \ldots, \boldsymbol{X}_{a1}^\top, \ldots, \boldsymbol{X}_{an_a}^\top)^\top,$$

where $\boldsymbol{X}_{ik} = (X_{i1k}, \ldots, X_{ibk})^\top$. Further let

$$\overline{\boldsymbol{X}} = (\overline{X}_{11}, \ldots, \overline{X}_{1b}, \ldots, \overline{X}_{a1}, \ldots, \overline{X}_{ab})^\top,$$

and $\overline{\boldsymbol{X}}_i = (\overline{X}_{i1}, \ldots, \overline{X}_{ib})^\top$, where $\overline{X}_{ij} = n_i^{-1} \sum_{k=1}^{n_i} X_{ijk}$. We assume $\boldsymbol{X}_{ik} \overset{\text{iid}}{\sim} \mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for $k = 1, \ldots, n_i$ and the a samples $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$ for $i = 1, \ldots, a$ are mutually independent. The usual setting gives the interpretation that X_{ijk} is the responses from the kth subject treated with the ith level of factor A and the jth level of factor B. The interaction effect will be denoted by AB. In this model X_{ijk} and $X_{i'j'k'}$ are assumed to be independent only if $i \neq i'$ or $k \neq k'$. Otherwise the dependence is completely unspecified.

Throughout this Chapter, $\mathbf{0}$ will denote a matrix of all zeros where the dimension will be clear from the context, and $\mathbf{1}_k$ denotes an k-dimensional vector $(1, ..., 1)^\top$ consisting of ones. The matrix $\boldsymbol{I}_k$ is the identity matrix, whereas $\boldsymbol{J}_k$ and $\boldsymbol{P}_k$ are defined as $\boldsymbol{J}_k = \mathbf{1}_k \mathbf{1}_k^\top$ and $\boldsymbol{P}_k = \boldsymbol{I}_k - k^{-1}\boldsymbol{J}_k$, respectively. We will use extensively the Kronecker (or direct) product $\boldsymbol{A} \otimes \boldsymbol{B}$ of matrices and the direct sum $\boldsymbol{A} \oplus \boldsymbol{B}$ of matrices. The symbol $\xrightarrow{D}$ stands as an abbreviation for "converges in distribution to", $\xrightarrow{P}$ for "converges in probability to" and acronym CMT for "Continuous Mapping Theorem". In estimating a sequence of parameters $\theta_b = O(1)$ by a sequence of estimators $T_{n,b}$, consistency is meant in the sense of $\mathrm{E}(T_{n,b} - \theta_b)^2 \to 0$ as (n, b) go to infinity.

Note that from the distributional assumption made above

$$\mathrm{E}[\boldsymbol{X}_{ik}] = \boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ib})^\top$$

and $\mathrm{Var}(\boldsymbol{X}_{ik}) = \boldsymbol{\Sigma}_i$ where $\boldsymbol{\Sigma}_i$ is a $b \times b$ positive definite matrix. Let

$$\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{1b}, \ldots, \mu_{a1}, \ldots, \mu_{ab})^\top$$

and $\widetilde{\boldsymbol{\Sigma}} = \bigoplus_{i=1}^{a} \boldsymbol{\Sigma}_i / n_i$. Then we have $\mathrm{E}[\overline{\boldsymbol{X}}] = \boldsymbol{\mu}$ and $\mathrm{Var}(\overline{\boldsymbol{X}}) = \widetilde{\boldsymbol{\Sigma}}$.

The three hypotheses of interest can be expressed as

$$\mathcal{H}_0^\phi : \boldsymbol{K}_\phi \boldsymbol{\mu} = \mathbf{0},$$

for $\phi \in \{AB, B, A\}$ with

$$\boldsymbol{K}_{AB} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b, \quad \boldsymbol{K}_B = \boldsymbol{J}_a \otimes \boldsymbol{P}_b \quad \text{and} \quad \boldsymbol{K}_A = \boldsymbol{D}_a \otimes b^{-1}\boldsymbol{J}_b,$$

where $\boldsymbol{D}_a = \mathrm{diag}\{n_1, \ldots, n_a\} - n^{-1}\boldsymbol{n}\boldsymbol{n}^\top$, $\boldsymbol{n} = (n_1, \ldots, n_a)^\top$ and $n = n_1 + n_2 + \cdots + n_a$. These null hypotheses correspond to no-interaction effects of levels of factor A with levels of factor B, no-main effects of factor B, and no-main effects of factor A, respectively. To see that the hypothesis of no interaction is equivalent to $\mathcal{H}^{AB}$, notice that no interaction means

$$\boldsymbol{C}_1(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_a) = \cdots = \boldsymbol{C}_1(\boldsymbol{\mu}_{a-1} - \boldsymbol{\mu}_a) = \mathbf{0}$$
$$\iff \boldsymbol{C}_1 \boldsymbol{M} \boldsymbol{C}_2^\top = \mathbf{0}_{(b-1)\times(a-1)} \iff (\boldsymbol{C}_2 \otimes \boldsymbol{C}_1)\boldsymbol{\mu} = \mathbf{0},$$

where $\boldsymbol{C}_1 = (\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1})$, $\boldsymbol{C}_2 = (\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1})$ and $\boldsymbol{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_a)$. The matrix $\boldsymbol{C} = \boldsymbol{C}_2 \otimes \boldsymbol{C}_1$ is a contrast matrix and is full row rank. Clearly, the hypothesis $\boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{0}$ is equivalent to $\boldsymbol{C}^\top(\boldsymbol{C}\boldsymbol{C}^\top)^{-1}\boldsymbol{C}\boldsymbol{\mu} = (\boldsymbol{P}_a \otimes \boldsymbol{P}_b)\boldsymbol{\mu} = \boldsymbol{0}$. The other two hypotheses for the main effects can also be expressed similarly.

Define,

$$\boldsymbol{S}_i = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i)(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i)^\top \quad \text{and} \quad \boldsymbol{S} = \frac{1}{n-a}\sum_{k=1}^{a}(n_i - 1)\boldsymbol{S}_i.$$

In this Chapter, we introduce test statistics for multi-group high-dimensional repeated measures analysis. Unlike likelihood ratio tests, our tests do not involve the inverse of the pooled sample covariance matrix $\boldsymbol{S}^{-1}$. In the high-dimensional case, more precisely when $b > n - a$, the sample covariance matrix $\boldsymbol{S}$ is not invertible, making the likelihood ratio tests inapplicable. Furthermore, $\boldsymbol{S}$ may not even converge to $\boldsymbol{\Sigma}$, the population covariance matrix (see, for example, Chen and Qin, 2010).

We derive the asymptotic distributions of our test statistics for equal covariance matrices as well as unequal covariance matrices. It should be noted that the results for the unequal covariance case do not necessarily reduce to the corresponding results for the equal covariance case by simply setting $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_a = \boldsymbol{\Sigma}$. There are some subtleties which warrant separate treatment of the two cases. First, the results for the equal covariance case are nice and clean. Instructively, it would make the results accessible if presented from the simpler to the more complex ones. Second, the proofs for the unequal covariance results build upon those for equal covariance. Third, the assumptions for the equal covariance case somewhat differ from those needed for the unequal covariance case. One assumption $A3'$, given in Section 2.4 on page 14, which requires proportional divergence of individual sample sizes with the dimension is not needed for the equal covariance case. The equal covariance case only requires the total sample sizes to grow with the dimension. Fourth, the constant c' (see Theorem 2.4.1) which contains unknown parameters in the unequal covariance case, will reduce to a known quantity c (see Theorem 2.3.1) in the equal covariance case. Estimation of c' is needed whereas no estimation of the analogous constant c, in the equal covariance case, is needed. In addition, the bounds for c' given in Theorem 2.4.1 do not quite reduce to those for c given in Theorem 2.3.1. The simplicity of the equal covariance situation affords us a more precise lower bound for c.

2.3 Tests under Equal Covariance Matrices

In this section, we assume that the covariance matrices $\boldsymbol{\Sigma}_i$ are equal and denote the common covariance matrix by $\boldsymbol{\Sigma}$. We will construct testing procedures under the following high-dimensional asymptotic frameworks:

$A1$: $c_j := \mathrm{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^j\}/b = O(1)$ as $b \to \infty$ for $j = 1, 2, 3, 4$.

$A2$: $n \to \infty$ and $b \to \infty$.

Note that $\beta_{\max}^j = O(1)$ for $j = 1, 2, 3, 4$ is sufficient for assumption $A1$ to hold where $\beta_{\max} = \max\{\beta_1, \ldots, \beta_b\}$ and $\beta_1, \ldots, \beta_b$ are the eigenvalues of $\boldsymbol{P}_b\boldsymbol{\Sigma}$. To elaborate on the significance of assumption $A1$, consider $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_b + \rho\boldsymbol{J}_b$ for $-1/(b-1) < \rho < 1$. This covariance structure is known, in multivariate statistics, as equi-correlation structure. For this covariance matrix, $A1$ holds because $\mathrm{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^j\} = (b-1)(1-\rho)^j$ for $j = 1, 2, 3, 4$. On the other hand, we can write $\boldsymbol{P}_b = \boldsymbol{Q}^\top \mathrm{diag}\{1, \ldots, 1, 0\}\boldsymbol{Q}$ where $\boldsymbol{Q}$ is an orthogonal matrix whose columns are the orthonormal eigenvectors of $\boldsymbol{P}_b$. The covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{Q}^\top \mathrm{diag}\{1, \ldots, b\}\boldsymbol{Q}$ doesn't satisfy $A1$ because $\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}) = b(b-1)/2$.

2.3.1 Test for interaction effect AB

We note that $\boldsymbol{K}_{AB}\boldsymbol{\mu} = \boldsymbol{0}$ if and only if $\boldsymbol{\mu}^\top\boldsymbol{K}_{AB}^\top\boldsymbol{K}_{AB}\boldsymbol{\mu} = \boldsymbol{\mu}^\top\boldsymbol{K}_{AB}\boldsymbol{\mu} = 0$, since $\boldsymbol{K}_{AB}$ is symmetric and idempotent matrix. Thus, the hypotheses for interaction effect AB is equivalent to

$$\mathcal{H}_0^{AB} : \boldsymbol{\mu}^\top\boldsymbol{K}_{AB}\boldsymbol{\mu} = 0 \quad \text{VS} \quad \mathcal{H}_1^{AB} : \boldsymbol{\mu}^\top\boldsymbol{K}_{AB}\boldsymbol{\mu} > 0.$$

Consider a reasonable estimator of $\boldsymbol{\mu}^\top\boldsymbol{K}_{AB}\boldsymbol{\mu}$ given by $H^{(AB)} = \overline{\boldsymbol{X}}^\top\boldsymbol{K}_{AB}\overline{\boldsymbol{X}}$. In Theorem 2.3.1 below asymptotic sampling distribution of a scaled and centered version of $H^{(AB)}$ is given.

Theorem 2.3.1. *If the null hypothesis* $\mathcal{H}_0^{AB}$ *holds, then*

$$U_{AB} := \frac{1}{\sqrt{b}} \left\{ \left(1 - \frac{1}{a}\right)^{-1} \left(\sum_{i=1}^{a} \frac{1}{n_i}\right)^{-1} H^{(AB)} - \operatorname{tr}(\boldsymbol{P}_b \boldsymbol{\Sigma}) \right\} \xrightarrow{D} \mathcal{N}(0, 2cc_2),$$

under the high-dimensional asymptotic frameworks A1 and A2, where

$$c = \frac{a(a-2)}{(a-1)^2} \sum_{i=1}^{a} \frac{1}{n_i^2} \Bigg/ \left(\sum_{i=1}^{a} \frac{1}{n_i}\right)^2 + \frac{1}{(a-1)^2} \in \left[\frac{1}{a-1}, 1\right].$$

The bounds given for c in Theorem 2.3.1, besides establishing that $c = O(1)$ as $b, n \to \infty$, provide insight into the influence of the value of a on the variance of U_{AB}. For example, $a = 2$ gives the largest possible variance. The variance could potentially decrease when a gets large. This is somewhat apparent in the simulation study Table 2.1.

The result of Theorem 2.3.1 depends on $bc_1 = \operatorname{tr}(\boldsymbol{P}_b \boldsymbol{\Sigma})$ and c_2 which are unknown quantities. For practical applications we need unbiased and consistent estimators of them. Define

$$\widehat{c}_1 = \frac{\operatorname{tr}(\boldsymbol{P}_b \boldsymbol{S})}{b} \quad \text{and } \widehat{c}_2 = \frac{(n-a)^2}{b(n-a-1)(n-a+2)} \left\{ \operatorname{tr}\{(\boldsymbol{P}_b \boldsymbol{S})^2\} - \frac{1}{n-a} \{\operatorname{tr}(\boldsymbol{P}_b \boldsymbol{S})\}^2 \right\}.$$

The next Theorem proves the unbiasedness and consistency of $\widehat{c}_i$ for $i = 1, 2$.

Theorem 2.3.2. *For* $i = 1, 2$, $\widehat{c}_i$ *is an unbiased and consistent estimator of* c_i *under the high-dimensional asymptotic frameworks A1 and A2. Moreover, we have* $\sqrt{b}(\widehat{c}_1 - c_1) \xrightarrow{P} 0$.

Using the results of Theorems 2.3.1 and 2.3.2, we propose a test statistic, namely $\widehat{T}_{AB}$, for testing $\mathcal{H}_0^{AB}$ and give its asymptotic null distribution in Corollary 2.3.3.

Corollary 2.3.3. *If the null hypothesis* $\mathcal{H}_0^{AB}$ *holds, then*

$$\widehat{T}_{AB} := \frac{1}{\sqrt{2bc\widehat{c}_2}} \left\{ \left(1 - \frac{1}{a}\right)^{-1} \left(\sum_{i=1}^{a} \frac{1}{n_i}\right)^{-1} H^{(AB)} - b\widehat{c}_1 \right\} \xrightarrow{D} \mathcal{N}(0, 1),$$

under the high-dimensional asymptotic frameworks A1 and A2.

For $a = 2$ the test statistic and results in Corollary 2.3.3 reduce to those of Theorem 2.1 of Takahashi and Shutoh (2016).

We close this section by mentioning that the proofs we provided do not require any relation in the rates of divergences of n and b. Please note that $\mathrm{Var}(\widehat{c}_2)$ goes to zero as long as both b and n tend to infinity even at a differing rate. We must acknowledge, though, that such an assumption is inevitable for the unequal covariance case.

2.3.2 Test for the main effect of factor B

We note that $\boldsymbol{K}_B\boldsymbol{\mu} = \boldsymbol{0}$ if and only if $\boldsymbol{\mu}^\top \boldsymbol{K}_B^\top \boldsymbol{K}_B \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{K}_B \boldsymbol{\mu} = 0$ since $\boldsymbol{K}_B$ is symmetric and idempotent matrix. The hypotheses for main effect of factor B are equivalent to

$$\mathcal{H}_0^B : \boldsymbol{\mu}^\top \boldsymbol{K}_B \boldsymbol{\mu} = 0 \quad \text{VS} \quad \mathcal{H}_1^B : \boldsymbol{\mu}^\top \boldsymbol{K}_B \boldsymbol{\mu} > 0.$$

Here also, a reasonable estimator of $\boldsymbol{\mu}^\top \boldsymbol{K}_B \boldsymbol{\mu}$ is $H^{(B)} = \overline{\boldsymbol{X}}^\top \boldsymbol{K}_B \overline{\boldsymbol{X}}$.

Theorem 2.3.4. *If the null hypothesis $\mathcal{H}_0^B$ holds, then*

$$U_B := \frac{1}{\sqrt{b}} \left\{ \left(\sum_{i=1}^{a} \frac{1}{n_i} \right)^{-1} H^{(B)} - \mathrm{tr}(\boldsymbol{P}_b \boldsymbol{\Sigma}) \right\} \xrightarrow{D} \mathcal{N}(0, 2c_2),$$

under the high-dimensional asymptotic frameworks A1 and A2.

Comparing the results in Theorems 2.3.1 and 2.3.4, the quantity U_{AB} is less variable than U_B.

A consistent estimator of c_1 and c_2 are given in Theorem 2.3.2. Corollary 2.3.5 proposes a test for the main effect of factor B and presents the asymptotic null distribution of the test statistic under the same asymptotic framework as in Corollary 2.3.3.

Corollary 2.3.5. *If the null hypothesis $\mathcal{H}_0^B$ holds, then*

$$\widehat{T}_B := \frac{1}{\sqrt{2b\widehat{c}_2}} \left\{ \left(\sum_{i=1}^{a} \frac{1}{n_i} \right)^{-1} H^{(B)} - b\widehat{c}_1 \right\} \xrightarrow{D} \mathcal{N}(0, 1),$$

under the high-dimensional asymptotic frameworks A1 and A2.

The testing problem in this subsection for $a = 1$ is covered by Pauly et al. (2015) when the contrast matrix is chosen to be $\boldsymbol{P}_b$, i.e., when C (in their notation) is replaced with $\boldsymbol{P}_b$. However, they use different consistent estimators for bc_1 and bc_2 but in the end our limiting distributions agree for the case $\mathrm{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^4\}/\mathrm{tr}^2\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^2\} \to 0$ as $b \to \infty$. It should be noted that assumption $A1$ implies that $\mathrm{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^4\}/\mathrm{tr}^2\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^2\} \to 0$. When $a = 2$, the test for no effect of levels of factor B in Takahashi and Shutoh (2016) were formulated in terms of the weighted group mean vectors where the weights are the sample sizes of the groups. Our hypothesis is formulated in terms of the simple average of the group means as a result of which, as one would naturally expect, the hypothesis does not depend on sample sizes. This difference resulted in different tests and asymptotic results.

2.3.3 Test for the main effect of factor A

We begin by establishing the equivalence of the hypotheses for the main effects of factor A expressed in a linear and quadratic forms.

Proposition 2.3.6. *The condition* $\boldsymbol{K}_A\boldsymbol{\mu} = \boldsymbol{0}$ *is equivalent to* $\boldsymbol{\mu}^\top\boldsymbol{K}_A\boldsymbol{\mu} = 0$.

According to Proposition 2.3.6, the hypothesis for the main effect of factor A is equivalent to

$$\mathcal{H}_0^A : \boldsymbol{\mu}^\top\boldsymbol{K}_A\boldsymbol{\mu} = 0 \quad \text{VS} \quad \mathcal{H}_1^A : \boldsymbol{\mu}^\top\boldsymbol{K}_A\boldsymbol{\mu} > 0.$$

It should also be noted that for any $\boldsymbol{x} = (x_1, \ldots, x_a)^\top \in \mathbb{R}^a$,

$$\boldsymbol{x}^\top\boldsymbol{D}_a\boldsymbol{x} = \sum_{i=1}^{a} n_i x_i^2 - n^{-1}(\sum_{i=1}^{a} n_i x_i)^2 = \sum_{i=1}^{a} n_i(x_i - \overline{x})^2 \geq 0,$$

where $\overline{x} = n^{-1}\sum_{i=1}^{a} n_i x_i$. Thus, $\boldsymbol{K}_A = \boldsymbol{D}_a \otimes \boldsymbol{J}_b/b$ is positive semidefinite.

Once again we will build our test from a reasonable estimator of $\boldsymbol{\mu}^\top\boldsymbol{K}_A\boldsymbol{\mu}$, namely $\overline{\boldsymbol{X}}^\top\boldsymbol{K}_A\overline{\boldsymbol{X}}$. It may seem that the hypothesis $\mathcal{H}_0^A$ depends on the sample sizes $n_1, \ldots, n_a$. Nevertheless, one can easily check that the hypothesis of no main effect of factor A is equivalent to $\boldsymbol{1}_b^\top\boldsymbol{\mu}_1 = \cdots = \boldsymbol{1}_b^\top\boldsymbol{\mu}_a$. This shows that the hypothesis $\mathcal{H}_0^A$ does not depend on the sample sizes. Furthermore, it is reasonable to use the between group sum of squares for the transformed random variables $Y_{ij} = \mathbf{1}_b^\top \boldsymbol{X}_{ij}$ to test this hypothesis. As it turns out $H^{(A)} = \overline{\boldsymbol{X}}^\top \boldsymbol{K}_A \overline{\boldsymbol{X}}$ is the between group sum of squares.

It is easy to show that $H^{(A)}/b$ is distributed as $d_1\chi^2_{a-1}$ under the null hypothesis, where $d_1 = \mathrm{tr}(\boldsymbol{J}_b\boldsymbol{\Sigma})/b^2$. Also one can see that $(n-a)\widehat{d}_1 = (n-a)\mathrm{tr}(\boldsymbol{J}_b\boldsymbol{S})/b^2$ is distributed as $d_1\chi^2_{n-a}$ and that $H^{(A)}$ is independent of $\widehat{d}_1$. The latter follows because $\overline{\boldsymbol{X}}$ is independent of $\boldsymbol{S}$. Thus, an exact test for $\mathcal{H}_0^A$ is

$$\widehat{T}_A = \frac{H^{(A)}/b(a-1)}{\widehat{d}_1},$$

which has an exact $\mathcal{F}_{a-1,n-a}$ distribution under the null hypothesis.

For $a = 2$, $\boldsymbol{D}_a = (1/n_1 + 1/n_2)^{-1}(1,-1)^\top(1,-1)$ and

$$\begin{aligned} bH^{(A)} &= \mathrm{vec}(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2)^\top(\boldsymbol{D}_a \otimes \boldsymbol{J}_b)\mathrm{vec}(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2) \\ &= \mathrm{vec}(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2)^\top(\boldsymbol{D}_a \otimes \mathbf{1}_b)(\boldsymbol{I}_a \otimes \mathbf{1}_b^\top)\mathrm{vec}(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2). \end{aligned}$$

Applying the identity $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^\top \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$, we get

$$H^{(A)} = \frac{1}{b}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}[(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)^\top \mathbf{1}_b]^2.$$

Therefore, our test statistic and that of Takahashi and Shutoh (2016) are equivalent.

2.4 Tests under Unequal Covariance Matrices

In this section, we do not assume that the a populations have equal covariances matrices. Relaxing the equal covariance matrices assumption necessitates adjustment of the asymptotic conditions. We will need the following assumptions to construct testing procedures under unequal covariance matrices.

A1′: $c'_j := (\sum_{i=1}^a 1/n_i)^{-j}\,\mathrm{tr}\{(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})^j\}/b = O(1)$ as $b \to \infty$ for $j = 1, 2, 3$.

A2′: $d'_1 := \mathrm{tr}(\boldsymbol{K}_A\widetilde{\boldsymbol{\Sigma}})/b = O(1)$ as $b \to \infty$.

A3′: $n_i \to \infty, b \to \infty$ and $b/n_i \to \xi_i \in (0, \infty)$ for $i = 1, \ldots, a$.

A4′: $c''_3 := (\sum_{i=1}^a 1/n_i)^{-3}\,\mathrm{tr}\{(\boldsymbol{K}_{AB}\widetilde{\boldsymbol{\Sigma}})^3\}/b = O(1)$ as $b \to \infty$.

A5′: $\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^4/b = O(1)$ as $b \to \infty$ for $i = 1, \ldots, a$.

While assumption $A3'$ require proportional divergence of the sample sizes $n_1, \ldots, n_a$ and the dimension b, assumptions $A1'$, $A2'$, $A4'$ and $A5'$ require regularity conditions on the covariance matrices. Some remarks are in order.

(i) Stronger but simpler assumptions which together with $A3'$ are sufficient for $A1'$ and $A4'$ are $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)/b = O(1)$, $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j)/b = O(1)$ and

$$\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j\boldsymbol{P}_b\boldsymbol{\Sigma}_k)/b = O(1)$$

as $b \to \infty$ for all $i, j, k \in \{1, \ldots, a\}$.

(ii) It can be seen that

$$c_1' = \frac{1}{b}\left(\sum_{i=1}^{n}\frac{1}{n_i}\right)^{-1}\sum_{i=1}^{a}\operatorname{tr}\left(\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right) = \frac{1}{b}\left(\sum_{i=1}^{n}\xi_i\right)^{-1}\sum_{i=1}^{a}\xi_i\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i).$$

Therefore, since $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i) \geq 0$ and $\xi_i > 0$ for $i = 1, \ldots, a$, the condition $c_1' = O(1)$ and assumption $A3'$ are sufficient for $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)/b = O(1)$ for $i = 1, \ldots, a$. Similarly,

$$c_2' = \left(\sum_{i=1}^{a}\xi_i\right)^{-2}\left[\frac{1}{b}\sum_{i=1}^{a}\xi_i^2\operatorname{tr}\left\{(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^2\right\} + \frac{1}{b}\sum_{i\neq j}^{a}\xi_i\xi_j\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j)\right].$$

Since $\xi_i > 0$ and $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j) = \operatorname{tr}\left\{(\boldsymbol{\Sigma}_j^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2})(\boldsymbol{\Sigma}_j^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2})'\right\} \geq 0$, it follows that $c_2' = O(1)$ and $A3'$ imply $\operatorname{tr}\left\{(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^2\right\}/b = O(1)$ for $i = 1, \ldots, a$. The manipulations for c_1' and c_2' above make it clear that the proportional-divergence assumption $A3'$ can be replaced with

$$A3'' : n_i \to \infty, b \to \infty \text{ and } n_i/n \to \tilde{\xi}_i \in (0,1) \quad \text{for} \quad i = 1, \ldots, a$$

without affecting the validity of the results.

(iii) As one can imagine, the assumptions needed for unequal covariance case are much more involved compared to the equal covariance case. For example, one can easily verify that

$$\operatorname{tr}\left\{(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})^j\right\} = \operatorname{tr}\left\{\left(\sum_{i=1}^{a}\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right)^j\right\},$$

for $j = 1, 2, 3$. With this simplification if the covariance matrices are equal, i.e., $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_a = \boldsymbol{\Sigma}$, then $c'_j = c_j$ for $j = 1, 2$. Furthermore, the assumptions $A1'$, $A4'$ and $A4$ reduce to $A1$. On the other hand, proportional divergence of individual sample sizes with the dimension (assumption $A3'$) is not needed for the equal covariance. It is adequate if the total sample size diverges with the dimension (see assumption $A2$).

To put the assumptions $A1'$, $A2'$, $A4'$ and $A5'$ in perspective, the covariance matrices $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$, for $i = 1, \ldots, a$ and any $-1/(b-1) < \rho_i < 1$, satisfy these assumptions because $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i) = (b-1)(1-\rho_i)$, $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j) = (b-1)(1-\rho_i)(1-\rho_j)$, $\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j\boldsymbol{P}_b\boldsymbol{\Sigma}_k) = (b-1)(1-\rho_i)(1-\rho_j)(1-\rho_k)$, $\operatorname{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^4\} = (b-1)(1-\rho_i)^4$, for all $i, j, k \in \{1, \ldots, a\}$, and $\operatorname{tr}(\boldsymbol{K}_A\widetilde{\boldsymbol{\Sigma}}) = \sum_{i=1}^a (1-n_i/n)(1+(b-1)\rho_i)$. On the contrary, $\boldsymbol{\Sigma}_i = \rho_i \boldsymbol{Q}^\top \operatorname{diag}\{1, \ldots, b\}\boldsymbol{Q}$ fail to satisfy the assumptions, where the columns of matrix $\boldsymbol{Q}$ are the orthonormal eigenvectors of $\boldsymbol{P}_b$ as equal case.

2.4.1 Test for interaction effect AB

Here also we start by presenting the asymptotic sampling distribution of a centered and scaled version of $H^{(AB)}$ when the covariance matrices are not necessarily equal.

Theorem 2.4.1. *If the null hypothesis $\mathcal{H}_0^{AB}$ holds, then*

$$U'_{AB} := \frac{1}{\sqrt{b}} \left(\sum_{i=1}^a \frac{1}{n_i} \right)^{-1} \left\{ \left(1 - \frac{1}{a}\right)^{-1} H^{(AB)} - \sum_{i=1}^a \frac{\operatorname{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)}{n_i} \right\} \xrightarrow{D} \mathcal{N}(0, 2c'c'_2),$$

under the high-dimensional asymptotic frameworks $A1'$, $A3'$ and $A4'$ where

$$c' = \frac{a(a-2)}{(a-1)^2} \sum_{i=1}^a \frac{\operatorname{tr}\left((\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^2\right)}{n_i^2} \Bigg/ \operatorname{tr}\left\{ \left(\sum_{i=1}^a \frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i} \right)^2 \right\} + \frac{1}{(a-1)^2} \in \left[\frac{1}{(a-1)^2}, 1 \right].$$

Notice, here also, that the bounds for c', besides establishing that $c' = O(1)$ under the asymptotic framework $A1'$ and $A3'$, give insight into how the variance of U'_{AB} is influenced by the value of a.

In what follows, unbiased and consistent estimators of the unknown quantities in the asymptotic sampling distribution in Theorem 2.4.1 will be given. To that end, let

$$c_{1i} = \frac{\text{tr}(\boldsymbol{P}_b \boldsymbol{\Sigma}_i)}{b}, \quad c_{2i} = \frac{\text{tr}\{(\boldsymbol{P}_b \boldsymbol{\Sigma}_i)^2\}}{b} \quad \text{and} \quad c_{2ii'} = \frac{\text{tr}(\boldsymbol{P}_b \boldsymbol{\Sigma}_i \boldsymbol{P}_b \boldsymbol{\Sigma}_{i'})}{b} \quad \text{for } i \neq i'.$$

Then we can see that

$$c_1' = \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-1} \frac{\text{tr}(\boldsymbol{K}_B \widetilde{\boldsymbol{\Sigma}})}{b} = \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-1} \sum_{i=1}^a \frac{c_{1i}}{n_i} \quad \text{and}$$
$$c_2' = \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \frac{\text{tr}\{(\boldsymbol{K}_B \widetilde{\boldsymbol{\Sigma}})^2\}}{b} = \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \sum_{i=1}^a \frac{c_{2i}}{n_i^2} + \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \sum_{i \neq i'} \frac{c_{2ii'}}{n_i n_{i'}}.$$

In view of assumption $A3'$, it suffices to find unbiased and consistent estimators of c_{1i}, c_{2i} and $c_{2ii'}$. Denote

$$\begin{aligned}
\widehat{c_{1i}} &= \frac{\text{tr}(\boldsymbol{P}_b \boldsymbol{S}_i)}{b}, \\
\widehat{c_{2i}} &= \frac{(n_i - 1)^2}{b(n_i+1)(n_i-2)} \left\{ \text{tr}\{(\boldsymbol{P}_b \boldsymbol{S}_i)^2\} - \frac{1}{n_i - 1} \{\text{tr}(\boldsymbol{P}_b \boldsymbol{S}_i)\}^2 \right\} \quad \text{and} \\
\widehat{c_{2ii'}} &= \frac{\text{tr}(\boldsymbol{P}_b \boldsymbol{S}_i \boldsymbol{P}_b \boldsymbol{S}_{i'})}{b} \quad \text{for} \quad i \neq i',
\end{aligned}$$

and define

$$\begin{aligned}
\widehat{c_1'} &= \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-1} \sum_{i=1}^a \frac{\widehat{c_{1i}}}{n_i}, \\
\widehat{c_2'} &= \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \sum_{i=1}^a \frac{\widehat{c_{2i}}}{n_i^2} + \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \sum_{i \neq i'} \frac{\widehat{c_{2ii'}}}{n_i n_{i'}} \quad \text{and} \\
\widehat{c'} &= \frac{a(a-2)}{(a-1)^2} \left(\sum_{i=1}^a \frac{1}{n_i}\right)^{-2} \sum_{i=1}^a \frac{\widehat{c_{2i}}}{n_i^2 \widehat{c_2'}} + \frac{1}{(a-1)^2}.
\end{aligned}$$

Unbiased and consistent estimators of c_i' for $i = 1, 2$ are given in Theorem 2.4.2 below.

Theorem 2.4.2. *$\widehat{c_1'}$, $\widehat{c_2'}$ and $\widehat{c'} \cdot \widehat{c_2'}$ are unbiased and consistent estimators of c_1', c_2' and $c' \cdot c_2'$, respectively, under the high-dimensional asymptotic frameworks $A1'$, $A3'$ and $A5'$. Moreover, we have $\sqrt{b}(\widehat{c_1'} - c_1') \xrightarrow{P} 0$.*

Asymptotic test for $\mathcal{H}^{AB}$ under unequal covariance assumption is devised in Corollary 2.4.3. The proof is analogous to that of Corollary 2.3.3.

Corollary 2.4.3. *If the null hypothesis* $\mathcal{H}_0^{AB}$ *holds, then*

$$\widehat{T}'_{AB} := \frac{1}{\sqrt{2b\,\widehat{c'c'_2}}}\left\{\left(1-\frac{1}{a}\right)^{-1}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1} H^{(AB)} - b\widehat{c}'_1\right\} \xrightarrow{D} \mathcal{N}(0,1),$$

under the high-dimensional asymptotic frameworks $A1'$, $A3'$, $A4'$ *and* $A5'$.

It is well known in the low-dimensional MANOVA that the effect of unequal covariance on tests that assume equal covariance is more pronounced when the sample sizes are different. In particular, the effect gets worse if smaller sample sizes are associated with large covariance matrices (large in the sense of the eigenvalues). Comparison of Corollaries 2.3.3 and 2.4.3 reveals that the same phenomena appears to exist in the high-dimensional tests of this Chapter.

2.4.2 Test for the main effect of factor B

As in the equal covariance case, here also the results for $H^{(B)}$ follow in a manner analogous to those of $H^{(AB)}$.

Theorem 2.4.4. *If the null hypothesis* $\mathcal{H}_0^{B}$ *holds, then*

$$U'_B := \frac{1}{\sqrt{b}}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1}\left\{H^{(B)} - \sum_{i=1}^{a}\frac{\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)}{n_i}\right\} \xrightarrow{D} \mathcal{N}(0, 2c'_2),$$

under the high-dimensional asymptotic frameworks $A1'$ *and* $A3'$.

Therefore, an asymptotic test for the main effects of factor B under unequal covariance matrices is as provided in Corollary 2.4.5.

Corollary 2.4.5. *If the null hypothesis* $\mathcal{H}_0^{B}$ *holds, then*

$$\widehat{T}'_B := \frac{1}{\sqrt{2b\widehat{c}'_2}}\left\{\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1} H^{(B)} - b\widehat{c}'_1\right\} \xrightarrow{D} \mathcal{N}(0,1),$$

under the high-dimensional asymptotic frameworks $A1'$, $A3'$ *and* $A5'$.

2.4.3 Test for the main effect of factor A

Notice that the rank of $\boldsymbol{K}_A$ is at most $a-1$. That is, it does not grow with b and, hence, it does not make sense to consider large-(b, n_i) asymptotic. Instead we consider an approximation based on matching moments that is known to work well in other situations (e.g., Brunner et al., 1997). Notice that $\mathrm{E}\left[H^{(A)}\right] = bd_1'$. An unbiased estimator of d_1' is

$$\widehat{d_1'} = \sum_{i=1}^{a} \left(1 - \frac{n_i}{n}\right) \widehat{d_{1i}},$$

where $\widehat{d_{1i}} = \mathrm{tr}(\boldsymbol{J}_b \boldsymbol{S}_i)/b^2$. Then a reasonable test statistic is

$$\widehat{T}_A' = \frac{H^{(A)}}{b\widehat{d_1'}}.$$

To get an approximate distribution of $\widehat{T}_A'$, we propose to approximate the distributions of $H^{(A)}$ and $b\widehat{d_1'}$ by constant multiples of chi-square distributions. More precisely, we assume approximately

$$H^{(A)} \sim g\chi_f^2 \quad \text{and} \quad b\widehat{d_1'} \sim g_0 \chi_{f_0}^2$$

where (g, f) and (g_0, f_0) are found by matching the first two moments. After some algebra, we get

$$f = \frac{2\{\mathrm{E}[H^{(A)}]\}^2}{\mathrm{Var}(H^{(A)})} = \frac{\sum_{i=1}^{a} \left(1 - \frac{n_i}{n}\right)^2 d_{1i}^2 + \sum_{i \neq i'} \left(1 - \frac{n_i}{n}\right)\left(1 - \frac{n_{i'}}{n}\right) d_{1i} d_{1i'}}{\sum_{i=1}^{a} \left(1 - \frac{n_i}{n}\right)^2 d_{1i}^2 + \sum_{i \neq i'} \frac{n_i n_{i'}}{n^2} d_{1i} d_{1i'}},$$

$$f_0 = \frac{2\{\mathrm{E}[bd_1']\}^2}{\mathrm{Var}(bd_1')} = \frac{\sum_{i=1}^{a} \left(1 - \frac{n_i}{n}\right)^2 d_{1i}^2 + \sum_{i \neq i'} \left(1 - \frac{n_i}{n}\right)\left(1 - \frac{n_{i'}}{n}\right) d_{1i} d_{1i'}}{\sum_{i=1}^{a} \left(1 - \frac{n_i}{n}\right)^2 \frac{d_{1i}^2}{n_i - 1}},$$

and $f_0 g_0 / fg = 1$, where $d_{1i} = \mathrm{tr}(\boldsymbol{J}_b \boldsymbol{\Sigma}_i)/b^2$.

Now, we approximate the distributions of $\widehat{T}_A'$ as

$$\widehat{T}_A' = \frac{H^{(A)}/fg}{b\widehat{d_1'}/f_0 g_0} \sim \mathcal{F}_{f, f_0},$$

under the null hypothesis $\mathcal{H}_0^A$. For practical application of the approximate degrees of freedoms, we need unbiased and consistent estimators of the quantity d_{1i} and its square d_{1i}^2 .

Theorem 2.4.6. *Under the asymptotic framework A2′ and A3′, for $i = 1, \dots, a$, unbiased and consistent estimators of d_{1i} and d_{1i}^2 are*

$$\widehat{d_{1i}} = \frac{\mathrm{tr}(\boldsymbol{J}_b \boldsymbol{S})}{b^2} \quad \textit{and} \quad \widehat{d_{1i}^2} = \frac{n_i - 1}{n_i + 1}(\widehat{d_{1i}})^2.$$

Although the estimators $\widehat{f}$ and $\widehat{f_0}$ of f and f_0, respectively, obtained by plugging-in the unbiased estimators of d_{1i} and d_{1i}^2 are not unbiased themselves, they will still be consistent and our numerical studies have shown that using such estimators give much improved accuracies.

2.5 Simulation Study

2.5.1 Size of the Tests

We generate 10,000 replications of data from $\boldsymbol{X}_{ik} \sim \mathcal{N}_b(\mu \mathbf{1}_b, \boldsymbol{\Sigma}_i)$ for $\mu \in \mathbb{R}$. Under this model, all the three null hypotheses $\mathcal{H}_0^{AB}$, $\mathcal{H}_0^{B}$ and $\mathcal{H}_0^{A}$ hold. Since all the three tests are invariant to the choice of μ, the particular value of μ used is immaterial. For covariance, we consider the structures $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$ and $\boldsymbol{\Sigma}_i$ which has ones for its diagonals and $\rho_i |j - j'|^{-1/4}$ for the (j, j') off-diagonal element. It should be noted that a covariance matrix $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$ will be positive definite if and only if $-1/(b-1) < \rho_i < 1$. We consider a range of values for ρ_i. We also consider random $\boldsymbol{\Sigma}_i$. Let $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i^\top \boldsymbol{\Lambda}_i \boldsymbol{Q}_i$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix with diagonal entries taken from $\mathcal{U}nif(0,1)$ and $\boldsymbol{Q}_i$ is a orthogonal matrix. Indeed, $\boldsymbol{Q}_i$ can be defined from the QR decomposition of a random matrix $\boldsymbol{Z}_i = (Z_{i,jj'})$ where $Z_{i,jj'}$ are iid random variables. Here, we consider three distributions for $Z_{i,jj'}$, namely $Z_{i,jj'} = 1_{\{j=j'\}}$ with probability 1, $Z_{i,jj'} \sim \mathcal{E}xp(1)$ and $Z_{i,jj'} \sim \mathcal{N}(0,1)$.

Although the assumed asymptotic frameworks stipulate n_i's to grow proportionally with b in the unequal covariance case, in reality the actual ratio varies from application to application. To investigate the effect of the various proportionality of growth, we look at values of several combinations of b, a and n_i's and take the desired (nominal) type I error rate $\alpha = 0.05$. For practical reasons, we also consider small b and large $n_1, \dots, n_a$ (and vice-versa) combinations with balanced as well as unbalanced designs. As far as the number of groups, we will consider $a = 2, 3, 4, 6$.

Tables 2.1 and 2.2 present actual type I error rates (test sizes) for the covariance structure $\boldsymbol{\Sigma}_i = \rho_i \boldsymbol{I}_b + (1-\rho_i)\boldsymbol{J}_b$ for equal and unequal group covariance cases, respectively. The empty cells in Table 1 and Table 3 correspond to the cases where b and ρ combinations do not yield positive definite covariance matrix. In the equal covariance case, since the test for the main effect of A is exact, it was not necessary to carry our simulation for this test. From Table 2.1 we see that, as n and b grow together, the performance of the tests in controlling type one error rates improve consistently for most of the cases. Table 2.2 seems to exhibit similar patterns in terms of the effects of the sizes of n and b. More noticeable is the test for the main effect of A appears to significantly improve as a gets larger. For the other covariance structure (Table 2.3 and 2.4) and the random covariance matrices (Tables 2.5 and 2.6), again similar patterns are observed with respect to n, b and a.

Table 2.1: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_b + \rho \boldsymbol{J}_b$.

		$\Pr(\widehat{T}_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$				$\Pr(\widehat{T}_{B} > z_\alpha)$ under $\mathcal{H}_0^{B}$			
a	$b, \boldsymbol{n}'$	$\rho = -0.01$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = -0.01$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
2	100,(26,26)	5.55	6.16	5.46	5.71	6.00	6.18	5.96	5.86
2	200,(51,51)		5.59	5.91	5.55		5.83	5.51	5.58
2	400,(51,51)		5.81	5.18	5.21		5.79	5.9	5.16
2	400,(101,101)		5.54	5.38	5.12		5.54	5.55	4.90
2	100,(12,25)	6.31	5.87	5.98	6.24	6.02	5.82	6.00	6.33
2	200,(25,50)		5.91	5.30	5.63		5.74	5.37	5.63
2	200,(50,100)		5.83	5.51	5.51		5.41	5.61	5.96
2	25,(50,100)	5.88	6.40	6.25	5.90	6.69	6.06	6.54	6.56
2	50,(50,100)	6.05	5.95	6.26	6.08	6.21	5.64	6.24	5.97
2	50,(100,100)	6.19	6.33	5.91	6.23	5.37	6.15	6.23	5.96
3	100,(18,18,17)	5.82	5.93	5.90	6.08	6.02	5.73	6.02	5.86
3	200,(35,34,34)		5.75	5.79	5.47		5.87	5.76	5.59
3	100,(12,13,25)	5.73	5.82	5.88	5.95	5.78	6.09	6.12	5.78
3	200,(25,25,50)		5.66	5.98	5.60		5.74	5.72	5.86
3	25,(50,100,100)	6.28	6.20	6.14	6.06	6.24	6.40	6.36	6.25
3	50,(50,100,100)	6.20	6.18	6.07	5.98	6.19	5.78	5.87	6.00
3	50,(100,100,100)	6.13	5.83	5.73	5.67	6.06	5.65	6.08	5.46
4	100,(14,14,13,13)	6.08	6.21	5.82	5.94	5.66	6.08	5.49	5.93
4	200,(26,26,26,26)		5.51	5.83	5.89		5.58	5.60	5.68
4	100,(12,13,25,25)	5.85	5.97	5.79	5.72	5.73	6.03	6.20	5.69
4	200,(25,25,50,50)		5.63	5.66	5.24		5.40	5.65	5.46
4	25,(50,50,100,100)	5.67	5.68	5.79	6.19	6.38	6.69	5.98	6.36
4	12,(100,100,100,100)	5.93	6.39	6.22	5.68	6.68	6.73	6.11	6.84
4	50,(100,100,100,100)	6.01	5.75	5.53	5.77	5.98	5.91	6.11	5.92
6	100,(10,10,9,9,9,9)	6.27	6.67	6.25	6.54	6.21	5.98	6.20	6.02

Table 2.2: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$.

a	$b, \boldsymbol{n}'$	ρ	$\Pr(\widehat{T}'_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$	$\Pr(\widehat{T}'_{B} > z_\alpha)$ under $\mathcal{H}_0^{B}$	$\Pr(\widehat{T}'_{A} > f_\alpha)$ under $\mathcal{H}_0^{A}$
2	100,(26,26)	(0,0.5)	5.85	5.61	4.89
2	200,(51,51)	(0,0.5)	5.37	5.53	5.21
2	400,(51,51)	(0,0.5)	5.29	5.64	4.76
2	400,(101,101)	(0,0.5)	5.39	5.53	4.97
2	100,(12,25)	(0,0.5)	6.11	6.18	4.92
2	200,(25,50)	(0,0.5)	6.23	5.72	4.62
2	200,(50,100)	(0,0.5)	5.98	6.12	4.97
2	25,(50,100)	(0,0.5)	6.81	6.33	5.11
2	50,(50,100)	(0,0.5)	6.01	6.04	4.71
2	50,(100,100)	(0,0.5)	6.09	5.95	5.07
3	100,(18,18,17)	(0,0.5,0.9)	5.68	6.57	4.60
3	200,(35,34,34)	(0,0.5,0.9)	5.44	5.98	4.99
3	100,(12,13,25)	(0,0.5,0.9)	6.79	6.61	4.31
3	200,(25,25,50)	(0,0.5,0.9)	5.52	5.60	4.81
3	25,(50,100,100)	(0,0.5,0.9)	6.69	6.40	4.80
3	50,(50,100,100)	(0,0.5,0.9)	6.30	6.22	4.58
3	50,(100,100,100)	(0,0.5,0.9)	5.94	6.00	4.38
4	100,(14,14,13,13)	(0,0.3,0.6,0.9)	6.55	5.73	5.30
4	200,(26,26,26,26)	(0,0.3,0.6,0.9)	5.70	5.81	4.62
4	100,(12,13,25,25)	(0,0.3,0.6,0.9)	6.36	6.44	4.25
4	200,(25,25,50,50)	(0,0.3,0.6,0.9)	5.96	5.82	5.01
4	25,(50,50,100,100)	(0,0.3,0.6,0.9)	6.24	6.27	4.94
4	12,(100,100,100,100)	(0,0.3,0.6,0.9)	6.27	6.63	4.85
4	50,(100,100,100,100)	(0,0.3,0.6,0.9)	5.63	5.83	5.00
6	100,(10,10,9,9,9,9)	(-0.01,0,0.2,0.4,0.6,0.8)	6.60	6.28	5.25

Table 2.3: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_{jj'})$, $\sigma_{jj} = 1$ and $\sigma_{jj'} = \rho/|j - j'|^{1/4}$ for $j \neq j'$.

		$\Pr(\widehat{T}_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$				$\Pr(\widehat{T}_{B} > z_\alpha)$ under $\mathcal{H}_0^{B}$			
a	$b, \boldsymbol{n}'$	$\rho = -0.01$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = -0.01$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
2	100,(26,26)	5.83	6.24	5.56	6.07	5.62	6.05	6.03	6.23
2	200,(51,51)	5.68	5.67	6.09	6.19	5.61	5.56	5.57	6.25
2	400,(51,51)		5.83	5.95	6.21		5.44	5.34	6.10
2	400,(101,101)		5.75	5.14	6.47		5.56	5.58	6.64
2	100,(12,25)	6.10	6.29	6.19	6.63	5.95	6.10	6.07	6.78
2	200,(25,50)	5.75	5.81	5.84	6.48	5.43	5.71	5.86	6.71
2	200,(50,100)		5.49	5.86	6.14		5.74	5.49	6.37
2	25,(50,100)	6.22	6.29	6.08	6.53	6.42	6.25	6.39	6.39
2	50,(50,100)	6.13	5.94	6.05	6.52	6.11	6.40	6.34	6.21
2	50,(100,100)	5.87	6.17	6.04	6.48	6.22	5.95	5.98	6.50
3	100,(18,18,17)	6.17	6.07	6.47	6.22	6.06	6.15	6.22	6.82
3	200,(35,34,34)		5.29	5.71	6.26		6.07	6.09	6.65
3	100,(12,13,25)	6.05	5.82	5.70	7.19	6.27	6.37	6.20	6.96
3	200,(25,25,50)	6.00	5.81	6.07	6.17	5.57	5.51	5.75	6.84
3	25,(50,100,100)	6.69	6.12	6.13	6.31	6.17	6.87	6.50	6.28
3	50,(50,100,100)	5.79	6.51	6.20	6.29	6.41	6.43	6.38	6.84
3	50,(100,100,100)	5.89	5.67	5.68	6.50	5.82	6.31	6.07	6.15
4	100,(14,14,13,13)	5.64	6.05	5.94	6.47	6.34	6.02	6.28	7.00
4	200,(26,26,26,26)	5.30	5.76	5.45	6.01	5.62	5.87	5.65	6.36
4	100,(12,13,25,25)	5.88	6.09	5.92	6.57	5.61	5.53	5.70	6.36
4	200,(25,25,50,50)	5.54	5.86	5.85	5.85	5.74	5.53	5.85	6.08
4	25,(50,50,100,100)	5.79	6.02	6.29	6.28	6.38	6.30	6.78	6.66
4	12,(100,100,100,100)	6.45	5.91	6.07	5.92	6.25	6.55	6.76	7.29
4	50,(100,100,100,100)	5.64	5.50	5.63	6.26	6.09	5.66	6.11	6.44
6	100,(10,10,9,9,9,9)	6.51	6.17	6.81	6.85	5.97	6.23	6.14	6.39

Table 2.4: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\Sigma}_i = (\sigma_{i,jj'})$, $\sigma_{i,jj} = 1$ and $\sigma_{i,jj'} = \rho_i/|j-j'|^{1/4}$ for $j \neq j'$.

a	$b, \boldsymbol{n}'$	ρ	$\Pr(\widehat{T}'_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$	$\Pr(\widehat{T}'_B > z_\alpha)$ under $\mathcal{H}_0^B$	$\Pr(\widehat{T}'_A > f_\alpha)$ under $\mathcal{H}_0^A$
2	100,(26,26)	(0,0.5)	6.49	6.17	5.22
2	200,(51,51)	(0,0.5)	6.02	6.11	4.78
2	400,(51,51)	(0,0.5)	5.93	6.00	5.29
2	400,(101,101)	(0,0.5)	5.71	5.54	5.03
2	100,(12,25)	(0,0.5)	6.60	6.58	4.79
2	200,(25,50)	(0,0.5)	6.19	5.80	4.60
2	200,(50,100)	(0,0.5)	5.57	5.93	5.23
2	25,(50,100)	(0,0.5)	7.00	6.55	4.57
2	50,(50,100)	(0,0.5)	6.25	6.19	5.04
2	50,(100,100)	(0,0.5)	5.99	5.91	5.27
3	100,(18,18,17)	(0,0.5,0.9)	6.39	6.19	4.68
3	200,(35,34,34)	(0,0.5,0.9)	6.45	6.32	4.78
3	100,(12,13,25)	(0,0.5,0.9)	6.69	6.18	4.41
3	200,(25,25,50)	(0,0.5,0.9)	6.18	5.87	4.97
3	25,(50,100,100)	(0,0.5,0.9)	6.28	6.39	4.80
3	50,(50,100,100)	(0,0.5,0.9)	6.03	6.29	4.97
3	50,(100,100,100)	(0,0.5,0.9)	5.74	6.26	4.81
4	100,(14,14,13,13)	(0,0.3,0.6,0.9)	6.80	6.77	5.14
4	200,(26,26,26,26)	(0,0.3,0.6,0.9)	6.42	6.37	5.04
4	100,(12,13,25,25)	(0,0.3,0.6,0.9)	6.68	5.66	5.04
4	200,(25,25,50,50)	(0,0.3,0.6,0.9)	6.11	6.40	5.04
4	25,(50,50,100,100)	(0,0.3,0.6,0.9)	6.20	6.15	4.99
4	12,(100,100,100,100)	(0,0.3,0.6,0.9)	6.71	7.14	4.80
4	50,(100,100,100,100)	(0,0.3,0.6,0.9)	6.17	6.03	4.89
6	100,(10,10,9,9,9,9)	(-0.01,0,0.2,0.4,0.6,0.8)	7.02	6.08	4.78

Table 2.5: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{Q}$, $\boldsymbol{Q}$ is defined from the QR decomposition of the random matrix $\boldsymbol{Z} = (Z_{jj'})$ and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal entries are drawn from $\mathcal{U}nif(0,1)$.

		$\Pr(\widehat{T}_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$			$\Pr(\widehat{T}_B > z_\alpha)$ under $\mathcal{H}_0^{B}$		
a	$b, \boldsymbol{n}'$	$Z_{jj'}$			$Z_{jj'}$		
		$1_{\{j=j'\}}$	$\mathcal{E}xp(1)$	$\mathcal{N}(0,1)$	$1_{\{j=j'\}}$	$\mathcal{E}xp(1)$	$\mathcal{N}(0,1)$
2	100,(26,26)	6.06	6.22	6.59	5.96	6.21	6.57
2	200,(51,51)	5.70	5.60	5.66	5.95	5.74	5.72
2	400,(51,51)	5.71	5.60	5.59	5.50	5.61	5.89
2	400,(101,101)	5.09	5.80	5.77	5.58	5.45	5.47
2	100,(12,25)	6.85	6.44	6.45	6.31	6.26	6.38
2	200,(25,50)	5.55	5.91	5.70	5.87	5.64	5.87
2	25,(50,100)	6.65	6.53	6.77	6.75	6.47	7.14
3	100,(12,13,25)	6.83	6.43	6.66	6.25	6.19	6.12
3	200,(25,25,50)	5.69	5.71	5.61	5.53	5.43	5.50
3	25,(50,100,100)	6.53	6.16	6.36	6.57	6.21	6.67
3	50,(50,100,100)	5.62	5.90	6.23	6.53	6.15	5.86
4	100,(14,14,13,13)	6.27	6.67	6.45	5.95	6.43	6.19
4	200,(26,26,26,26)	5.20	5.92	5.79	5.72	6.02	5.69
4	100,(12,13,25,25)	5.84	6.40	6.28	5.99	5.78	6.33
4	200,(25,25,50,50)	5.73	5.52	5.65	5.64	6.15	5.95
4	25,(50,50,100,100)	6.23	6.43	5.87	6.25	6.24	6.60
4	12,(100,100,100,100)	6.07	6.14	6.58	7.17	7.12	6.75
4	50,(100,100,100,100)	5.52	6.17	6.31	6.16	6.60	6.08
6	100,(10,10,9,9,9,9)	6.30	6.38	6.24	6.43	5.61	6.11

Table 2.6: Achieved Type I error rate ($\times 100\%$) for testing interaction effect AB and main effect B when sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i^\top \boldsymbol{\Lambda}_i \boldsymbol{Q}_i$, $\boldsymbol{Q}_i$ is defined from the QR decomposition of the random matrix $\boldsymbol{Z}_i = (Z_{i,jj'})$ and $\boldsymbol{\Lambda}_i$ is a diagonal matrix whose diagonal entries are drawn from $\mathcal{U}nif(0,1)$.

		$\Pr(\widehat{T}'_{AB} > z_\alpha)$ under $\mathcal{H}_0^{AB}$			$\Pr(\widehat{T}'_{B} > z_\alpha)$ under $\mathcal{H}_0^{B}$			$\Pr(\widehat{T}'_{A} > f_\alpha)$ under $\mathcal{H}_0^{B}$		
a	$b, \boldsymbol{n}'$	$Z_{jj'}$			$Z_{jj'}$			$Z_{jj'}$		
		$1_{\{j=j'\}}$	$\mathcal{E}xp(1)$	$\mathcal{N}(0,1)$	$1_{\{j=j'\}}$	$\mathcal{E}xp(1)$	$\mathcal{N}(0,1)$	$1_{\{j=j'\}}$	$\mathcal{E}xp(1)$	$\mathcal{N}(0,1)$
2	100,(26,26)	6.43	6.07	6.00	6.15	6.23	6.07	4.91	4.75	4.97
2	200,(51,51)	5.68	6.79	5.71	6.02	6.09	5.85	4.57	5.04	5.03
2	400,(51,51)	5.67	5.40	5.54	5.20	5.54	5.73	4.89	5.06	5.02
2	400,(101,101)	5.69	5.39	5.33	5.26	5.73	5.73	4.97	4.81	5.15
2	100,(12,25)	6.50	6.59	6.12	6.38	6.49	6.52	4.55	5.02	5.46
2	200,(25,50)	5.66	5.68	5.74	6.11	6.17	6.05	4.78	4.77	4.57
2	25,(50,100)	6.62	6.23	6.42	6.74	6.47	6.62	4.91	5.20	4.61
3	100,(12,13,25)	6.58	7.10	6.26	6.42	6.15	6.11	4.58	4.98	4.72
3	200,(25,25,50)	5.84	5.96	6.01	6.05	5.79	5.91	4.82	4.71	4.74
3	25,(50,100,100)	6.35	6.51	6.52	6.44	6.24	6.46	5.09	5.12	4.97
3	50,(50,100,100)	6.11	6.26	6.13	6.28	6.25	5.97	4.53	5.19	5.08
4	100,(14,14,13,13)	6.43	6.46	6.22	6.64	6.26	6.04	5.07	4.71	4.93
4	200,(26,26,26,26)	5.97	5.51	5.95	5.55	5.55	5.54	4.96	4.75	4.96
4	100,(12,13,25,25)	6.51	5.80	6.16	6.37	5.94	6.02	5.34	5.11	4.80
4	200,(25,25,50,50)	5.71	6.06	5.78	5.77	5.29	5.47	4.98	4.99	5.44
4	25,(50,50,100,100)	6.04	6.15	6.15	6.23	6.11	6.36	5.28	4.89	5.21
4	12,(100,100,100,100)	6.35	6.59	6.74	6.58	6.61	6.79	5.43	4.62	5.12
4	50,(100,100,100,100)	5.83	5.65	5.67	6.41	6.29	5.99	5.01	5.11	5.14
6	100,(10,10,9,9,9,9)	6.49	6.53	6.67	5.93	5.79	6.01	4.92	4.73	5.08

For the covariance structure used in Table 2.1 the assumptions are satisfied uniformly in ρ because $c_i = (1-\rho)^i(1-1/b)$. Furthermore, as $\rho \to 1$, the sampling variability of T_{AB} goes to 0 which could lead to better control of size. On the other hand, for the covariance structure in Table 2,

$$c'_j = \{\sum_{i=1}^{a}(1-\rho_i)/n_i\}^j(\sum_{i=1}^{a}1/n_i)^{-j}(1-1/b) \quad \text{for } j = 1, 2, 3$$

$$c''_3 = \left(\sum_{i=1}^{a}1/n_i\right)^{-3}\Big\{\left(\frac{a-1}{a}\right)^3\sum_{i=1}^{a}\frac{(1-\rho_i)^3}{n_i^3} + \frac{6(a-1)}{a^3}\sum_{i<j}\frac{(1-\rho_i)^2(1-\rho_j)}{n_i^2 n_j} - \frac{36}{a^3}\sum_{i<j<k}\frac{(1-\rho_i)(1-\rho_j)(1-\rho_k)}{n_i n_j n_k}\Big\}(1-1/b)$$

which are all $O(1)$. What is more, $c'_2 \to 0$ as $\rho_i \to 1$. So does the variance of T'_{AB} and T'_B. On the contrary, our numerical calculations of the ratio of the traces to dimensions for the covariances used in the Table 2.3 show that c_3 and c_4 diverge as $b \to \infty$. For example, we obtained the sequence $(c_3, c_4) = (.333, .310), (0.649, 0.942), (1.295, 3.045), (2.729, 10.700), (6.001, 39.616), (13.589, 151.361)$ for $b = 12, 25, 50, 100, 200, 400$, respectively. Similar pattern should exist for the covariances in Table 2.4 as well. Regardless, it is reassuring to see that the effect of the divergence is negligible on the quality of approximation of our results. As a result of this one may conjecture to drop these assumptions. There is also another supporting evidence of this phenomena in Tables 2.5 and 2.6 where random covariance matrices is used. So this again shows that the choice of the covariance matrix seems to have no effect.

2.5.2 Power Comparison

An approximate yet popular method for repeated measures analysis when the covariances are unequal and unstructured is Huynh (1978), later corrected by Lecoutre (1991). This method performs well when n_i are all large. In this section, we compare the power of this method with the methods proposed in this Chapter. For both methods, the respective approximate null distributions are used to determine the critical values.

To keep the comparison manageable, we fix $a = 3$, $b \in \{10, 100, 200\}$ and $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$ where $\rho_1 = 0.2$, $\rho_2 = 0.5$ and $\rho_3 = 0.9$. In regards to sample sizes and dimension, we use the combinations $(b; n_1, n_2, n_3) = (10; 50, 100, 100), (100; 25, 12, 13)$ and $(200; 50, 25, 25)$. For the alternative hypotheses, we take $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$ and consider two structures for $\boldsymbol{\mu}_1$, namely $\boldsymbol{\mu}_{1i} = (1+\delta)$ for i odd, $\boldsymbol{\mu}_{1i} = (1-\delta)$ for i even, and $\boldsymbol{\mu}_1 = (1+\delta, \mathbf{1}_{b-1}^\top)^\top$ as δ varies from 0 to 1. It should be noted that the later structure represents a departure that approaches to the null hypothesis at a rate of $b^{-1/2}$. More precisely, the scaled departure from the null $||\boldsymbol{\mu}_1 - \mathbf{1}_b||/(\text{tr}(\boldsymbol{\Sigma}))^{1/2}$ are $|\delta|$ and $|\delta|/\sqrt{b}$, respectively.

Figure 2.2: Power comparison of the proposed methods and the methods by Huynh (1978), later corrected by Lecoutre (1991). Data is generated from $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. In the plots, $a = 3$, $\boldsymbol{\Sigma}_i = (1-\rho_i)\boldsymbol{I}_b + \rho_i \boldsymbol{J}_b$, $\rho_1 = 0.2$, $\rho_2 = 0.5$ and $\rho_3 = 0.9$ are used. In both panels, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$. In the left panel $\boldsymbol{\mu}_{1i} = (1+\delta)$ for i odd, $\boldsymbol{\mu}_{1i} = (1-\delta)$ for i even, and in the right panel $\boldsymbol{\mu}_1 = (1+\delta, \mathbf{1}_{b-1}^\top)^\top$.

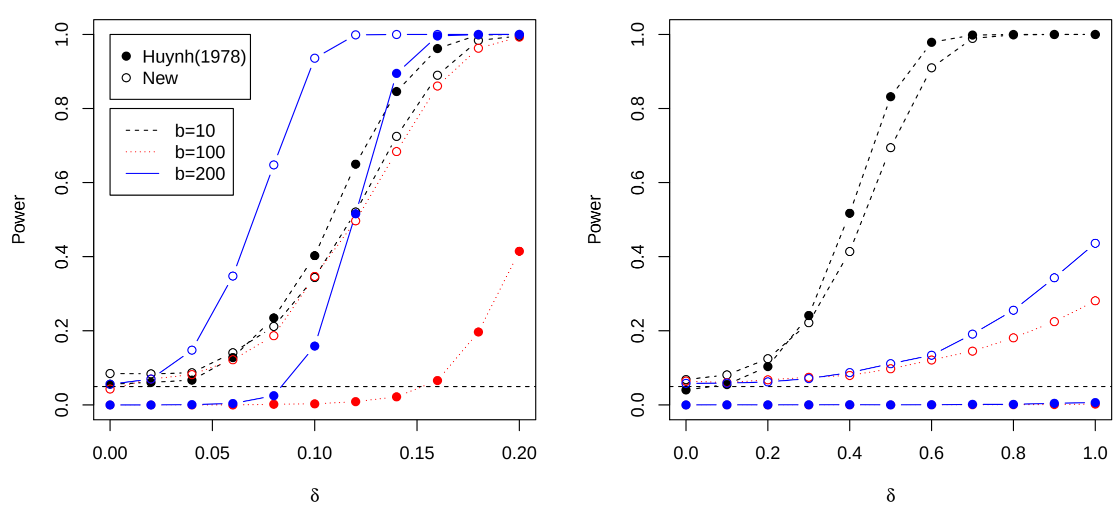
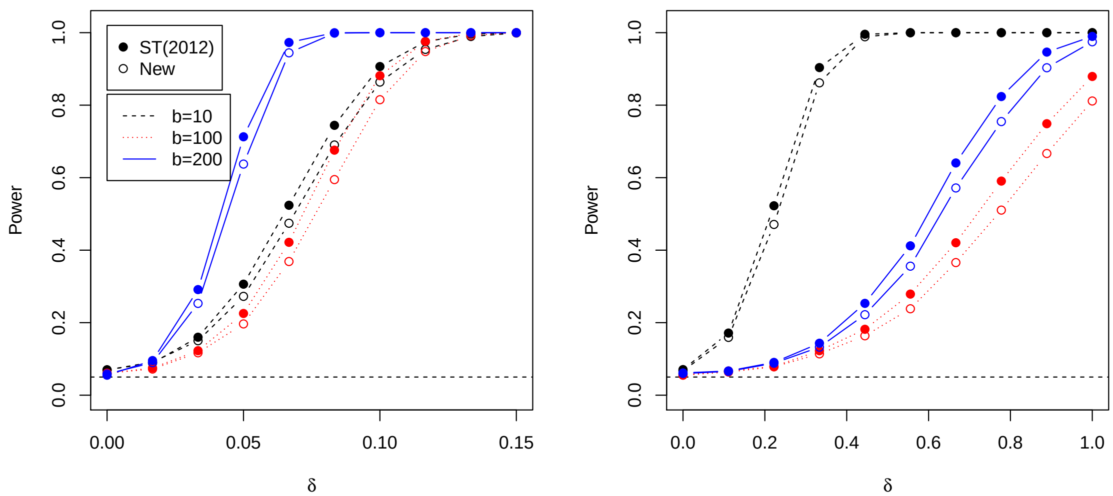


Figure 2.2 shows results for testing the interaction effect. The results for the main effect of B are similar. The power curves clearly demonstrates that as the dimension b gets large, the methods proposed in this Chapter show unequivocal superiority over those of Huynh (1978). The superiority clearly holds when the departure from the null is very mild. However, for small b and large sample sizes (n_1, n_2, n_3), Huynh (1978) methods have clear edge which fade away as b gets larger. It is interesting to observe

Figure 2.3: Power comparison of the test for main effect of factor B (Test for Flatness) in Section 2.3 and the test by Takahashi and Shutoh (2016) for $a = 2$. Data is generated from $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. In the plots, $\boldsymbol{\Sigma} = 0.8\boldsymbol{I}_b + 0.2\boldsymbol{J}_b$ is used. In the left panel, $\boldsymbol{\mu}_1 = \delta \mathbf{1}_{b/2} \otimes (1, -1)^\top$ and in the right panel $\boldsymbol{\mu}_1 = \delta(1, 0, \ldots, 0)^\top$. In both panels, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_b$.

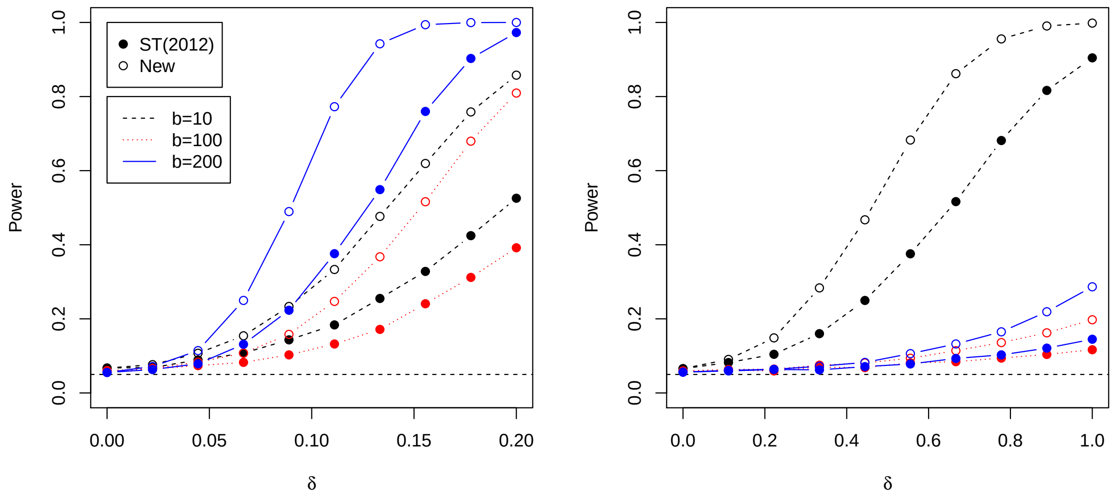


that Hynuh's (Huynh, 1978) methods are very conservative when the dimension is large. Also the power of our method tends to get bigger as b gets larger even when the scaled departure from the null does not change with b.

As discussed in Section 2.2, when $a = 2$ and the covariance matrices are equal, the tests for interaction effect and main effects of A in the current manuscript and those in Takahashi and Shutoh (2016) agree. However, the tests for main effects of B are different. In the remainder of this section, we present simulation results to compare the powers of these two test.

First, we consider the case where the parameter space is constrained by parallelism hypothesis. We look at two forms of departure from the null hypothesis (flatness), namely $\boldsymbol{\mu}_1 = \delta \mathbf{1}_{b/2} \otimes (1, -1)^\top$ and $\boldsymbol{\mu}_1 = \delta(1, 0, \ldots, 0)^\top$ for values of δ between 0 and 1. In both cases, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_b$. For dimension and sample size combinations, we use $(b; n_1, n_2) = (10; 50, 100), (100; 12, 25)$ and $(200; 25, 50)$. For simplicity, we keep the covariance matrix $\boldsymbol{\Sigma} = 0.8\boldsymbol{I}_p + 0.2\boldsymbol{J}_p$. From the power curves in Figure 2.3, the test of Takahashi and Shutoh (2016) seems to have a slight edge over ours.

Figure 2.4: Power comparison of the test for main effect of factor B (Test for Flatness) in Section 2.3 and the test by Takahashi and Shutoh (2016) for $a = 2$. Data is generated from $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. In the plots, $\boldsymbol{\Sigma} = 0.8\boldsymbol{I}_b + 0.2\boldsymbol{J}_b$ is used. In the left panel, $\boldsymbol{\mu}_1 = \delta \mathbf{1}_{b/2} \otimes (1, -1)^\top + \mathbf{1}_b$ and in the right panel $\boldsymbol{\mu}_1 = \delta(1, 0, \ldots, 0)^\top + \mathbf{1}_b$. In both panels, $\boldsymbol{\mu}_2 = \mathbf{0}$.



Another common departure from flatness in two group clinical trials is when the control group maintains flat mean profile but the treatment group may have fluctuating mean profile. Under this alternative, parallelism is obviously violated but the researcher may be interested in testing average flatness versus non flatness. One can make the argument that the test statistic for flatness presented in this Chapter and that of Takahashi and Shutoh (2016) can be used to detect lack of average flatness when parallelism is not known a priori. To investigate the powers of our test and that of Takahashi and Shutoh (2016) for this type of alternative, we consider two forms of departure from flatness, namely $\boldsymbol{\mu}_1 = \delta \mathbf{1}_{b/2} \otimes (1, -1)^\top + \mathbf{1}_b$ and $\boldsymbol{\mu}_1 = \delta(1, 0, \ldots, 0)^\top + \mathbf{1}_b$ for values of δ between 0 and 1. In both cases, $\boldsymbol{\mu}_2 = \mathbf{0}$. We keep the the other parameters of the simulation parameters the same as in the previous paragraph. From Figure 2.4, our test has clear edge over that of Takahashi and Shutoh (2016).

2.6 Real Data Analysis

We analyze a publicly available data obtained from the University of California-Irvine Machine Learning Repository.[1] The data arose from a large study to examine Electroencephalograph (EEG) correlates of genetic predisposition to alcoholism. Measurements from 64 electrodes placed on subject's scalps recorded 256 times for 1 second. The study involved two groups of subjects: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from a picture set. In this section, we analyze the data only for the single stimulus (S1) exposure. The outcome measurements are Event-Related Potentials (ERP) indicating the level of electrical activity (in volts) in the region of the brain where each of the electrodes is placed.

We analyze the data from each electrode (location of the brain) separately and adjust the resulting p-values for multiplicity so that False Discovery Rate (FDR), the expected proportion of false rejections, is controlled (Benjamini and Hochberg, 1995). In the notations of this Chapter, this data set has $a = 2$, $b = 256$, $n_1 = 77$, $n_2 = 45$ and $n = 122$. Factor A is the group factor and factor B (the within-subject factor) is the time.

Indeed, the main hypotheses of interest are whether ERP profiles are similar between the alcoholic and control groups. If different, to identify for which electrode (which part of the brain) dissimilarity occurs. In other words, interest lies in knowing at which electrodes do time and alcoholism interact on ERP outcome. Table 2.7 shows FDR adjusted p-values for testing group-by-time interaction for each of the 64 channels (at FDR = 0.05). The columns in the table contain channel names (Ch), p-values based on analysis assuming equal covariance (E) and p-values without assuming equal covariance (U).

Each channel (electrode) has names identifying the location of the electrode on the scalp. The names are made up of a letter identifying the anatomical location of the placement of the electrode (F-frontal lobe, T-temporal lobe, P-parietal lobe and

[1] web address: https://archive.ics.uci.edu/ml/datasets/EEG+Database

O-occipital lobe) and a number identifying the hemisphere of the brain (odd number– the left hemisphere and even number – the right hemisphere and letter z (zero) is used for the mid-line) . The name A and Fp identify the earlobe and frontal polar sites, respectively, whereas C identifies the central location between the frontal and parietal lobes). Combinations of two letters indicates intermediate locations, e.g., FC: in between frontal and central electrode locations (see Figure 2.5 for details).

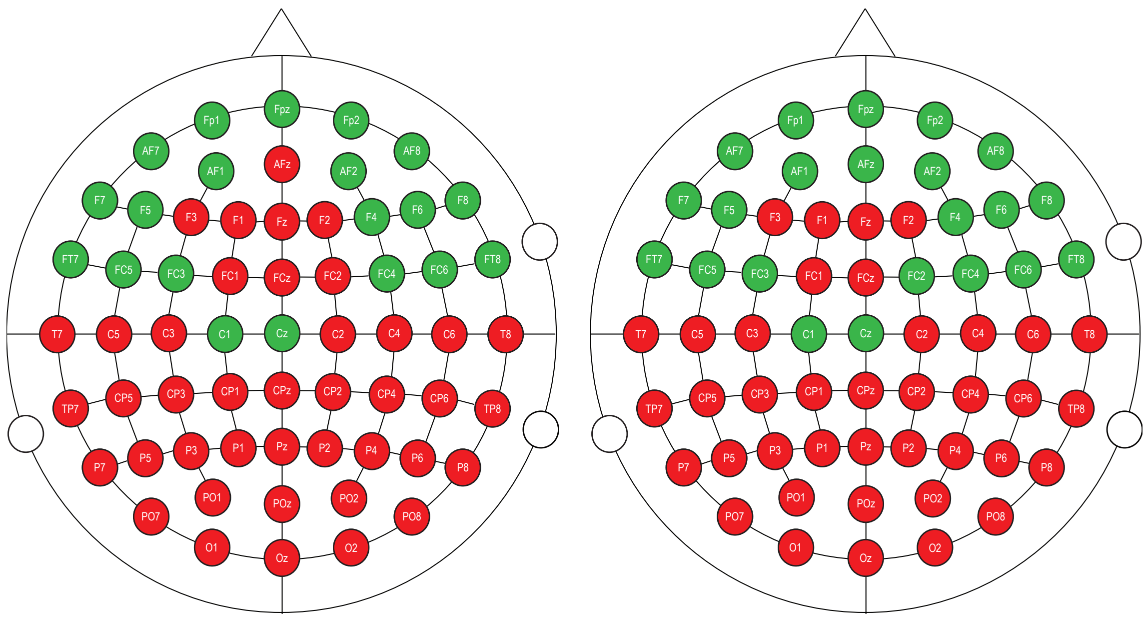
Table 2.7: False Discovery Rate (FDR) adjusted p-values for testing time$\times$group interaction for Electroencephalograph (EEG) experiment involving Alcoholic and Control subjects. In the table, the columns are channel label (Ch), p-value based on equal covariance assumption (E) and p-value based on unequal covariance assumption (U).

Ch	E	U	Ch	E	U	Ch	E	U	Ch	E	U
AF1	0.054	0.086	CP6	0.000	0.000	FC6	0.129	0.236	P5	0.000	0.000
AF2	0.058	0.098	CPZ	0.000	0.000	FCZ	0.000	0.000	P6	0.000	0.000
AF7	0.716	0.737	CZ	0.595	0.641	FP1	0.668	0.701	P7	0.000	0.000
AF8	0.716	0.741	F1	0.000	0.000	FP2	0.716	0.744	P8	0.000	0.000
AFZ	0.007	0.016	F2	0.000	0.001	FPZ	0.716	0.737	PO1	0.000	0.000
C1	0.157	0.147	F3	0.003	0.005	FT7	0.595	0.689	PO2	0.000	0.000
C2	0.013	0.007	F4	0.120	0.212	FT8	0.034	0.095	PO7	0.000	0.000
C3	0.000	0.000	F5	0.356	0.429	FZ	0.000	0.000	PO8	0.000	0.000
C4	0.000	0.000	F6	0.469	0.544	nd	0.001	0.002	POZ	0.000	0.000
C5	0.000	0.000	F7	0.595	0.641	O1	0.000	0.000	PZ	0.000	0.000
C6	0.000	0.000	F8	0.716	0.744	O2	0.000	0.000	T7	0.000	0.000
CP1	0.000	0.000	FC1	0.001	0.002	OZ	0.000	0.000	T8	0.000	0.000
CP2	0.000	0.000	FC2	0.079	0.048	P1	0.000	0.000	TP7	0.000	0.000
CP3	0.000	0.000	FC3	0.286	0.410	P2	0.000	0.000	TP8	0.000	0.000
CP4	0.000	0.000	FC4	0.716	0.769	P3	0.000	0.000	X	0.716	0.744
CP5	0.000	0.000	FC5	0.536	0.640	P4	0.000	0.000	Y	0.046	0.034

The channel-by-channel decisions based on our method and that of Huynh (1978) are displayed in Figure 2.5. The figure depicts the scalp of a human being viewed from the top, the triangle marking the nose. The locations of the electrodes are indicated by bubbles. The color of the bubbles indicates whether the brain activity pattern for that channel is significantly dissimilar (red) or not significantly dissimilar (green).

It is clear from the table that there are no evidence in the data to show difference in the electrical activity patterns between the two groups in the frontal regions of the brain. Most of the significant differences occur in the central, parietal and occipital regions. More precisely, there is similarity in the activity patterns for the alcoholic

Figure 2.5: Channel-by-Channel results for EEG data analysis on testing the similarity in brain activity between alcoholic and control subjects. Left panel contains results based on our method and right panel contains results from the methods by Huynh (1978). Red means brain activity patterns are significantly dissimilar at $\alpha = 0.05$. Green means that the similarity hypothesis cannot be rejected.



(a) New Method (b) Huynh (1978) Method

and control groups in the all of frontal electrodes with the exception of FC1, FCZ, FZ, F1, F2, F3, FC2 and AFZ. On the other hand, all the other data from all the other electrodes in the non-frontal with the exception of C1 and CZ. The results distinctly demarcate contagious similar and non-similar activity regions. In particular, FC1, FCZ, FZ, F1, F2, F3 and AFZ form a isolated region dissimilarity with the frontal region of the brain (see Figure 2.5). Our method has found two more significant channels (AFZ, FC2, and Y) than the method by Huynh (1978).

2.7 Discussion and Conclusion

Tests for repeated measures design are introduced when both sample size and the dimension are large. The tests allow the covariance to be equal or unequal but otherwise unstructured. To the best of our knowledge, tests for high-dimensional repeated measures with unequal group covariance matrices did not exist prior to the current manuscript.

The equal covariance case has seen some recent advancement in the one-group (Pauly et al., 2015) and two-group (Takahashi and Shutoh, 2016) cases. The test statistic in Pauly et al. (2015) essentially differ from ours in the way the asymptotic variance of U_B is estimated. We make efficient use of all observations which resulted in the additional assumptions $\text{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma})^4\}/b = O(1)$. Our results for $a = 2$, however, corroborate with those of Takahashi and Shutoh (2016) for testing the interaction (parallelism hypothesis) and the group effect or effect of factor A (coincidence hypothesis). Nevertheless, our tests for the within subject factor or effect of factor B (flatness hypothesis) differ.

In practical application, one needs to know whether covariances are equal or not to choose which test to use. This choice could be informed by testing equality of covariance matrices. Tests for equality of covariance matrix in high dimensional framework are given, among others, by Schott (2007b) and Srivastava and Yanagihara (2010). Heuristically, one can also conduct the tests in this Chapter with and without the assumption of equality of the covariance matrices. If the decisions agree, the burden of testing equality of variance is removed altogether. Otherwise, caution has to be exercises to decide which results to use. Cursory inspection of the empirical covariance matrices could also be a useful guide in some cases. It should, however, be stressed that the consequence of the assumption of equal variance could be substantial when the group sample sizes differ largely.

The simulation results suggest that the approximations work only when b is very large. This may not be practical in some applications. Second order asymptotic that includes terms of order $b^{-1/2}$, b^{-1} and so forth have the potential to improve the approximations. Also discernible in the simulation is that the quality of the approximation depends on the value of a, the larger the better. Intuitively, this makes sense because large a means more data under the null hypothesis. This prompts consideration of an asymptotic framework that allows a to grow together with the sample sizes and the number of repeated measurements. We defer consideration of these problems to future manuscripts.

In the proofs, multivariate normality of the repeated measures is mostly needed for its nice property of independence up to correlation. Results under statistical models that include this later assumption has been derived in Chen and Qin (2010) for two-sample and Srivastava and Kubokawa (2013) for multiple-sample comparison of mean vectors. In the interest of space, we opted to relegate the investigation of this model in the Chapter 4.

2.8 Appendix: Proofs

Proof of Theorem 2.3.1. Note that

$$U_{AB} = \frac{1}{\sqrt{b}}\Big(\{\mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K})\}^{-1}\overline{\boldsymbol{X}}^\top(\boldsymbol{P}_a \otimes \boldsymbol{P}_b)\overline{\boldsymbol{X}} - \mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma})\Big),$$

where $\boldsymbol{K} = \mathrm{diag}\{1/n_1, \ldots, 1/n_a\}$. Obviously, $\mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K}) = (1-1/a)\sum_{i=1}^{a} 1/n_i$. We can write $\boldsymbol{P}_a = \boldsymbol{L}\boldsymbol{L}^\top$ and $\boldsymbol{P}_b = \boldsymbol{M}\boldsymbol{M}^\top$ where $\boldsymbol{L}_{a\times(a-1)}$ and $\boldsymbol{M}_{b\times(b-1)}$ are full rank matrices with ranks $a-1$ and $b-1$, respectively. Under the null hypothesis $\mathcal{H}_0^{AB}$, $(\boldsymbol{L}^\top \otimes \boldsymbol{M}^\top)\overline{\boldsymbol{X}} \sim \mathcal{N}_{ab}(\boldsymbol{0}, (\boldsymbol{L}^\top\boldsymbol{K}\boldsymbol{L}) \otimes (\boldsymbol{M}^\top\boldsymbol{\Sigma}\boldsymbol{M}))$, whence $\overline{\boldsymbol{X}}^\top(\boldsymbol{P}_a \otimes \boldsymbol{P}_b)\overline{\boldsymbol{X}} = \{(\boldsymbol{L}^\top \otimes \boldsymbol{M}^\top)\overline{\boldsymbol{X}}\}^\top\{(\boldsymbol{L}^\top \otimes \boldsymbol{M}^\top)\overline{\boldsymbol{X}}\}$. Notice that the eigenvalues of $\boldsymbol{M}^\top\boldsymbol{\Sigma}\boldsymbol{M}$ and $\boldsymbol{\Sigma}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}^{1/2}$ are the same. So are that of $\boldsymbol{L}^\top\boldsymbol{K}\boldsymbol{L}$ and $\boldsymbol{K}^{1/2}\boldsymbol{P}_a\boldsymbol{K}^{1/2}$. Therefore, under the null hypothesis $\mathcal{H}_0^{AB}$,

$$U_{AB} = \frac{1}{\sqrt{b}}\left\{\left(1-\frac{1}{a}\right)^{-1}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j Y_{ij} - \mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma})\right\},$$

where α_i's are the eigenvalues of $\boldsymbol{K}^{1/2}\boldsymbol{P}_a\boldsymbol{K}^{1/2}$, β_j's are the eigenvalues of $\boldsymbol{\Sigma}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}^{1/2}$, and Y_{ij}'s are independently and identically distributed as chi-square distribution with 1 degree of freedom for $i = 1, \ldots, a$, and $j = 1, \ldots, b$. So the characteristic function of U_{AB} is

$$\varphi(t) = \mathrm{E}[\exp(\imath\, tU_{AB})] = \prod_{i=1}^{a}\prod_{j=1}^{b}\left(1 - \frac{2\imath\, t\alpha_i\beta_j}{\sqrt{b}\,\mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K})}\right)^{-1/2}\exp\left[-\frac{\imath\, t\,\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma})}{\sqrt{b}}\right],$$

where $\imath = \sqrt{-1}$. Therefore, we have

$$\ln\{\varphi(t)\} = -\frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b}\ln\left(1 - \frac{2\imath\, t\alpha_i\beta_j}{\sqrt{b}\,\mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K})}\right) - \frac{\imath\, t\,\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma})}{\sqrt{b}}$$

$$= \frac{(\imath\, t)^2}{2} \cdot \frac{\mathrm{tr}\{(\boldsymbol{P}_a \boldsymbol{K})^2\}}{\{\mathrm{tr}(\boldsymbol{P}_a \boldsymbol{K})\}^2} \cdot 2c_2 + \mathrm{O}\left(\frac{|t|^3 \mathrm{tr}\{(\boldsymbol{P}_a \boldsymbol{K})^3\}}{\sqrt{b}\, \{\mathrm{tr}(\boldsymbol{P}_a \boldsymbol{K})\}^3} c_3 \right),$$

by applying Taylor expansion to $\ln[1 - (2\imath\, t\alpha_i\beta_j)/\{\sqrt{b}\ \mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K})\}]$. Here

$$\frac{\mathrm{tr}\{(\boldsymbol{P}_a \boldsymbol{K})^2\}}{\{\mathrm{tr}(\boldsymbol{P}_a \boldsymbol{K})\}^2} = \frac{a(a-2)}{(a-1)^2} \sum_{i=1}^{a} \frac{1}{n_i^2} \Bigg/ \left(\sum_{i=1}^{a} \frac{1}{n_i} \right)^2 + \frac{1}{(a-1)^2} = c.$$

Since we know that $\boldsymbol{K}^{1/2}$ is symmetric and positive definite matrix, and $\boldsymbol{P}_a$ is positive semidefinite and symmetric, so does $\boldsymbol{K}^{1/2}\boldsymbol{P}_a\boldsymbol{K}^{1/2}$. Thus $\alpha_i \geq 0$, and at least one α_i bigger than 0. For any $k \geq 2$, we have

$$\frac{\mathrm{tr}\{(\boldsymbol{P}_a \boldsymbol{K})^k\}}{\{\mathrm{tr}(\boldsymbol{P}_a \boldsymbol{K})\}^k} = \frac{\sum_{i=1}^a \alpha_i^k}{(\sum_{i=1}^a \alpha_i)^k} \leq 1.$$

Moreover, by using Hölder's inequality we have

$$\sum_{i=1}^{a} \alpha_i \leq \left(\sum_{i=1}^{a} \alpha_i^k \right)^{1/k} \cdot \left(\sum_{i=1}^{a} 1^{k/(k-1)} \right)^{(k-1)/k} = \left(a^{(k-1)} \sum_{i=1}^{a} \alpha_i^k \right)^{1/k}.$$

This proves that the ratio $\mathrm{tr}\{(\boldsymbol{P}_a\boldsymbol{K})^k\}/\{\mathrm{tr}(\boldsymbol{P}_a\boldsymbol{K})\}^k \geq 1/a^{k-1}$. So under the assumption $A1$ and $A2$, the remainder term of the Taylor series converges to 0 as b goes to infinity. Also by Hölder's inequality, $(\sum_{i=1}^a 1/n_i^2)/(\sum_{i=1}^a 1/n_i)^2 \geq 1/a$ which establishes the lower bound for c. □

Proof of Theorem 2.3.2. Although a much easier proof will be given in Chapter 3 (Kong and Harrar, 2017), we still include this one since it gives more idea on how to construct the unbiased estimators. It is obvious that $(n_i-1)\boldsymbol{S}_i \sim \mathcal{W}_b(\boldsymbol{\Sigma}, n_i-1)$, where $\mathcal{W}_p(\boldsymbol{\Omega}, \nu)$ stands for p dimensional Wishart distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Omega}$. Thus $(n-a)\boldsymbol{S}$, the sum of a independent $W_b(\boldsymbol{\Sigma}, n_i - 1)$, has distribution $\mathcal{W}_b(\boldsymbol{\Sigma}, n-a)$. The reminder of the proof is similar to that of Lemma 2.1 of Srivastava (2005). We give sketchy detail below. (See also the Proof of Theorem 2.4.2.)

Let $\boldsymbol{\Gamma}$ be the orthogonal matrix such that

$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}^\top = \boldsymbol{\Lambda}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1 \ldots, \lambda_b)$ and λ_i's are the eigenvalues of $\boldsymbol{\Sigma}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}^{1/2}$. Also let $\boldsymbol{w}_i$ be iid $\mathcal{N}_{n-a}(\boldsymbol{0}, \boldsymbol{I}_{n-a})$ for $i = 1, \ldots, b$. Then

$$(n-a)\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}) = \sum_{i=1}^{b} \lambda_i v_{ii},$$

$$(n-a)^2\{\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S})\}^2 = \sum_{i=1}^{b} \lambda_i^2 v_{ii}^2 + \sum_{i\neq j} \lambda_i\lambda_j v_{ii}v_{jj} \quad \text{and}$$

$$(n-a)^2\mathrm{tr}\{(\boldsymbol{P}_b\boldsymbol{S})^2\} = \sum_{i=1}^{b} \lambda_i^2 v_{ii}^2 + \sum_{i\neq j} \lambda_i\lambda_j v_{ij}^2,$$

where $v_{ii} = \boldsymbol{w}_i^\top\boldsymbol{w}_i$ and $v_{ij} = \boldsymbol{w}_i^\top\boldsymbol{w}_j = \boldsymbol{w}_j^\top\boldsymbol{w}_i$ for $i \neq j$. Thus,

$$\widehat{c}_1 = \frac{1}{b(n-a)} \sum_{i=1}^{b} \lambda_i v_{ii} \quad \text{and}$$

$$\begin{aligned}
\widehat{c}_2 =& \frac{1}{b(n-a-1)(n-a+2)} \left\{ \sum_{i=1}^{b} \lambda_i^2 v_{ii}^2 \right. \\
&\left. + \sum_{i\neq j} \lambda_i\lambda_j v_{ij}^2 - \frac{1}{(n-a)} \left(\sum_{i=1}^{b} \lambda_i^2 v_{ii}^2 + \sum_{i\neq j} \lambda_i\lambda_j v_{ii}v_{jj} \right) \right\} \\
=& \frac{1}{b(n-a)(n-a+2)} \sum_{i=1}^{b} \lambda_i^2 v_{ii}^2 \\
&+ \frac{1}{b(n-a-1)(n-a+2)} \sum_{i\neq j} \lambda_i\lambda_j \left(v_{ij}^2 - \frac{1}{n-a} v_{ii}v_{jj} \right).
\end{aligned}$$

Finally, we have $\mathrm{E}[\widehat{c}_i] = c_i$, for $i = 1, 2$. Furthermore, $\mathrm{Var}(\widehat{c}_1) = 2c_2/\{b(n-a)\}$ and

$$\begin{aligned}
\mathrm{Var}(\widehat{c}_2) =& \frac{4c_2^2}{(n-a)^2} - \frac{4c_2^2}{(n-a)^3} + \frac{12c_2^2}{(n-a)^4} - \frac{20c_2^2}{(n-a)^5} + \frac{8c_4}{b(n-a)} + \\
& \frac{4c_4}{b(n-a)^2} - \frac{12c_4}{b(n-a)^3} + \frac{20c_4}{b(n-a)^4} + \mathrm{O}(n^{-6})c_2^2 + \mathrm{O}(b^{-1})\mathrm{O}(n^{-5})c_4,
\end{aligned}$$

under the high-dimensional asymptotic frameworks $A1$ and $A2$. Applying Chebyshev's inequality completes the proof. □

Proof of Corollary 2.3.3. From Theorem 2.3.2, we know that $\widehat{c}_2/c_2 \xrightarrow{P} 1$ and $\sqrt{b}(\widehat{c}_1 - c) \xrightarrow{P} 0$. Then the desired result follows by applying CMT and Slutsky's Theorem to

$$\widehat{T}_{AB} = \frac{1}{\sqrt{2bcc_2(\widehat{c}_2/c_2)}} \left(\left(1 - \frac{1}{a}\right)^{-1} \left(\sum_{i=1}^{a} \frac{1}{n_i} \right)^{-1} H^{(AB)} - bc_1 \right) - \frac{\sqrt{b}}{\sqrt{2cc_2(\widehat{c}_2/c_2)}} (\widehat{c}_1 - c_1).$$

$\square$

Proof of Theorem 2.3.4. Obviously, $\mathrm{tr}(\boldsymbol{J}_a\boldsymbol{K}) = \sum_{i=1}^{a} 1/n_i$. The remainder of the proof is analogous to that of Theorem 2.3.1. $\square$

Proof of Proposition 2.3.6. Obviously, $\boldsymbol{K}^{1/2} \otimes \boldsymbol{I}_b$ is nonsingular. Thus $b\boldsymbol{K}_A\boldsymbol{\mu} = (\boldsymbol{D}_a \otimes \boldsymbol{J}_b)\,\boldsymbol{\mu} = \boldsymbol{0}$ if and only if

$$\left(\boldsymbol{K}^{1/2} \otimes \boldsymbol{I}_b\right)(\boldsymbol{D}_a \otimes \boldsymbol{J}_b)\,\boldsymbol{\mu} = \left(\boldsymbol{K}^{1/2}\boldsymbol{D}_a \otimes \boldsymbol{J}_b\right)\boldsymbol{\mu} = \boldsymbol{0},$$

if and only if

$$b^{-1}\boldsymbol{\mu}^\top \left(\boldsymbol{K}^{1/2}\boldsymbol{D}_a \otimes \boldsymbol{J}_b\right)^\top \left(\boldsymbol{K}^{1/2}\boldsymbol{D}_a \otimes \boldsymbol{J}_b\right)\boldsymbol{\mu} = \boldsymbol{\mu}^\top (\boldsymbol{D}_a\boldsymbol{K}\boldsymbol{D}_a \otimes \boldsymbol{J}_b)\,\boldsymbol{\mu} = b\boldsymbol{\mu}^\top\boldsymbol{K}_A\boldsymbol{\mu} = 0,$$

since $\boldsymbol{D}_a = \boldsymbol{D}_a\boldsymbol{K}\boldsymbol{D}_a$. $\square$

Proof of Theorem 2.4.1. It is easy to see, under the null hypothesis $\mathcal{H}_0^{AB}$

$$U'_{AB} = \frac{1}{\sqrt{b}}\left\{\left(1-\frac{1}{a}\right)^{-1}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1}\sum_{j=1}^{ab}\alpha_j Y_j - \left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-1}\mathrm{tr}(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})\right\},$$

where α_j's are the eigenvalues of $\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}_{AB}\widetilde{\boldsymbol{\Sigma}}^{1/2}$, and Y_j's are independently and identically distributed as chi-square distribution with 1 degree of freedom for $j = 1, \ldots, ab$. So the characteristic function of U'_{AB} is

$$\varphi(t) = \mathrm{E}[\exp(\imath\, U'_{AB})] = \prod_{j=1}^{ab}\left(1 - \frac{2\imath\, t\alpha_i}{\sqrt{b}\ (1-1/a)\sum\limits_{i=1}^{a}\frac{1}{n_i}}\right)^{-1/2}\exp\left[-\frac{\imath\ t\ \mathrm{tr}(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})}{\sqrt{b}\ \sum\limits_{i=1}^{a}\frac{1}{n_i}}\right].$$

Therefore, we have

$$\begin{aligned}\ln\{\varphi(t)\} &= -\frac{1}{2}\sum_{j=1}^{ab}\ln\left(1 - \frac{2\imath\ t\alpha_i}{\sqrt{b}\ (1-1/a)\sum\limits_{i=1}^{a}\frac{1}{n_i}}\right) - \frac{\imath\ t\ \mathrm{tr}(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})}{\sqrt{b}\ \sum\limits_{i=1}^{a}\frac{1}{n_i}}\\ &= \frac{(\imath\ t)^2}{2}\cdot 2c'c'_2 + \mathrm{O}\left(\frac{|t|^3 c''_3}{\sqrt{b}\ (1-1/a)^3}\right)\end{aligned}$$

by applying Taylor expansion to $\ln[1 - 2\imath\, t\alpha_i/\{\sqrt{b}\ (1-1/a)\sum_{i=1}^{a} 1/n_i\}]$. Here note that

$$\begin{aligned}
&(1-1/a)^{-2}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-2}\sum_{i=1}^{ab}\frac{\alpha_i^2}{b}\\
=&(1-1/a)^{-2}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-2}\operatorname{tr}\left\{(\boldsymbol{K}_{AB}\widetilde{\boldsymbol{\Sigma}})^2\right\}\\
=&(1-1/a)^{-2}\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-2}\left[\left(1-\frac{2}{a}\right)\sum_{i=1}^{a}\operatorname{tr}\left\{\left(\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right)^2\right\}+\frac{1}{a^2}\operatorname{tr}(\boldsymbol{K}_B\widetilde{\boldsymbol{\Sigma}})^2\right]\\
=&\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^{-2}\frac{a(a-2)}{(a-1)^2}\sum_{i=1}^{a}\operatorname{tr}\left\{\left(\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right)^2\right\}+\frac{c_1'}{(a-1)^2}=c'c_2',
\end{aligned}$$

where $c_2' = O(1)$ by assumption $A1'$. Next we will determine bounds for c'. We know that

$$\operatorname{tr}\left\{\left(\sum_{i=1}^{a}\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right)^2\right\}=\sum_{i=1}^{a}\frac{\operatorname{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^2\}}{n_i^2}+\operatorname{tr}\left(\sum_{i\neq j}\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_j}{n_in_j}\right).$$

Thus, we have

$$\sum_{i=1}^{a}\frac{\operatorname{tr}\{(\boldsymbol{P}_b\boldsymbol{\Sigma}_i)^2\}}{n_i^2}\Bigg/\operatorname{tr}\left\{\left(\sum_{i=1}^{a}\frac{\boldsymbol{P}_b\boldsymbol{\Sigma}_i}{n_i}\right)^2\right\}\in(0,1]$$

which implies $c' \in [(a-1)^{-2}, 1]$. Note that the lower end of the interval is close because the first term in c' will be 0 when $a = 2$. Otherwise, the interval is open on the left side for $a \geq 3$. Finally, under assumption $A4'$, the remainder term of the Taylor series converges to 0 as b goes to infinity. □

Proof of Theorem 2.4.2. In light of remark (ii) in Section 2.3, we see from Theorem 2.3.2 that $\widehat{c_{1i}}$ and $\widehat{c_{2i}}$ are unbiased and consistent estimator of c_{1i} and c_{2i}, respectively, by taking $a = 1$ (see Srivastava, 2005). Also we have $\sqrt{b}(\widehat{c_{1i}} - c_{1i}) \xrightarrow{P} 0$. Thus, under the high-dimensional asymptotic frameworks $A1'$, $A3'$ and $A5'$, $\widehat{c}_1'$ is an unbiased and consistent estimator of c_1' and $\sqrt{b}(\widehat{c}_1 - c_1') \xrightarrow{P} 0$ by Slutsky's Theorem.

Next we will prove that $\widehat{c_{2ii'}}$ is an unbiased and consistent estimator of $c_{2ii'}$. It is obvious that $(n_i - 1)\boldsymbol{S}_i \sim \mathcal{W}_b(\boldsymbol{\Sigma}_i, n_i - 1)$. Thus there exists $\boldsymbol{Y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i-1})$, where $\boldsymbol{y}_{ij}$ are iid $\mathcal{N}_b(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, such that $(n_i - 1)\boldsymbol{S}_i = \boldsymbol{Y}_i\boldsymbol{Y}_i^\top$. Then $\boldsymbol{Y}_i = \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{U}_i$, where

$\boldsymbol{U}_i = (\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{in})$, and $\boldsymbol{u}_{ij}$ are iid $\mathcal{N}_b(\mathbf{0}, \boldsymbol{I}_b)$. Let $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Gamma}_{i'}$ be two orthogonal matrices such that $\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2} = \boldsymbol{\Gamma}_i^\top \boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_{i'}$ where $\boldsymbol{\Lambda}_{ii'} = \mathrm{diag}(\lambda_{1ii'}, \ldots, \lambda_{bii'})$ and $\lambda_{jii'}$, for $j = 1, \ldots, b$ are the square root of eigenvalues of

$$\left(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2}\right)\left(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2}\right)^\top = \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2}.$$

Obviously, we have

$$\boldsymbol{\Sigma}_{i'}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2} = \left(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2}\right)^\top = \left(\boldsymbol{\Gamma}_i^\top\boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_{i'}\right)^\top = \boldsymbol{\Gamma}_{i'}^\top\boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_i.$$

Thus,

$$\begin{aligned}
&(n_i - 1)(n_{i'} - 1)\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}_i\boldsymbol{P}_b\boldsymbol{S}_{i'}) = \mathrm{tr}\left(\boldsymbol{P}_b\boldsymbol{Y}_i\boldsymbol{Y}_i^\top\boldsymbol{P}_b\boldsymbol{Y}_{i'}\boldsymbol{Y}_{i'}^\top\right)\\
=&\mathrm{tr}\left(\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2}\boldsymbol{U}_{i'}\boldsymbol{U}_{i'}^\top\boldsymbol{\Sigma}_{i'}^{1/2}\right)\\
=&\mathrm{tr}\left(\boldsymbol{U}_i^\top\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'}^{1/2}\boldsymbol{U}_{i'}\boldsymbol{U}_{i'}^\top\boldsymbol{\Sigma}_{i'}^{1/2}\boldsymbol{P}_b\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{U}_i\right)\\
=&\mathrm{tr}\left(\boldsymbol{U}_i^\top\boldsymbol{\Gamma}_i^\top\boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_{i'}\boldsymbol{U}_{i'}\boldsymbol{U}_{i'}^\top\boldsymbol{\Gamma}_{i'}^\top\boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_i\boldsymbol{U}_i\right)\\
=&\mathrm{tr}\left(\sum_{j=1}^b \lambda_{jii'}\boldsymbol{w}_{ji}\boldsymbol{w}_{ji'}^\top \cdot \sum_{j=1}^b \lambda_{jii'}\boldsymbol{w}_{ji'}\boldsymbol{w}_{ji}^\top\right)\\
=&\mathrm{tr}\left(\sum_{j=1}^b \lambda_{jii'}^2\boldsymbol{w}_{ji}\boldsymbol{w}_{ji'}^\top\boldsymbol{w}_{ji'}\boldsymbol{w}_{ji}^\top + \sum_{j\neq j'}\lambda_{jii'}\lambda_{j'ii'}\boldsymbol{w}_{ji}\boldsymbol{w}_{ji'}^\top\boldsymbol{w}_{j'i'}\boldsymbol{w}_{j'i}^\top\right),
\end{aligned}$$

where $\boldsymbol{U}_i^\top\boldsymbol{\Gamma}_i^\top = (\boldsymbol{w}_{1i}, \ldots, \boldsymbol{w}_{bi})$ and $\boldsymbol{w}_{ji}$ are iid $\mathcal{N}_{n_i-1}(\mathbf{0}, \boldsymbol{I}_{n_i-1})$ and $\boldsymbol{U}_{i'}^\top\boldsymbol{\Gamma}_{i'}^\top = (\boldsymbol{w}_{1i'}, \ldots, \boldsymbol{w}_{bi'})$ and $\boldsymbol{w}_{ji'}$ are iid $\mathcal{N}_{n_{i'}-1}(\mathbf{0}, \boldsymbol{I}_{n_{i'}-1})$. Using cyclic commutativity of the trace operator

$$\begin{aligned}
&(n_i - 1)(n_{i'} - 1)\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}_i\boldsymbol{P}_b\boldsymbol{S}_{i'})\\
=&\sum_{j=1}^b \lambda_{jii'}^2\boldsymbol{w}_{ji}^\top\boldsymbol{w}_{ji}\boldsymbol{w}_{ji'}^\top\boldsymbol{w}_{ji'} + \sum_{j\neq j'}\lambda_{jii'}\lambda_{j'ii'}\boldsymbol{w}_{j'i}^\top\boldsymbol{w}_{ji}\boldsymbol{w}_{ji'}^\top\boldsymbol{w}_{j'i'}\\
=&\sum_{j=1}^b \lambda_{jii'}^2 v_{jj}^{(i)} v_{jj}^{(i')} + \sum_{j\neq j'}\lambda_{jii'}\lambda_{j'ii'} v_{jj'}^{(i)} v_{jj'}^{(i')},
\end{aligned}$$

where $v_{jj}^{(i)} := \boldsymbol{w}_{ji}^\top\boldsymbol{w}_{ji}$ are iid chi-square random variables with $n_i - 1$ degrees of freedom, and $v_{jj'}^{(i)} := \boldsymbol{w}_{ji}^\top\boldsymbol{w}_{j'i} = \boldsymbol{w}_{j'i}^\top\boldsymbol{w}_{ji}$ and $v_{jj'}^{(i')} := \boldsymbol{w}_{ji'}^\top\boldsymbol{w}_{j'i'} = \boldsymbol{w}_{j'i'}^\top\boldsymbol{w}_{ji'}$, for any $j \neq j'$, are independent and each has mean 0. Thus,

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}_i\boldsymbol{P}_b\boldsymbol{S}_{i'})] = \sum_{j=1}^b \lambda_{jii'}^2 = \mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'})$$

and

$$
\begin{aligned}
&\mathrm{E}\left[(\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}_i\boldsymbol{P}_b\boldsymbol{S}_{i'})^2\right]\\
=&\frac{1}{(n_i-1)^2(n_{i'}-1)^2}\mathrm{E}\left[\left(\sum_{j=1}^{b}\lambda_{jii'}^2 v_{jj}^{(i)}v_{jj}^{(i')}+\sum_{j\neq j'}\lambda_{jii'}\lambda_{j'ii'}v_{jj'}^{(i)}v_{jj'}^{(i')}\right)^2\right]\\
=&\frac{1}{(n_i-1)^2(n_{i'}-1)^2}\mathrm{E}\left[\sum_{j=1}^{b}\lambda_{jii'}^4 v_{jj}^{(i)2}v_{jj}^{(i')2}+\sum_{j\neq j'}\lambda_{jii'}^2\lambda_{j'ii'}^2 v_{jj}^{(i)}v_{jj}^{(i')}v_{j'j'}^{(i)}v_{j'j'}^{(i')}\right.\\
&+\sum_{j\neq j'}2\lambda_{jii'}^2\lambda_{j'ii'}^2 v_{jj'}^{(i)2}v_{jj'}^{(i')2}+\sum_{j\neq j'}2\lambda_{jii'}^3\lambda_{j'ii'}v_{jj}^{(i)}v_{jj}^{(i')}v_{jj'}^{(i)}v_{jj'}^{(i')}\\
&\left.+\sum_{j\neq j'}2\lambda_{jii'}\lambda_{j'ii'}^3 v_{j'j'}^{(i)}v_{j'j'}^{(i')}v_{jj'}^{(i)}v_{jj'}^{(i')}\right]\\
=&\frac{(n_i+1)(n_{i'}+1)}{(n_i-1)(n_{i'}-1)}\sum_{j=1}^{b}\lambda_{jii'}^4+\left(1+\frac{2}{(n_i-1)(n_{i'}-1)}\right)\sum_{j\neq j'}\lambda_{jii'}^2\lambda_{j'ii'}^2,
\end{aligned}
$$

since

$$
\mathrm{E}[v_{jj}^{(i)}]=(n_i-1),\quad \mathrm{E}[v_{jj}^{(i)2}]=(n_i-1)(n_i+1),\quad \mathrm{E}[v_{jj'}^{(i)2}]=(n_i-1)
$$

and the last two terms have mean zero. Note that we omitted the terms with three and four different j's in our calculations because the means for these terms are zeros. Thus,

$$
\begin{aligned}
&\mathrm{Var}\left(\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{S}_i\boldsymbol{P}_b\boldsymbol{S}_{i'})\right)\\
=&\frac{(n_i+1)(n_{i'}+1)}{(n_i-1)(n_{i'}-1)}\sum_{j=1}^{b}\lambda_{jii'}^4+\left(1+\frac{2}{(n_i-1)(n_{i'}-1)}\right)\sum_{j\neq j'}\lambda_{jii'}^2\lambda_{j'ii'}^2-\left(\sum_{j=1}^{b}\lambda_{jii'}^2\right)^2\\
=&\frac{2(n_i+n_{i'})}{(n_i-1)(n_{i'}-1)}\left(\sum_{j=1}^{b}\lambda_{jii'}^2\right)^2+\frac{2(1-n_i-n_{i'})}{(n_i-1)(n_{i'}-1)}\sum_{j\neq j'}\lambda_{jii'}^2\lambda_{j'ii'}^2\\
\leq&\frac{2}{(n_i-1)(n_{i'}-1)}\left\{\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'})\right\}^2.
\end{aligned}
$$

The last inequality follows from

$$
\sum_{j\neq j'}\lambda_{jii'}^2\lambda_{j'ii'}^2\leq\sum_{jj'}\lambda_{jii'}^2\lambda_{j'ii'}^2=\left(\sum_{j=1}^{b}\lambda_{jii'}\right)^2=\left\{\mathrm{tr}(\boldsymbol{P}_b\boldsymbol{\Sigma}_i\boldsymbol{P}_b\boldsymbol{\Sigma}_{i'})\right\}^2.
$$

Finally, we have $\mathrm{E}[\,\widehat{c_{2ii'}}] = c_{2ii'}$, and

$$
\mathrm{Var}(\widehat{c_{2ii'}})\leq\frac{2}{(n_i-1)(n_{i'}-1)}c_{2ii'}^2\leq\frac{2}{(n_i-1)(n_{i'}-1)}c_2'^2=\mathrm{O}(n^{-2}),
$$

under the high-dimensional asymptotic frameworks $A1'$ and $A3'$. This with Chebyshev's inequality, proves that $\widehat{c_{2ii'}}$ is an unbiased and consistent estimator of $c_{2ii'}$. Therefore, $\widehat{c_2'}$ and $\widehat{c'} \cdot \widehat{c_2'}$ are unbiased and consistent estimator of c_2' and $c' \cdot c_2'$ respectively. $\square$

Proof of Theorem 2.4.6. Observe that

$$\widehat{d_{1i}} = \frac{\text{tr}(\boldsymbol{J}_b \boldsymbol{S})}{b^2} = \frac{1}{n_i - 1} \frac{1}{b} \text{tr}(\boldsymbol{J}_b \boldsymbol{\Sigma}_i) \boldsymbol{Y}_i,$$

where $\boldsymbol{Y}_i \sim \chi^2_{n_i - 1}$. Thus,

$$\begin{aligned}
\text{E}(\widehat{d_{1i}}) =& \frac{1}{b^2} \text{tr}(\boldsymbol{J}_b \boldsymbol{\Sigma}_i) = d_{1i} \\
\text{E}[(\widehat{d_{1i}})^2] =& \left(\frac{1}{b^2} \text{tr}(\boldsymbol{J}_b \boldsymbol{\Sigma}_i) \right)^2 \frac{(n_i - 1)(n_i + 1)}{(n_i - 1)^2} = \frac{n_i + 1}{n_i - 1} d_{1i}^2 \\
\text{E}[(\widehat{d_{1i}})^4] =& \frac{(n_i + 1)(n_i + 3)(n_i + 5)}{(n_i - 1)^3} d_{1i}^4 \quad \text{and} \\
\text{Var}[(\widehat{d_{1i}})^2] =& \frac{(n_i + 1)(10 n_i + 14)}{(n_i - 1)^3} d_{1i}^4 = \text{O}(n_i^{-1}) d_{1i}^4.
\end{aligned}$$

$\square$

Chapter 3 Accurate Inference for Repeated Measures in High Dimensions

3.1 Introduction

As we know, repeated measures data arise in various disciplines of the sciences, social sciences, engineering and humanities. Study designs such as time course studies, cross-over designs and split-plot designs naturally lead to repeated measures data. The distinctive feature of repeated measures data is that the observations from the same study unit (observational or experimental) are commensurate and exhibit correlations. Analysis of continuous repeated measures data to make inference on the effects of one or many between- or within-subject crossed or nested factor effects fall into three broad categories: multivariate analysis, univariate analysis and mixed model analyses. Mixed model analyses involve some assumption concerning the correlations of the repeated measures. Despite its generality in modeling the correlation and leading to exact inference, the multivariate method is not applicable when the number of repeated measures is larger than the error degrees of freedom.

Univariate methods on the other hand focus on adjusting the analysis of variance (ANOVA) for the within-unit correlation. It is well know that when all observations are independent, ANOVA test statistics have exact F distribution. In the presence of the within-unit correlation, the ANOVA tests are valid only if these correlations satisfy a condition known as *sphericity* (Bock, 1963; Huynh and Feldt, 1970). Box (1954) suggested a correction which involves adjustment of the numerator and denominator degrees of freedom of the F-distribution by a constant multiplying factor, commonly referred to as Box's ε. Since the constant factor ε depends on the unknown within-unit covariance matrix, solution such as using lower bound for ε (Geisser and Greenhouse, 1958) or estimates of the within-unit covariance matrix in the calculation of ε (Huynh and Feldt, 1976; Huynh, 1978; Lecoutre, 1991) have been implemented in practical applications. These solutions have been shown to work satisfactorily in terms of controlling Type I error rate when the number of repeated measures is low compared to the degrees of freedom for estimating the covariance matrix. However, the univariate approach is obviously approximate and Rencher and Christensen (2012, p. 219) argue that it has no power advantage over the exact multivariate test. They continue to say "... The only case in which we need to fall back on univariate test is when there are insufficient degrees of freedom to perform multivariate test...", i.e., when the number of repeated measures is larger than the error degrees of freedom to estimate the within-unit covariance.

Well before most researchers embarked on the high-dimension-low-sample-size (HDLSS) problem, Collier et al. (1967), Stoloff (1970) and Maxwell and Arvey (1982) have numerically demonstrated that the univariate approaches for repeated measures tend to be very conservative. In an attempt to improve the estimation of ε in the high dimensional situation, Chi et al. (2012) used "dual" forms of the within sum-of-squares and cross-products matrices . They claim that, besides giving stable estimates of ε, the use of the "dual" version has computational advantage. The approaches of Brunner et al. (2012) and Happ et al. (2015), on the other hand, overcome the high dimensional problem by using the so-called ANOVA-type statistic (Brunner et al., 1997, 1999) and then use F-approximation to their null distribution by matching first two moments of the numerator and denominator quadratic forms with that of a scaled-gamma distribution, an approach shown to be successful in a related problem by Brunner et al. (1997). Again, these approaches although were shown to be numerically satisfactory, they are only approximate solutions. On the other hand, by deriving asymptotic distributions of some suitable statistics in the high-dimensional asymptotic framework, Pauly et al. (2015), Takahashi and Shutoh (2016) and Harrar and Kong (2016) (as given in Chapter 2) devised asymptotically-valid tests. Pauly et al. (2015) consider high-dimensional repeated measures analysis for one sample situation but with the possibility of several within subject factors. The two-sample situation was considered by Takahashi and Shutoh (2016) assuming equal covariance matrices for the two populations. More generally, Harrar and Kong (2016) addressed the multi-group as well as the unequal covariance cases. Other works such as Wang and Akritas (2010a,b) and Wang et al. (2010) are also high-dimensional asymptotic results applicable for repeated measures but assume that the repeated measurements are inherently ordered and the dependence between the measurements decays as the separation between them increases. High-dimensional asymptotic mean vector comparisons have recently received attention in the statistics literature (see Bai and Saranadasa, 1996; Chen and Qin, 2010; Katayama et al., 2013; Cai et al., 2014, and the references there in) under assumptions different from that of Pauly et al. (2015), Takahashi and Shutoh (2016), and Harrar and Kong (2016). However, these recent results are asymptotic and are not applicable in the repeated measures setting.

More specifically, Harrar and Kong (2016) have proven asymptotic normality for their test statistics under certain assumptions on the covariances. In their simulation study, Harrar and Kong (2016) noticed that the error of approximation from these asymptotic distributions could be considerable unless both the number of repeated measurements and replication sizes are large. The present Chapter aims to derive second order asymptotics for the tests considered in Harrar and Kong (2016). In addition, the results in the current Chapter are more general in the sense that they are applicable in situations where there are multiple within and/or between unit crossed and/or nested factors.

This Chapter is organized as follows. Section 3.2 introduces the statistical model, hypotheses and notations used in the remainder of the Chapter. Test statistics for the various effects are presented in Section 3.3 together with asymptotic expansions for their null distributions. The asymptotic power are derived in Section 3.4. Numerical studies are carried out in Section 3.5. First, Monte Carlo simulations are used to show the gain in accuracy from the asymptotic expansions for a selection of covariance matrices and wide choices of values for the number of repeated measures and replication sizes. Data from a large Electroencephalogram (EEG) study of alcoholic and control subjects is used to illustrate the application of the results in Section 3.6. Also, simulation results by generating data with similar design parameters as the real data is considered later in the section. Section 3.7 contains discussions and conclusions. All proofs and preliminary results are placed in the Appendix.

3.2 Model and Hypotheses

Suppose n_i independent b-dimensional observations; for $i = 1, \ldots, a$; are available from multivariate normal populations $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ denoted by $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$ and assume that the a samples are mutually independent. The aim of this Chapter is to derive second order asymptotic result for testing hypotheses in repeated measures analysis when both the total sample size and the number of repeated measurements tend to infinity.

Let

$$\boldsymbol{X} = (\boldsymbol{X}_{11}^\top, \ldots, \boldsymbol{X}_{1n_1}^\top, \boldsymbol{X}_{21}^\top, \ldots, \boldsymbol{X}_{2n_2}^\top, \ldots, \boldsymbol{X}_{a1}^\top, \ldots, \boldsymbol{X}_{an_a}^\top)^\top,$$

where $\boldsymbol{X}_{ik} = (X_{i1k}, \ldots, X_{ibk})^\top$. Further, let

$$\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_1^\top, \ldots, \overline{\boldsymbol{X}}_a^\top)^\top,$$

where $\overline{\boldsymbol{X}}_i = n_i^{-1} \sum_{k=1}^{n_i} \boldsymbol{X}_{ik}$. The usual setting gives the interpretation that X_{ijk} is the responses from the kth subject treated with the ith level of factor A and the jth level of factor B. In this model X_{ijk} and $X_{i'j'k'}$ are assumed to be independent only if $i \neq i'$ or $k \neq k'$. Otherwise the dependence is completely unspecified.

Note that from the distributional assumption made above

$$\mathrm{E}[\boldsymbol{X}_{ik}] = \boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ib})^\top$$

and $\mathrm{Var}(\boldsymbol{X}_{ik}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a $b \times b$ positive definite matrix. Let

$$\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{1b}, \ldots, \mu_{a1}, \ldots, \mu_{ab})^\top$$

and $\widetilde{\boldsymbol{\Sigma}} = \bigoplus_{i=1}^{a} \boldsymbol{\Sigma}/n_i = \boldsymbol{D} \otimes \boldsymbol{\Sigma}$ where $\boldsymbol{D} = \mathrm{diag}(1/n_1, \ldots, 1/n_a)$. Then we have $\mathrm{E}[\overline{\boldsymbol{X}}] = \boldsymbol{\mu}$ and $\mathrm{Var}(\overline{\boldsymbol{X}}) = \widetilde{\boldsymbol{\Sigma}}$.

The hypotheses of interest can be expressed as

$$\mathcal{H}_0 : \boldsymbol{K}\boldsymbol{\mu} = \boldsymbol{0} \ \text{ VS } \ \mathcal{H}_1 : \boldsymbol{K}\boldsymbol{\mu} \neq \boldsymbol{0} \tag{3.1}$$

with $\boldsymbol{K} = \boldsymbol{T}_1 \otimes \boldsymbol{T}_2$, where $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are $a \times a$ and $b \times b$ matrices respectively. We require that the two matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are symmetric and there exist positive definite matrices $\boldsymbol{G}_i$, such that $\boldsymbol{T}_i\boldsymbol{G}_i\boldsymbol{T}_i = \boldsymbol{T}_i$ ($\boldsymbol{G}_i = \boldsymbol{I}$ if $\boldsymbol{T}_i$ is idempotent). Actually, we can apply the linear transformation $\boldsymbol{G}_1^{-1/2} \otimes \boldsymbol{G}_2^{-1/2}$ on the data $\overline{\boldsymbol{X}}$ and use the symmetric and idempotent matrix $\boldsymbol{G}_i^{1/2}\boldsymbol{T}_i\boldsymbol{G}_i^{1/2}$ instead of $\boldsymbol{T}_i$. With this manipulation, the new $\boldsymbol{K}$ still defines the same hypotheses as in (3.1). Therefore, without loss of generality, we can assume that $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are symmetric and idempotent matrices directly. For such $\boldsymbol{T}_i$, $\boldsymbol{K}$ is positive semidefinite matrix. The Rank($\boldsymbol{T}_2$) may be finite or grow with b at the rate O(b). In the following Theorem we establish an equivalent quadratic form expressions for the hypotheses (3.1).

Theorem 3.2.1. *The null hypotheses (3.1) are equivalent to*

$$\mathcal{H}_0' : \boldsymbol{\mu}^\top \boldsymbol{K} \boldsymbol{\mu} = 0 \quad \textit{vs.} \quad \mathcal{H}_1' : \boldsymbol{\mu}^\top \boldsymbol{K} \boldsymbol{\mu} > 0.$$

The above setup may give the impression that this Chapter is dealing with one between-subject and one within-subject factor with levels a and b, respectively. However, the indices $i = 1, \ldots, a$ and $j = 1, \ldots, b$ are to be viewed as lexicographic order of the between-subject factor level combinations and within-subject factor level combinations, respectively. Therefore, the setup covers repeated measures in factorial designs with crossed and nested factors. The factors $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ of matrix $\boldsymbol{K}$ can be viewed as parts of the contrast matrix concerning the between-subject factors and the within-subject factors, respectively. More specifically, suitable choices of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can allow between-subject and within-subject mean comparisons. For a concrete example, consider a factorial design in which there are two between-subject crossed factors, say A and C with a and c levels, respectively, and two within-subject factors, say B and D, where the levels of D are nested within that of B (see also other specific designs considered in Section 3.5). Suppose B has b levels and the jth level of B has d_j levels of D nested within it. The mean vector in this set up would be $\boldsymbol{\mu} = (\boldsymbol{\mu}_{11}^\top, \ldots, \boldsymbol{\mu}_{1c}^\top, \ldots, \boldsymbol{\mu}_{a1}^\top, \ldots, \boldsymbol{\mu}_{ac}^\top)^\top$ where $\boldsymbol{\mu}_{ik} = (\boldsymbol{\mu}_{ik11}^\top, \ldots, \boldsymbol{\mu}_{ik1d_1}^\top, \ldots, \boldsymbol{\mu}_{ikb1}^\top, \ldots, \boldsymbol{\mu}_{ikbd_b}^\top)^\top$. To test the interaction effect of A and B, for instance, we would use $\boldsymbol{T}_1 = \boldsymbol{P}_a \otimes c^{-1}\boldsymbol{J}_c$ and $\boldsymbol{T}_2 = \boldsymbol{Q}(\boldsymbol{Q}^\top\boldsymbol{Q})^-\boldsymbol{Q}^\top$ where $\boldsymbol{Q} = (\bigoplus_{j=1}^b d_j^{-1}\mathbf{1}_{d_j}) \otimes \boldsymbol{P}_b$ and $(\boldsymbol{Q}^\top\boldsymbol{Q})^-$ is the generalized inverse of $\boldsymbol{Q}^\top\boldsymbol{Q}$. Further, the set up above can be reset accordingly. For example a would be replaced by ac, and $\boldsymbol{D} = \mathrm{diag}\{n_{11}, \ldots, n_{1c}, \ldots, n_{a1}, \ldots, n_{ac}\}^{-1}$.

3.3 Higher-Order Asymptotic Tests

We have seen in Theorem 3.2.1 that $\boldsymbol{K\mu} = 0$ if and only if $\boldsymbol{\mu}^\top \boldsymbol{K\mu} = 0$. A reasonable estimator of $\boldsymbol{\mu}^\top \boldsymbol{K\mu}$ is given by $H = \overline{\boldsymbol{X}}^\top \boldsymbol{K}\overline{\boldsymbol{X}}$. Now, define

$$c_k := \mathrm{tr}(\boldsymbol{T}_2\boldsymbol{\Sigma})^k = \mathrm{tr}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{T}_2\boldsymbol{\Sigma}^{1/2})^k,$$

for $k = 1, \ldots, 8$ and $n = \sum_{i=1}^{a} n_i - a$. In this section, we will devise a test for

$$\mathcal{H}_0 : \boldsymbol{K\mu} = \boldsymbol{0}$$

under the following high-dimensional asymptotic framework, when the rank of $\boldsymbol{T}_2$ grows with b:

$B1$: $c_8/c_2^4 = \mathrm{O}(b^{-3})$ as $b \to \infty$.

$B2$: $n \to \infty$ and $b \to \infty$ such that $b/n = \gamma \to \gamma_0 \in (0, \infty)$.

It is well known that (e.g. Yang et al., 2001)

$$\mathrm{tr}(AB)^m \leq \{\mathrm{tr}(A^{2m})\mathrm{tr}(B^{2m})\}^{1/2},$$

for any positive semidefinite matrices A and B. Assumption $B1$ is a sparsity condition on the covariance matrix. Using the trace inequality above we have

$$c_k/c_2^{k/2} = \mathrm{O}(b^{-k/2+1}) \text{ as } b \to \infty \text{ for any } 1 \leq k \leq 7.$$

For example, if $k = 4$, then $c_4 \leq (c_8 \cdot b)^{1/2}$ and if $k = 6$, $c_6 \leq (c_8 c_4)^{1/2}$. Therefore, $c_4/c_2^2 = \mathrm{O}(b^{-1})$ and $c_6/c_2^3 = \mathrm{O}(b^{-2})$. Assumption $B2$ is weaker than the usual requirement that each of sample sizes to diverge and have the same relation with b.

First, we assume $\boldsymbol{\Sigma}$ is known. A centered and suitably-scaled version of H given by

$$T = \frac{H - \mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})c_1}{\sqrt{2\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})^2 c_2}}$$

yields a reasonable test statistic for testing $\mathcal{H}_0$. Let $\delta_k = \mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})^k/\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^k$ for $k = 2, 3, 4$. Since $\boldsymbol{T}_1$ is a symmetric and idempotent matrix, one can see that $0 < \delta_k < 1$ and, hence, $\delta_k = \mathrm{O}(1)$ as $n \to \infty$. In order to generalize the test statistic T for the unknown covariance case, we need to estimate c_1 and c_2 to the appropriate order. The estimators $\widehat{c}_1$ and $\widehat{c}_2$ defined by

$$\widehat{c}_1 = \mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S}) \quad \text{and} \quad \widehat{c}_2 = \frac{n^2}{(n-1)(n+2)}\left\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^2 - \frac{1}{n}\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})\}^2\right\}, \tag{3.2}$$

where

$$\boldsymbol{S} = \frac{1}{n}\sum_{k=1}^{a}(n_i - 1)\boldsymbol{S}_i \quad \text{and} \quad \boldsymbol{S}_i = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i)(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i)^\top,$$

have desirable asymptotic properties given in Theorem 3.3.1. Here, it should be noted that $\widehat{c}_1$ and $\widehat{c}_2$ are unbiased estimators of c_1 and c_2, respectively (Srivastava, 2005; Harrar and Kong, 2016).

Theorem 3.3.1. *Under the high-dimensional asymptotic frameworks $B1$ and $B2$, the estimators $\widehat{c}_1$ and $\widehat{c}_2$ have the following asymptotic properties:*

(i) Asymptotic equivalence: $(\widehat{c}_1 - c_1)/\sqrt{c_2} = O_p(b^{-1/2})$ and $(\widehat{c}_2 - c_2)/c_2 = O_p(b^{-1})$.

(ii) Ratio consistency: $\widehat{c}_2/c_2 \xrightarrow{p} 1$.

Next we study the asymptotic sampling distribution of the test-statistic,

$$\widehat{T} = \frac{\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1} H - \widehat{c}_1}{\sqrt{2\delta_2\widehat{c}_2}}$$

which is obtained from T by replacing c_1 and c_2 by their empirical counterparts.

It is shown in the Appendix that $\widehat{T}$ can be expanded as

$$\widehat{T} = T - \frac{V}{\sqrt{b}} - \frac{1}{b}\frac{TW}{2} + \mathrm{O}_p(b^{-3/2}),$$

where

$$V = \frac{\sqrt{b}(\widehat{c}_1 - c_1)}{\sqrt{2\delta_2 c_2}} \quad \text{and} \quad W = b(\widehat{c}_2 - c_2)/c_2,$$

are $\mathrm{O}_p(1)$ by Theorem 3.3.1. The characteristic function of $\widehat{T}$ can be expanded as given in the following Theorem.

Theorem 3.3.2. *If the null hypothesis $\mathcal{H}_0$ holds, then under the high-dimensional asymptotic frameworks B1 and B2, the characteristic function of $\widehat{T}$ can be expanded as*

$$\phi_{\widehat{T}}(t) = e^{\frac{1}{2}\imath^2 t^2}\left\{1 + \frac{1}{\sqrt{b}}\imath^3 t^3 \eta_3 + \frac{1}{b}(\imath^2 t^2 \frac{\gamma}{2\delta_2} + \imath^4 t^4 \eta_4 + \imath^6 t^6 \frac{\eta_3^2}{2}) + O(b^{-3/2})\right\},$$

where $\eta_3 = \dfrac{4b^{1/2}\delta_3 c_3}{3(2\delta_2 c_2)^{3/2}}$ *and* $\eta_4 = \dfrac{2b\delta_4 c_4}{(2\delta_2 c_2)^2}$.

Note that, by Assumption $B1$, η_3 and η_4 are O(1). Inverting the characteristic function term by term, we get asymptotic expansion for the distribution function of $\widehat{T}$ as follows.

Theorem 3.3.3. *If the null hypothesis $\mathcal{H}_0$ holds, then under the high-dimensional asymptotic frameworks B1 and B2, the distribution function of $\widehat{T}$ can be expanded as*

$$F_{\widehat{T}}(x) = G_{\widehat{T}}(x) + O(b^{-3/2}),$$

uniformly in x where

$$G_{\widehat{T}}(x) = \Phi(x) - \frac{1}{\sqrt{b}}\eta_3\Phi^{(3)}(x) + \frac{1}{b}\{\frac{\gamma}{2\delta_2}\Phi^{(2)}(x) + \eta_4\Phi^{(4)}(x) + \frac{\eta_3^2}{2}\Phi^{(6)}(x)\}$$

and $\Phi^{(j)}(x)$ is the jth derivative of the standard normal cumulative distribution function $\Phi(x)$.

More specifically, Theorem 3.3.3 states that $\sup_{x\in\mathbb{R}} |F_{\widehat{T}}(x) - G_{\widehat{T}}(x)| = \mathrm{O}(b^{-3/2})$. The function $G_{\widehat{T}}(x)$ can alternatively be expressed as

$$G_{\widehat{T}}(x) = \Phi(x) - \phi(x)\left[\frac{1}{\sqrt{b}}\eta_3 h_2(x) + \frac{1}{b}\{\frac{\gamma}{2\delta_2}h_1(x) + \eta_4 h_3(x) + \frac{\eta_3^2}{2}h_5(x)\}\right]$$

where $\phi(x)$ is the standard normal density functions and $h_i(x)$ is the ith Hermite polynomial. The first five Hermite polynomials are:

$$h_1(x) = x,\ h_2(x) = x^2 - 1,\ h_3(x) = x^3 - 3x,$$
$$h_4(x) = x^4 - 6x^2 + 3 \text{ and } h_5(x) = x^5 - 10x^3 + 15x.$$

It should be emphasized that when the terms of orders $b^{-1/2}$ and b^{-1} are ignored, assumptions $B1$ and $B2$ can be relaxed as: (i) the assumption of proportional divergence of n and b in $B2$ is not needed (Harrar and Kong, 2016) and (ii) the sparsity condition on the covariance matrix (assumption $B1$) is needed only for $c_4/c_2^2 = \mathrm{o}(1)$, in which case, the assumption reduces to that of Chen and Qin (2010).

Let $u(z)$ be defined by $P(\widehat{T} \leq u(z)) = P(Z \leq z)$ where Z is a standard normal random variable. In what follows, asymptotic expansion of $u(z)$ in terms of z known as Cornish-Fisher expansion (Hill and Davis, 1968) is given in Corollary 3.3.4.

Corollary 3.3.4. *If the null hypothesis $\mathcal{H}_0$ holds, then under the high-dimensional asymptotic frameworks $B1$ and $B2$, $u(z) = u_A(z) + O(b^{-3/2})$ where*

$$u_A(z) = z + \frac{1}{\sqrt{b}}\eta_3 h_2(z) + \frac{1}{b}\Big\{\frac{\gamma}{2\delta_2}h_1(z) + \eta_4 h_3(z) + \frac{\eta_3^2}{2}h_5(z) - z\eta_3^2 h_2(z)(\frac{1}{2}h_2(z) - 2)\Big\}.$$

The expansions $G_{\widehat{T}}(x)$ and $u_A(z)$ are approximations for the CDF and quantile, respectively, of $\widehat{T}$ under the null hypothesis. In these approximations, η_3 and η_4 depend on c_2 , c_3, and c_4 which are unknown quantities. Therefore, for practical applications, we need an estimated version of the expansions which is uniformly correct up to $\mathrm{O}_p(b^{-3/2})$ in the sense that $\sup_{x\in\mathbb{R}} |F_{\widehat{T}}(x) - \widehat{F}_{\widehat{T}}(x)| = \mathrm{O}_p(b^{-3/2})$ where $\widehat{F}_{\widehat{T}}(x)$ is the estimated version of $F_{\widehat{T}}(x)$. To that end, let $\widehat{c}_1$ and $\widehat{c}_2$ be as defined in (3.2) and define $\widehat{c}_3$ and $\widehat{c}_4$ as

$$\begin{aligned}
\widehat{c}_3 =& \frac{n^4}{m_1}\left[\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^3 - \frac{3}{n}\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^2\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S}) + \frac{2}{n^2}\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})\}^3\right] \quad \text{and} \\
\widehat{c}_4 =& \frac{n^5(n^2+n+2)}{m_2}\left[\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^4 - \frac{4}{n}\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^3\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})\right. \\
&- \frac{2n^2+3n-6}{n(n^2+n+2)}\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^2\}^2 + \frac{2(5n+6)}{n(n^2+n+2)}\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^2\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})\}^2 \\
&\left. - \frac{5n+6}{n^2(n^2+n+2)}\{\mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})\}^4\right],
\end{aligned}$$

where $m_1 = (n-2)(n-1)(n+2)(n+4)$ and $m_2 = m_1(n+1)(n-3)(n+6)$. These estimators are unbiased and enjoy some higher order asymptotic properties that makes them suitable for use in asymptotic expansions.

Theorem 3.3.5. *Under the high-dimensional asymptotic frameworks $B1$ and $B2$, the estimators $\widehat{c}_2, \widehat{c}_3$ and $\widehat{c}_4$ have the following properties:*

(i) Unbiasedness: $E[\widehat{c}_i] = c_i$ for $i = 2, 3, 4$.

(ii) Asymptotic equivalence: $\dfrac{b^{1/2}\widehat{c}_3}{c_2^{3/2}} = \dfrac{b^{1/2}c_3}{c_2^{3/2}} + O_p(b^{-1})$ *and* $\dfrac{b\widehat{c}_4}{c_2^2} = \dfrac{bc_4}{c_2^2} + O_p(b^{-1})$

Now, by using Theorems 3.3.1 and 3.3.5, we know that

$$\widehat{\eta}_3 = \frac{4b^{1/2}\delta_3\widehat{c}_3}{3(2\delta_2\widehat{c}_2)^{3/2}} = \eta_3 + \mathrm{O}_p(b^{-1}), \quad \widehat{\eta}_3^2 = \eta_3^2 + \mathrm{O}_p(b^{-1})$$

$$\text{and} \ \ \widehat{\eta}_4 = \frac{2b\delta_4\widehat{c}_4}{(2\delta_2\widehat{c}_2)^2} = \eta_4 + \mathrm{O}_p(b^{-1}).$$

Therefore, we can define the estimated version $\widehat{G}_{\widehat{T}}(x)$ of $G_{\widehat{T}}(x)$ of Theorem 3.3.3 by replacing η_3 and η_4 with $\widehat{\eta}_3$ and $\widehat{\eta}_4$, respectively.

Before closing this section, we provide an approximate solution in the situation where the rank of $\boldsymbol{T}_2$ does not grow with b. Note that if $\boldsymbol{T}_2$ has finite rank, under the null hypothesis, we may use the approximations

$$H \overset{\text{approx}}{\sim} \frac{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\delta_2 c_2}{c_1}\chi^2_{c_1^2/(c_2\delta_2)} \quad \text{and} \quad \widehat{c}_1 \overset{\text{approx}}{\sim} \frac{n^{-1}c_2}{c_1}\chi^2_{nc_1^2/c_2}.$$

These approximations are obtained by matching the first two moments with that of a scaled Chi-Square distribution. Further, it is known that H is independent of $\widehat{c}_1$. Thus, a test statistic for $\mathcal{H}_0$ is

$$\widehat{T} = \frac{H}{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\widehat{c}_1},$$

and its distribution may be approximated by $\mathcal{F}_{c_1^2/(c_2\delta_2),\ nc_1^2/c_2}$ distribution under the null hypothesis. In the case when the rank of $\boldsymbol{T}_2$ is 1, it turns out that $c_2 = c_1^2$ and the distribution of $\widehat{T}$ can be approximated by $\mathcal{F}_{1/\delta_2,\ n}$. A matrix of special interest in testing the equality of mean vectors given that they are parallel is $\boldsymbol{T}_1 = \mathrm{diag}(n_1, \ldots, n_a) - (n_1, \ldots, n_a)^\top(n_1, \ldots, n_a)/(n+a)$. In this case, $\widehat{T}$ has exact $\mathcal{F}_{a-1,\ n}$ distribution (Harrar and Kong, 2016). In the more general case, we need to estimate c_1^2/c_2 consistently under the asymptotic frameworks $B1$ and $B2$. From Theorem 3.3.1, we know that $\widehat{c}_1^2/\widehat{c}_2 = c_1^2/c_2 + \mathrm{O}_p(b^{-1/2})$. So we can use $\widehat{c}_1^2/\widehat{c}_2$ to estimate it.

3.4 Asymptotic Power

In this section, the asymptotic powers are derived using the methods similar to Chen and Qin (2010). Under the alternative hypothesis $\mathcal{H}_1$, the expectation and variance of H are

$$\mathrm{E}(H) = \mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})c_1 + \boldsymbol{\mu}^\top \boldsymbol{K}\boldsymbol{\mu} \qquad \text{and}$$

$$\mathrm{Var}(H) = 2\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})^2 c_2 + 4\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1) \otimes (\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu},$$

respectively. We derive the asymptotic power under the local alternatives

$$B3 \,:\, \{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1) \otimes (\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu} = \mathrm{o}(\delta_2 c_2),$$

$$B4 \,:\, \delta_2 c_2 = \mathrm{o}(\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1) \otimes (\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu}).$$

A standardized version of H is

$$T_1 = \frac{H - \mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})c_1 - \boldsymbol{\mu}^\top \boldsymbol{K}\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})^2 c_2 + 4\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1) \otimes (\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu}}}.$$

Asymptotic distribution of T_1 is established in Theorem 3.4.1.

Theorem 3.4.1. *Under the high-dimensional asymptotic framework $B1$ and $B2$, and either the assumption $B3$ or $B4$, $T_1 \xrightarrow{D} \mathcal{N}(0,1)$.*

It turns out that the power functions under the local alternatives $B3$ and $B4$ depend on the mean vectors through $\Delta = \boldsymbol{\mu}^\top \boldsymbol{K}\boldsymbol{\mu}$. Specifically, define the power function of $\widehat{T}$ by

$$\beta(\Delta) = \mathrm{P}(\widehat{T} > z_\alpha).$$

Then, by using Theorem 3.4.1, we can obtain the power functions under the local alternative $B3$ and $B4$ as given in Corollary 3.4.2.

Corollary 3.4.2. *Under the assumption $B1$ and $B2$,*

(a) if $B3$ holds, the power function is

$$\beta(\Delta) = G\left(\frac{\{tr(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\Delta}{\sqrt{2\delta_2 c_2}} - z_\alpha\right)$$

(b) if $B4$ holds, the power function is

$$\beta(\Delta) = \Phi\left(\frac{\Delta}{\sqrt{4\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1) \otimes (\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu}}}\right) \to 1.$$

The power under $B4$ in (b) of Corollary 3.4.2 tends to 1 because

$$\boldsymbol{\mu}^\top(\boldsymbol{T}_1\boldsymbol{D}\boldsymbol{T}_1)\otimes(\boldsymbol{T}_2\boldsymbol{\Sigma}\boldsymbol{T}_2)\boldsymbol{\mu} \le \{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}\boldsymbol{\mu}^\top\boldsymbol{K}\boldsymbol{\mu}\sqrt{\delta_2 c_2} = \{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}\Delta\sqrt{\delta_2 c_2}.$$

Note that when $a = 2, \boldsymbol{T}_1 = \boldsymbol{P}_2$, $\boldsymbol{T}_2 = \boldsymbol{I}_b$ and n_1 and n_2 are of same order, the power functions have the same form as in Bai and Saranadasa (1996) and Chen and Qin (2010).

3.5 Simulation Study

To exhibit the improvement resulting from the asymptotic expansion and, hence, facilitate comparison with the limiting distributions in Harrar and Kong (2016), the simulation study will mainly focus on the model where there is one between- and one within-subject factors. We generate 10,000 replications of data from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Although the assumed asymptotic frameworks stipulate n to grow proportionally with b, in reality the actual ratio varies from application to application. To investigate the effect of various proportionality of growth, we look at values of several combinations of a, b and $n_i's$. For practical reasons, we also consider small b and large $n_1, \dots, n_a$ (and vice-versa) combinations with balanced as well as unbalanced designs. For the number of groups (number of levels of factor A), we will consider $a = 2, 3, 4, 6.$ and we set α at 0.01 and 0.05.

Tables 3.1–3.5 present actual Type I error rates (test sizes) for the covariance structures $\boldsymbol{\Sigma} = \rho\boldsymbol{I}_b + (1-\rho)\boldsymbol{J}_b$, $\boldsymbol{\Sigma} = (\rho^{|j-j'|})$ and $\boldsymbol{\Sigma} = (\rho/(j-j')^{1/4})$, respectively. We consider a range of values for ρ. For the first covariance structure, the assumptions in $B2$ are satisfied uniformly in ρ because $c_i = (1-\rho)^i(b-1)$. However, for the other two covariance structures these assumptions do not hold except for $\rho = 0$. Especially, when ρ is close to 1, the quantities b^2c_6/c_2^3 and b^3c_8/c_2^4 diverge very fast. To see the extent of the violation of the assumption for the third covariance structure, for example, $b^{3/2}c_5/c_2^{5/2} = 3.0, 6.6, 16.6, 46.1, 133.5, 392.6$ for $b = 12, 25, 50, 100, 200, 400$ and $\rho = 0.5$. These numbers for b^3c_8/c_2^4 are $302.7, 1349.7, 6311.9, 33055.1, 190299.6, 1182934$ for $\rho = 0.9$. It should be noted that the covariance matrix structure $\boldsymbol{\Sigma} = \rho\boldsymbol{I}_b + (1-\rho)\boldsymbol{J}_b$ will be positive definite if and only if $-1/(b-1) < \rho < 1$. The empty cells in Tables

3.1 and 3.5 correspond to the cases where b and ρ combinations do not yield positive definitive covariance matrices. In all the three tables we consider the contrast matrices $\boldsymbol{T}_1 = \boldsymbol{P}_a$ or $\boldsymbol{J}_a/a$ and $\boldsymbol{T}_2 = \boldsymbol{P}_b$. Another contrast matrix of particular interest in repeated measures analysis is $\boldsymbol{T}_2 = \boldsymbol{J}_b/b$. However, the distribution of $\widehat{T}$ in this case does not depend on b. Hence, we do not carry out simulation for this contrast matrix.

First and foremost, comparing Table 3.1, 3.3 and 3.5 (results for $\alpha = 0.05$) with the results in Harrar and Kong (2016), one can clearly see a marked gain in accuracy resulting from the inclusion of higher-order terms in the asymptotic expansion. We can see in Table 3.1 that for both tests (i.e. $\boldsymbol{T}_1 = \boldsymbol{P}_a$ and $\boldsymbol{J}_a/a$), a large number of the achieved error rates are within a tenth of the actual values. This phenomena happens more for weaker correlations than for stronger ones. Further, it is clear from the Table that the performance of the tests in controlling Type I error rates is excellent when either the sample sizes or the dimension is large. It seems also the case that when n_i's are small, the tests control Type I error rate better as a gets larger. For example, looking at the rows for $a = 6$, performance appear to be satisfactory for small sample sizes but large dimension. Tables 3.3 and 3.5 seem to exhibit similar patters and behaviors. Likewise, for $\alpha = 0.01$, the asymptotic expansion provides a gain in accuracy in controlling Type I error rates (see Tables 3.2, 3.4 and 3.6).

Table 3.1: Achieved Type I error rates ($\times 100\%$) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_b + \rho \boldsymbol{J}_b$. The nominal size is $\alpha = 0.05$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$ $\rho =$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$ $\rho =$			
a	b	$(n_1, \ldots, n_a)$	-0.01	0	0.2	0.5	-0.01	0	0.2	0.5
2	12	(50,100)	5.19	4.98	4.84	4.99	5.10	5.12	5.35	5.05
2	12	(100,100)	4.96	4.94	5.12	4.91	4.65	5.43	4.78	4.98
2	25	(50,100)	5.12	4.83	5.04	4.90	5.02	4.81	4.84	4.80
2	25	(100,100)	5.20	4.95	5.08	5.14	5.25	5.16	4.93	5.25
2	50	(50,100)	4.73	5.25	5.06	5.30	5.01	4.97	5.41	5.14
2	50	(100,100)	4.82	4.89	5.04	4.96	4.65	5.00	4.97	4.33
2	100	(12,13)	5.11	5.56	5.08	5.22	5.30	4.94	5.31	5.58
2	100	(12,25)	5.02	5.10	5.11	5.42	5.08	5.10	5.02	4.97
2	100	(25,25)	4.96	5.35	4.75	4.83	4.89	4.85	5.06	4.79
2	100	(25,50)	5.28	5.31	5.20	5.21	4.61	5.18	5.28	4.97
2	200	(25,25)		4.87	4.95	5.06		5.36	5.33	5.04
2	200	(25,50)		5.24	5.23	4.69		4.98	4.92	4.95
2	200	(50,50)		5.08	4.99	4.98		4.98	5.45	4.96
2	200	(50,100)		5.20	4.97	4.89		4.96	4.91	5.13
2	400	(50,50)		4.98	4.98	4.71		4.79	5.01	4.81
2	400	(50,100)		4.81	4.96	4.95		4.93	4.74	4.95
2	400	(100,100)		5.04	4.91	5.07		5.36	5.53	5.19
2	400	(100,200)		5.04	4.57	5.35		5.07	5.19	5.31
3	12	(50,100,100)	5.15	5.08	4.82	4.92	4.81	4.61	4.73	5.19
3	12	(100,100,100)	4.80	4.90	5.05	5.15	4.88	4.92	5.02	4.85
3	25	(50,100,100)	5.15	5.05	5.11	5.20	5.14	5.20	4.84	5.16
3	25	(100,100,100)	4.79	4.87	4.94	5.20	4.87	4.87	5.50	4.92
3	50	(50,100,100)	5.06	5.63	5.03	5.18	4.82	4.78	4.90	4.80
3	50	(100,100,100)	5.15	5.18	5.15	4.72	4.66	5.13	5.22	4.97
3	100	(16,17,17)	4.95	5.27	4.87	5.19	4.99	5.32	5.23	5.07
3	100	(16,17,33)	4.86	5.21	4.85	5.46	4.70	5.18	5.17	5.01
3	200	(33,33,34)		5.18	5.20	5.29		5.09	5.09	5.32
3	200	(33,34,67)		4.53	5.10	5.17		4.99	5.10	5.19
3	200	(50,50,50)		4.62	4.88	4.86		5.02	4.92	5.08
3	200	(50,50,100)		4.85	5.03	5.10		4.60	4.68	5.20
4	12	(50,50,100,100)	5.13	5.20	5.32	4.96	4.93	5.05	5.12	5.57
4	12	(100,100,100,100)	5.01	5.04	4.94	5.08	5.03	4.77	5.23	5.06
4	25	(50,50,100,100)	4.81	4.64	4.99	4.65	5.26	5.27	5.01	5.11
4	25	(100,100,100,100)	5.19	5.14	4.89	5.05	4.99	5.14	4.70	5.33
4	50	(50,50,100,100)	5.36	5.10	4.70	5.06	5.22	4.92	5.04	5.08
4	50	(100,100,100,100)	5.11	5.32	5.29	5.08	5.06	4.95	4.94	4.93
4	100	(12,12,13,13)	5.02	5.38	5.63	4.97	5.36	5.16	5.20	5.43
4	100	(12,13,25,25)	5.02	5.29	4.98	5.27	5.11	5.33	5.03	5.12
4	200	(25,25,25,25)		5.19	5.17	4.34		4.82	4.72	4.61
4	200	(25,25,50,50)		4.96	4.89	5.17		4.89	5.04	4.52
4	200	(50,50,50,50)		5.31	4.84	5.66		5.07	5.09	4.91
4	200	(50,50,100,100)		4.97	4.95	4.76		4.71	4.74	5.14
6	100	(8,8,8,8,9,9)	4.98	5.57	5.7	5.34	5.07	5.46	5.22	4.83
6	100	(8,8,9,16,17,17)	5.13	5.08	4.78	4.74	4.88	4.89	5.07	5.18
6	200	(16,16,17,17,17,17)		4.97	5.13	5.15		5.39	4.74	5.25

Table 3.2: Achieved Type I error rates ($\times 100\%$) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_b + \rho\boldsymbol{J}_b$. The nominal size is $\alpha = 0.01$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$ $\rho =$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$ $\rho =$			
a	b	$(n_1, \ldots, n_a)$	-0.01	0	0.2	0.5	-0.01	0	0.2	0.5
2	12	(50,100)	1.00	0.99	1.00	1.13	1.03	1.27	0.81	0.95
2	12	(100,100)	1.16	0.91	1.01	1.07	1.06	0.92	1.00	1.05
2	25	(50,100)	0.99	1.06	1.10	1.10	1.04	1.08	0.99	1.01
2	25	(100,100)	0.95	0.76	0.96	0.96	1.10	1.14	1.12	1.00
2	50	(50,100)	1.27	1.02	1.03	1.04	0.84	1.01	0.94	1.01
2	50	(100,100)	1.04	1.09	0.84	1.05	0.96	0.89	1.04	0.89
2	100	(12,13)	1.17	1.02	1.22	1.07	1.24	1.01	0.99	1.21
2	100	(12,25)	1.06	1.02	0.93	0.94	1.14	0.98	1.00	1.16
2	100	(25,25)	1.07	1.04	0.95	1.01	1.15	1.14	1.07	1.04
2	100	(25,50)	0.91	1.03	0.9	0.98	1.12	0.93	1.01	1.08
2	200	(25,25)		1.17	0.85	1.07		1.00	1.06	1.11
2	200	(25,50)		0.95	0.85	1.00		1.00	0.99	0.94
2	200	(50,50)		1.03	0.92	1.03		0.98	1.05	1.06
2	200	(50,100)		1.02	1.30	1.07		0.95	0.95	0.88
2	400	(50,50)		1.01	1.22	0.94		0.95	0.94	0.88
2	400	(50,100)		1.06	0.95	1.10		1.07	1.04	0.98
2	400	(100,100)		0.92	1.03	0.92		1.05	0.94	0.85
2	400	(100,200)		1.08	1.08	0.82		1.11	1.04	0.92
3	12	(50,100,100)	1.08	1.13	1.19	1.04	1.09	1.17	0.97	0.96
3	12	(100,100,100)	0.94	1.11	1.10	1.03	0.92	1.00	0.99	0.99
3	25	(50,100,100)	1.07	0.94	0.96	1.14	0.93	1.12	1.02	1.05
3	25	(100,100,100)	1.07	1.13	0.79	1.05	1.08	0.81	1.04	0.84
3	50	(50,100,100)	1.00	0.78	0.92	0.97	1.19	1.02	0.82	1.00
3	50	(100,100,100)	1.16	0.94	1.12	1.10	0.94	1.04	0.76	0.93
3	100	(16,17,17)	1.06	0.92	1.13	1.13	1.09	0.94	1.10	0.95
3	100	(16,17,33)	1.17	1.16	0.92	0.91	1.17	1.04	1.03	1.06
3	200	(33,33,34)		0.91	1.17	0.98		0.97	1.00	1.13
3	200	(33,34,67)		0.94	0.91	0.84		1.08	0.99	1.15
3	200	(50,50,50)		0.86	1.07	1.09		0.92	0.85	1.01
3	200	(50,50,100)		0.98	1.01	1.02		1.16	0.95	1.12
4	12	(50,50,100,100)	0.71	0.97	0.92	1.09	0.99	1.00	1.01	0.96
4	12	(100,100,100,100)	1.06	1.04	1.03	1.07	0.98	1.09	1.09	0.90
4	25	(50,50,100,100)	1.00	1.08	1.01	1.21	0.89	1.07	0.90	0.87
4	25	(100,100,100,100)	0.88	0.91	0.72	0.95	0.92	1.09	1.07	0.89
4	50	(50,50,100,100)	1.00	0.95	1.08	1.09	1.11	0.92	1.12	0.96
4	50	(100,100,100,100)	1.04	1.02	1.07	1.24	0.93	1.11	1.00	0.90
4	100	(12,12,13,13)	1.04	0.86	1.14	1.16	1.05	1.03	0.80	0.97
4	100	(12,13,25,25)	1.07	0.99	1.04	1.12	1.02	1.03	1.03	1.03
4	200	(25,25,25,25)		1.04	0.95	1.02		0.84	0.96	1.10
4	200	(25,25,50,50)		1.02	1.06	1.07		0.95	1.19	1.09
4	200	(50,50,50,50)		1.06	0.95	1.02		0.98	1.00	1.25
4	200	(50,50,100,100)		0.95	1.11	1.12		0.96	0.91	0.98
6	100	(8,8,8,8,9,9)	1.03	0.91	1.21	1.23	0.99	1.07	0.99	0.98
6	100	(8,8,9,16,17,17)	0.86	1.12	1.26	1.16	1.01	1.08	0.98	0.95
6	200	(16,16,17,17,17,17)		1.00	1.11	1.07		0.99	0.94	0.98

Table 3.3: Achieved Type I error rates ($\times 100\%$) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\rho^{|j-j'|})$. The nominal size is $\alpha = 0.05$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$ $\rho =$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$ $\rho =$			
a	b	$(n_1, \ldots, n_a)$	-0.01	0	0.2	0.5	-0.01	0	0.2	0.5
2	12	(50,100)	4.78	5.10	5.08	4.69	4.89	5.72	5.33	4.69
2	12	(100,100)	4.84	5.04	5.29	5.34	5.00	5.07	4.94	5.01
2	25	(50,100)	4.83	4.92	5.31	4.79	5.23	5.08	5.07	5.31
2	25	(100,100)	5.27	4.89	5.51	5.04	4.89	5.28	5.09	4.87
2	50	(50,100)	4.48	4.82	5.34	5.41	4.98	4.85	4.86	4.85
2	50	(100,100)	4.93	4.48	5.31	5.20	4.81	5.34	4.73	5.12
2	100	(12,13)	5.24	5.04	5.04	5.46	5.61	5.18	5.29	5.30
2	100	(12,25)	4.99	5.47	4.93	5.08	5.11	5.28	4.88	5.19
2	100	(25,25)	5.01	5.13	4.67	5.12	5.56	5.12	4.88	4.78
2	100	(25,50)	5.03	5.11	4.92	4.97	5.23	4.77	4.93	5.19
2	200	(25,25)	5.65	5.33	4.66	4.98	4.96	5.33	5.36	4.58
2	200	(25,50)	4.53	5.00	4.72	4.89	5.01	5.18	5.05	4.94
2	200	(50,50)	5.17	5.09	5.12	4.81	5.22	4.92	5.22	5.12
2	200	(50,100)	5.20	4.67	4.85	5.29	5.16	4.90	5.28	4.82
2	400	(50,50)	4.86	4.74	5.03	5.15	4.92	4.94	4.99	4.99
2	400	(50,100)	4.91	5.03	5.02	4.41	5.10	4.97	5.08	5.07
2	400	(100,100)	4.83	5.10	5.37	4.87	4.76	5.29	4.69	5.19
2	400	(100,200)	5.05	5.08	5.42	4.60	5.18	4.74	4.51	5.19
3	12	(50,100,100)	4.95	5.32	4.63	4.61	4.79	5.14	4.90	4.89
3	12	(100,100,100)	5.10	5.11	5.05	4.75	5.18	4.62	4.79	4.70
3	25	(50,100,100)	4.95	4.78	4.58	5.24	5.13	5.04	5.30	4.73
3	25	(100,100,100)	4.97	4.93	5.17	5.01	5.28	5.03	5.17	5.06
3	50	(50,100,100)	5.03	4.66	4.86	5.09	5.08	5.23	4.80	5.05
3	50	(100,100,100)	5.23	4.79	4.78	5.09	5.32	5.02	4.79	5.24
3	100	(16,17,17)	5.35	5.12	5.04	5.15	4.90	5.02	4.93	5.35
3	100	(16,17,33)	5.39	5.12	5.27	4.83	5.26	5.03	4.98	5.09
3	200	(33,33,34)	5.43	5.23	5.19	5.39	5.16	5.12	5.16	4.80
3	200	(33,34,67)	4.97	5.18	5.32	5.23	5.04	4.95	5.12	4.97
3	200	(50,50,50)	4.74	5.11	4.92	5.35	4.98	5.34	5.08	4.98
3	200	(50,50,100)	5.01	5.18	5.39	5.00	5.38	5.01	5.06	4.98
4	12	(50,50,100,100)	5.02	4.97	5.03	4.95	4.87	4.83	4.68	5.22
4	12	(100,100,100,100)	5.17	5.34	4.49	5.16	4.75	4.83	4.66	4.62
4	25	(50,50,100,100)	5.06	5.10	5.16	4.67	5.47	5.03	5.13	5.13
4	25	(100,100,100,100)	5.38	5.52	4.38	5.13	4.74	4.99	5.22	4.77
4	50	(50,50,100,100)	5.02	5.11	4.76	5.12	5.15	5.15	5.03	5.05
4	50	(100,100,100,100)	4.88	4.83	5.01	4.78	4.76	4.94	4.96	5.07
4	100	(12,12,13,13)	5.31	5.26	4.64	4.99	5.12	5.17	4.91	5.35
4	100	(12,13,25,25)	5.27	5.24	5.45	5.2	5.36	5.38	4.89	5.11
4	200	(25,25,25,25)	5.15	5.26	4.88	5.01	5.18	5.07	5.03	5.15
4	200	(25,25,50,50)	4.98	5.26	4.93	5.00	5.03	4.90	4.93	5.53
4	200	(50,50,50,50)	5.26	5.11	4.70	5.02	5.30	4.92	4.78	4.94
4	200	(50,50,100,100)	5.58	5.10	4.84	5.30	4.81	5.13	4.93	5.22
6	100	(8,8,8,8,9,9)	5.54	4.96	5.29	5.38	5.37	5.02	5.29	5.38
6	100	(8,8,9,16,17,17)	5.28	4.76	5.01	5.04	4.80	5.11	5.17	5.23
6	200	(16,16,17,17,17,17)	5.12	4.97	5.28	5.08	4.92	5.20	4.75	5.15

Table 3.4: Achieved Type I error rates (×100%) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\rho^{|j-j'|})$. The nominal size is $\alpha = 0.01$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$ $\rho =$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$ $\rho =$			
a	b	$(n_1, \ldots, n_a)$	−0.01	0	0.2	0.5	−0.01	0	0.2	0.5
2	12	(50,100)	1.12	0.96	1.11	1.07	1.07	1.00	0.95	1.05
2	12	(100,100)	1.23	1.03	0.94	0.89	0.98	1.08	1.04	0.99
2	25	(50,100)	1.01	1.05	1.12	1.22	1.04	1.01	0.99	1.03
2	25	(100,100)	1.00	1.20	0.95	0.86	0.93	1.00	1.01	1.02
2	50	(50,100)	0.79	0.86	0.93	1.00	1.13	1.01	0.96	0.79
2	50	(100,100)	0.98	0.81	1.23	0.96	0.99	1.00	1.04	0.78
2	100	(12,13)	1.17	1.17	1.01	1.26	1.17	1.17	1.14	0.97
2	100	(12,25)	1.18	1.16	0.91	1.36	1.13	1.22	1.02	1.05
2	100	(25,25)	1.07	0.95	1.18	0.98	0.91	0.90	1.15	0.98
2	100	(25,50)	1.00	0.92	0.93	1.19	1.15	0.90	0.81	1.04
2	200	(25,25)	0.95	0.95	1.08	0.87	1.12	1.01	0.95	0.97
2	200	(25,50)	1.11	1.04	0.92	1.07	1.00	0.89	1.14	0.85
2	200	(50,50)	1.04	0.94	1.02	0.92	1.01	1.04	0.98	0.86
2	200	(50,100)	0.85	1.13	1.02	1.06	1.01	0.85	0.93	1.05
2	400	(50,50)	1.09	1.09	1.04	1.08	0.98	1.05	1.12	0.86
2	400	(50,100)	1.07	1.02	1.04	0.83	1.10	1.05	0.87	0.92
2	400	(100,100)	1.01	1.01	1.12	1.07	1.00	1.05	1.03	1.12
2	400	(100,200)	0.97	0.96	0.88	1.06	1.06	0.98	1.02	1.04
3	12	(50,100,100)	0.92	0.98	1.18	1.09	0.93	1.07	1.09	1.05
3	12	(100,100,100)	0.82	1.24	0.99	0.94	1.10	0.89	0.99	0.92
3	25	(50,100,100)	0.90	0.95	1.05	0.99	1.09	0.98	0.84	0.96
3	25	(100,100,100)	1.00	1.25	1.08	1.04	1.00	1.15	0.91	1.06
3	50	(50,100,100)	1.05	1.12	0.92	0.97	1.02	0.82	1.04	1.08
3	50	(100,100,100)	0.85	0.91	1.11	0.90	1.01	0.88	1.06	1.04
3	100	(16,17,17)	1.11	0.92	1.29	1.02	1.07	0.84	0.93	0.98
3	100	(16,17,33)	1.14	1.28	0.98	0.95	1.03	1.01	1.07	1.28
3	200	(33,33,34)	1.26	1.11	0.97	1.12	1.07	0.99	1.04	0.93
3	200	(33,34,67)	1.06	1.04	0.99	1.05	0.99	0.99	1.02	0.99
3	200	(50,50,50)	1.01	0.95	1.04	1.09	1.12	1.05	1.10	1.12
3	200	(50,50,100)	0.99	1.03	0.87	1.26	1.11	1.06	0.98	1.08
4	12	(50,50,100,100)	0.79	1.02	1.10	0.85	1.08	1.03	1.17	1.03
4	25	(50,50,100,100)	1.08	0.90	1.36	0.92	1.12	0.87	1.09	0.95
4	25	(100,100,100,100)	0.89	1.23	1.11	1.08	0.99	1.14	1.12	0.96
4	50	(50,50,100,100)	1.02	0.97	0.87	0.96	1.22	0.98	1.01	0.95
4	50	(100,100,100,100)	0.97	1.03	1.04	0.92	0.98	1.13	1.29	0.92
4	100	(12,12,13,13)	1.06	1.07	1.16	1.03	1.09	1.03	1.08	1.01
4	100	(12,13,25,25)	0.97	1.04	1.10	1.12	1.19	0.93	1.03	0.91
4	200	(25,25,25,25)	1.01	0.99	1.00	0.99	0.90	0.95	0.96	1.00
4	200	(25,25,50,50)	0.91	1.05	0.90	1.04	0.72	1.08	0.94	0.95
4	200	(50,50,100,100)	0.93	0.88	1.04	1.08	0.84	0.87	1.06	0.88
6	100	(8,8,8,8,9,9)	1.11	1.24	1.14	1.33	0.99	0.92	1.14	1.00
6	100	(8,8,9,16,17,17)	1.16	1.05	0.95	1.29	1.04	0.91	1.04	1.19
6	200	(16,16,17,17,17,17)	1.12	1.05	0.90	1.05	0.99	1.05	0.99	0.90

Table 3.5: Achieved Type I error rates (×100%) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\rho/(j-j')^{1/4})$. The nominal size is $\alpha = 0.05$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$			
			$\rho =$				$\rho =$			
a	b	$(n_1, \ldots, n_a)$	-0.01	0	0.2	0.5	-0.01	0	0.2	0.5
2	12	(50,100)	5.04	5.06	5.28	5.41	5.40	4.95	5.63	4.92
2	12	(100,100)	5.12	4.83	5.55	4.88	5.13	4.82	5.29	4.84
2	25	(50,100)	5.24	5.20	4.90	5.02	5.23	5.25	4.89	4.42
2	25	(100,100)	5.15	5.00	4.97	4.63	5.05	4.68	5.26	4.48
2	50	(50,100)	5.24	4.95	5.06	4.97	4.80	4.97	4.83	4.70
2	50	(100,100)	4.85	4.69	5.42	5.15	5.06	5.46	5.13	5.11
2	100	(12,13)	4.88	5.46	5.05	5.58	5.51	5.07	5.25	5.23
2	100	(12,25)	4.94	4.72	5.08	5.15	4.44	5.00	5.24	4.64
2	100	(25,25)	4.94	4.57	5.01	4.45	5.90	5.18	5.18	5.12
2	100	(25,50)	5.32	4.85	5.19	4.98	4.76	5.59	4.87	5.28
2	200	(25,25)	5.12	5.45	5.19	4.88	5.05	4.99	5.18	5.03
2	200	(25,50)	5.07	4.90	4.90	5.21	5.28	5.20	5.12	4.99
2	200	(50,50)	5.24	5.32	5.20	4.60	4.93	5.46	4.42	5.01
2	200	(50,100)	4.53	5.32	5.14	4.97	4.99	5.18	4.91	4.46
2	400	(50,50)		5.45	4.74	4.73		4.70	5.13	4.33
2	400	(50,100)		5.10	4.85	4.65		5.50	5.05	4.84
2	400	(100,100)		4.93	4.88	4.87		4.91	4.87	4.94
2	400	(100,200)		4.77	5.17	5.09		4.61	4.83	4.68
3	12	(50,100,100)	4.97	5.10	4.89	4.71	4.87	5.18	5.21	5.43
3	12	(100,100,100)	4.83	4.61	5.18	5.41	4.84	4.75	4.88	4.84
3	25	(50,100,100)	5.42	4.71	5.12	5.21	4.92	4.64	5.35	4.97
3	25	(100,100,100)	5.13	5.05	5.05	4.69	5.29	4.72	4.83	4.77
3	50	(50,100,100)	5.12	4.76	5.26	4.96	5.02	5.21	4.89	5.12
3	50	(100,100,100)	4.89	5.11	5.06	5.06	5.36	5.12	5.41	4.63
3	100	(16,17,17)	4.70	5.15	5.03	5.33	4.69	5.15	4.97	4.90
3	100	(16,17,33)	5.11	5.18	5.26	5.35	4.99	4.74	5.05	5.07
3	200	(33,33,34)	5.00	4.91	5.12	5.04	4.86	5.15	5.01	4.88
3	200	(33,34,67)	4.80	4.91	5.00	4.41	5.29	5.14	5.03	4.78
3	200	(50,50,50)	5.22	4.65	4.67	4.99	5.23	5.10	5.35	4.75
3	200	(50,50,100)	4.93	4.92	4.75	5.29	5.32	4.98	5.13	4.81
4	12	(50,50,100,100)	4.72	5.04	4.86	5.26	5.14	5.05	4.64	5.05
4	12	(100,100,100,100)	5.41	5.01	5.30	4.90	4.97	4.74	5.15	4.85
4	25	(50,50,100,100)	4.98	4.97	4.81	5.46	5.38	4.93	4.66	4.75
4	25	(100,100,100,100)	5.19	4.57	5.16	4.75	4.79	5.05	5.24	4.81
4	50	(50,50,100,100)	4.87	5.13	4.74	4.94	5.25	5.18	5.38	4.75
4	50	(100,100,100,100)	5.41	5.63	5.32	5.20	5.54	5.03	4.80	4.89
4	100	(12,12,13,13)	5.43	5.39	5.22	4.93	5.13	5.24	5.07	5.07
4	100	(12,13,25,25)	5.21	4.97	5.07	5.16	5.24	5.12	4.97	4.76
4	200	(25,25,25,25)	4.78	5.34	4.86	5.07	5.14	5.08	5.32	5.10
4	200	(25,25,50,50)	4.97	4.83	4.90	4.83	4.85	5.11	4.83	4.72
4	200	(50,50,50,50)	4.96	4.68	5.23	5.00	4.93	5.33	4.97	5.12
4	200	(50,50,100,100)	4.95	4.92	4.88	4.76	4.95	5.15	5.01	4.87
6	100	(8,8,8,8,9,9)	5.27	4.76	5.19	5.57	5.35	5.33	5.12	5.25
6	100	(8,8,9,16,17,17)	5.27	4.91	4.90	5.30	5.02	5.17	5.01	4.92
6	200	(16,16,17,17,17,17)	5.27	5.05	4.93	5.36	4.78	5.20	4.72	4.84

Table 3.6: Achieved Type I error rates ($\times 100\%$) for the testing procedures when $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and sampling from $\mathcal{N}_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\rho/(j-j')^{1/4})$. The nominal size is $\alpha = 0.01$.

			$\boldsymbol{T}_1 = \boldsymbol{P}_a$				$\boldsymbol{T}_1 = \boldsymbol{J}_a/a$			
			$\rho =$				$\rho =$			
a	b	$(n_1, \ldots, n_a)$	-0.01	0	0.2	0.5	-0.01	0	0.2	0.5
2	12	(50,100)	1.03	1.10	0.97	1.01	1.17	0.93	1.02	0.90
2	12	(100,100)	0.91	0.86	1.02	0.89	1.26	1.11	0.93	1.12
2	25	(50,100)	1.32	1.13	0.86	0.99	1.14	1.08	1.15	0.92
2	25	(100,100)	0.93	1.04	0.95	1.12	1.04	1.08	1.07	0.97
2	50	(50,100)	0.99	0.91	0.94	1.04	1.06	1.07	0.97	1.09
2	50	(100,100)	1.11	0.97	1.13	1.06	1.13	1.00	1.02	0.78
2	100	(12,13)	1.10	1.09	1.15	1.13	1.04	1.22	1.19	1.24
2	100	(12,25)	1.23	1.04	1.01	1.02	0.93	1.02	1.01	1.26
2	100	(25,25)	1.03	1.05	0.93	1.11	1.08	0.97	1.08	1.00
2	100	(25,50)	0.96	1.08	0.98	1.03	1.00	0.92	1.14	1.21
2	200	(25,25)	0.96	1.09	0.71	0.90	1.02	1.02	1.07	1.12
2	200	(25,50)	0.94	1.03	1.02	0.95	1.13	1.05	1.09	1.08
2	200	(50,50)	1.00	1.02	0.84	1.02	1.03	0.98	1.14	1.06
2	200	(50,100)	0.88	1.20	0.98	0.94	1.21	0.99	0.99	1.20
2	400	(50,50)		1.14	1.02	0.87		1.14	1.06	0.91
2	400	(50,100)		0.98	1.04	1.00		0.85	1.02	0.94
2	400	(100,100)		1.10	0.91	1.07		0.90	1.11	0.86
2	400	(100,200)		0.93	0.97	0.78		0.87	1.04	0.89
3	12	(50,100,100)	1.01	1.22	0.83	1.00	1.05	0.92	1.00	0.86
3	12	(100,100,100)	0.86	1.26	1.09	1.12	0.96	0.84	1.08	0.87
3	25	(50,100,100)	1.03	1.16	1.15	1.06	1.00	0.94	0.94	1.02
3	25	(100,100,100)	1.09	1.02	0.93	0.99	1.12	1.06	0.99	0.87
3	50	(50,100,100)	0.98	1.05	1.11	1.06	0.99	1.06	0.92	0.99
3	50	(100,100,100)	1.13	1.22	1.18	1.20	1.05	1.02	1.10	0.63
3	100	(16,17,17)	1.13	1.21	1.11	1.08	1.04	0.98	0.97	0.96
3	100	(16,17,33)	1.12	1.05	1.14	1.23	1.11	0.91	0.96	0.90
3	200	(33,33,34)	1.15	1.02	1.10	0.90	0.98	0.95	1.22	0.90
3	200	(33,34,67)	0.97	1.04	0.98	1.03	0.98	0.92	1.06	1.06
3	200	(50,50,50)	1.19	0.83	1.11	0.95	0.98	0.98	1.19	0.97
3	200	(50,50,100)	0.84	1.10	1.03	1.09	1.06	1.18	1.19	1.07
4	12	(50,50,100,100)	0.98	1.02	1.10	1.07	1.05	1.05	0.94	0.89
4	12	(100,100,100,100)	0.89	0.95	1.02	1.13	1.06	1.01	1.16	1.07
4	25	(50,50,100,100)	1.05	1.02	0.97	0.99	1.16	0.95	1.00	0.97
4	25	(100,100,100,100)	1.15	1.00	1.09	1.10	1.15	0.93	1.01	0.90
4	50	(50,50,100,100)	0.98	0.97	1.11	1.10	1.00	1.12	0.95	0.87
4	50	(100,100,100,100)	1.04	1.08	0.85	0.94	1.17	0.85	0.97	0.91
4	100	(12,12,13,13)	0.91	1.16	1.02	1.36	1.07	0.96	1.21	1.20
4	100	(12,13,25,25)	0.85	1.15	0.91	1.08	1.10	1.01	1.07	1.09
4	200	(25,25,25,25)	1.12	1.01	1.00	1.08	1.20	1.02	1.12	1.03
4	200	(25,25,50,50)	1.07	0.82	1.21	0.90	1.02	1.04	1.16	0.85
4	200	(50,50,50,50)	0.89	0.77	1.00	0.99	1.00	0.92	0.85	0.83
4	200	(50,50,100,100)	1.05	1.06	1.03	1.07	0.98	1.03	0.89	0.95
6	100	(8,8,8,8,9,9)	0.96	1.08	1.09	1.22	1.02	0.91	1.01	1.18
6	100	(8,8,9,16,17,17)	1.11	0.97	1.15	0.97	0.89	0.98	1.02	1.04
6	200	(16,16,17,17,17,17)	0.98	1.05	1.02	1.33	1.20	1.09	1.04	0.93

To investigate performance in terms of power, we compare the power of the method by Chi et al. (2012) with the methods proposed in this Chapter taking $\boldsymbol{T}_2 = \boldsymbol{P}_b$ and setting $\boldsymbol{T}_1$ to either $\boldsymbol{P}_a$ or $\boldsymbol{J}_a/a$. To keep the comparison manageable, we fix $a = 3$ and $\boldsymbol{\Sigma} = \rho \boldsymbol{I}_b + (1-\rho)\boldsymbol{J}_b$ where $\rho = 0.2$. In regards to sample sizes and dimension, we use the combinations $(b; n_1, n_2, n_3) = (10; 5, 10, 10), (10; 50, 100, 100), (100; 5, 10, 10)$ and $(100; 50, 100, 100)$. For the alternative hypotheses, when $\boldsymbol{T}_1 = \boldsymbol{P}_a$, we take $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{0}$ and consider two structures for $\boldsymbol{\mu}_1$. The first one represents a dense alternative, namely $\boldsymbol{\mu}_{1i} = (1+\delta)$ for i odd and $\boldsymbol{\mu}_{1i} = (1-\delta)$ for i even, and the other one represents a sparse alternative, namely $\boldsymbol{\mu}_1 = (1+\delta, \boldsymbol{1}_{b-1}^\top)^\top$. In both cases δ is made to vary from 0 to 1. When $\boldsymbol{T}_1 = \boldsymbol{J}_a/a$, we take $\boldsymbol{\mu}_1 = \boldsymbol{1}_b + \boldsymbol{\mu}_2$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ and consider two structures of $\boldsymbol{\mu}_2$ representing dense and sparse alternatives. For the first one, we take $\boldsymbol{\mu}_{2i} = \delta$ for i odd and $\boldsymbol{\mu}_{2i} = -\delta$ for i even, and for the second one we take $\boldsymbol{\mu}_2 = (\boldsymbol{0}_{b-1}^\top, \delta)^\top$. Here also, δ varies from 0 to 1. The later structure for both values of $\boldsymbol{T}_1$ represent departures that approach to the null hypotheses at the rate $b^{1/2}$. More precisely, the scaled departure from the null $||\boldsymbol{\mu}_1 - \boldsymbol{1}_b||/\mathrm{tr}(\boldsymbol{\Sigma})^{1/2}$ are δ and $|\delta|/\sqrt{b}$, respectively. Figures 3.1 and 3.2 show power results for $\boldsymbol{T}_1 = \boldsymbol{P}_a$ and $\boldsymbol{T}_1 = \boldsymbol{J}_a/a$, respectively. For dense alternatives (left panels), our methods has a clear advantage in all cases. More pronounced dominance is observed, in particular, when n is small. Both methods perform comparably well for sparse alternatives (right panels) except Chi et al. (2012) has an edge when b is large.

Figure 3.1: Power comparison of the proposed methods and the test by Chi et al. (2012) for $\boldsymbol{T}_1 = \boldsymbol{P}_a$ and $\boldsymbol{T}_2 = \boldsymbol{P}_b$. Data is generated from $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = 0.8\boldsymbol{I}_b + 0.2\boldsymbol{J}_b$. In the both panel of the plot, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$. In the left panel $\boldsymbol{\mu}_{1i} = (1+\delta)$ for i odd, $\boldsymbol{\mu}_{1i} = (1-\delta)$ for i even and in the right panel $\boldsymbol{\mu}_1 = (1+\delta, \mathbf{1}_{b-1}^\top)^\top$.

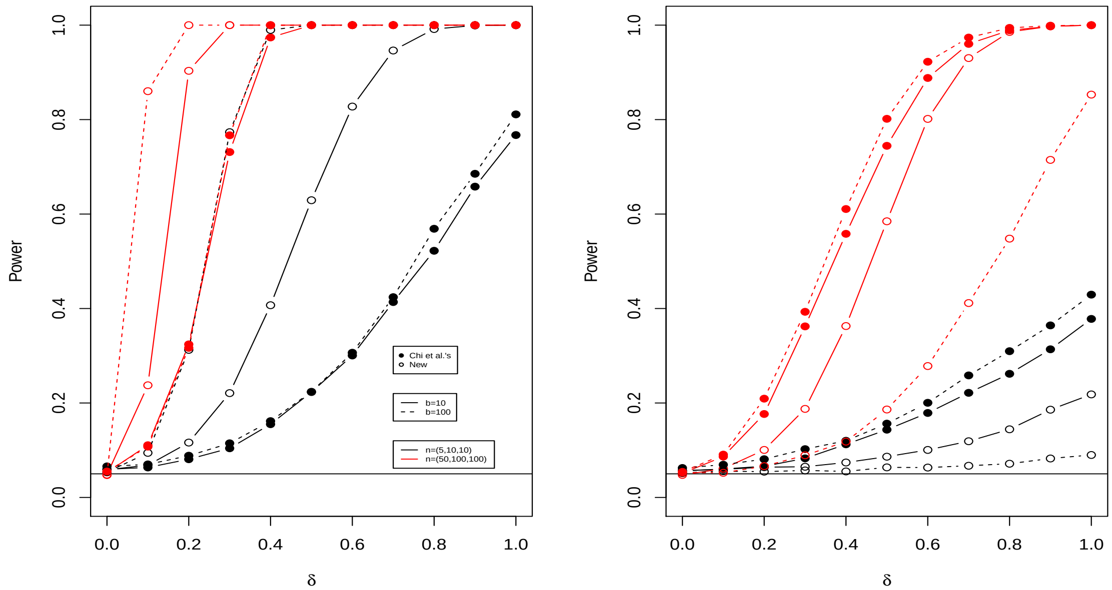
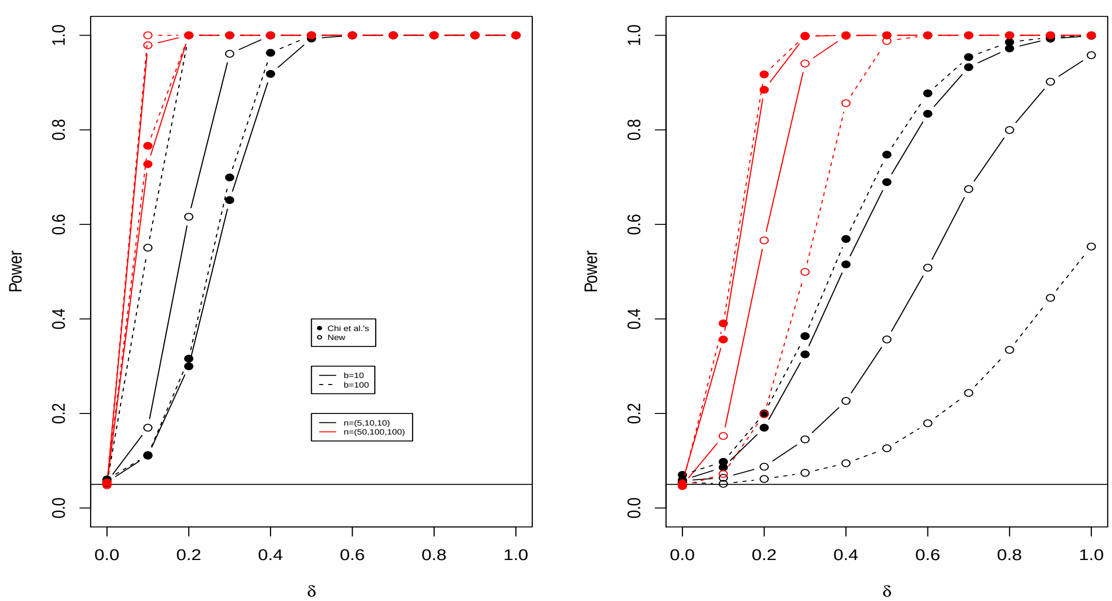


Figure 3.2: Power comparison of the proposed methods and the test by Chi et al. (2012) for $\boldsymbol{T}_1 = \boldsymbol{J}_a/a$ and $\boldsymbol{T}_2 = \boldsymbol{P}_b$. Data is generated from $\mathcal{N}_b(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = 0.8\boldsymbol{I}_b + 0.2\boldsymbol{J}_b$. In the both panel of the plot, $\boldsymbol{\mu}_1 = \mathbf{1}_b + \boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$. In the left panel $\boldsymbol{\mu}_{2i} = \delta$ for i odd and $\boldsymbol{\mu}_{2i} = -\delta$ for i even and in the right panel $\boldsymbol{\mu}_2 = (\mathbf{0}_{b-1}^\top, \delta)^\top$.



3.6 Real Data Analysis

In this section, we analyze a publicly available data obtained from the University of California-Irvine Machine Learning Repository[1]. The data arose from a large study to examine Electroencephalograph (EEG) correlates of genetic predisposition to alcoholism. Measurements from 64 electrodes placed on subject's scalps were recorded 256 times for 1 second. The study involved two groups of subjects: alcoholic ($n_1 = 77$) and control ($n_2 = 45$). Each subject was exposed to either a single stimulus (S1) or two stimuli (S1 and S2) which were pictures of objects chosen from a picture set. The sixty-four electrodes (channels) are divided into groups based on their location on the scalp (frontal, temporal, parietal and occipital lobes). To illustrate the application of the methods concisely, we focus the analysis on data from the stimulus S1 and the seventeen frontal-lobe channels. The outcome measurements are Event-Related Potentials (ERP) indicating the level of electrical activity (in volts) in the region of the brain where each of the electrodes is placed. This repeated measures data has two within-subject factors (time and channels) and one between-subject factor (alcohol use). The within-subject factors time and channel have 256 and 17 levels, respectively.

The main research questions of interest are: ($\mathcal{H}_{01}$) whether the ERP profiles over time differ between channels and groups (three-way interaction: alcohol $\times$ time $\times$ channel); ($\mathcal{H}_{02}$) whether ERP profiles are similar between the channels when averaged over groups (similar time profiles for all the channels); ($\mathcal{H}_{03}$) if the time profiles of ERP are similar between the two groups averaged over channels; ($\mathcal{H}_{04}$) whether the ERP profiles are constant (flat) when averaged over channels and groups. For describing the contrast matrices, we assume the data vectors from each subject are arranged by grouping the 17 channels within each time point, i.e. the data vector from the jth subject in the ith group is $\boldsymbol{X}_{ij} = (X_{ij11}, \ldots, X_{ij1,17}, \ldots, X_{ij,256,1}, \ldots, X_{ij,256,17})^\top$. In the notations of this Chapter, the four hypotheses of interest, viz. $\mathcal{H}_{0i}$ for $i = 1, 2, 3, 4$, can be tested by using the contrast matrices $\boldsymbol{T}_1 = \boldsymbol{P}_2$ and $\boldsymbol{T}_2 = \boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$;

[1]web address: https://archive.ics.uci.edu/ml/datasets/EEG+Database accessed on May 5, 2016.

$\boldsymbol{T}_1 = \boldsymbol{J}_2/2$ and $\boldsymbol{T}_2 = \boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$; $\boldsymbol{T}_1 = \boldsymbol{P}_2$ and $\boldsymbol{T}_2 = \boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$; and $\boldsymbol{T}_1 = \boldsymbol{J}_2/2$ and $\boldsymbol{T}_2 = \boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$, respectively. The results of the analysis are presented in Table 3.7. Overall time-profile similarity across groups (averaged over channels) cannot be rejected (p- value = 0.205). In fact, channel-by-channel similarity of time profiles of ERP across groups cannot be rejected (p-value = 0.196). However, the flatness over time is rejected overall for all channels as well as channel-by-channel.

Table 3.7: Analysis for EEG data in frontal channels ($a = 2$, $b = 256$ and $d = 17$).

Hypothesis	$\boldsymbol{T}_1$	$\boldsymbol{T}_2$	$\widehat{T}$	p-value
$\mathcal{H}_{01}$	$\boldsymbol{P}_2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$	0.535	0.196
$\mathcal{H}_{02}$	$\boldsymbol{J}_2/2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$	26.252	0
$\mathcal{H}_{03}$	$\boldsymbol{P}_2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$	0.489	0.205
$\mathcal{H}_{04}$	$\boldsymbol{J}_2/2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$	42.430	0

As a way of ascertaining the reproducibility and reliability of the results in Table 3.7, we conducted a simulation study using parameters similar to that of the EEG data. For table 3.8, we generate 1000 replications of data from $\mathcal{N}_b(\mathbf{0}, \boldsymbol{\Sigma}_i)$. We look at values of $b = 256$, $d = 17$, $a = 2$ and $n_1 = 77$ $n_2 = 45$ and take $\alpha = 0.05$. Table 3.8 present actual Type I error rates (test sizes) for the covariance structures $\boldsymbol{\Sigma}_1 = \rho \boldsymbol{I}_b + (1-\rho)\boldsymbol{J}_b$, for $\rho = 0.2$ and random matrices $\boldsymbol{\Sigma}_i$ for $i = 2, 3, 4$ defined as follows. Let $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i^\top \boldsymbol{\Lambda}_i \boldsymbol{Q}_i$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix with diagonal entries taken from $\mathcal{U}nif(0,1)$ and $\boldsymbol{Q}_i$ is orthogonal matrix. Indeed, $\boldsymbol{Q}_i$ can be defined from the QR decomposition of a random matrix $\boldsymbol{Z}_i = (Z_{i,jj'})$ where $Z_{i,jj'}$ are iid random variables. Here, we consider three distributions for $Z_{i,jj'}$, namely $Z_{2,jj'} = 1_{\{j=j'\}}$ with probability 1, $Z_{3,jj'} \sim \mathcal{E}xp(1)$ and $Z_{4,jj'} \sim \mathcal{N}(0,1)$.

Table 3.8: Achieved Type I error rates ($\times 100\%$) for the testing procedures with parameters similar to EEG data, i.e. $a = 2, b = 256, d = 17, n_1 = 77, n_2 = 45$.

$\boldsymbol{T}_1$	$\boldsymbol{T}_2$	Σ_1	Σ_2	Σ_3	Σ_4
$\boldsymbol{P}_2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$	5.0	5.9	4.5	5.8
$\boldsymbol{J}_2/2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{P}_{17}$	4.5	4.0	5.4	5.8
$\boldsymbol{P}_2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$	4.6	4.6	5.4	5.4
$\boldsymbol{J}_2/2$	$\boldsymbol{P}_{256} \otimes \boldsymbol{J}_{17}/17$	5.6	4.8	5.2	4.6

It is clear from Table 3.8 that the achieved Type I error rates are satisfactorily close to 5% regardless of the covariance matrix assumed.

3.7 Discussion and Conclusion

This Chapter derives approximations for the null distributions and quantiles of some test statistics in repeated measures. The approximations ensure the errors to be of order $O(b^{-3/2})$ where b is the dimension, i.e. the number of repeated measures. Factorial designs are treated in a unified manner where multiple between- and within-subjects factors which may be crossed or nested are allowed. General covariance structure is allowed where no pre-determined sequence is assumed among the repeated measurements. Therefore, the repeated measurements could be over time or under different treatment conditions.

The asymptotic results require some regularity condition on the covariance matrix. Such assumption appears to be inevitable as long as one prefers to consider unstructured covariance matrix. Our observation from the simulation is that this assumption does not appear to restrict the utility of the results for application in more general situations. Nevertheless, we made somewhat milder requirements compared to similar works (see, for example, Bai and Saranadasa, 1996; Takahashi and Shutoh, 2016). Indeed, one may conjecture to drop these assumptions. This Chapter also assumes proportional divergence of the sample size and dimension, i.e. $n/b \to \gamma_0 \in (0, \infty)$, but otherwise either one can be larger than the other. We should point out that this assumption can be relaxed to cover other cases, namely $n = O(b^{\epsilon})$ for $\epsilon > 1$ or $\epsilon < 1$. However, the expanded cumulative distribution function may have terms with order different from $b^{-j/2}$ for $j = 1, 2, \ldots$ in which case the standard Cornish-Fisher formula for the quantile will not apply. Non standard expansions will need to be derived for the quantiles. Regardless, our impression from the simulation is that the effect of these terms may be insignificant. In the interest of avoiding complications we did not pursue these cases further.

The development of this Chapter is under normality. We recommend testing the validly of this assumption before applying the methods. Transformation that improve normality could also be attempted in the event non-normality is detected or suspected. In the proofs, multivariate normality of the repeated measures is mostly needed for its nice property of independence up to correlation and independence of some linear and quadratic forms. Limiting distribution results under statistical models that include independence up to correlation assumption have been derived in Bai and Saranadasa (1996) and Chen and Qin (2010) for two-sample and Srivastava and Kubokawa (2013) for multiple-sample comparison of mean vectors. In the interest of space, we opted to relegate the investigation of these models for limiting distribution as well as asymptotic expansion to a follow-up manuscript.

3.8 Appendix: Proofs

Lemma 3.8.1. *If the null hypothesis $\mathcal{H}_0$ holds, then under the high-dimensional asymptotic frameworks B1 and B2, the characteristic function of T can be expanded as*

$$\phi_T(t) = e^{\frac{1}{2}\imath^2 t^2}\left\{1 + \frac{1}{\sqrt{b}}\imath^3 t^3 \eta_3 + \frac{1}{b}(\imath^4 t^4 \eta_4 + \imath^6 t^6 \eta_3^2/2) + O(b^{-3/2})\right\}.$$

Proof of Lemma 3.8.1. Let $\boldsymbol{Z} = \widetilde{\boldsymbol{\Sigma}}^{-1/2}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$. Then

$$T = \frac{\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\boldsymbol{Z}^\top\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z} - c_1}{\sqrt{2\delta_2 c_2}}.$$

The characteristic function of T is

$$\begin{aligned}\phi_T(t) &= \exp\big(-\frac{\imath t c_1}{\sqrt{2\delta_2 c_2}}\big)\mathrm{E}\Big[\exp\big\{\frac{\imath t}{\sqrt{2\delta_2 c_2}}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\boldsymbol{Z}^\top\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z}\big\}\Big]\\ &= \exp\big(-\frac{\imath t c_1}{\sqrt{2\delta_2 c_2}}\big)\int_{\boldsymbol{Z}}(2\pi)^{-ab/2}\exp\big\{-\frac{1}{2}\boldsymbol{Z}^\top\boldsymbol{M}\boldsymbol{Z}\big\}d\boldsymbol{Z}\\ &= \exp\big(-\frac{\imath t c_1}{\sqrt{2\delta_2 c_2}}\big)|\boldsymbol{M}|^{-1/2},\end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{I} - \frac{2\imath t}{\sqrt{2\delta_2 c_2}}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}$ and $|\boldsymbol{M}|$ is the determinant of $\boldsymbol{M}$. Let α_i's be the eigenvalues of $\boldsymbol{T}_1\boldsymbol{D}$, β_j's be the eigenvalues of $\boldsymbol{\Sigma}^{1/2}\boldsymbol{T}_2\boldsymbol{\Sigma}^{1/2}$, then

$$|\boldsymbol{M}| = \prod_{i=1}^{a}\prod_{j=1}^{b}\big(1 - \frac{2\imath t}{\sqrt{2\delta_2 c_2}}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\alpha_i\beta_j\big).$$

Thus by Taylor's series expansion, we have

$$\begin{aligned}\log|\boldsymbol{M}|^{-1/2} =& -\frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b}\log\left(1-\frac{2\imath t}{\sqrt{2\delta_2 c_2}}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\alpha_i\beta_j\right)\\ =&\sum_{k=1}^{\infty}\frac{2^{k/2-1}}{k}(\imath t)^k\frac{\delta_k c_k}{\delta_2^{k/2}c_2^{k/2}}\\ =&\frac{\imath t c_1}{\sqrt{2\delta_2 c_2}}+\frac{1}{2}\imath^2 t^2+\frac{1}{\sqrt{b}}\imath^3 t^3\eta_3+\frac{1}{b}\imath^4 t^4\eta_4+\mathrm{O}(b^{-3/2}).\end{aligned}$$

So the characteristic function of T can be expanded as

$$\phi_T(t)=e^{\varphi_T(t)}=e^{\frac{1}{2}\imath^2 t^2}\left\{1+\frac{1}{\sqrt{b}}\imath^3 t^3\eta_3+\frac{1}{b}(\imath^4 t^4\eta_4+\imath^6 t^6\eta_3^2/2)+\mathrm{O}(b^{-3/2})\right\}.$$

□

Proof of Theorem 3.3.2. Denote $H^*=\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}H$. Note that

$$\frac{\widehat{c}_2}{c_2}=1+\frac{1}{b}W \quad \text{and} \quad \frac{\widehat{c}_1-c_1}{\sqrt{2\delta_2 c_2}}=\frac{V}{\sqrt{b}}.$$

By Taylors' expansion, we have

$$\left(\frac{\widehat{c}_2}{c_2}\right)^{-1/2}=1-\frac{1}{2b}W+\mathrm{O}_p(b^{-2}).$$

Then

$$\begin{aligned}\widehat{T}=&\frac{H^*-\widehat{c}_1}{\sqrt{2\delta_2\widehat{c}_2}}=\frac{(H^*-c_1)-(\widehat{c}_1-c_1)}{\sqrt{2\delta_2 c_2}\sqrt{\widehat{c}_2/c_2}}\\ =&\frac{(H^*-c_1)-(\widehat{c}_1-c_1)}{\sqrt{2\delta_2 c_2}}\left\{1-\frac{1}{2b}W+\mathrm{O}_p(b^{-2})\right\}\\ =&T-\frac{V}{\sqrt{b}}-\frac{TW}{2b}+\mathrm{O}_p(b^{-3/2}),\end{aligned}$$

where $T=\dfrac{(H^*-c_1)}{\sqrt{2\delta_2 c_2}}$. So the characteristic function of $\widehat{T}$ is

$$\begin{aligned}\phi_{\widehat{T}}(t)=&\mathrm{E}[e^{\imath t\widehat{T}}]=\mathrm{E}\left[e^{\imath tT}\cdot e^{\imath t\left(-\frac{V}{\sqrt{b}\sqrt{2\delta_2 c_2}}-\frac{TW}{2b}+\mathrm{O}_p(b^{-3/2})\right)}\right]\\ =&\mathrm{E}\left[e^{\imath tT}\cdot\left\{1-\frac{\imath tV}{\sqrt{b}}-\frac{\imath tTW}{2b}+\frac{\imath^2 t^2 V^2}{b}+\mathrm{O}_p(b^{-3/2})\right\}\right]\\ =&\mathrm{E}[e^{\imath tT}]+\frac{\imath^2 t^2}{2b}\mathrm{E}[e^{\imath tT}]\mathrm{E}[V^2]+\mathrm{O}(b^{-3/2}),\end{aligned}$$

since T is independent of with both (V, W) and $\mathrm{E}[V] = \mathrm{E}[W] = 0$.

Finally, using Lemma 3.8.1 and the fact that

$$\mathrm{E}[V^2] = \frac{b}{2\delta_2 c_2}\mathrm{Var}(\widehat{c}_1) = \frac{b}{n\delta_2} = \frac{\gamma}{\delta_2}$$

(by the Proof of Theorem 3.3.1), we have the desired result. □

Proof of Theorem 3.3.1 and Theorem 3.3.5. We know that $n\boldsymbol{S} \sim \mathcal{W}_b(\boldsymbol{\Sigma}, n)$, where $\mathcal{W}_b(\boldsymbol{\Sigma}, n)$ stands for b-dimensional Wishart distribution with degrees of freedom n and scale matrix $\boldsymbol{\Sigma}$. Denote $a_k = \mathrm{tr}(\boldsymbol{T}_2\boldsymbol{S})^k$ for $k = 1, \ldots, 8$ and define

$$b_{\boldsymbol{T}_2\boldsymbol{\Sigma}} = (c_1^k, c_1^{k-2}c_2, \ldots, c_k)^\top \quad \text{and} \quad b_{\boldsymbol{T}_2\boldsymbol{S}} = (a_1^k, a_1^{k-2}a_2, \ldots, a_k)^\top$$

to be the vector of traces of the kth order moments. That means, for each partition of k, for example $k = \nu_1 + \cdots \nu_q$, where $\nu_1 \leq \cdots \leq \nu_q$ and $q \leq k$, we include $c_{\nu_1} \ldots c_{\nu_q}$ and $a_{\nu_1} \ldots a_{\nu_q}$ to the vectors $b_{\boldsymbol{T}_2\boldsymbol{\Sigma}}$ and $b_{\boldsymbol{T}_2\boldsymbol{S}}$, respectively, at same position. It is known (Fujikoshi, 1973) that $\mathrm{E}(b_{\boldsymbol{T}_2\boldsymbol{S}}) = \boldsymbol{F}_k b_{\boldsymbol{T}_2\boldsymbol{\Sigma}}$, where the matrices $\boldsymbol{F}_k$ have been calculated by Fujikoshi (1973) up to $k = 6$ and by Watamori (1990) for $k = 7, 8$. Using these it can be shown that

$$\mathrm{E}[\widehat{c}_k] = c_k, \text{for } k = 1, 2, 3, 4.$$

Further, under the high-dimensional asymptotic frameworks $B1$ and $B2$ and after lengthy algebra, it can be seen that

$$\begin{aligned}
\mathrm{Var}\Big(\frac{\widehat{c}_1}{\sqrt{c_2}}\Big) &= \frac{2}{n} = \mathrm{O}(n^{-1}), \\
\mathrm{Var}\Big(\frac{\widehat{c}_2}{c_2}\Big) &= \frac{4}{n(n-1)(n+2)c_2^2}\Big[nc_2^2 + (2n^2+3n-6)c_4\Big] = \mathrm{O}(n^{-2}), \\
\mathrm{Var}\Big(\frac{\sqrt{b}\widehat{c}_3}{c_2^{3/2}}\Big) &= \frac{6b}{nm_1c_2^3}\Big[n^2c_2^3 + 3n(n-1)(n+4)c_3^2 + 3n(n^2+3n-12)c_2c_4 + \\
&\quad (3n^4 + 15n^3 - 20n^2 - 120n + 160)c_6\Big] = \mathrm{O}(n^{-2}) \quad \text{and} \\
\mathrm{Var}\Big(\frac{b\widehat{c}_4}{c_2^2}\Big) &= \frac{8b^2}{nm_2c_2^4}\Big[f_1c_2^4 + f_2c_2c_3^2 + f_3c_2^2c_4 + f_4c_4^2 + f_5c_3c_5 + f_6c_2c_6 + f_7c_8\Big] \\
&= \mathrm{O}(n^{-2}),
\end{aligned}$$

where

$$
\begin{aligned}
&f_1 = n^2(n^2+n+2),\ \ f_2 = 8n^2(n+1)(n-3)(n+6),\\
&f_3 = 2n^2(2n^3+11n^2-47n+54),\\
&f_4 = n(6n^5+40n^4-85n^3-631n^2+726n+1224),\\
&f_5 = 8n(n+1)(n-3)(n+6)(n^2+4n-16),\\
&f_6 = 4n(n^5+10n^4-11n^3-220n^2+276n+480) \quad \text{and}\\
&f_7 = 2(2n^7+23n^6+38n^5-423n^4-992n^3+4066n^2-420n-5040).
\end{aligned}
$$

See also Srivastava (2005) and Hyodo et al. (2014). □

Proof of Theorem 3.4.1. Using similar techniques as in the proof of Lemma 3.8.1,

$$
\begin{aligned}
T_1 &= \frac{H - \operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})c_1 - \boldsymbol{\mu}^\top \boldsymbol{K}\boldsymbol{\mu}}{\sqrt{2\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})^2 c_2 + 4\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu}}}\\
&= \frac{\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}(\boldsymbol{Z}^\top \widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z} + 2\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z}) - c_1}{\sqrt{2\delta_2 c_2 + 4\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu}}}.
\end{aligned}
$$

Denote $\sigma_1^2 = 2\delta_2 c_2 + 4\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu}$. The characteristic function of T_1 is

$$
\begin{aligned}
\phi_{T_1}(t) =& \exp\big(-\frac{\imath t c_1}{\sigma_1}\big)\mathrm{E}\Big[\exp\big\{\frac{\imath t}{\sigma_1}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}(\boldsymbol{Z}^\top \widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z} + 2\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\big\}\Big]\\
=& \exp\big(-\frac{\imath t c_1}{\sigma_1}\big)|\boldsymbol{M_1}|^{-1/2}\exp\Big\{\frac{2\imath^2 t^2}{\sigma_1^2}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top \boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{M_1}^{-2}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\boldsymbol{\mu}\Big\},
\end{aligned}
$$

where $\boldsymbol{M}_1 = \boldsymbol{I} - \frac{2\imath t}{\sigma_1}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}$. Further,

$$
\log|\boldsymbol{M}_1|^{-1/2} = -\frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b}\log\left(1 - \frac{2\imath t}{\sigma_1}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\alpha_i\beta_j\right) = \frac{\imath t c_1}{\sigma_1} + \frac{\imath^2 t^2 \delta_2 c_2}{\sigma_1^2} + \mathrm{o}(1).
$$

Under assumption $B1, B2$, $B3$ or $B4$, We can prove that:

$$
\frac{1}{\sigma_1^2}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}'\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{M_1}^{-2}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\boldsymbol{\mu}\Big\} = \mathrm{o}(1),
$$

Since

$$
\boldsymbol{M_1}^{-1} = \boldsymbol{I} + \frac{2\imath t}{\sigma_1}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2} + \frac{4\imath^2 t^2}{\sigma_1^2}\{\operatorname{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2} + \cdots,
$$

$$\boldsymbol{M}_{\mathbf{1}}^{-2} = \boldsymbol{I} + \frac{4\imath t}{\sigma_1}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-1}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2} + \frac{12\imath^2t^2}{\sigma_1^2}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2} + \cdots .$$

Now, under either $B3$ or $B4$, we have

$$\frac{1}{\sigma_1^3}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-3}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu} \leq \frac{\sqrt{\delta_2 c_2}}{\sigma_1^3}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu} = \mathrm{o}(1),$$

and

$$\frac{1}{\sigma_1^4}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-4}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu} \leq \frac{\delta_2 c_2}{\sigma_1^4}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu} = \mathrm{o}(1).$$

Therefore,

$$\begin{aligned}&\exp\left\{\frac{2\imath^2t^2}{\sigma_1^2}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{M}_{\mathbf{1}}^{-2}\widetilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{K}\boldsymbol{\mu}\right\}\\ =&\exp\left\{\frac{2\imath^2t^2}{\sigma_1^2}\{\mathrm{tr}(\boldsymbol{T}_1\boldsymbol{D})\}^{-2}\boldsymbol{\mu}^\top\boldsymbol{K}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{K}\boldsymbol{\mu}\right\} + \mathrm{o}(1) = 1,\end{aligned}$$

and the characteristic function of T_1 is $\phi_{T_1}(t) = \exp(\frac{\imath^2t^2}{2}) + \mathrm{o}(1)$. □

Chapter 4 High-Dimensional Inference Under Non-normality

4.1 Introduction

In recent years, there is an increasing demand for effectively analyzing of high-dimensional data. In the bid to cope with this rising demand, comparison of high-dimensional mean vectors has received a renewed attention in the last two decades. High-dimensional means both the sample size and dimension are large but one could be much larger than the other without any restriction. Early theoretical attempts for analyzing high-dimensional data in the context of the sample size being larger than the dimension date back to the late fifties (Dempster, 1958, 1960).

For the sake of simplicity, we introduce the problem in the simplest case of comparing two mean vectors and extensions to more general cases will be outlined later. Consider two mutually independent random samples $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i} \in \mathbb{R}^p$ for $i = 1, 2$, which have means $\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1p})^\top$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \dots, \mu_{2p})^\top$ and positive definite covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Other than existence of the first two moments, no parametric structure is assumed among or within the mean vectors nor the covariances of the populations. Define the sample summary statistics as

$$\overline{\boldsymbol{X}}_i = \frac{1}{n_i}\sum_{i=1}^{n_i} \boldsymbol{X}_{ij} \quad \text{and} \quad \boldsymbol{S}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\overline{\boldsymbol{X}}_{ij} - \overline{\boldsymbol{X}}_i)(\overline{\boldsymbol{X}}_{ij} - \overline{\boldsymbol{X}}_i)^\top$$

for $i = 1, 2$.

Consider testing the following high dimensional hypotheses:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \quad \text{VS} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{4.1}$$

Bai and Saranadasa (1996) proposed high-dimensional test for (4.1) assuming $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. More recently, Chen and Qin (2010) proposed and studied the test statistic

$$\begin{aligned} T_{\mathrm{CQ}}(\boldsymbol{X}) &= (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)^\top(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - n_1^{-1}\mathrm{tr}(\boldsymbol{S}_1) - n_2^{-1}\mathrm{tr}(\boldsymbol{S}_2) \\ &= \frac{\sum_{i\neq j}^{n_1} \boldsymbol{X}_{1i}^\top \boldsymbol{X}_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} \boldsymbol{X}_{2i}^\top \boldsymbol{X}_{2j}}{n_2(n_2-1)} - \frac{2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \boldsymbol{X}_{1i}^\top \boldsymbol{X}_{2j}}{n_1 n_2}, \end{aligned}$$

(see also Hu et al. (2017) for multi-group test) weakening the equality as well as other regularity assumptions on the covariance matrices. The first two moments of the statistics T_{CQ} are

$$\mathrm{E}\{T_{\mathrm{CQ}}(\boldsymbol{X})\} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad \text{and} \quad \mathrm{Var}\{T_{\mathrm{CQ}}(\boldsymbol{X})\} = \sigma_n^2 + \sigma_{n2}^2,$$

where

$$\sigma_n^2 = \frac{2}{n_1(n_1-1)}\mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \frac{2}{n_2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_2^2) + \frac{4}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) \quad \text{and}$$
$$\sigma_{n2}^2 = 4n_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_1(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + 4n_2^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_2(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

Many papers (one, two or multiple groups) have investigated this test or its modified version (e.g., Chen et al., 2010; Wang et al., 2015; Ghosh and Biswas, 2016; Hu et al., 2017; Zhou et al., 2017). There are also other mean-based tests that assume weak dependence (e.g. Srivastava and Kubokawa, 2013; Cai et al., 2014; Cai and Xia, 2014; Feng et al., 2015; Gregory et al., 2015).

Chen and Qin (2010) (also, Bai and Saranadasa, 1996; Hu et al., 2017) required the following conditions.

- $C1$: For $i = 1$ or 2, assume $\boldsymbol{X}_{ij} = \boldsymbol{\Gamma}_i\boldsymbol{Z}_{ij} + \boldsymbol{\mu}_i$, for $j = 1, \ldots, n_i$, where $\boldsymbol{\Gamma}_i$ is a $p \times m$ matrix for some $m \geq p$ such that $\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i^\top = \boldsymbol{\Sigma}_i$ and $\boldsymbol{Z}_{ij}$ are m-variate identically and independently distributed random vectors.

- $C2$: The components of $\boldsymbol{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})^\top$ satisfy $\mathrm{E}(\boldsymbol{Z}_{ij}) = \boldsymbol{0}$, $\mathrm{Var}(\boldsymbol{Z}_{ij}) = \boldsymbol{I}_m$, and $\mathrm{E}[Z_{ijk}^4] = 3 + \triangle_i < \infty$ and

$$\mathrm{E}(Z_{ijl_1}^{\alpha_1} \cdots, Z_{ijl_q}^{\alpha_q}) = \mathrm{E}(Z_{ijl_1}^{\alpha_1}) \cdots \mathrm{E}(Z_{ijl_q}^{\alpha_q})$$

  for a positive integers q such that $1 \leq l_1 < l_2 < \cdots < l_q \leq p$ and $\sum_{i=1}^q \alpha_i \leq 8$.

- $C3$: The sample sizes diverge proportionally, i.e. $n_1/n \to \kappa \in (0, 1)$, where $n = n_1 + n_2$.

- $C4$: The covariance matrices satisfy the regularity condition

$$\mathrm{tr}(\boldsymbol{\Sigma}_{i_1}\boldsymbol{\Sigma}_{i_2}\boldsymbol{\Sigma}_{i_3}\boldsymbol{\Sigma}_{i_4}) = o\left[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}\right] \text{ for } i_1, i_2, i_3, i_4 \in \{1, 2\}.$$

$C5$: The mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ satisfy $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_i (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o\,[\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2]$.

Whereas there is essentially no restriction imposed by conditions $C1$, condition $C3$ is mild and reasonable. The local alternative condition $C5$ is automatically satisfied under the null hypothesis. There is, however, redundancy in condition $C4$. Indeed, we only need

$$\mathrm{tr}(\boldsymbol{\Sigma}_i^4) = o\left[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}\right], \text{ for } i = 1, 2,$$

while others can be derived using that. For example, from the result in Yang et al. (2001), it follows that

$$\mathrm{tr}(\boldsymbol{\Sigma}_1^2 \boldsymbol{\Sigma}_2^2) \le \{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)\mathrm{tr}(\boldsymbol{\Sigma}_2^4)\}^{1/2} = o[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}],$$

$$\mathrm{tr}\{(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)^2\} \le \{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)\mathrm{tr}(\boldsymbol{\Sigma}_2^4)\}^{1/2} = o[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}],$$

and that

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^3) &\le \left[\mathrm{tr}\{(\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{1/2})^2\}\mathrm{tr}(\boldsymbol{\Sigma}_2^4)\}\right]^{1/2} \\ &= \left[\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\}\mathrm{tr}(\boldsymbol{\Sigma}_2^4)\right]^{1/2} = o[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}].\end{aligned}$$

Condition $C2$ requires existence and factoring of mixed moments up to the eighth order. This requirement is unnecessarily strong. Indeed, this condition is close to the assumption of normality. For example, it does exclude *spherically-contoured models* for $\boldsymbol{Z}_{ij}$. Spherically-contoured distribution is a popular semi-parametric model that covers the multivariate normal as a special case (Fang and Zhang, 1990). It also covers models that are lighter and heaver tailed than the normal distribution. Examples include the multivariate t, multivariate Laplace and multivariate Logistic distributions, to mention a few. When $\boldsymbol{Z}_{ij}$ is spherically distributed, it can be shown, for example, that

$$\frac{E(Z_{ijk}^2 Z_{ijl}^2)}{E(Z_{ijk}^2)E(Z_{ijl}^2)} = \frac{cp^2}{p(p+2)}$$

for $k < l$ provided the expectations exist (see Fang and Zhang, 1990; Anderson, 2003) where c depends on the specific spherical distribution. For example, $c = p(p+1)/p^2$ for multivariate normal distribution and factoring the expectation $E(Z_{ijk}^2 Z_{ijl}^2) =$

$E(Z_{ijk}^2)E(Z_{ijl}^2)$ happens only for this value of c. It should also be noted that the lack of factoring can occur also for mixed moments of order six and eight.

The two-sample results of Chen and Qin (2010) were recently extended to the one-way MANOVA layout in Hu et al. (2017). The assumptions are essentially the same as $C1$–$C5$ except that the indices i, i_1, i_2, i_3, i_4 run from 1 to a where a is the number of samples. The test statistic considered is a formal extensions of T_{CQ} given by

$$\begin{aligned} T_{\text{HBWW}}(\boldsymbol{X}) &= \sum_{i<k}^{a}(\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{X}}_k)^\top(\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{X}}_k) - (a-1)\sum_{i=1}^{a} n_i^{-1}\text{tr}\,(\boldsymbol{S}_i) \\ &= \text{tr}\left(a\overline{\boldsymbol{X}}^\top \boldsymbol{P}_a \overline{\boldsymbol{X}} - (a-1)\sum_{i=1}^{a} n_i^{-1}\boldsymbol{S}_i\right) \end{aligned} \tag{4.2}$$

where $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_1, \ldots, \overline{\boldsymbol{X}}_a)^\top$ and $\boldsymbol{P}_a = \boldsymbol{I}_a - a^{-1}\mathbf{1}_a\mathbf{1}_a^\top$. The later form hints a formal extension of the statistic for a general factorial design where the appropriate projection matrix that target the hypothesis of interest should be used in place of $\boldsymbol{P}_a$. To see why this may work in the general case, denote

$$\boldsymbol{H} = \overline{\boldsymbol{X}}^\top \boldsymbol{P}_a \overline{\boldsymbol{X}} = \sum_{i=1}^{a}(\overline{\boldsymbol{X}}_i - \widetilde{\boldsymbol{X}})(\overline{\boldsymbol{X}}_i - \widetilde{\boldsymbol{X}})^\top \quad \text{and} \quad \boldsymbol{G} = (a-1)\sum_{i=1}^{a} n_i^{-1}\boldsymbol{S}_i$$

where $\widetilde{\boldsymbol{X}} = a^{-1}\sum_{i=1}^{a}\overline{\boldsymbol{X}}_i$. Here, $\boldsymbol{H}$ may be viewed as the between sum of squares and crossproducts matrix in MANOVA but using unweighted average for the overall mean (Harrar and Bathke, 2008). Similarly, $\boldsymbol{G}$ can be viewed as the within sum of squares and crossproducts matrix. It is easy to verify that

$$E(\boldsymbol{H}) - E(\boldsymbol{G}) = \mathbf{0} \quad \text{if and only if} \quad \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_a.$$

Therefore $T_{\text{HBWW}}(\boldsymbol{X})$ defines a reasonable test statistic.

The aim of this Chapter is to broaden the scope of applicability of pertinent high-dimensional tests for mean vectors by replacing stringent assumptions with realistic ones. This allows, for example, application of the theory for rank-based methods (Kong and Harrar, 2018b, as given in Chapter 5), where the assumption of existence of higher order moments are not needed. To that end, this Chapter is organized as follows. Some preliminary results on the order of quadratic form of high dimensional random vectors are presented in Section 4.2. This section also states general regularity conditions on the dependence among the multiple variables and points out some realistic scenarios that lead to the assumed dependence models. Section 4.3 deals with asymptotic distributions of the test statistics under the null as well as local alternatives. Also considered in Section 4.3 are various consistent estimators of the asymptotic variance. Section 4.4 provides some details on the multi-group extension. The numerical accuracy of the results are investigated in Section 4.5 with simulation studies that mimic realistic data generation mechanisms. The methods are applied to an Electroencephalograph (EEG) dataset in Section 4.6. Discussions and concluding remarks will be the provided in Section 4.7. All proofs and technical details are placed in the Appendix.

4.2 Model for Dependence

In this section, we present model on the dependence of the multivariate data that improves condition $C2$ in two important ways. First, the assumption of factoring of mixed moments up to the eighth order are removed and only the fourth order mixed moments are regulated without factoring requirement. Our condition is significantly milder than $C2$ in that it is satisfied by popular multivariate distributions (Elliptically-Contoured) and fairly weak but realistic model for dependence such as α-mixing (strong-mixing). Another significant improvement pertains to making the regularity condition on the original variables rather than on the normalized versions ($\boldsymbol{Z}_{ij}$). The significance of this improvement is that the model in $C1$ is not quite natural for common type of dependence conditions (e.g., mixing condition) or they are not convenient for rank-based applications (Kong and Harrar, 2018b) because original observations are ranked rather than their normalized versions. To facilitate ease of presentation, in this section we drop the subscripts (i and j) that identify the sample and the subject to which the vectors belong.

$C6$: Suppose $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$ be a centered p-variate random vector. Let $\{\varphi_k\}_{k=1}^{\infty}$ be a non-increasing sequence of nonnegative number, such that, for all $k_1 < k_2 <$

$k_3 < k_4$,

$$|\mathrm{Cov}(Z_{k_1}, Z_{k_2}Z_{k_3}Z_{k_4})| \leq \varphi_{k_2-k_1},\ |\mathrm{Cov}(Z_{k_1}Z_{k_2}Z_{k_3}, Z_{k_4})| \leq \varphi_{k_4-k_3}, \\ |\mathrm{Cov}(Z_{k_1}Z_{k_2}, Z_{k_3}Z_{k_4})| \leq \varphi_{k_3-k_2},\ \text{and } |\mathrm{Cov}(Z_{k_1}, Z_{k_2})| \leq \varphi_{k_2-k_1}. \tag{4.3}$$

Also, let $\{\phi_k\}_{k=1}^{\infty}$ be a sequence of nonnegative number, such that, for all $k_1 < k_2$

$$\mathrm{Cov}(Z_{k_1}^2, Z_{k_2}^2) \leq \phi_{k_2-k_1}. \tag{4.4}$$

$C7$: Assume Φ_0, Φ_1 and Φ_2 are bounded where

$$\Phi_0 = \sup_k\{\mathrm{E}[Z_k^4]\}, \quad \Phi_1 = \sum_{k=1}^{\infty} k\varphi_k, \quad \text{and} \quad \Phi_2 = \sum_{k=1}^{\infty} \phi_k.$$

For simplicity, we can further assume that $E(Z_i^2) \leq 1$, as this can always be achieved by rescaling the variables. These assumptions deal with covariances of Z_i's products up to the fourth order only and are closely related to the classical fourth-order cumulant condition for a stationary time series (see Theorem V.4 in Hannan (1970) and Assumption A in Andrews (1991)). Clearly, condition $C2$ implies condition $C6$ and $C7$, with $\varphi_k = \phi_k = 0$ for $k = 1, 2, \ldots$. We give two examples in the remarks below to illustrate that conditions $C6$ and $C7$ taken together are much milder than condition $C2$.

Remark 1. *Suppose $\boldsymbol{Z}$ has spherical distribution with finite fourth moment (see, for example Fang and Zhang, 1990). Conditions $C6$ and $C7$ hold automatically by the symmetry of the distribution with $\varphi_k = 0$ and proper ϕ_k since*

$$\mathrm{Cov}(Z_{k_1}^2, Z_{k_2}^2) = O(p^{-2}).$$

Remark 2. *Suppose the component random variables in $\boldsymbol{Z}$ with zero mean, and bounded moments of order 4δ for some $\delta > 1$, constitute an α-mixing sequence with mixing coefficients $\{\alpha_k, k = 1, 2, \ldots\}$, as p tends to infinity, that means,*

$$\sup_{A\in\mathcal{A}_l, B\in\mathcal{B}_{l,k}\in\mathbb{Z}^+} |P(A\cap B) - P(A)P(B)| \leq \alpha_k \ \text{as } k \to \infty,$$

where

$$\mathcal{A}_l = \sigma\{Z_1, \ldots, Z_l\}, \quad \mathcal{B}_{l,k} = \sigma\{Z_{l+k}, Z_{l+k+1}, \ldots\}$$

and $\sigma(\cdot)$ denotes the σ-field generated by the random variables. This model for dependence is particularly attractive for repeated measures data. In this case, α-mixing condition basically requires the dependence between observations from the same subject to decay as the separation between the observations increases. With the coefficients α_k (nonincreasing) condition $C6$ holds for $\varphi_k = \phi_k = D\alpha_k^{(\delta-1)/\delta}$ and large enough $D > 0$ (see, for example, Corollary A.2 in Hall and Heyde, 1980; Yaskov, 2015). Condition $C7$ is satisfied for some α_k, for example $\alpha_k = O(k^{-5})$ when $\delta > 5/4$.

Conditions $C6$ and $C7$ afford us an inequality on the variance of quadratic forms which was established in Yaskov (2015).

Theorem 4.2.1. *(Theorem 2.2 of Yaskov, 2015) Under condition $C6$, there is a universal constant $C > 0$ such that for all $p \times p$ matrices $\boldsymbol{A}$,*

$$\mathrm{Var}(\boldsymbol{Z}^\top \boldsymbol{A}\boldsymbol{Z}) \leq C(\Phi_0 + \Phi_1 + \Phi_2)\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^\top).$$

When the component random variables in $\boldsymbol{Z}$ are uncorrelated, it can easily be verified that condition (4.3) in $C6$ can be reduced to

$$|\mathrm{E}[Z_{k_1}Z_{k_2}Z_{k_3}Z_{k_4}]| \leq \min\{\varphi_{k_2-k_1}, \varphi_{k_3-k_2}, \varphi_{k_4-k_3}\}, \quad k_1 < k_2 < k_3 < k_4. \tag{4.5}$$

A version of the inequality in Theorem 4.2.1, which is convenient in light of condition $C1$ together with condition $C6$ is as follows.

Corollary 4.2.2. *Let $\boldsymbol{Z}$ be centered orthonormal random variables satisfy condition $C6$. There is a universal constant $C > 0$ such that for all $p \times p$ matrices $\boldsymbol{A}$,*

$$\mathrm{Var}(\boldsymbol{Z}^\top \boldsymbol{\Gamma}^\top \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{Z}) \leq C(\Phi_0 + \Phi_1 + \Phi_2)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top), \text{ for } i = 1, 2,$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$.

4.3 Main Results

4.3.1 Asymptotic Results

For the two-sample testing problem, it was proved in Chen and Qin (2010) that under conditions $C1$–$C5$,

$$\frac{T_{\text{CQ}}(\boldsymbol{X}) - ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{\sigma_n} \xrightarrow{D} \mathcal{N}(0,1), \text{ as } p, n \to \infty. \tag{4.6}$$

In order to formulate a test procedure based on (4.6), σ_n^2 needs to be consistently estimated. A few unbiased and ratio-consistent estimators of $\text{tr}(\boldsymbol{\Sigma}_i^2)$ and $\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$, and hence of σ_n^2 are available in the literature (Bai and Saranadasa, 1996; Chen and Qin, 2010; Chen et al., 2010; Li and Chen, 2012). While the estimator of Bai and Saranadasa (1996) is designed for $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ case, it has the advantage that it is uniformly minimum variance unbiased under normality and is easy to compute. The other estimators (Chen and Qin, 2010; Li and Chen, 2012) are asymptotically equivalent and designed for the unequal covariance case.

The estimators of Chen and Qin (2010) are

$$\widetilde{\text{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{1}{n_i(n_i-1)}\text{tr}\Big\{\sum_{k\neq l}^{n_i}(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i(k,l)})\boldsymbol{X}_{ik}^\top(\boldsymbol{X}_{il} - \overline{\boldsymbol{X}}_{i(k,l)})\boldsymbol{X}_{il}^\top\Big\} \tag{4.7}$$

and

$$\widetilde{\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \frac{1}{n_1n_2}\text{tr}\Big\{\sum_{k=1}^{n_1}\sum_{l=1}^{n_2}(\boldsymbol{X}_{1k} - \overline{\boldsymbol{X}}_{1(k)})\boldsymbol{X}_{1k}^\top(\boldsymbol{X}_{2l} - \overline{\boldsymbol{X}}_{2(l)})\boldsymbol{X}_{2l}^\top\Big\} \tag{4.8}$$

where $\overline{\boldsymbol{X}}_{i(k,l)}$ is the ith sample mean after excluding $\boldsymbol{X}_{ik}$ and $\boldsymbol{X}_{il}$ and $\overline{\boldsymbol{X}}_{i(k)}$ is the ith sample mean after excluding $\boldsymbol{X}_{ik}$. Under conditions $C1$–$C5$,

$$\frac{\widetilde{\text{tr}(\boldsymbol{\Sigma}_i^2)}}{\text{tr}(\boldsymbol{\Sigma}_i^2)} \xrightarrow{P} 1 \quad \text{and} \quad \frac{\widetilde{\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}}{\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} \xrightarrow{P} 1, \text{ as } n, p \to \infty. \tag{4.9}$$

Therefore, a ratio-consistent estimator of σ_n^2 under H_0 is

$$\widehat{\sigma}_n^2 = \frac{2}{n_1(n_1-1)}\widetilde{\text{tr}(\boldsymbol{\Sigma}_1^2)} + \frac{2}{n_2(n_2-1)}\widetilde{\text{tr}(\boldsymbol{\Sigma}_2^2)} + \frac{4}{n_1n_2}\widetilde{\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}.$$

The estimators in Li and Chen (2012) (see also, Chen et al., 2010, for $a = 1$) also satisfy (4.9) under $C1$–$C5$. These estimators can be conveniently expressed as

U-statistics:

$$\widehat{\operatorname{tr}(\mathbf{\Sigma}_i^2)} = \frac{1}{(n_i)_4} \sum_{k_1 \neq k_2 \neq l_1 \neq l_2}^{n_i} \operatorname{tr}\Big\{(\boldsymbol{X}_{ik_1} - \boldsymbol{X}_{ik_2})(\boldsymbol{X}_{ik_1} - \boldsymbol{X}_{ik_2})^\top (\boldsymbol{X}_{il_1} - \boldsymbol{X}_{il_2})(\boldsymbol{X}_{il_1} - \boldsymbol{X}_{il_2})^\top\Big\}, \tag{4.10}$$

and

$$\operatorname{tr}(\widehat{\mathbf{\Sigma}_1\mathbf{\Sigma}_2}) = \frac{1}{(n_1)_2(n_2)_2} \sum_{k_1 \neq k_2}^{n_1} \sum_{l_1 \neq l_2}^{n_2} \operatorname{tr}\Big\{(\boldsymbol{X}_{1k_1} - \boldsymbol{X}_{1k_2})(\boldsymbol{X}_{1k_1} - \boldsymbol{X}_{1k_2})^\top (\boldsymbol{X}_{2l_1} - \boldsymbol{X}_{2l_2})(\boldsymbol{X}_{2l_1} - \boldsymbol{X}_{2l_2})^\top\Big\}, \tag{4.11}$$

where $(n_i)_k = n_i!/(n_i - k)!$.

In the following theorems, we establish that (4.6) and (4.9) hold when assumption $C2$ is replaced by the weaker assumptions $C6$ and $C7$.

Theorem 4.3.1. *Under conditions $C1$, $C3$– $C7$, the asymptotic normality result (4.6) holds.*

Theorem 4.3.2. *Under the conditions $C1$, $C3$–$C7$, the consistency result (4.9) holds for the estimators defined from (4.7) and (4.8) or for (4.10) and (4.11).*

4.3.2 Computational Formulae for the Ratio Consistent Estimators

For ease of proving Theorem 4.3.2, we can rewrite $\widehat{\operatorname{tr}(\mathbf{\Sigma}_i^2)}$ and $\operatorname{tr}(\widehat{\mathbf{\Sigma}_1\mathbf{\Sigma}_2})$ as follows:

$$\begin{aligned}\widehat{\operatorname{tr}(\mathbf{\Sigma}_i^2)} =& \frac{1}{(n_i)_2} \sum_{k_1 \neq k_2} (\boldsymbol{X}_{ik_1}^\top \boldsymbol{X}_{ik_2})^2 - \frac{2}{(n_i)_3} \sum_{k_1 \neq k_2 \neq k_3} (\boldsymbol{X}_{ik_1}^\top \boldsymbol{X}_{ik_2})(\boldsymbol{X}_{ik_1}^\top \boldsymbol{X}_{ik_3}) \\ &+ \frac{1}{(n_i)_4} \sum_{k_1 \neq k_2 \neq k_3 \neq k_4} (\boldsymbol{X}_{ik_1}^\top \boldsymbol{X}_{ik_2})(\boldsymbol{X}_{ik_3}^\top \boldsymbol{X}_{ik_4}).\end{aligned}$$

$$\begin{aligned}\operatorname{tr}(\widehat{\mathbf{\Sigma}_1\mathbf{\Sigma}_2}) =& \frac{1}{n_1 n_2} \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} (\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_2})^2 - \frac{1}{(n_1)(n_2)_2} \sum_{k_1=1}^{n_1} \sum_{k_2 \neq k_3}^{n_2} (\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_2})(\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_3}) \\ &- \frac{1}{(n_1)_2(n_2)} \sum_{k_1 \neq k_2}^{n_2} \sum_{k_3=1}^{n_1} (\boldsymbol{X}_{2k_3}^\top \boldsymbol{X}_{1k_1})(\boldsymbol{X}_{2k_3}^\top \boldsymbol{X}_{1k_2}) \\ &+ \frac{1}{(n_1)_2(n_2)_2} \sum_{k_1 \neq k_2}^{n_1} \sum_{k_3 \neq k_4}^{n_2} (\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_3})(\boldsymbol{X}_{1k_2}^\top \boldsymbol{X}_{2k_4}).\end{aligned}$$

These forms of the estimators have also been used in elsewhere (Chen et al., 2010; Li and Chen, 2012; Zhang et al., 2018, etc...). From Hu et al. (2017) and Zhang et al. (2018), the estimators can be further rewritten as

$$\widehat{\text{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{(n_i-1)_2}{(n_i)_4}|||\boldsymbol{\Theta}_i|||_2^2 - \frac{2(n_i-1)}{(n_i)_4}|||\boldsymbol{\Theta}_i|||_{1,2}^2 + \frac{1}{(n_i)_4}|||\boldsymbol{\Theta}_i|||_1^2 \tag{4.12}$$

and

$$\widehat{\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \text{tr}(\boldsymbol{S}_1\boldsymbol{S}_2),$$

where $\boldsymbol{\Theta}_i = \boldsymbol{X}_i^\top\boldsymbol{X}_i - \text{diag}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)$, and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i})$ be a $p \times n_1$ matrix. For any matrix $\boldsymbol{A} = (a_{ij})_{m\times n}$, we denote

$$|||\boldsymbol{A}|||_q = \Big\{\sum_{i=1}^m\sum_{j=1}^n(a_{ij})^q\Big\}^{1/q}$$

and

$$|||\boldsymbol{A}|||_{p,q} = \Big[\sum_{i=1}^m\Big\{\sum_{j=1}^n(a_{ij})^q\Big\}^{q/p}\Big]^{1/q}.$$

Note that if all $a_{ij} \geq 0$ or for even number q, these are entrywise norm and $L_{p,q}$ norm of $\boldsymbol{A}$.

Another expression of the estimator (4.10) of $\text{tr}(\boldsymbol{\Sigma}_i^2)$ was given in Himeno and Yamada (2014) as,

$$\widehat{\text{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{n_1-1}{n_1(n_i-2)(n_i-3)}\{(n_i-1)(n_i-2)\text{tr}(\boldsymbol{S}_i^2) + \text{tr}^2(\boldsymbol{S}) - n_iQ\},$$

where $Q = (n_i-1)^{-1}\sum_{i=1}^{n_i}||\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i||^4$.

Using the same method, we also give the simple form of estimators (4.7-4.8) in Chen and Qin (2010), which can be rewritten as:

$$\begin{aligned}\widetilde{\text{tr}(\boldsymbol{\Sigma}_i^2)} =& \frac{1}{(n_i)_2}\sum_{k_1\neq k_2}(\boldsymbol{X}_{ik_1}^\top\boldsymbol{X}_{ik_2})^2 - \frac{2n_i-5}{(n_i)(n_i-1)(n_i-2)^2}\sum_{k_1\neq k_2\neq k_3}(\boldsymbol{X}_{ik_1}^\top\boldsymbol{X}_{ik_2})(\boldsymbol{X}_{ik_1}^\top\boldsymbol{X}_{ik_3})\\ &+ \frac{1}{(n_i)(n_i-1)(n_i-2)^2}\sum_{k_1\neq k_2\neq k_3\neq k_4}(\boldsymbol{X}_{ik_1}^\top\boldsymbol{X}_{ik_2})(\boldsymbol{X}_{ik_3}^\top\boldsymbol{X}_{ik_4})\end{aligned}$$

and

$$\widetilde{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \frac{1}{n_1 n_2}\sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2}(\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_2})^2 - \frac{1}{(n_1)(n_2)_2}\sum_{k_1=1}^{n_1}\sum_{k_2\neq k_3}^{n_2}(\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_2})(\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_3})$$
$$- \frac{1}{(n_1)_2(n_2)}\sum_{k_1\neq k_2}^{n_2}\sum_{k_3=1}^{n_1}(\boldsymbol{X}_{2k_3}^\top \boldsymbol{X}_{1k_1})(\boldsymbol{X}_{2k_3}^\top \boldsymbol{X}_{1k_2})$$
$$+ \frac{1}{(n_1)_2(n_2)_2}\sum_{k_1\neq k_2}^{n_1}\sum_{k_3\neq k_4}^{n_2}(\boldsymbol{X}_{1k_1}^\top \boldsymbol{X}_{2k_3})(\boldsymbol{X}_{1k_2}^\top \boldsymbol{X}_{2k_4}).$$

After much simplification and rearrangement, these estimators have simplified forms:

$$\widetilde{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{(n_i-1)}{n_i(n_i-2)^2}|||\boldsymbol{\Theta}_i|||_2^2 - \frac{2n_i-1}{(n_i)_3(n_i-2)}|||\boldsymbol{\Theta}_i|||_{1,2}^2 + \frac{1}{(n_i)_3(n_i-2)}|||\boldsymbol{\Theta}_i|||_1^2. \quad (4.13)$$

and

$$\widetilde{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \mathrm{tr}(\boldsymbol{S}_1\boldsymbol{S}_2),$$

respectively. Examining (4.12) and (4.13), it is easy to see that the simulation computation will be substantially improved by using the rewritten form for both the estimators of Chen and Qin (2010) and Li and Chen (2012). The calculations for the original forms cost $O(n_i^4)$, but for the simplified forms cost only $O(n_i)$. Comparing (4.12) and (4.13), the simplified forms of the two estimators of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ share the same leading order term which was also noted by Li and Chen (2012), and the estimators of $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$ are the the same.

4.3.3 Other Conditions

4.3.3.1 Assumptions on Original Observations

In Theorem 4.3.1 and Theorem 4.3.2, the dependence conditions $C6$ and $C7$ are assumed on $\boldsymbol{Z}_{ij}$, which is defined in condition $C1$. For some type of dependence (such as α-mixing) or for some applications (Kong and Harrar, 2018b), making assumptions on $\boldsymbol{Z}_{ij}$ may not be realistic. For these situations we require two new conditions.

$C8$: The covariance matrix $\boldsymbol{\Sigma}_i$, $i = 1, 2$, satisfies $\mathrm{tr}\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2)\} \to \infty$ as $p \to \infty$.

$C9$: $p/n \to \eta \in (0, \infty)$ as $n, p \to \infty$.

Condition $C8$ is rather mild. We know that $\mathrm{tr}(\mathbf{\Sigma}^2) = \sum_{i=1}^{n} \lambda_i^2$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{\Sigma}$. Then, condition $C8$ holds, if infinite number of the eigenvalues remain bounded away from zero as $p \to \infty$. If condition $C8$ holds, it is easy to verify that

$$n^2\sigma_n^2 \geq \mathrm{tr}(\mathbf{\Sigma}_1^2) + \mathrm{tr}(\mathbf{\Sigma}_2^2) + \mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2) = \mathrm{tr}\{(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^2)\} \to \infty.$$

Condition $C9$, which stipulates the same rate of growth for n and b, is not new (e.g., Bai and Saranadasa, 1996).

The following theorems state asymptotic results for $T_{CQ}(\boldsymbol{X})$ by using conditions $C8$ and $C9$ instead of $C1$. More precisely, the dependence is assumed only on the centered original variables, i.e. we assume conditions $C6$–$C7$ on $\boldsymbol{Z}_{ij}$ where $\boldsymbol{X}_{ij} = \boldsymbol{Z}_{ij} + \boldsymbol{\mu}_i$.

Theorem 4.3.3. *Under conditions $C3$–$C9$, the asymptotic normality result (4.6) holds.*

Theorem 4.3.4. *Under conditions $C3$–$C9$, the consistency result (4.9) holds for the estimators defined from (4.7) and (4.8) or (4.10) and (4.11).*

Note that Theorems 4.3.3 and 4.3.4 can be directly applied by assuming the sequences $\{X_{ij1}, X_{ij2}, \ldots\}$ to be α-mixing sequences for all i and j with some dependence coefficients α_k such that conditions $C6$ and $C7$ are satisfied.

4.3.3.2 Assumptions on Quadratic Forms

Throughout the proofs of Theorem 4.3.1 and Theorem 4.3.2, we note that Corollary 4.2.2 plays a crucial rule. Apparently, condition $C10$ is sufficient to prove Theorem 4.3.1 and Theorem 4.3.2 instead of the conditions $C1$, $C6$ and $C7$.

$C10$: There are universal positive constants C and D, such that for all $p \times p$ symmetric real matrix $\boldsymbol{A}$,

$$\mathrm{Var}(\boldsymbol{X}_{ij}^\top \boldsymbol{A}\boldsymbol{X}_{ij}) \leq C\mathrm{tr}\{(\boldsymbol{A}\mathbf{\Sigma}_i)^2\} + D\mathrm{tr}^2(\boldsymbol{A}\mathbf{\Sigma}_i).$$

Conditions $C1$ and $C2$ imply $C10$, since (see Auxiliary results in Zhang et al., 2018)

$$\mathrm{Var}(\boldsymbol{X}_{ij}^{\top}\boldsymbol{A}\boldsymbol{X}_{ij}) = 2\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Sigma}_i)^2\} + \triangle_i ||\mathrm{diag}(\boldsymbol{\Gamma}_i\boldsymbol{A}\boldsymbol{\Gamma}_i)||^2 \leq (2+\triangle_i)\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Sigma}_i)^2\}.$$

The beauty of this condition, besides being weaker than $C2$, is that it doesn't regulate the dependence based on the separation between the variables within the observation vectors. For example, Elliptically-Contoured populations satisfy condition $C10$, since (Mathai et al., 1995)

$$\mathrm{Var}(\boldsymbol{X}_{ij}^{\top}\boldsymbol{A}\boldsymbol{X}_{ij}) = 2(\kappa_i+1)\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Sigma}_i)^2\} + \kappa_i \mathrm{tr}^2(\boldsymbol{A}\boldsymbol{\Sigma}_i),$$

for $\kappa_i < \infty$, where $\kappa_i = p^{-1}(p+2)^{-1}\mathrm{E}(\boldsymbol{X}_{ij}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{ij})^2 - 1$.

Theorem 4.3.5. *Under conditions $C3$– $C5$ and $C10$, the asymptotic normality result (4.6) holds.*

Theorem 4.3.6. *Under conditions $C3$– $C5$ and $C10$, the consistency result (4.9) holds for the estimators defined from (4.7) and (4.8) or from (4.10) and (4.11).*

4.3.4 Test and Asymptotic Power

Theorems 4.3.1 (4.3.3, 4.3.5) and Theorem 4.3.2 (4.3.4, 4.3.6) lead to the test statistic

$$Q_n = T_{CQ}(\boldsymbol{X})/\widehat{\sigma}_n \xrightarrow{D} \mathcal{N}(0,1), \text{ as } n, p \to \infty,$$

under H_0 and assumptions $C1$, $C3$–$C7$ (or $C3$–$C9$, or $C3$–$C5$ and $C10$), where a ratio-consistent estimator of σ_n^2 is defined to be

$$\widehat{\sigma}_n^2 = \frac{2}{n_1(n_1-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)} + \frac{4}{n_1 n_2}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}, \tag{4.14}$$

by the estimators $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}$ defined from (4.7) and (4.8) or (4.10) and (4.11). Our proposed test with an α level of significance rejects H_0 if $Q_n > \xi_\alpha$, where ξ_α is the upper α quantile of standard normal distribution. Also Theorems 4.3.1 (4.3.3, 4.3.5) and Theorems 4.3.2 (4.3.4, 4.3.6) allow us to discuss the power properties of the proposed test under assumption $C5$. The power under the local alternative $C5$ is

$$\beta_{n1}(||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2) = \Phi\left(-\xi_\alpha + \frac{n\kappa(1-\kappa)||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{\sqrt{2\text{tr}\{(1-\kappa)\boldsymbol{\Sigma}_1 + \kappa\boldsymbol{\Sigma}_2)\}^2}}\right),$$

where Φ is the cumulative function of standard normal distribution. This indicates that the proposed test has nontrivial power under the alternative hypothesis under assumption $C5$ as long as

$$\frac{n\kappa(1-\kappa)||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{\sqrt{2\text{tr}\{(1-\kappa)\boldsymbol{\Sigma}_1 + \kappa\boldsymbol{\Sigma}_2\}^2}}$$

does not vanish to 0 as $n, p \to \infty$.

4.4 Extensions

4.4.1 Extension to Multi-Group Equality Test

Suppose there are $a(> 2)$ groups and, for $i = 1, \ldots, a$, let the ith sample $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$ be iid with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. The test statistic T_{CQ} was generalized to multiple groups by Hu et al. (2017) for testing the hypotheses:

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_a, \quad \text{VS} \quad H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i_1} \text{ for some } i \neq i_1. \tag{4.15}$$

In order to generalize the asymptotic results in Section 4.3 to the multi-group case, we first reformulate the assumptions by making the necessary notational modifications.

$C1'$: For $i = 1, \ldots, a$, assume $\boldsymbol{X}_{ij} = \boldsymbol{\Gamma}_i\boldsymbol{Z}_{ij} + \boldsymbol{\mu}_i$, for $j = 1, \ldots, n_i$, where $\boldsymbol{\Gamma}_i$ is a $p \times m$ matrix for some $m \geq p$ such that $\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i^\top = \boldsymbol{\Sigma}_i$ and $\boldsymbol{Z}_{ij}$ are m-variate identically and independently distributed random vectors.

$C3'$: The sample sizes diverge proportionally, i.e. $n_i/n \to \kappa_i \in (0, 1)$ for $i = 1, \ldots, a$ where $n = n_1 + \cdots + n_a$.

$C4'$: The covariance matrices satisfy the regularity condition

$$\text{tr}(\boldsymbol{\Sigma}_{i_1}\boldsymbol{\Sigma}_{i_2}\boldsymbol{\Sigma}_{i_3}\boldsymbol{\Sigma}_{i_4}) = o\left[\text{tr}^2\{(\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_a)^2\}\right]$$

for $i_1, i_2, i_3, i_4 \in \{1, \ldots, a\}$.

$C5'$: For any $i, i_1, i_2 \in \{1, \ldots, a\}$, $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1})^\top \boldsymbol{\Sigma}_{i_2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1}) = o\left[\mathrm{tr}(\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_a)^2\right]$,

$C8'$: The covariance matrix $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, a$, satisfies $\mathrm{tr}\{(\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_a)^2)\} \to \infty$ as $p \to \infty$.

The test statistic in (4.2) can be expressed as (Hu et al., 2017),

$$T_{\mathrm{HBWW}}(\boldsymbol{X}) = (a-1)\sum_{i=1}^{a} \frac{\sum_{k \neq k_1}^{n_i} \boldsymbol{X}_{ik}^\top \boldsymbol{X}_{ik_1}}{n_i(n_i - 1)} - 2\sum_{i<i_1}^{a} \frac{\sum_{k=1}^{n_i} \sum_{k_1=1}^{n_{i_1}} \boldsymbol{X}_{ik}^\top \boldsymbol{X}_{i_1 k_1}}{n_i n_{i_1}}.$$

Note that when $a = 2$, this statistic reduces to T_{CQ}. The mean and variance of T_{HBWW} are

$$E\{T_{\mathrm{HBWW}}(\boldsymbol{X})\} = \sum_{i<i_1}^{a} ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1}||^2, \quad \text{and} \quad \mathrm{Var}\{T_{\mathrm{HBWW}}(\boldsymbol{X})\} = \sigma_{na}^2 + \sigma_{na2}^2,$$

where

$$\sigma_{na}^2 = \sum_{i=1}^{a} \frac{2(a-1)^2}{n_i(n_i-1)} \mathrm{tr}(\boldsymbol{\Sigma}_i^2) + \sum_{i<i_1}^{a} \frac{4}{n_i n_{i_1}} \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i_1}) \quad \text{and}$$
$$\sigma_{na2}^2 = 4\sum_{i=1}^{a} n_i^{-1} \Big(\sum_{i_1=1}^{a} \boldsymbol{\mu}_{i_1} - a\boldsymbol{\mu}_i\Big)^\top \boldsymbol{\Sigma}_i \Big(\sum_{i_1=1}^{a} \boldsymbol{\mu}_{i_1} - a\boldsymbol{\mu}_i\Big).$$

Using the same technique as Theorem 4.3.1, we can prove the following theorem.

Theorem 4.4.1. *Under the assumption $C1'$, $C3'$–$C5'$ $C6$ and $C7$,*

$$\frac{T_{HBWW}(\boldsymbol{X}) - \sum_{i<i_1}^{a} ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1}||^2}{\sigma_{na}} \xrightarrow{D} \mathcal{N}(0,1), \text{ as } p, n \to \infty.$$

The asymptotic normality in Theorem 4.4.1 can also be proved under the conditions $C3'$–$C5'$ $C6, C7, C8'$ and $C9$, where $C6$ and $C7$ are made on $\boldsymbol{Z}_{ij} = \boldsymbol{X}_{ij} - \boldsymbol{\mu}_i$, or under conditions $C3'$–$C5'$ and $C10$.

In multiple groups case, we can similarly construct the ratio-consistent estimator of σ_{na}^2 as follows:

$$\widehat{\sigma}_{na}^2 = \sum_{i=1}^{a} \frac{2(a-1)^2}{n_i(n_i-1)} \widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} + \sum_{i<i_1}^{a} \frac{4}{n_i n_{i_1}} \widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i_1})},$$

where $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i_1})}$ are defined in (4.7) and (4.8) or as in (4.10) and (4.11) for $i \neq i_1 \in \{1, \ldots, a\}$. The proofs are exactly the same. These lead to the test statistic

$$Q_{na}(\boldsymbol{X}) = T_{\text{HBWW}}(\boldsymbol{X})/\widehat{\sigma}_{na} \xrightarrow{D} \mathcal{N}(0,1), \text{ as } n, p \to \infty, \tag{4.16}$$

under assumptions $C1'$, $C3'$–$C5'$ $C6$ and $C7$ (or under $C3'$–$C5'$ $C6, C7, C8'$ and $C9$, or under $C3'$–$C5'$ and $C10$). Our proposed test for an α level of significance rejects H_0 if $Q_{na} > \xi_\alpha$.

The power of the proposed test for the multi-group case can also be derived under assumption $C5'$. From the above discussions, under the local alternative $C5'$, the power function is

$$\beta_{n1}(\sum_{i<i_1}^{a} ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1}||^2) = \Phi\left(-\xi_\alpha + \frac{\sum_{i<i_1}^{a} ||\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i_1}||^2}{\sigma_{na}}\right) + o(1).$$

4.4.2 Extension to Multi-Group Parallelism Test

In related works, Harrar and Kong (2016) and Hyodo (2017) considered comparison of mean profiles in multiple groups for normal and Elliptical populations, respectively, under high-dimensional frameworks. The test statistic investigated in this Chapter can also be manipulated for use in testing parallelism of the mean profiles in different groups. The hypothesis of parallelism is

$$H_0^{(P)}: \forall i, \boldsymbol{\mu}_i - \boldsymbol{\mu}_a = \gamma_i \mathbf{1}_p \text{ for } \gamma_i \in \mathbb{R}, \text{ VS } H_1^{(P)}: \exists\, i, \boldsymbol{\mu}_i - \boldsymbol{\mu}_a \neq \gamma_i \mathbf{1}_p, \forall \gamma_i \in \mathbb{R}. \tag{4.17}$$

To deal with that parallelism test, we can transform the random vector first in such a way that the parallelism hypothesis reduces to equality of mean vectors for the transformed random vectors. The parallelism hypothesis can equivalently be stated in terms of equality of $\boldsymbol{P}\boldsymbol{\mu}_i$, where $\boldsymbol{P} = \boldsymbol{I}_p - \boldsymbol{J}_p/p$ (e.g., Harrar and Kong, 2016; Hyodo, 2017),

$$H_0: \boldsymbol{P}\boldsymbol{\mu}_1 = \cdots = \boldsymbol{P}\boldsymbol{\mu}_a, \text{ VS } H_1: \exists\, i \neq i_1, \boldsymbol{P}\boldsymbol{\mu}_i \neq \boldsymbol{P}\boldsymbol{\mu}_{i_1}. \tag{4.18}$$

Therefore, we can transform the data by setting $\boldsymbol{X}'_{ij} = \boldsymbol{P}\boldsymbol{X}_{ij}$, which has the mean $\boldsymbol{P}\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{P}\boldsymbol{\Sigma}_i\boldsymbol{P}^\top$. Then we can use the test statistic $Q_{na}(\boldsymbol{X}')$ in (4.16) defined on $\boldsymbol{X}'_{ij}$ to test the hypotheses (4.18). It can be shown that Hyodo

(2017) studied the same test statistic as Chen and Qin (2010) (or Hu et al., 2017) but under Elliptical populations. It is well noted that estimators used in Hyodo (2017) are exactly the same as the estimators (4.10) and (4.11) defined on $\boldsymbol{X}'_{ij}$.

4.5 Simulation Study

Numerical performance of Chen and Qin (2010)'s method has been investigated in many papers (e.g. Srivastava et al., 2013; Cai et al., 2014; Feng and Sun, 2015; Feng et al., 2016; Hu et al., 2017). Here, we focus the simulation on the test for parallelism to evaluate and compare numerical performances of the tests in Chen and Qin (2010) with asymptotic variance estimators constructed from (4.7) and (4.8) (hereinafter referred to as CQ); the test Chen and Qin (2010) with asymptotic variance estimator constructed from (4.10) and (4.11) (referred to as CQ1) and the test in Harrar and Kong (2016) (referred to as HK). Note that, CQ1 and the test in Hyodo (2017) are exactly the same. For the simulations, we generate data from

(i) Multivariate normal distribution with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

(ii) Multivariate t distribution with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ and degrees of freedom $\nu_1 = 6$ and $\nu_2 = 8$.

(iii) Multivariate contaminated-normal distribution which has density function

$$f_i(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \alpha_i, \eta_i) = \alpha_i \phi(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + (1-\alpha_i)\phi(\boldsymbol{x}|\boldsymbol{\mu}_i, \eta_i \boldsymbol{\Sigma}_i)$$

with parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ where $\phi(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of the multivariate normal $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For the other parameters, we fix $\eta_1 = 5$, $\alpha_1 = 0.5$, $\eta_2 = 3$, and $\alpha_2 = 0.1$.

Note that populations (ii) and (iii) do not satisfy conditions $C2$. However, since these populations are Elliptically contoured, they satisfy condition $C10$.

The empirical size of CQ, CQ1 and HK are presented in Tables 4.1–4.3 where we set $\boldsymbol{\mu}_1 = (\mu_{11}, \ldots, \mu_{1p})^\top$, $\mu_{1j} \overset{iid}{\sim} \text{Uniform}(0,1)$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_p$. We investigate the effects of there different types of $\boldsymbol{\Sigma}_i$:

(1) $\boldsymbol{\Sigma}_{11} = 0.5\boldsymbol{I}_p + 0.5\boldsymbol{J}_p$ and $\boldsymbol{\Sigma}_{12} = 0.9\boldsymbol{I}_p + 0.1\boldsymbol{J}_p$,

(2) $\boldsymbol{\Sigma}_{21} = (0.5^{|j-j_1|})$ and $\boldsymbol{\Sigma}_{22} = (0.1^{|j-j_1|})$ and

(3) $\boldsymbol{\Sigma}_{31} = (0.5|j-j_1|^{-1/2})$ and $\boldsymbol{\Sigma}_{32} = (0.1|j-j_1|^{-1/2})$.

The sizes are calculated with 10,000 replications for the significance level $\alpha = 0.05$.

Table 4.1: Achieved Type I error rate for multivariate normal distribution with $\boldsymbol{\mu}_1 = (\mu_1, \ldots, \mu_p)^\top$, where $\mu_{ij} \overset{iid}{\sim}$ Uniform$(0,1)$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_p$, and three different pairs of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	CQ1	HK	CQ	CQ1	HK	CQ	CQ1	HK
	(50, 90)	0.056	0.056	0.058	0.061	0.061	0.062	0.061	0.061	0.062
50	(100,150)	0.056	0.056	0.056	0.063	0.063	0.063	0.060	0.060	0.060
	(200,240)	0.060	0.060	0.060	0.060	0.060	0.060	0.058	0.058	0.058
	(50, 90)	0.058	0.058	0.059	0.062	0.062	0.063	0.061	0.061	0.062
100	(100,150)	0.059	0.059	0.060	0.061	0.061	0.061	0.061	0.061	0.061
	(200,240)	0.059	0.059	0.059	0.057	0.057	0.057	0.060	0.060	0.060
	(50, 90)	0.057	0.058	0.059	0.057	0.058	0.058	0.059	0.059	0.060
200	(100,150)	0.054	0.054	0.055	0.058	0.058	0.059	0.059	0.059	0.060
	(200,240)	0.053	0.053	0.053	0.056	0.056	0.056	0.062	0.062	0.062
	(50, 90)	0.056	0.056	0.058	0.059	0.059	0.060	0.060	0.060	0.061
400	(100,150)	0.052	0.052	0.053	0.060	0.060	0.060	0.059	0.059	0.060
	(200,240)	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057

Table 4.2: Achieved Type I error rate for multivariate t distribution with $\boldsymbol{\mu}_1 = (\mu_1, \ldots, \mu_p)^\top$, where $\mu_{ij} \overset{iid}{\sim}$ Uniform$(0,1)$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_p$, degrees freedom $\nu_1 = 6$, $\nu_2 = 8$, and three different pairs of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	CQ1	HK	CQ	CQ1	HK	CQ	CQ1	HK
	(50, 90)	0.059	0.059	0.031	0.064	0.064	0.048	0.062	0.062	0.050
50	(100,150)	0.055	0.055	0.036	0.059	0.059	0.050	0.062	0.062	0.053
	(200,240)	0.065	0.065	0.051	0.066	0.066	0.060	0.060	0.060	0.055
	(50, 90)	0.057	0.057	0.020	0.058	0.058	0.035	0.059	0.059	0.037
100	(100,150)	0.058	0.058	0.026	0.058	0.058	0.043	0.059	0.059	0.045
	(200,240)	0.055	0.055	0.036	0.057	0.057	0.047	0.057	0.057	0.049
	(50, 90)	0.055	0.055	0.004	0.057	0.058	0.020	0.061	0.061	0.026
200	(100,150)	0.054	0.054	0.013	0.057	0.057	0.031	0.059	0.059	0.036
	(200,240)	0.053	0.053	0.023	0.064	0.064	0.045	0.058	0.058	0.042
	(50, 90)	0.056	0.056	0.002	0.058	0.058	0.007	0.060	0.060	0.014
400	(100,150)	0.055	0.055	0.005	0.055	0.055	0.013	0.059	0.059	0.024
	(200,240)	0.053	0.053	0.011	0.057	0.057	0.027	0.055	0.055	0.034

Table 4.3: Achieved Type I error rate for multivariate contaminate normal distribution with $\boldsymbol{\mu}_1 = (\mu_1, \ldots, \mu_p)^\top$, where $\mu_{ij} \overset{iid}{\sim}$ Uniform$(0,1)$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{1}_p$, $\eta_1 = 5$, $\alpha_1 = 0.5$, $\eta_2 = 3$, $\alpha_2 = 0.1$, and three different pairs of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	CQ1	HK	CQ	CQ1	HK	CQ	CQ1	HK
50	(50, 90)	0.061	0.061	0.057	0.066	0.066	0.064	0.062	0.062	0.061
	(100,150)	0.060	0.060	0.059	0.065	0.065	0.065	0.062	0.062	0.061
	(200,240)	0.061	0.061	0.060	0.056	0.056	0.056	0.060	0.060	0.059
100	(50, 90)	0.059	0.059	0.052	0.058	0.058	0.054	0.062	0.062	0.060
	(100,150)	0.061	0.061	0.057	0.056	0.056	0.054	0.057	0.057	0.056
	(200,240)	0.058	0.058	0.056	0.056	0.056	0.055	0.063	0.063	0.062
200	(50, 90)	0.060	0.060	0.045	0.056	0.056	0.047	0.061	0.061	0.055
	(100,150)	0.055	0.055	0.047	0.060	0.060	0.055	0.062	0.062	0.059
	(200,240)	0.054	0.054	0.050	0.059	0.059	0.055	0.057	0.057	0.055
400	(50, 90)	0.053	0.054	0.029	0.055	0.056	0.039	0.062	0.062	0.049
	(100,150)	0.053	0.053	0.036	0.057	0.057	0.045	0.059	0.059	0.052
	(200,240)	0.053	0.053	0.043	0.055	0.055	0.049	0.058	0.058	0.054

From Table 4.1, we note that the performances of the three tests are about the same under normality. For the heavier tailed populations (Tables 4.2 and 4.3), HK is not performing well as expected. In particular, it is too conservative for large p but its performance improves as n increases. The tests CQ and CQ1, which are designed to work under non-normality, perform well for all the three distributions. In fact, CQ and CQ1 are nearly identical. As the asymptotic framework suggests, the quality of the asymptotic approximation substantially improves when both n and p are large.

4.6 Real Data Application

In this section, we analyze the Electroencephalograph (EEG) data for the single stimulus (S1) exposure only to compare the results with that of Harrar and Kong (2016). Event-Related Potential (ERP) measures the level of brain activity. The EEG data[1] found at the University of California-Irvine Machine Learning Repository was from a large study to examine EEG correlates of genetic predisposition to alcoholism. Sixty-four electrodes were used to measure ERP and recorded 256 times for in one second. Each channel (electrode) has names identifying the location of the electrode on the scalp. The names are made up of a letter identifying the anatomical location

[1]Web Address: https://archive.ics.uci.edu/ml/datasets/EEG%2BDatabase

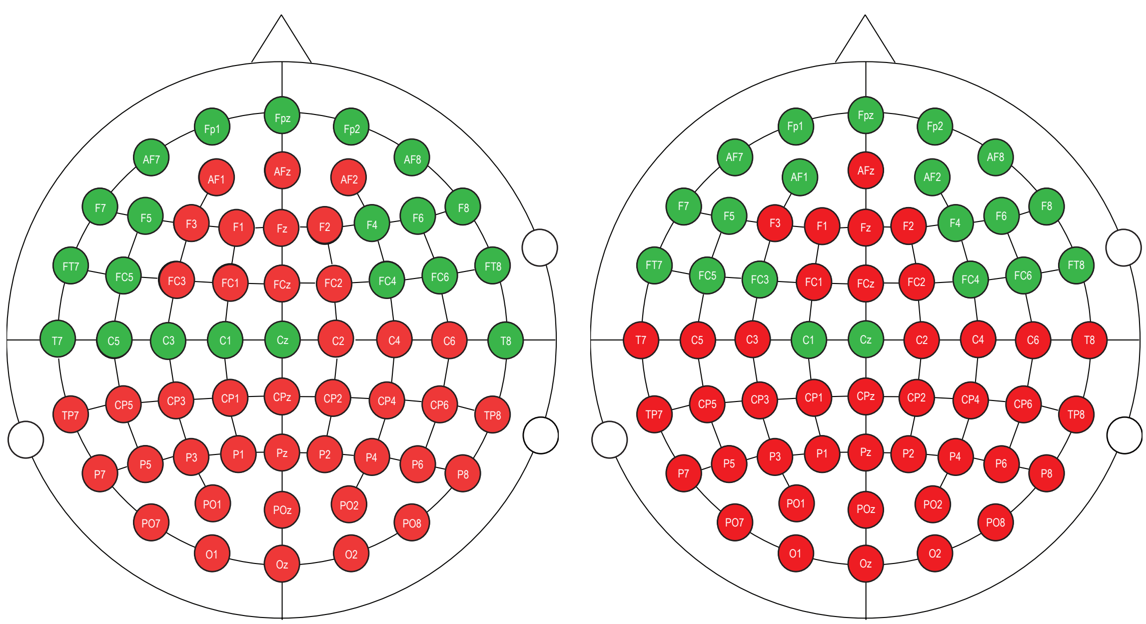
of the placement of the electrode (F–frontal lobe, T–temporal lobe, P–parietal lobe and O–occipital lobe) and a number identifying the hemisphere of the brain (odd number – the left hemisphere and even number – the right hemisphere and letter z (zero) is used for the mid-line). The exception to this naming rule is that, due to their placement and depending on the individual, the "C" electrodes can exhibit/represent EEG activity more typical of Frontal, Temporal, and some Parietal-Occipital activity. ERP reading from an electrode indicates the level of electrical activity (in volts) in the region of the brain where the electrode is placed. There are two groups of subjects in the study: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from a picture set. In this Chapter, we analyze the data only for the single stimulus (S1) exposure. For a more detailed account of the EEG data, see Harrar and Kong (2016). The main objective is to compare CQ1 and HK for testing whether Event-Relates Potential (ERP) mean profiles are similar between the alcoholic and control groups.

Table 4.4 shows FDR adjusted p–values for testing equality and similarity (parallelism) of ERP mean profiles for each of the 64 channels (at FDR = 0.05). The columns in the table contain channel names (Ch) and p-values for testing equality (E) and Similarity (P). The channel-by-channel decisions based on the CQ1 method for similarity (parallelism) are displayed in Figure 4.1 panel (b). The figure depicts the scalp of a human viewed from the top, the triangle marking the nose. The locations of the electrodes are indicated by bubbles. The colors of the bubbles indicate whether the brain activity pattern for that channel is significantly dissimilar (red) or not significantly dissimilar (green). Also shown in Figure 4.1 (panel (a)) is the significance results for equality hypothesis, i.e. whether the mean activity levels are equal between the two groups.

Table 4.4: False Discovery Rate (FDR) adjusted p–values for testing equality and parallelism of mean profiles for Electroencephalograph (EEG) experiment involving alcoholic and control subjects. In the table, the columns are channel label (Ch) and the p-value for equality (E) and parallelism (P) group mean profiles.

Ch	E	P	Ch	E	P	Ch	E	P	Ch	E	P
AF1	0.007	0.071	CP6	0	0	FC6	0.370	0.236	P5	0	0
AF2	0.018	0.08	CPz	0	0	FCz	0	0	P6	0	0
AF7	0.805	0.761	Cz	0.468	0.643	FP1	0.786	0.716	P7	0	0
AF8	0.805	0.763	F1	0	0	FP2	0.812	0.768	P8	0	0
AFz	0	0.010	F2	0	0.001	FPz	0.787	0.761	PO1	0	0
C1	0.672	0.140	F3	0	0.005	FT7	0.747	0.693	PO2	0	0
C2	0.040	0.005	F4	0.057	0.208	FT8	0.449	0.092	PO7	0	0
C3	0.367	0	F5	0.197	0.424	Fz	0	0	PO8	0	0
C4	0.002	0	F6	0.439	0.543	nd	0.064	0.002	POz	0	0
C5	0.094	0	F7	0.468	0.643	O1	0	0	Pz	0	0
C6	0	0	F8	0.770	0.768	O2	0	0	T7	0.217	0
CP1	0	0	FC1	0	0.001	Oz	0	0	T8	0.065	0
CP2	0	0	FC2	0.003	0.037	P1	0	0	TP7	0	0
CP3	0	0	FC3	0.018	0.407	P2	0	0	TP8	0	0
CP4	0	0	FC4	0.805	0.768	P3	0	0	X	0.770	0.768
CP5	0	0	FC5	0.568	0.640	P4	0	0	Y	0	0.028

Figure 4.1: Channel-by-Channel results for EEG data analysis on testing the equality (panel (a)) or similarity (panel (b)) in brain activity between alcoholic and control subjects. Red means brain activity patterns are significantly unequal or dissimilar at $\alpha = 0.05$. Green means that the similarity hypothesis cannot be rejected.



(a) Equality hypothesis (b) Similarity hypothesis

Comparing the significance results in panel (b) of Figure 4.1 with the corresponding Figure in Harrar and Kong (2016), we note that the results for testing similarity in mean brain activity levels between alcoholic and control subjects for CQ1 are same as HK test. It is clear from both panels (a) and (b) of Figure 4.1 that there are no evidence in the data to show difference in the mean electrical activity nor in the patterns between the two groups in the central frontal region of the brain. Most of the significant differences occur in the temporal, parietal and occipital lobes. Of note, the results clearly demarcate contagious similar and non-similar activity regions of the brain.

4.7 Discussion and Conclusion

Recent high-dimensional tests for mean vectors in two or multiple groups are examined. In particular, several realistic and milder conditions are provided to replace existing conditions and the entire theory is reproved under these new conditions. Specifically, the standard assumptions do not cover common multivariate models for dependent data. For example, the simple and popular models for dependent data such as elliptically-contoured and α-mixing are excluded. Further, the methods impose near-independence conditions on the normalized versions of the observations. Some authors refer to these restrictive conditions as pseudo-independence. Besides being strong, making assumptions on normalized versions may not be realistic.

The simulation study investigated the empirical sizes by generating data that violate dependence assumptions imposed in the existing works. No prior simulation study investigated the tests under these models. The numerical results suggest that the finite-sample approximations of high-dimensional asymptotic results are excellent. The application of the results for conducting profile analysis are illustrated via simulated as well as real data set. The formal extension of the methods to a factorial design is also indicated in the manuscript. The detail of such extension is postponed for a future investigation.

4.8 Appendix: Proofs

Conditions $C6$ and $C7$ have implications in regulating the trace of the powers of covariance matrices.

Lemma 4.8.1. *Assume $C6$ and $C7$ hold on $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, where $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are independent and centered p-variate random vectors with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ respectively. Then* $\mathrm{tr}(\boldsymbol{\Sigma}_i^2) = O(p)$ *for $i = 1, 2$ and* $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) = O(p)$.

Proof. Note that $\mathrm{Var}(Y_{ik})$ is bounded by $\varphi_0 = \max\{1 + \Phi_0, \varphi_1\}$. By condition (2), $|\mathrm{Cov}(Y_{ik}, Y_{il})| \leq \varphi_{|k-l|}$, for $i = 1, 2$. Then we have

$$\begin{aligned}\frac{1}{p}\mathrm{tr}(\boldsymbol{\Sigma}_i^2) =& \frac{1}{p}\sum_{k=1}^{p}\mathrm{Var}(Y_{ik})^2 + \frac{1}{p}\sum_{k=1}^{p}\sum_{l=1}^{p}|\mathrm{Cov}(Y_{ik}, Y_{il})|^2 \\ \leq& \frac{1}{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\varphi_{|k-l|}^2 = \frac{1}{p}\sum_{k=0}^{p-1}(p-k)\varphi_k^2 \leq \sum_{k=0}^{p-1}\varphi_k^2 \leq \infty.\end{aligned}$$

It is easy to see that $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) = O(p)$ since $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) \leq \left\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\right\}^{1/2}$. □

4.8.1 Proof of Theorem 4.3.1 and Theorem 4.3.3

Note that

$$T_n(\boldsymbol{X}) - ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2 = T_n(\boldsymbol{X}^c) + T_{n2}(\boldsymbol{X}^c),$$

where

$$T_{n2}(\boldsymbol{X}^c) = \frac{2\sum_{i=1}^{n_1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{X}_{1i}^c}{n_1} + \frac{2\sum_{i=1}^{n_2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \boldsymbol{X}_{2i}^c}{n_2},$$

and $\boldsymbol{X}_{ij}^c = \boldsymbol{X}_{ij} - \boldsymbol{\mu}_i$, for $i = 1, 2$ and $j = 1, \ldots, n_i$. It is easy to show that

$$\mathrm{E}[T_n(\boldsymbol{X}^c)] = \mathrm{E}[T_{n2}(\boldsymbol{X}^c)] = 0,$$

and

$$\mathrm{Var}(T_n(\boldsymbol{X}^c)) = \sigma_n^2 \quad \mathrm{Var}(T_{n2}(\boldsymbol{X}^c)) = \sigma_{n2}^2.$$

Thus, under assumption $C3$,

$$\frac{T_n(\boldsymbol{X}) - ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{\sqrt{\mathrm{Var}(T_n(\boldsymbol{X})}} = \frac{T_n(\boldsymbol{X}^c)}{\sigma_n} + o_P(1).$$

Next, we will prove the asymptotic normality of $T_n(\boldsymbol{X})$ under assumption $C1, C3$-$C7$. Without loss of generality, we assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$. Let $\boldsymbol{Y}_j = \boldsymbol{X}_{1j}$, for $j = 1, \ldots, n_1$, and $\boldsymbol{Y}_{n_1+j} = \boldsymbol{X}_{2j}$, for $j \in 1, \ldots, n_2$. Let $g(i) = 1$, if $1 \le i \le n_1$, or 2 if $n_1 < i \le n$, and let $\phi_{ij} = c_{ij}\boldsymbol{Y}_i^\top \boldsymbol{Y}_j$, for $1 \le i < j \le n$, where

$$c_{ij} = \begin{cases} 2n_1^{-1}(n_1-1)^{-1}, & \text{if } 0 < i < j \le n_1 \\ 2n_2^{-1}(n_2-1)^{-1}, & \text{if } n_1 < i < j \le n_1 + n_2 \\ -2n_1^{-1}n_2^{-1} & \text{if } 0 < i \le n_1 < j \le n_1 + n_2 \end{cases}.$$

For $m = 2, \ldots, n$, define $V_{nm} = \sum_{i=1}^{m-1} \phi_{im}$ and define $\mathcal{F}_{n0} = \{\emptyset, \Omega\}$ and $\mathcal{F}_{nm} = \sigma\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m\}$, which is the σ algebra generated by $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m\}$. Then, it is clear that

$$\mathcal{F}_{n0} \subseteq \mathcal{F}_{n1} \subseteq \cdots \subseteq \mathcal{F}_{nn}.$$

Note that

$$\mathrm{E}[V_{nm}|\mathcal{F}_{nm-1}] = \sum_{i=1}^{m-1} \mathrm{E}[\phi_{im}|\mathcal{F}_{nm-1}] = \sum_{i=1}^{m-1} c_{im}\boldsymbol{Y}_i^\top \mathrm{E}[\boldsymbol{Y}_m] = 0,$$

$$\mathrm{E}[V_{nm}^2] = \sum_{i=1}^{m-1} c_{im}^2 \mathrm{E}[\boldsymbol{Y}_i^\top \boldsymbol{Y}_m \boldsymbol{Y}_m^\top \boldsymbol{Y}_i] = \sum_{i=1}^{m-1} c_{im}^2 \mathrm{tr}\Big\{\mathrm{E}[\boldsymbol{Y}_m\boldsymbol{Y}_m^\top]\mathrm{E}[\boldsymbol{Y}_i\boldsymbol{Y}_i^\top]\Big\}.$$

When $m \le n_1$,

$$\mathrm{E}[V_{nm}^2] = \frac{2(m-1)}{n_1^2(n_1-1)^2}\mathrm{tr}(\boldsymbol{\Sigma}_1^2),$$

and when $n_1 < m \le n$,

$$\mathrm{E}[V_{nm}^2] = \frac{4}{n_1 n_2^2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) + \frac{4(m-n_1-1)}{n_2^2(n_2-1)^2}\mathrm{tr}(\boldsymbol{\Sigma}_2^2).$$

So $\{V_{nm}, \mathcal{F}_{nm}\}_{m=1}^n$ is the sequence of square-integrable martingale difference. Furthermore, we have

$$T_n(\boldsymbol{X}) = \sum_{m=2}^n V_{nm}.$$

To prove Theorem 4.3.1, we only need to prove the following Lemma 4.8.2 and the Lindeberg condition Lemma 4.8.3. Then, the desired result follows by the martingale difference central limit theorem, see Shiryaev (2015) or Hall and Heyde (1980).

Lemma 4.8.2. *Under the assumptions $C1$, $C3$– $C7$,*

$$\sigma_n^{-2}\sum_{m=2}^{n}\mathrm{E}[V_{nm}^2|\mathcal{F}_{n,m-1}] \xrightarrow{P} 1, \text{ as } n\to\infty.$$

Proof. First,

$$\sum_{m=2}^{n}\mathrm{E}[V_{nm}^2|\mathcal{F}_{nm-1}]-\sigma_n^2=\sum_{m=2}^{n}\sum_{i,j=1}^{m-1}c_{im}c_{jm}\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_{g(m)}\boldsymbol{Y}_j-\sigma_n^2=\sum_{k=1}^{4}B_k,$$

where

$$\begin{aligned}
B_1=&\sum_{i=1}^{n_1}\frac{4(n_1-i)}{n_1^2(n_1-1)^2}\{\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_1\boldsymbol{Y}_i-\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\}+\sum_{j=n_1+1}^{n}\frac{4(n-j)}{n_2^2(n_2-1)^2}\{\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_2\boldsymbol{Y}_j-\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\},\\
B_2=&\sum_{i=1}^{n_1}\frac{4}{n_1^2n_2}\{\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_2\boldsymbol{Y}_i-\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\}+\sum_{i=2}^{n_1}\sum_{j=1}^{i-1}\frac{8}{n_1^2n_2}\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_2\boldsymbol{Y}_j,\\
B_3=&\sum_{i=2}^{n_1}\sum_{j=1}^{i-1}\frac{8(n_1-i)}{n_1^2(n_1-1)^2}\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_1\boldsymbol{Y}_j+\sum_{i=n_1+2}^{n}\sum_{j=n_1+1}^{i-1}\frac{8(n-i)}{n_2^2(n_2-1)^2}\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_2\boldsymbol{Y}_j,\quad\text{and}\\
B_4=&\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\frac{4(n-j)}{n_1n_2^2(n_2-1)}\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_2\boldsymbol{Y}_j.
\end{aligned}$$

From this decomposition, it is easy to check $\sum_{k=1}^{4}\mathrm{E}[B_i]=0$. By Holder's inequality,

$$\mathrm{Var}\Big(\sum_{m=2}^{n}\mathrm{E}[V_{nm}^2|\mathcal{F}_{n,m-1}]\Big)=\mathrm{E}\Big[\sum_{k=1}^{4}B_k\Big]^2\le 4\sum_{k=1}^{4}\mathrm{E}[B_i^2].$$

Therefore, we only need to prove $\sigma_n^{-4}\sum_{k=1}^{4}\mathrm{E}[B_i^2]=o(1)$. Under assumptions $C1$ and $C6$, and by Corollary 4.2.2, for $i,i_1\in\{1,2\}$ and $j\neq j_1\in\{1\ldots,n\}$, we have

$$\mathrm{Var}(\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_i\boldsymbol{Y}_j)\le C(\Phi_1+\Phi_2+\Phi_3)\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{g(j)})^2\}$$

and

$$\mathrm{Var}(\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_i\boldsymbol{Y}_{j_1})=\mathrm{E}(\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_i\boldsymbol{Y}_{j_1}\boldsymbol{Y}_{j_1}^\top\boldsymbol{\Sigma}_i\boldsymbol{Y}_j)=\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{g(j)}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{g(j_1)}).$$

Consequently, under assumptions $C3, C_4$ and $C7$, we have

$$\begin{aligned}
\mathrm{E}[B_1^2] =& \sum_{i=1}^{n_1} \frac{16(n_1-i)^2}{n_1^4(n_1-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_1 \boldsymbol{Y}_i) + \sum_{i=n_1+1}^{n} \frac{16(n-i)^2}{n_2^4(n_2-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_i) \\
\leq& \sum_{i=1}^{2} \frac{4(2n_i-1)}{n_i^3(n_i-1)^3} C(\Phi_1+\Phi_2+\Phi_3)\mathrm{tr}(\boldsymbol{\Sigma}_i^4) = o(\sigma_n^4/n), \\
\mathrm{E}[B_2^2] =& \sum_{i=1}^{n_1} \frac{16}{n_1^4 n_2^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_i) + \sum_{i=2}^{n_1}\sum_{j=1}^{i-1} \frac{64}{n_1^4 n_2^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) \\
\leq& \frac{16}{n_1^3 n_2^2} C(\Phi_1+\Phi_2+\Phi_3)\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + \frac{32(n_1-1)}{n_1^3 n_2^2}\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} = o(\sigma_n^4/n), \\
\mathrm{E}[B_3^2] =& \sum_{i=2}^{n_1}\sum_{j=1}^{i-1} \frac{64(n_1-i)^2}{n_1^4(n_1-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_1 \boldsymbol{Y}_j) + \sum_{i=n_1+2}^{n}\sum_{j=n_1+1}^{i-1} \frac{64(n-i)^2}{n_2^4(n_2-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) \\
=& \sum_{i=1}^{2} \frac{16(n_i-2)}{3n_i^3(n_i-1)^2}\mathrm{tr}(\boldsymbol{\Sigma}_i^4) = o(\sigma_n^4) \quad \text{and} \\
\mathrm{E}[B_4^2] =& \sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} \frac{16(n-j)^2}{n_1^2 n_2^4(n_2-1)^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) = \frac{8(2n_2-1)}{3n_1 n_2^3(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^3) = o(\sigma_n^4).
\end{aligned}$$

That finishes the proof. □

Lemma 4.8.3. *Under the assumptions $C1$, $C3$– $C7$,*

$$\sigma_n^{-4}\sum_{m=2}^{n} \mathrm{E}\{V_{nm}^2 I(|V_{nm}|>\epsilon)|\mathcal{F}_{n,m-1}\} \xrightarrow{P} 0, \text{ as } n\to\infty, \ \forall\varepsilon>0.$$

Proof. Since

$$\sum_{m=2}^{n} \mathrm{E}\{V_{nm}^2 I(|V_{nm}|>\epsilon)|\mathcal{F}_{n,m-1}\} \leq \epsilon^{2-q}\sum_{m=2}^{n}\mathrm{E}(V_{nm}^q|\mathcal{F}_{nm-1}),$$

for some $q>2$. By choosing $q=4$ the conclusion of the lemma is true if we can prove

$$\mathrm{E}\Big\{\sum_{m=2}^{n}\mathrm{E}(V_{nm}^4|\mathcal{F}_{n,m-1})\Big\} = \sum_{m=2}^{n}\mathrm{E}[V_{nm}^4] = o(\sigma_n^4).$$

Note that

$$\mathrm{E}[V_{nm}^4] = \mathrm{E}\Big\{\Big(\sum_{i=1}^{m-1}\phi_{im}\Big)^4\Big\} = \sum_{i=1}^{m-1}\mathrm{E}(\phi_{im}^4) + 6\sum_{i<j}^{m-1}\mathrm{E}(\phi_{im}^2\phi_{jm}^2).$$

Now, for $1 \le i \ne j < m$, under assumption $C1$ and $C6$

$$
\begin{aligned}
\mathrm{E}(\phi_{im}^4) =& c_{im}^4 \mathrm{E}\Big[\mathrm{E}\big\{(\boldsymbol{Y}_i^\top \boldsymbol{Y}_m)^4 | \mathcal{F}_{nm-1}\big\}\Big] = c_{im}^4 \mathrm{E}\Big[\mathrm{E}\big\{(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m)^2 | \mathcal{F}_{nm-1}\big\}\Big] \\
=& c_{im}^4 \mathrm{E}\Big[\mathrm{Var}\big(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m | \mathcal{F}_{nm-1}) + \mathrm{tr}^2(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)})\Big] \\
\le& c_{im}^4 \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \mathrm{E}\Big\{\big(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)} \boldsymbol{Y}_i\big)^2\Big\} \\
\le& c_{im}^4 \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \Big\{\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)} \boldsymbol{Y}_i) + \mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)} \boldsymbol{\Sigma}_{g(m)})\Big\} \\
\le& c_{im}^4 \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \Big[C(\Phi_0 + \Phi_1 + \Phi_2) \mathrm{tr}\big\{(\boldsymbol{\Sigma}_{g(i)} \boldsymbol{\Sigma}_{g(m)})^2\big\} \\
& + \mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)} \boldsymbol{\Sigma}_{g(m)})\Big],
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}(\phi_{im}^2 \phi_{jm}^2) =& \mathrm{E}\Big\{\mathrm{E}(\phi_{im}^2 \phi_{jm}^2 | \mathcal{F}_{nm-1})\Big\} \\
=& c_{im}^2 c_{jm}^2 \mathrm{E}\big\{(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m)(\boldsymbol{Y}_m^\top \boldsymbol{Y}_j \boldsymbol{Y}_j^\top \boldsymbol{Y}_m) | \mathcal{F}_{nm-1}\big\} \\
=& c_{im}^2 c_{jm}^2 \mathrm{E}\Big\{\mathrm{Cov}(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m, \boldsymbol{Y}_m^\top \boldsymbol{Y}_j \boldsymbol{Y}_j^\top \boldsymbol{Y}_m) | \mathcal{F}_{nm-1}) \\
& + \mathrm{tr}(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)}) \mathrm{tr}(\boldsymbol{Y}_j \boldsymbol{Y}_j^\top \boldsymbol{\Sigma}_{g(m)})\Big\} \\
\le& c_{im}^2 c_{jm}^2 \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \mathrm{E}\Big\{(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)} \boldsymbol{Y}_i)(\boldsymbol{Y}_j^\top \boldsymbol{\Sigma}_{g(m)} \boldsymbol{Y}_j)\Big\} \\
\le& c_{im}^2 c_{jm}^2 \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \mathrm{tr}(\boldsymbol{\Sigma}_{g(i)} \boldsymbol{\Sigma}_{g(m)}) \mathrm{tr}(\boldsymbol{\Sigma}_{g(j)} \boldsymbol{\Sigma}_{g(m)}).
\end{aligned}
$$

Thus, under assumption $C3$, $C4$ and $C7$, we have

$$
\begin{aligned}
\sum_{m=2}^{n} \mathrm{E}[V_{nm}^4] =& \sum_{m=2}^{n} \sum_{i=1}^{m-1} \mathrm{E}(\phi_{im}^4) + 6 \sum_{m=2}^{n} \sum_{i<j}^{m-1} \mathrm{E}(\phi_{im}^2 \phi_{jm}^2) \\
\le& \big\{1 + C(\Phi_0 + \Phi_1 + \Phi_2)\big\} \left\{\sum_{i=1}^{2} \frac{2n_i - 3}{2n_i^3 (n_i - 1)^3} \mathrm{tr}^2(\boldsymbol{\Sigma}_i^2) \right. \\
& \left. + \frac{3n_1 - 2}{n_1^3 n_2^3} \mathrm{tr}^2(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) + \frac{3}{n_1 n_2^3 (n_2 - 1)} \mathrm{tr}^2(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) \mathrm{tr}(\boldsymbol{\Sigma}_2^2) \right\} \\
& + \sum_{i=1}^{2} O\Big\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^4)}{n_i^3 (n_i - 1)^3}\Big\} + O\Big\{\frac{\mathrm{tr}\{(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)^2\}}{n_1^3 n_2^3}\Big\} = o(\sigma_n^4).
\end{aligned}
$$

□

Theorem 4.3.3, we just need to prove that Lemma 4.8.2 and Lemma 4.8.3 hold under assumptions $C3$–$C9$. From the above discussions, it suffices to prove

$$\sum_{k=1}^{4} \mathrm{E}[B_i^2] = o(\sigma_n^4), \quad \text{and} \quad \sum_{m=2}^{n} \mathrm{E}[V_{nm}^4] = o(\sigma_n^4).$$

In fact, by Theorem 4.2.1, under assumption $C3$–$C9$, we have

$$\begin{aligned}
\mathrm{E}[B_1^2] =& \sum_{i=1}^{n_1} \frac{16(n_1-i)^2}{n_1^4(n_1-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_1 \boldsymbol{Y}_i) + \sum_{i=n_1+1}^{n} \frac{16(n-i)^2}{n_2^4(n_2-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_i) \\
\leq& \sum_{i=1}^{2} \frac{4(2n_i-1)}{n_i^3(n_i-1)^3} C(\Phi_1+\Phi_2+\Phi_3)\mathrm{tr}(\boldsymbol{\Sigma}_i^2) = o(\sigma_n^4), \\
\mathrm{E}[B_2^2] =& \sum_{i=1}^{n_1} \frac{16}{n_1^4 n_2^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_i) + \sum_{i=2}^{n_1}\sum_{j=1}^{i-1} \frac{64}{n_1^4 n_2^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) \\
\leq& \frac{16}{n_1^3 n_2^2} C(\Phi_1+\Phi_2+\Phi_3)\mathrm{tr}(\boldsymbol{\Sigma}_2^2) + \frac{32(n_1-1)}{n_1^3 n_2^2}\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} = o(\sigma_n^4), \\
\mathrm{E}[B_3^2] =& \sum_{i=2}^{n_1}\sum_{j=1}^{i-1} \frac{64(n_1-i)^2}{n_1^4(n_1-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_1 \boldsymbol{Y}_j) + \sum_{i=n_1+2}^{n}\sum_{j=n_1+1}^{i-1} \frac{64(n-i)^2}{n_2^4(n_2-1)^4}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) \\
=& \sum_{i=1}^{2} \frac{16(n_i-2)}{3n_i^3(n_i-1)^2}\mathrm{tr}(\boldsymbol{\Sigma}_i^4) = o(\sigma_n^4) \quad \text{and} \\
\mathrm{E}[B_4^2] =& \sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} \frac{16(n-j)^2}{n_1^2 n_2^4(n_2-1)^2}\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_2 \boldsymbol{Y}_j) = \frac{8(2n_2-1)}{3n_1 n_2^3(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^3) = o(\sigma_n^4).
\end{aligned}$$

Further, for $1 \leq i \neq j < m$, under assumption $C6$,

$$\begin{aligned}
\mathrm{E}(\phi_{im}^4) =& c_{im}^4 \mathrm{E}\Big[\mathrm{E}\big\{(\boldsymbol{Y}_i^\top \boldsymbol{Y}_m)^4 | \mathcal{F}_{nm-1}\big\}\Big] = c_{im}^4 \mathrm{E}\Big[\mathrm{E}\big\{(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m)^2 | \mathcal{F}_{nm-1}\big\}\Big] \\
=& c_{im}^4 \mathrm{E}\Big[\mathrm{Var}\big(\boldsymbol{Y}_m^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{Y}_m | \mathcal{F}_{nm-1}\big) + \mathrm{tr}^2(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)})\Big] \\
\leq& c_{im}^4 \mathrm{E}\Big\{C(\Phi_0+\Phi_1+\Phi_2)\big(\boldsymbol{Y}_i^\top \boldsymbol{Y}_i\big)^2 + \big(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}_{g(m)} \boldsymbol{Y}_i\big)^2\Big\} \\
\leq& c_{im}^4 \Big[C(\Phi_0+\Phi_1+\Phi_2)\big\{\mathrm{Var}(\boldsymbol{Y}_i^\top \boldsymbol{Y}_i) + \mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)})\big\} \\
&+ \mathrm{Var}(\boldsymbol{Y}_m^\top \boldsymbol{\Sigma}_{g(i)} \boldsymbol{Y}_m) + \mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)}\boldsymbol{\Sigma}_{g(m)})\Big] \\
\leq& c_{im}^4 C^2(\Phi_0+\Phi_1+\Phi_2)^2 p + c_{im}^4 C(\Phi_0+\Phi_1+\Phi_2)\Big[\mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)}) + \mathrm{tr}\big\{(\boldsymbol{\Sigma}_{g(i)}\boldsymbol{\Sigma}_{g(m)})^2\big\}\Big] \\
&+ c_{im}^4 \mathrm{tr}^2(\boldsymbol{\Sigma}_{g(i)}\boldsymbol{\Sigma}_{g(m)}),
\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{E}(\phi_{im}^2\phi_{jm}^2) =&\mathrm{E}\Big\{\mathrm{E}(\phi_{im}^2\phi_{jm}^2|\mathcal{F}_{nm-1})\Big\} = c_{im}^2c_{jm}^2\mathrm{E}\{(\boldsymbol{Y}_m^\top\boldsymbol{Y}_i\boldsymbol{Y}_i^\top\boldsymbol{Y}_m)(\boldsymbol{Y}_m^\top\boldsymbol{Y}_j\boldsymbol{Y}_j^\top\boldsymbol{Y}_m)|\mathcal{F}_{nm-1}\} \\
=&c_{im}^2c_{jm}^2\mathrm{E}\Big\{\mathrm{Cov}(\boldsymbol{Y}_m^\top\boldsymbol{Y}_i\boldsymbol{Y}_i^\top\boldsymbol{Y}_m, \boldsymbol{Y}_m^\top\boldsymbol{Y}_j\boldsymbol{Y}_j^\top\boldsymbol{Y}_m)|\mathcal{F}_{nm-1}) \\
&+\mathrm{tr}(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_{g(m)})\mathrm{tr}(\boldsymbol{Y}_j\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_{g(m)})\Big\} \\
\le&c_{im}^2c_{jm}^2\mathrm{E}\Big\{C(\Phi_0+\Phi_1+\Phi_2)(\boldsymbol{Y}_i^\top\boldsymbol{Y}_i)(\boldsymbol{Y}_j^\top\boldsymbol{Y}_j)+(\boldsymbol{Y}_i^\top\boldsymbol{\Sigma}_{g(m)}\boldsymbol{Y}_i)(\boldsymbol{Y}_j^\top\boldsymbol{\Sigma}_{g(m)}\boldsymbol{Y}_j)\Big\} \\
\le&c_{im}^2c_{jm}^2\Big\{C(\Phi_0+\Phi_1+\Phi_2)\mathrm{tr}(\boldsymbol{\Sigma}_{g(i)})\mathrm{tr}(\boldsymbol{\Sigma}_{g(j)})+\mathrm{tr}(\boldsymbol{\Sigma}_{g(i)}\boldsymbol{\Sigma}_{g(m)})\mathrm{tr}(\boldsymbol{\Sigma}_{g(j)}\boldsymbol{\Sigma}_{g(m)}\Big\}.
\end{aligned}
$$

Thus under assumption $C3$–$C9$, we have

$$
\begin{aligned}
\sum_{m=2}^n\mathrm{E}[V_{nm}^4] =& \sum_{m=2}^n\sum_{i=1}^{m-1}\mathrm{E}(\phi_{im}^4)+6\sum_{m=2}^n\sum_{i<j}^{m-1}\mathrm{E}(\phi_{im}^2\phi_{jm}^2) \\
\le&\sum_{i=1}^2\frac{2n_i-3}{2n_i^3(n_i-1)^3}\mathrm{tr}^2(\boldsymbol{\Sigma}_i^2)+\frac{3n_1-2}{n_1^3n_2^3}\mathrm{tr}^2(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) \\
&+\frac{3}{n_1n_2^3(n_2-1)}\mathrm{tr}^2(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)+\sum_{i=1}^2O\Big\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^4)}{n_i^3(n_i-1)^3}\Big\} \\
&+O\Big\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1)\mathrm{tr}(\boldsymbol{\Sigma}_2)}{n_1^3n_2^3}\Big\}+\sum_{i=1}^2O\Big\{\frac{\mathrm{tr}^2(\boldsymbol{\Sigma}_i)}{n^5}\Big\}+O\Big\{\frac{p}{n^6}\Big\} \\
=&o(\sigma_n^4).
\end{aligned}
$$

The last step is following from condition $C8$ and $C9$ as

$$
\begin{aligned}
&\mathrm{tr}(\boldsymbol{\Sigma}_i)\mathrm{tr}(\boldsymbol{\Sigma}_j)\le\frac{1}{2}\{\mathrm{tr}^2(\boldsymbol{\Sigma}_i)+\mathrm{tr}^2(\boldsymbol{\Sigma}_j)\}\le p[\mathrm{tr}\{(\boldsymbol{\Sigma}_1)^2\}+\mathrm{tr}\{(\boldsymbol{\Sigma}_2)^2\}] \\
&\le p\mathrm{tr}\{(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2)^2\}=o(n^5\sigma_n^4),
\end{aligned}
$$

for any $i,j\in\{1,2\}$, and $p=o(n^5\sigma_n^4)$.

4.8.2 Proof of Theorem 4.3.2 and Theorem 4.3.4

For Theorem 4.3.2, we will only present the proof of the ratio-consistency of $\mathrm{tr}(\boldsymbol{\Sigma}_1^2)$ in (4.10) under assumptions $C1, C3$-$C7$ as the proofs of the others follow along the same lines. The proof of Theorem 4.3.4 proceeds along the same steps and, therefore, is omitted. For notational convenience, we denote $\boldsymbol{X}_{1j}-\boldsymbol{\mu}_1$ as $\boldsymbol{X}_j$, $\boldsymbol{\mu}_1$ as $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1$ as $\boldsymbol{\Sigma}$ and n_1 as n, since we are effectively in the one-sample situation.

Proof of Theorem 4.3.2. It suffices to show that

$$\mathrm{E}\{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\} = \mathrm{tr}(\boldsymbol{\Sigma}^2) \quad \text{and} \quad \mathrm{Var}\{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\} = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)).$$

Note that

$$\begin{aligned}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)} =& \frac{1}{(n)_2}\sum_{i\neq j}(\boldsymbol{X}_i^\top \boldsymbol{X}_j)^2 - \frac{2}{(n)_3}\sum_{i\neq j\neq k}(\boldsymbol{X}_i^\top \boldsymbol{X}_j)(\boldsymbol{X}_i^\top \boldsymbol{X}_k)\\ &+ \frac{1}{(n)_4}\sum_{i\neq j\neq k\neq l}(\boldsymbol{X}_i^\top \boldsymbol{X}_j)(\boldsymbol{X}_k^\top \boldsymbol{X}_l) = A_1 + A_2 + A_3,\end{aligned}$$

where

$$A_1 = \frac{1}{(n)_2}\sum_{i\neq j}^{n} \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{X}_i,$$

and $A_i, i = 2, 3$ are defined to be the corresponding terms in the equation.

It is obvious that $\mathrm{E}\{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\} = \mathrm{tr}(\boldsymbol{\Sigma}^2)$, since the $\mathrm{E}(A_2) = \mathrm{E}(A_3) = 0$ and $\mathrm{E}(A_1) = \mathrm{tr}(\boldsymbol{\Sigma}^2)$. To prove $\mathrm{Var}\{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\} = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2))$, it is enough to show that $\mathrm{E}(A_1^2) = \mathrm{tr}^2(\boldsymbol{\Sigma}^2)(1 + o(1))$ and $\mathrm{E}(A_i^2) = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2))$ for $i = 2, 3$. Indeed, under conditions $C1$, $C3$-$C7$,

$$\begin{aligned}\mathrm{E}(A_1^2) =& \frac{1}{n^2(n-1)^2}\mathrm{E}\Big(\sum_{i\neq j}^{n} \boldsymbol{X}_j^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{X}_j\Big)^2\\ =& \frac{1}{n^2(n-1)^2}\mathrm{E}\Big[\sum_{i\neq j}^{n} 2(\boldsymbol{X}_j^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{X}_j)^2 + \sum_{i\neq j\neq k}^{n} 4\boldsymbol{X}_j^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{X}_k \boldsymbol{X}_k^\top \boldsymbol{X}_j\\ &+ \sum_{i\neq j\neq k\neq l}^{n} \boldsymbol{X}_j^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_l^\top \boldsymbol{X}_k \boldsymbol{X}_k^\top \boldsymbol{X}_l\Big]\\ =& O\Big\{\frac{2n-5}{n(n-1)}\Big\}\mathrm{tr}^2(\boldsymbol{\Sigma}^2) + O\Big\{\frac{1}{n(n-1)}\Big\}\mathrm{tr}(\boldsymbol{\Sigma}^4) + \frac{(n-2)(n-3)}{n(n-1)}\mathrm{tr}^2(\boldsymbol{\Sigma}^2)\\ =& \mathrm{tr}^2(\boldsymbol{\Sigma}^2)(1 + o(1)),\end{aligned}$$

$$\begin{aligned}\mathrm{E}(A_2^2) =& \frac{4}{\{(n)_3\}^2}\mathrm{E}\Big(\sum_{i\neq j\neq k}^{n} \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_k^\top \boldsymbol{X}_i\Big)^2 \leq \frac{4}{\{(n)_3\}^2}\mathrm{E}\Big\{\sum_{i}^{n} n\Big(\sum_{j\neq k(\neq i)}^{n} \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_k^\top \boldsymbol{X}_i\Big)^2\Big\}\\ =& \frac{8n}{\{(n)_3\}^2}\sum_{i}^{n}\sum_{j\neq k(\neq i)}^{n} \mathrm{E}\Big(\boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_k^\top \boldsymbol{X}_i)^2\Big)\\ =& O\Big\{\frac{1}{(n-1)(n-2)}\Big\}\mathrm{tr}^2(\boldsymbol{\Sigma}^2) = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)),\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{E}(A_3^2) =& \frac{1}{\{(n)_4\}^2}\mathrm{E}\Big(\sum_{i\neq j\neq k\neq l}^{n}\boldsymbol{X}_i^\top\boldsymbol{X}_j\boldsymbol{X}_k^\top\boldsymbol{X}_l\Big)^2\\
\leq& \frac{1}{\{(n)_4\}^2}\mathrm{E}\Big\{\sum_{i\neq j}^{n} n(n-1)\Big(\sum_{j\neq k(\neq i,j)}^{n}\boldsymbol{X}_i^\top\boldsymbol{X}_j\boldsymbol{X}_k^\top\boldsymbol{X}_l\Big)^2\Big\}\\
=& \frac{2n(n-1)}{\{(n)_4\}^2}\sum_{i\neq j}^{n}\sum_{j\neq k(\neq i,j)}^{n}\mathrm{E}\Big(\boldsymbol{X}_i^\top\boldsymbol{X}_j\boldsymbol{X}_k^\top\boldsymbol{X}_l)^2\Big)\\
=& O\Big\{\frac{1}{(n-2)(n-3)}\Big\}\mathrm{tr}^2(\boldsymbol{\Sigma}^2) = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)),
\end{aligned}
$$

the inequalities in the proof arise from application of Hölder's inequality. □

Chapter 5 High-Dimensional Rank-Based Inference

5.1 Introduction

High-dimensional data have been the subject of theoretical and applied investigations in the last few decades sparked by advance in technology that allowed large number of observations to be collected from each analysis unit (subject). For example, genomic studies, satellite imaging, modern diagnostic and intervention modalities generate high-dimensional data. To analyze these data, in particular in the context of group comparison or treatment efficacy, the asymptotic theory requires both the sample size and dimension to diverge. Sparsity conditions that characterize the nature of the within-unit dependence are also needed to establish the results. Some of these results assume multivariate normality (Dempster, 1958, 1960; Fujikoshi et al., 2004; Schott, 2007a; Srivastava and Du, 2008; Yamada and Srivastava, 2012; Dong et al., 2017), while others assume existence of higher-order moments and pseudo-independence in the sense that higher order mixed moments factor into the product of the corresponding univariate moments (Bai and Saranadasa, 1996; Chen and Qin, 2010; Srivastava and Kubokawa, 2013; Hu et al., 2017) or other forms weaker dependence (Cai et al., 2014; Cai and Xia, 2014; Feng et al., 2015; Gregory et al., 2015). The nonparametric methods (Wang et al., 2015; Ghosh and Biswas, 2016) are essentially mean based and also assume (generalized) elliptical populations.

In this Chapter, we pursue a fully nonparametric approach for high-dimensional comparison of population or treatment groups in which neither existence of moments nor pseudo independence assumptions are required. For brevity, we focus on the two-group situation and extensions to more general cases will be outlined later.

For each $i = 1, 2$, $j = 1, \ldots, n_i$, assume $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^\top$ to be identically and independently distributed observations with marginal distributions $X_{ijk} \sim F_{ik}$, which are assumed to be non-degenerate. Denote the total sample size by $n = n_1 + n_2$ and let $N = np$ be the total number of observations. Here, we are using the normalized version of distribution function, which is defined by

$$F_{ik}(x) = \frac{1}{2}\{F_{ik}^{+}(x) + F_{ik}^{-}(x)\} = P(X_{i1k} < x) + \frac{1}{2}P(X_{i1k} = x)$$

where $F_{ik}^{+}(x) = P(X_{i1k} \leq x)$ and $F_{ik}^{-}(x) = P(X_{i1k} < x)$ are, respectively, the right- and left-continuous versions of the distribution function. Using normalized distribution function allows us to treat the discrete and continuous cases in a unified manner (e.g., Akritas et al., 1997). Define the nonparametric relative summary effect by

$$\omega_{ik} = \mathrm{E}[H(X_{i1k})] = \int H dF_{ik},$$

for $i = 1, 2$ and $k = 1, \ldots, p$, where $H(x) = N^{-1}\sum_{k=1}^{p}\{n_1 F_{1k}(x) + n_2 F_{2k}(x)\}$. Let $Z \sim H$ be a random variable, then it is an easy matter to show that

$$\omega_{ik} = P(X_{i1k} > Z) + \frac{1}{2}P(X_{i1k} = Z).$$

Therefore, ω_{ik} indicates the tendency of observations on the k-th variables from group i to be larger (smaller) than observations on a random variables from the average distribution according as $\omega_{ik} > 1/2 (< 1/2)$. In view of this, we consider the testing problem

$$H_0 : \boldsymbol{\omega}_1 = \boldsymbol{\omega}_2 \quad \text{VS} \quad H_1 : \boldsymbol{\omega}_1 \neq \boldsymbol{\omega}_2, \tag{5.1}$$

where $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{ip})^\top$ for $i = 1, 2$. That is, the hypothesis is stated in terms of ω_{ik}, a quantity that does not involve any parameter nor require existence of any moment. Besides these obvious advantages, the hypothesis in terms of the relative summary effects does not impose equality of marginal nor joint distribution under the null hypothesis. The significance of this versatility is that the treatment groups could be different but in ways that are not interesting to the researcher. It can easily be seen that hypotheses of equality of all marginal distributions or joint distributions between the two treatment groups imply $\omega_{1k} = \omega_{2k}$ for $k = 1, \ldots, p$. Mean vectors are not well defined when some of the variable are measured in ordinal scales, rendering recent high-dimensional parametric tests for comparing mean vectors: Bai and Saranadasa (1996); Chen and Qin (2010); Srivastava and Kubokawa (2013); Ahmad (2014); Cai et al. (2014); Feng and Sun (2015); Feng et al. (2015); Gregory et al. (2015); Xu et al. (2016) and nonparametric ones (Wang et al., 2015; Ghosh and Biswas, 2016) inappropriate.

This Chapter is organized as follows. Asymptotic equivalence theory for quadratic forms in ranks is developed in Section 5.2. Also in this section, a result on the transfer of dependence on original observations to Asymptotic Rank Transforms (main asymptotic tool of this Chapter) is stated. Section 5.3 introduces the test statistic and develops asymptotic normality result for it. The extension of the methods to mulit-group case is taken up in Section 5.4. A simulation study is carried out in Section 5.5 to numerically show the performance of the new test in comparison with competing methods. Section 5.6 uses high-dimensional Electroencephalograph (EEG) data to illustrate the application of the new rank-method. Discussions and some conclusions are summarized in Section 5.7. All proofs and technical details are placed in the Appendix.

5.2 High-Dimensional Quadratic Forms in Ranks

The approach we follow in this Chapter seeks rank-based estimate of the relative summary effects and uses these estimates to construct a high-dimensional asymptotic test. It would be natural to estimate ω_{ik} by

$$\widehat{\omega}_{ik} = \int \widehat{H} d\widehat{F}_{ik}$$

where $\widehat{F}_{ik} = \frac{1}{2}(\widehat{F}^{+}_{ik} + \widehat{F}^{-}_{ik})$ where $\widehat{F}^{+}_{ik}$ and $\widehat{F}^{-}_{ik}$ are the right- and left-continuous versions of the empirical distribution function and $\widehat{H}(x) = N^{-1}\sum_{k=1}^{p}[n_1\widehat{F}_{1k}(x) + n_2\widehat{F}_{2k}(x)]$. More specifically, $\widehat{F}_{ik}(x) = n_i^{-1}\sum\limits_{j=1}^{n_i} c(x - X_{ijk})$ where $c(u) = \{c^{-}(u) + c^{+}(u)\}/2$, $c^{-}(u) = I(u > 0)$ and $c^{+}(u) = I(u \geq 0)$ are the normalized, left-continuous and right-continuous, respectively, versions of the counting function.

Define the asymptotic rank transforms (ART) $\boldsymbol{Y}_{ij}$ and rank transform (RT) $\widehat{\boldsymbol{Y}}_{ij}$ by $Y_{ijk} = H(X_{ijk})$ and $\widehat{Y}_{ijk} = \widehat{H}(X_{ijk})$, respectively. Let R_{ijk} be the (mid-) rank of X_{ijk} among all the N observations

$$\{X_{111}, \ldots, X_{1n_11}, X_{211}, \ldots, X_{2n_21}, \ldots, X_{11p}, \ldots, X_{1n_1p}, X_{21p}, \ldots, X_{2n_2p}\}.$$

It is easy to see that the rank transforms are related to the ranks by the relation $\widehat{Y}_{ijk} = N^{-1}(R_{ijk} - \frac{1}{2})$. After some simplification, the estimator $\widehat{\omega}_{ik}$ can be expressed as

$$\widehat{\omega}_{ik} = n_i^{-1} \sum_{j=1}^{n_i} \widehat{Y}_{ijk} = \frac{1}{N}(\overline{R}_{i\cdot k} - 1/2)$$

where $\overline{R}_{i\cdot k} = n_i^{-1} \sum_{j=1}^{n_i} R_{ijk}$. It can be shown that $\widehat{\omega}_{ik}$ is L_2 consistent for ω_{ik} under the asymptotic framework that n_1 and n_2 diverge but p is fixed.

Suppose $T(\mathbf{R})$ is a test statistic defined in terms of the ranks R_{ijk} and let $T(\mathbf{Y})$ and $T(\widehat{\mathbf{Y}})$ be the same test statistic calculated based on the asymptotic rank transforms Y_{ijk} and rank transforms $\widehat{Y}_{ijk}$, respectively. For our purpose, the test statistic will be a quadratic form in $\mathbf{R}$. To achieve weak convergence, we need the within-subject dependence to be regulated. Using a regularity condition on dependence, we prove a general result, which is useful to establish asymptotic equivalence of quadratic forms in rank transforms $\widehat{Y}_{ijk}$ and the analogous quadratic forms in asymptotic rank transforms Y_{ijk} under the high-dimensional asymptotic framework.

5.2.1 Regularity Condition on Dependence

A sequence of random variables $X_1, X_2, \ldots$ is said to be an α-mixing (strong mixing) sequence (process) with mixing coefficients $\{\alpha_k, k = 1, 2, \ldots\}$, if

$$\sup_{A \in \mathcal{A}_l, B \in \mathcal{B}_{k,l}, l \in \mathbb{Z}^+} |P(A \cap B) - P(A)P(B)| \leq \alpha_k, \text{ as } k \to \infty.$$

where

$$\mathcal{A}_l = \sigma\{X_1, \ldots, X_l\}, \quad \mathcal{B}_{k,l} = \sigma\{X_{l+k}, X_{l+k+1}, \ldots\}$$

and $\sigma(\cdot)$ denotes the σ-field generated by the random variables. Although this model for dependence is particularly suitable for repeated measures data, it can also be motivated for more general data (Xu et al., 2016). In repeated measures data, measurements corresponding to different subjects are independent and those corresponding to the same subject are assumed to satisfy an α-mixing condition. The α-mixing condition basically requires the dependence between two observations from the same subject to decay as the separation between the observations (k) increases.

Assumed α-mixing property on the process $\{X_{ijk}, k = 1, \ldots,\}$ for any i, j automatically transfers over to the process $\{Y_{ijk}, k = 1, \ldots,\}$. This fact, proved in Bradley (2005), is summarized in Lemma 5.2.1 for convenience. As multiple α-mixing sequences are involved in the lemma, we use identifier of the sequence in the notation for the mixing coefficient α_k as $\alpha(\boldsymbol{X}, k)$ for a given α-mixing sequence $\boldsymbol{X} = \{X_k, k \in \mathcal{Z}\}$, where $\mathcal{Z}$ is a countable index set.

Lemma 5.2.1. *(Theorem 5.2 in Bradley, 2005) Suppose that $\boldsymbol{X}_i = \{X_{ik}, k \in \mathcal{Z},\}$ is a sequence of α-mixing random variables, for each $i = 1, 2, 3, \ldots$, and they are independent of each other. Suppose that for each $k \in \mathcal{Z}$, $h_k : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \cdots \to \mathbb{R}$ is a Borel function. Define the sequence $\boldsymbol{U} = \{U_k, k \in \mathcal{Z}\}$ of random variables by $U_k = h_k(X_{1k}, X_{2k}, X_{3k}, \ldots), k \in \mathcal{Z}$. Then for each $m \geq 1$, the sequence $\mathbf{U}$ is α-mixing with mixing coefficient $\alpha(\boldsymbol{U}, m) \leq \sum_{i=1}^{\infty} \alpha(\boldsymbol{X}_i, m)$.*

In applying this lemma to our situation, by its almost everywhere continuity, the function $H(x)$ is a Borel functions with a single argument. Therefore, the sequence of ARTs $\{Y_{ijk}, k = 1, \ldots,\}$ is an α-mixing process with the same mixing coefficients as $\{X_{ijk}, k = 1, \ldots,\}$ for each $i = 1, 2$ and $j = 1, \ldots, n_i$.

5.2.2 Asymptotic Equivalence of Quadratic Forms

Due to their simple linear relationship, a test statistic in terms of ranks can be equivalently expressed in terms of the rank transforms (RT). The ART are asymptotic versions of RT at least for large n but fixed p in the sense that the two are asymptotically close in probability (e.g., Brunner et al., 1999). The next theorem will characterize the closeness (in the sense of L_2-norm) between quadratic forms in ART and RT for the high-dimensional situation. The principal utility of this result is that studying the asymptotic properties of quadratic forms in ARTs is relatively less involved because they are independent for different units (subjects).

Lemma 5.2.2. *Suppose that for each $i = 1, 2, 3, \ldots$, $\boldsymbol{X}_i = \{X_{ik}, k = 1, \ldots p\}$ is a sequence of α-mixing random variables with $\alpha_k = O(k^{-5})$. Suppose these sequences $\boldsymbol{X}_i, i = 1, \ldots n$ are independent of each other with marginal distribution $X_{ik} \sim F_{ik}$,* *for $k = 1, \ldots, p$. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^\top$ be a $n \times p$ matrix and $\boldsymbol{Y}$ and $\widehat{\boldsymbol{Y}}$ be the corresponding matrix of same dimension whose components are the asymptotic rank transforms and rank transforms, respectively, defined in Section 5.2. Let $\boldsymbol{C} = (c_{ik})$ be an $n \times n$ symmetric matrix with diagonals $c_{ii} = 0$. And let*

$$D_{\boldsymbol{C}} = \sum_{i=1}^{n} \sum_{k=1}^{n} |c_{ik}| \quad \textit{and} \quad S_{\boldsymbol{C}} = \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} |c_{ik} c_{il}| + \sum_{i \neq k}^{n} c_{ik}^2.$$

Furthermore, let

$$T_N = \mathrm{tr}(\widehat{\boldsymbol{Y}}_\omega^\top \boldsymbol{C} \widehat{\boldsymbol{Y}}_\omega) = \mathrm{Vec}(\widehat{\boldsymbol{Y}}_\omega)^\top (\boldsymbol{I}_p \otimes \boldsymbol{C}) \mathrm{Vec}(\widehat{\boldsymbol{Y}}_\omega)$$

and

$$V_N = \mathrm{tr}(\boldsymbol{Y}_\omega^\top \boldsymbol{C} \boldsymbol{Y}_\omega) = \mathrm{Vec}(\boldsymbol{Y}_\omega)^\top (\boldsymbol{I}_p \otimes \boldsymbol{C}) \mathrm{Vec}(\boldsymbol{Y}_\omega)$$

be two traces of $p \times p$-matrices of quadratic forms generated by matrix $\boldsymbol{C}$, where $\boldsymbol{Y}_\omega = \boldsymbol{Y} - \boldsymbol{\omega}$, $\widehat{\boldsymbol{Y}}_\omega = \widehat{\boldsymbol{Y}} - \boldsymbol{\omega}$ and $\boldsymbol{\omega}$ is the $n \times p$ matrix of the expectations $\omega_{ik} = \mathrm{E}(Y_{ik})$ for $i = 1, \ldots, n$, and $k = 1, \ldots, p$. Then,

$$\mathrm{E}\{(T_N - V_N)^2\} = O(D_{\boldsymbol{C}}^2/n^2) + O(S_{\boldsymbol{C}}/n),$$

as $n, p \to \infty$, while $p/n \to \eta \in (0, \infty)$.

The zero diagonal condition on the quadratic matrix C typically holds for quadratic forms arising from asymptotic manipulations of MANOVA and ANOVA decompositions (e.g., Akritas and Arnold, 2000). Asymptotic equivalence between quadratic forms of RT and ART has been considered by Bathke and Lankowski (2005) for univariate case and Bathke and Harrar (2008) in the multivariate case but small and bounded p. The major improvement established in Lemma 5.2.2 is covering the high-dimensional situation.

5.3 Test Statistic

A popular nonparametric test statistic literature (Brunner et al., 1997, 1999) is the ANOVA-type statistic. The test considers $N||\widehat{\boldsymbol{\omega}}_1 - \widehat{\boldsymbol{\omega}}_2||^2$. More precisely, the ANOVA-type statistic is defined by

$$Q = \frac{1}{N} ||\overline{\boldsymbol{R}}_{1\cdot} - \overline{\boldsymbol{R}}_{2\cdot}||^2$$

where $\overline{\boldsymbol{R}}_{i\cdot} = (\overline{R}_{i\cdot 1}, \ldots, \overline{R}_{i\cdot p})^\top$. Besides its simplicity, the ANOVA-type test has favorable small sample properties in terms of controlling Type I error rate and having power advantage over the popular Wald-type test (Brunner et al., 1997). When the dimension p is fixed, asymptotic theory seeks to establish root-n asymptotic equivalence between centered versions of $\widehat{\boldsymbol{\omega}}$ and averages of independent random variables where standard limit theorem are applied on the later (e.g., Akritas et al., 1997; Brunner et al., 1997, 1999). This manipulation is not applicable when the dimension as well as sample size go to infinity.

In the high-dimensional inference, the ANOVA-type test as defined above is not particularly convenient to work with. Let $\boldsymbol{R} = (\boldsymbol{R}_{11}, \ldots, \boldsymbol{R}_{1n_1}, \boldsymbol{R}_{21}, \ldots, \boldsymbol{R}_{2n_2})$ where $\boldsymbol{R}_{ij} = (R_{ij1}, \ldots, R_{ijp})^\top$. Harrar and Bathke (2008) studied the difference between the rank-based quadratic forms $\boldsymbol{H}(\boldsymbol{R})$ and $\boldsymbol{G}(\boldsymbol{R})$, where

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{R}) &= \boldsymbol{R}\left[\left(\bigoplus_{i=1}^{2}\frac{1}{n_i}\mathbf{1}_{n_i}\right)\boldsymbol{P}_2\left(\bigoplus_{i=1}^{2}\frac{1}{n_i}\mathbf{1}_{n_i}^\top\right)\right]\boldsymbol{R}^\top =: \boldsymbol{R}\boldsymbol{C}_1\boldsymbol{R}^\top \quad \text{and} \\ \boldsymbol{G}(\boldsymbol{R}) &= \frac{1}{2}\boldsymbol{R}\left[\bigoplus_{i=1}^{2}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}\right]\boldsymbol{R}^\top =: \boldsymbol{R}\boldsymbol{C}_2\boldsymbol{R}^\top \end{aligned} \tag{5.2}$$

to construct a valid nonparametric test. Here, $\bigoplus_{i=1}^{r} A_i$ is the block-diagonal matrix whose diagonal blocks are $A_1, \ldots, A_r$. Interestingly, $2\text{tr}(\boldsymbol{H}(\boldsymbol{R}) - \boldsymbol{G}(\boldsymbol{R}))$ is the same as the test statistic

$$T_n(\boldsymbol{R}) = \frac{\sum_{i\neq j}^{n_1}\boldsymbol{R}_{1i}^\top\boldsymbol{R}_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2}\boldsymbol{R}_{2i}^\top\boldsymbol{R}_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\boldsymbol{R}_{1i}^\top\boldsymbol{R}_{2j}}{n_1 n_2},$$

studied by Chen and Qin (2010) but defined on (mid-) ranks. When applied to the original data set $\boldsymbol{X}$, Chen and Qin (2010) noted that the L_2 based statistic of Bai and Saranadasa (1996) and the ANOVA-type statistic contain terms that are not useful for testing mean differences in high dimensions but rather complicate theoretical derivations. Therefore, the difference in $2\text{tr}(H - G)$ exclude terms that are asymptotically negligible, whereby making the asymptotic manipulation tractable without adverse consequence on performance. Furthermore, $E(\boldsymbol{H}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})) = \mathbf{0}$ if and only if $E(\boldsymbol{X}_{1i}) = E(\boldsymbol{X}_{2i})$. Motivated by these, we adapt the test statistic of

Chen and Qin (2010) and define it in terms of (mid-) ranks to make it useful for the nonparametric hypothesis.

Unlike most high-dimensional tests, no assumption on the covariances nor higher moments of the data are required in this Chapter. Indeed, none of the moments have to exist for the validity of the asymptotic results derived in this Chapter. To establish the weak convergence theory, however, we require the dependence among the different variables to be regulated by imposing a local dependence structure in the form of a strong mixing condition. Similar assumptions are also required, for example, in Cai and Xia (2014); Gregory et al. (2015); Xu et al. (2016) among others. Of particular note, Xu et al. (2016) provide a motivation for such type of condition in context of Genome-Wide Association Studies (GWAS).

Using the assumed regularity condition on the dependence, we establish a general result useful to establish asymptotic equivalence of quadratic forms in terms of the RT $\widehat{Y}_{ijk}$ and the analogous quadratic forms in ART Y_{ijk} under the high-dimensional asymptotic framework. This result will be instrumental because it allows us to establish the equivalence between the rank-based statistics $T_n(\boldsymbol{R})$ and its analog based on the asymptotic rank transform $T_n(\boldsymbol{Y})$. Note that the later one is a function $H(X_{ijk})$, for all i, j, k which are independent over for different values of i or j. Furthermore, H is the average of all marginal distribution functions. That is, it is uniformly bounded by 1, which guarantees the existence of all its moments. These two facts make $T_n(\boldsymbol{Y})$ amenable for treatment by existing high-dimensional results (e.g., Chen and Qin, 2010; Hu et al., 2017) under the relaxed conditions given in Chapter 4, cited as Kong and Harrar (2018a).

Denote $\boldsymbol{\omega}_i = E(\boldsymbol{Y}_{i1})$ and $\boldsymbol{\Sigma}_i = \mathrm{Var}(\boldsymbol{Y}_{i1})$ for $i = 1, 2$. Defining $\boldsymbol{Y}_{ij}^c = \boldsymbol{Y}_{ij} - \boldsymbol{\omega}_i$, $\widehat{\boldsymbol{Y}}_{ij}^c = \widehat{\boldsymbol{Y}}_{ij}^c - \boldsymbol{\omega}_i$ and $\boldsymbol{R}_{ij}^c = \boldsymbol{R}_{ij} - \boldsymbol{\omega}_{R,i}$, where $\boldsymbol{\omega}_{R,i} = \mathrm{E}[\boldsymbol{R}_{ij}] = N\boldsymbol{\omega}_i - \frac{1}{2}\mathbf{1}_p$. Let

$$\sigma_n^2 = \mathrm{Var}(T_n(\boldsymbol{Y}^c)) = \frac{2}{n_1(n_1-1)}\mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \frac{2}{n_2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_2^2) + \frac{4}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2).$$

It is obviously that

$$\frac{T_n(\boldsymbol{R}^c)}{N^2} = T_n(\widehat{\boldsymbol{Y}}^c) = T_n(\widehat{\boldsymbol{Y}}), \text{under } H_0.$$

Before we state the asymptotic results for the rank-based test, we list some conditions below for ease of reference.

$D1$: As $p \to \infty$, $\mathbf{X}_{ij} = \{X_{ijk}, k = 1, \ldots, p\}$ is an α-mixing random vector with $\alpha_m = O(m^{-5})$ as $p \to \infty$.

$D2$: The sample sizes diverge proportionally, i.e. $n_1/n \to \kappa \in (0, 1)$, where $n = n_1 + n_2$.

$D3$: The covariance matrices satisfy the regularity condition $\mathrm{tr}(\boldsymbol{\Sigma}_{i_1}\boldsymbol{\Sigma}_{i_2}\boldsymbol{\Sigma}_{i_3}\boldsymbol{\Sigma}_{i_4}) = o\left[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}\right]$ for $i_1, i_2, i_3, i_4 \in \{1, 2\}$.

$D4$: The mean vectors $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ satisfy $(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2)^\top \boldsymbol{\Sigma}_i (\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2) = o\left[\mathrm{tr}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\right]$.

$D5$: The covariance matrix $\boldsymbol{\Sigma}_i$, $i = 1, 2$, satisfies $\mathrm{tr}\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2)\} \to \infty$ as $p \to \infty$.

$D6$: $p/n \to \eta \in (0, \infty)$ as $n, p \to \infty$.

As it turns out, assumption $D1$ and $D6$ are sufficient to establish high-dimensional asymptotic equivalence between the test statistics defined in terms of ranks and defined in terms of ARTs.

Theorem 5.3.1. *Under assumptions $D1$ and $D6$,*

$$T_n(\widehat{\boldsymbol{Y}}^c) - T_n(\boldsymbol{Y}^c) = o_p(\sigma_n), \text{ as } n, p \to \infty.$$

Recall that, for each $i = 1, 2$, $\mathbf{Y}_{ij}$ are iid with mean $\boldsymbol{\omega}_i$ and covariance $\boldsymbol{\Sigma}_i$ for $j = 1, \ldots, n_i$. The components of $\boldsymbol{Y}_{ij}$ are bounded random variables and, hence, moments of any order exist. Therefore, by applying Theorem 4.3.3 (Theorem 3.3 of Kong and Harrar, 2018a), we get asymptotic normal distribution for $T(\mathbf{Y}^c)$.

Theorem 5.3.2. *Under assumptions $D1$–$D6$,*

$$\sigma_n^{-1} T_n(\boldsymbol{Y}^c) \xrightarrow{D} \mathcal{N}(0, 1), \text{ as } n, p \to \infty.$$

Theorems 5.3.1 and 5.3.2 afford us asymptotic normal distribution for the rank based test statistic $T(\boldsymbol{R})$.

Theorem 5.3.3. *Under the null hypothesis and assumptions D1–D6,*

$$\frac{T_n(\widehat{\boldsymbol{Y}})}{\sigma_n} \xrightarrow{D} \mathcal{N}(0,1), \text{ as } n, p \to \infty.$$

In terms of the ART vectors $\boldsymbol{Y}_{ij}$, unbiased and consistent estimator of σ_n^2 can be constructed by consistently estimating $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$. These estimators can be obtained by using the results in Kong and Harrar (2018a). Define

$$\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{1}{(n_i)_4} \sum_{k_1 \neq k_2 \neq l_1 \neq l_2}^{n_i} \mathrm{tr}\Big\{(\boldsymbol{Y}_{ik_1} - \boldsymbol{Y}_{ik_2})(\boldsymbol{Y}_{ik_1} - \boldsymbol{Y}_{ik_2})^\top (\boldsymbol{Y}_{il_1} - \boldsymbol{Y}_{il_2})(\boldsymbol{Y}_{il_1} - \boldsymbol{Y}_{il_2})^\top\Big\}, \tag{5.3}$$

and

$$\mathrm{tr}\widehat{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \frac{1}{(n_1)_2(n_2)_2} \sum_{k_1 \neq k_2}^{n_1} \sum_{l_1 \neq l_2}^{n_2} \mathrm{tr}\Big\{(\boldsymbol{Y}_{1k_1} - \boldsymbol{Y}_{1k_2})(\boldsymbol{Y}_{1k_1} - \boldsymbol{Y}_{1k_2})^\top (\boldsymbol{Y}_{2l_1} - \boldsymbol{Y}_{2l_2})(\boldsymbol{Y}_{2l_1} - \boldsymbol{Y}_{2l_2})^\top\Big\}, \tag{5.4}$$

where $(n_i)_k = n_i!/(n_i - k)!$.

Theorem 5.3.4. *An unbiased and, under assumptions D1–D6, a ratio-consistent estimator of σ_n^2 is*

$$\widehat{\sigma}_n^2 = \frac{2}{n_1(n_1-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)} + \frac{4}{n_1 n_2}\mathrm{tr}\widehat{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}.$$

The fact that the ART random vectors $\boldsymbol{Y}_{ij}$ are unobservable limits the application of this estimator in practice. A reasonable approach to fill the gap is to replace the ARTs $\boldsymbol{Y}_{ij}$ by the RT $\widehat{\boldsymbol{Y}}_{ij}$, their empirical version, to get a computable estimator. The resulting estimator of σ_n^2 will be denoted by $\widehat{\sigma}_n^2(\widehat{\boldsymbol{Y}}))$. Therefore, for an approximate size α test, we propose the test that rejects H_0 if $T_n(\widehat{\boldsymbol{Y}}))/\widehat{\sigma}_n(\widehat{\boldsymbol{Y}})) > z_\alpha$ where z_α is the $1 - \alpha$ quantile of the standard normal distribution.

5.4 Multi-group Test Statistic

To facilitate a formal extension of the two-group test to a multi-group situation, we recall that the two-group test statistic can be expressed as $T_n(\mathbf{R}) = 2\ tr(H(\mathbf{R}) -$

$G(\mathbf{R}))$ where $\boldsymbol{H}(\boldsymbol{R})$ and $\boldsymbol{G}(\boldsymbol{R})$ are as defined in (5.2). The matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ can be formally extended to the a-group situation as

$$\boldsymbol{C}_1 = \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right) \boldsymbol{P}_a \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right)^{\top} \text{ and } \boldsymbol{C}_2 = \left(1 - \frac{1}{a}\right) \bigoplus_{i=1}^{a} \frac{1}{n_i(n_i - 1)} \boldsymbol{P}_{n_i}. \quad (5.5)$$

Now let $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^{\top}$ be the jth observation vector in the ith group and assume $\boldsymbol{X}_{ij}$ are iid with joint distribution F_i for $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$. With the matrices in (5.5), it can be shown that a formal extension of the test statistic to the a-group situation is

$$\begin{aligned} T_{a,n}(\boldsymbol{R}) &= a \ \mathrm{tr}(\boldsymbol{H}(\boldsymbol{R}) - \boldsymbol{G}(\boldsymbol{R})) \\ &= (a-1) \sum_{i=1}^{a} \frac{\sum_{k \neq l}^{n_i} \boldsymbol{R}_{ik}^{\top} \boldsymbol{R}_{il}}{n_i(n_i - 1)} - 2 \sum_{i < i_1}^{a} \frac{\sum_{k=1}^{n_i} \sum_{l=1}^{n_{i_1}} \boldsymbol{R}_{ik}^{\top} \boldsymbol{R}_{i_1 l}}{n_i n_{i_1}} \end{aligned}$$

where $\boldsymbol{R}_{ij} = (R_{ij1}, \ldots, R_{ijp})^{\top}$ and R_{ijk} is the (mid-) rank of $\mathbf{X}_{ijk}$ among all $N = n \times p$ observations and $n = n_1 + \cdots + n_a$. Along the same lines we get asymptotic normal distribution analogous to Theorem 5.3.2 in the a-group situation with σ_n^2 defined by

$$\sigma_n^2 = \sum_{i=1}^{a} \frac{2(a-1)^2}{n_i(n_i - 1)} \mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \sum_{i<j}^{a} \frac{4}{n_1 n_2} \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j).$$

where $\boldsymbol{\Sigma}_i = \mathrm{Var}(\boldsymbol{Y}_{i1})$, $\boldsymbol{Y}_{i1} = (H(X_{i11}), \ldots, H(X_{i1p}))^{\top}$ and

$$H(x) = N^{-1} \sum_{k=1}^{p} \sum_{i=1}^{a} n_i F_{ik}(x).$$

With estimator of σ_n^2 also defined analogously, an approximate size α test can be constructed.

5.5 Simulation Study

In this section, we report results of simulation study intended to compare the empirical sizes and powers of the rank-based test, the CQ test defined in Chen and Qin (2010) and the SK test defined in Srivastava and Kubokawa (2013). Data for the ith group is generated from:

(i) Multivariate normal distribution with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$, $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

(ii) Multivariate t distribution with location vector $\boldsymbol{\mu}_i$, and scale matrix $\boldsymbol{\Sigma}_i$ but degrees of freedom fixed at $\nu_1 = 6$ and $\nu_2 = 8$ in groups 1 and 2, respectively.

(iii) Contaminated multivariate normal distribution with pdf

$$f_i(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \alpha_i, \eta_i) = \alpha_i \phi(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + (1 - \alpha_i)\phi(\boldsymbol{x}|\boldsymbol{\mu}_i, \eta_i \boldsymbol{\Sigma}_i)$$

where $\phi(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of the multivariate normal $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The other parameters are fixed at $\eta_1 = 5$, $\alpha_1 = 0.5$, $\eta_2 = 3$, and $\alpha_2 = 0.1$

(iv) Multivariate Cauchy distribution with location vector $\boldsymbol{\mu}_i$ and scale matrix $\boldsymbol{\Sigma}_i$

These distributions represent light, moderately-heavy and very-heavy tailed distributions with the possibility of getting outliers.

The empirical size of the rank-based, CQ and SK tests are presented in Tables 5.1–5.4 in which we fix $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$ and consider three settings for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, here in after denoted by $\boldsymbol{\Sigma}_{l1}$ and $\boldsymbol{\Sigma}_{l2}$, respectively, for $l = 1, 2, 3$. The three settings are:

$l = 1$**:** $\boldsymbol{\Sigma}_{11} = 0.5\boldsymbol{I}_p + 0.5\boldsymbol{J}_p$ and $\boldsymbol{\Sigma}_{12} = 0.9\boldsymbol{I}_p + 0.1\boldsymbol{J}_p$,

$l = 2$**:** $\boldsymbol{\Sigma}_{21} = (0.5^{|j-j_1|})$ and $\boldsymbol{\Sigma}_{22} = (0.1^{|j-j_1|})$,

$l = 3$**:** $\boldsymbol{\Sigma}_{31} = (0.5|j - j_1|^{-1/2})$ and $\boldsymbol{\Sigma}_{32} = (0.1|j - j_1|^{-1/2})$.

The sizes as well as powers are calculated for $10,000$ replications and the actual level of significance is set at $\alpha = 0.05$.

From Tables 5.1–5.4, the performance of CQ and the new rank-based method are about the same for all covariance structures and population distributions except that the rank method show a liberal tendency when the covariance between observations is constant or decays at a slow rate ($k^{-1/2}$). Regardless, the quality of approximation generally improves as the sample size and dimension increase. The hypothesis for the CQ method is equality of mean vectors where means do not even exist for the Cauchy distribution. Despite all the strong moment assumption made, its performance in terms of the empirical size is comparably well to the rank-based method. SK is designed for equal covariance situation. However, it showed reasonable performance

Table 5.1: Achieved Type I error rate for multivariate normal distribution with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$ and three different pairs of $(\boldsymbol{\Sigma}_{l1}, \boldsymbol{\Sigma}_{l2})$ for $l = 1, 2, 3$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	SK	Rank	CQ	SK	Rank	CQ	SK	Rank
50	(50, 90)	0.074	0.089	0.077	0.061	0.062	0.063	0.062	0.069	0.061
	(100,150)	0.068	0.059	0.071	0.059	0.055	0.057	0.068	0.068	0.067
	(200,240)	0.072	0.042	0.078	0.055	0.047	0.057	0.063	0.058	0.062
100	(50, 90)	0.074	0.074	0.075	0.057	0.060	0.060	0.059	0.069	0.061
	(100,150)	0.072	0.056	0.076	0.061	0.061	0.063	0.065	0.070	0.064
	(200,240)	0.066	0.028	0.070	0.059	0.054	0.059	0.060	0.057	0.059
200	(50, 90)	0.072	0.057	0.075	0.058	0.061	0.061	0.066	0.082	0.067
	(100,150)	0.071	0.040	0.074	0.058	0.058	0.057	0.063	0.070	0.062
	(200,240)	0.070	0.023	0.074	0.057	0.053	0.058	0.068	0.067	0.069
400	(50, 90)	0.077	0.049	0.079	0.056	0.057	0.055	0.064	0.082	0.062
	(100,150)	0.073	0.031	0.077	0.055	0.056	0.055	0.057	0.069	0.059
	(200,240)	0.070	0.017	0.074	0.057	0.053	0.057	0.060	0.062	0.059
800	(50, 90)	0.073	0.032	0.076	0.052	0.050	0.053	0.060	0.075	0.058
	(100,150)	0.066	0.021	0.070	0.055	0.054	0.057	0.060	0.071	0.060
	(200,240)	0.069	0.010	0.073	0.056	0.052	0.054	0.061	0.063	0.061
1600	(50, 90)	0.077	0.024	0.080	0.053	0.047	0.052	0.059	0.075	0.061
	(100,150)	0.075	0.015	0.077	0.053	0.050	0.051	0.061	0.072	0.061
	(200,240)	0.072	0.007	0.076	0.049	0.045	0.049	0.059	0.059	0.057

Table 5.2: Achieved Type I error rate for multivariate t distribution with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$, degrees freedom $\nu_1 = 6$, $\nu_2 = 8$, and three different pairs of $(\boldsymbol{\Sigma}_{l1}, \boldsymbol{\Sigma}_{l2})$ for $l = 1, 2, 3$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	SK	Rank	CQ	SK	Rank	CQ	SK	Rank
50	(50, 90)	0.073	0.091	0.077	0.060	0.064	0.061	0.069	0.076	0.065
	(100,150)	0.071	0.065	0.075	0.064	0.064	0.065	0.065	0.068	0.064
	(200,240)	0.069	0.041	0.073	0.064	0.057	0.064	0.063	0.057	0.063
100	(50, 90)	0.077	0.078	0.075	0.063	0.065	0.062	0.068	0.079	0.066
	(100,150)	0.071	0.053	0.078	0.061	0.061	0.061	0.059	0.067	0.059
	(200,240)	0.072	0.036	0.081	0.059	0.051	0.058	0.061	0.055	0.062
200	(50, 90)	0.069	0.055	0.074	0.059	0.048	0.057	0.064	0.071	0.064
	(100,150)	0.070	0.041	0.076	0.059	0.052	0.060	0.061	0.062	0.059
	(200,240)	0.070	0.025	0.073	0.059	0.046	0.062	0.065	0.057	0.066
400	(50, 90)	0.073	0.048	0.076	0.057	0.033	0.057	0.068	0.060	0.063
	(100,150)	0.071	0.031	0.080	0.056	0.039	0.059	0.059	0.055	0.061
	(200,240)	0.071	0.018	0.074	0.055	0.036	0.051	0.059	0.049	0.062
800	(50, 90)	0.073	0.033	0.074	0.056	0.013	0.054	0.058	0.036	0.058
	(100,150)	0.071	0.022	0.074	0.053	0.018	0.054	0.058	0.042	0.056
	(200,240)	0.067	0.010	0.073	0.055	0.018	0.054	0.056	0.037	0.056
1600	(50, 90)	0.077	0.023	0.080	0.052	0.004	0.052	0.060	0.020	0.060
	(100,150)	0.075	0.013	0.075	0.050	0.005	0.048	0.063	0.028	0.062
	(200,240)	0.068	0.007	0.075	0.052	0.006	0.053	0.059	0.024	0.058

Table 5.3: Achieved Type I error rate for multivariate contaminate normal distribution with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$, $\eta_1 = 5$, $\alpha_1 = 0.5$, $\eta_2 = 3$, $\alpha_2 = 0.1$, and three different pairs of $(\boldsymbol{\Sigma}_{l1}, \boldsymbol{\Sigma}_{l2})$ for $l = 1, 2, 3$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	SK	Rank	CQ	SK	Rank	CQ	SK	Rank
	(50, 90)	0.072	0.110	0.078	0.063	0.276	0.066	0.073	0.219	0.079
50	(100,150)	0.069	0.075	0.078	0.066	0.189	0.067	0.067	0.151	0.073
	(200,240)	0.068	0.043	0.080	0.066	0.102	0.064	0.062	0.085	0.071
	(50, 90)	0.071	0.096	0.075	0.065	0.391	0.065	0.065	0.284	0.070
100	(100,150)	0.073	0.067	0.082	0.056	0.243	0.056	0.067	0.187	0.073
	(200,240)	0.069	0.036	0.081	0.060	0.120	0.060	0.060	0.093	0.069
	(50, 90)	0.071	0.079	0.079	0.057	0.543	0.058	0.066	0.393	0.071
200	(100,150)	0.072	0.050	0.079	0.059	0.356	0.058	0.062	0.253	0.073
	(200,240)	0.071	0.028	0.082	0.060	0.144	0.058	0.064	0.113	0.073
	(50, 90)	0.079	0.069	0.087	0.054	0.739	0.054	0.062	0.538	0.070
400	(100,150)	0.073	0.040	0.081	0.057	0.521	0.058	0.063	0.344	0.070
	(200,240)	0.070	0.020	0.076	0.059	0.189	0.057	0.061	0.135	0.068
	(50, 90)	0.073	0.047	0.079	0.053	0.896	0.057	0.059	0.711	0.065
800	(100,150)	0.072	0.028	0.077	0.049	0.702	0.052	0.064	0.482	0.070
	(200,240)	0.070	0.011	0.078	0.054	0.247	0.054	0.064	0.168	0.072
	(50, 90)	0.072	0.033	0.078	0.047	0.978	0.049	0.061	0.876	0.065
1600	(100,150)	0.071	0.016	0.081	0.054	0.882	0.053	0.060	0.654	0.066
	(200,240)	0.066	0.007	0.076	0.052	0.316	0.049	0.059	0.209	0.065

Table 5.4: Achieved Type I error rate (×100%) for multivariate Cauchy distribution with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$, and three different pairs of $(\boldsymbol{\Sigma}_{l1}, \boldsymbol{\Sigma}_{l2})$ for $l = 1, 2, 3$.

		$\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{12}$			$\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$			$\boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_{32}$		
p	(n_1, n_2)	CQ	SK	Rank	CQ	SK	Rank	CQ	SK	Rank
	(50, 90)	0.071	0.010	0.070	0.056	0.000	0.060	0.057	0.001	0.066
50	(100,150)	0.071	0.006	0.070	0.051	0.000	0.065	0.056	0.000	0.066
	(200,240)	0.070	0.003	0.073	0.053	0.000	0.066	0.060	0.000	0.062
	(50, 90)	0.071	0.007	0.070	0.051	0.000	0.060	0.053	0.000	0.063
100	(100,150)	0.073	0.005	0.075	0.049	0.000	0.061	0.056	0.000	0.059
	(200,240)	0.073	0.003	0.074	0.051	0.000	0.059	0.054	0.000	0.063
	(50, 90)	0.075	0.004	0.075	0.048	0.000	0.057	0.054	0.000	0.057
200	(100,150)	0.075	0.002	0.073	0.049	0.000	0.057	0.054	0.000	0.058
	(200,240)	0.074	0.001	0.075	0.048	0.000	0.056	0.055	0.000	0.063
	(50, 90)	0.072	0.004	0.074	0.046	0.000	0.055	0.054	0.000	0.061
400	(100,150)	0.074	0.001	0.072	0.049	0.000	0.058	0.054	0.000	0.064
	(200,240)	0.073	0.000	0.069	0.048	0.000	0.056	0.049	0.000	0.055
	(50, 90)	0.069	0.001	0.076	0.045	0.000	0.052	0.051	0.000	0.063
800	(100,150)	0.079	0.001	0.074	0.046	0.000	0.052	0.051	0.000	0.058
	(200,240)	0.071	0.001	0.073	0.046	0.000	0.056	0.054	0.000	0.056
	(50, 90)	0.075	0.001	0.075	0.046	0.000	0.053	0.051	0.000	0.060
1600	(100,150)	0.072	0.000	0.072	0.044	0.000	0.053	0.051	0.000	0.059
	(200,240)	0.068	0.000	0.071	0.046	0.000	0.049	0.054	0.000	0.057

for the lighter tail (multivariate normal and t with df ≥ 6) distributions but can not be recommended for heavy tailed populations.

The power plots of the rank-based test, CQ test and SK test are displayed in Figures 5.1–5.3. To keep the investigation manageable, while showing the essential features, we limit the power plots to multivariate t, multivariate contaminated-normal and multivariate Cauchy populations. For the alternative points, we consider $\boldsymbol{\mu}_1 = \mathbf{0}_p$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \ldots, \mu_{2p})^\top$ where μ_{2k} are iid Uniform$(0, \delta)$ for $\delta \in \{0, 0.1, 0.2, \ldots, 1\}$. In the left panel of each figure, the covariance structure $l = 2$ ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{22}$) are used and in the right panel the covariance structure $l = 3$ ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_{31}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{32}$) are used. The sample sizes will be fixed at $n_1 = 100$ and $n_2 = 150$, but three different dimensions $p = 50, 100$ or 200 are investigated.

Figure 5.1: Power comparison of the test for the locations when $a = 2$, for $p = 50, 100$ or 200, and $n_1 = 100$ and $n_2 = 150$. Data are generated from multivariate t distribution with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ with degrees of freedom ν_i for $i = 1, 2$. In both plots, $\boldsymbol{\mu}_1 = \mathbf{0}_p$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \ldots, \mu_{2p})^\top$ where μ_{2k} are iid Uniform$(0, \delta)$; $\nu_1 = 6$ and $\nu_2 = 8$. In the left panel, $\boldsymbol{\Sigma}_1 = (0.5^{|j-j_1|})$ and $\boldsymbol{\Sigma}_2 = (0.1^{|j-j_1|})$ are used. In the right panel, $\boldsymbol{\Sigma}_1 = (0.5|j - j_1|^{-1/2})$ and $\boldsymbol{\Sigma}_2 = (0.1|j - j_1|^{-1/2})$ are used.

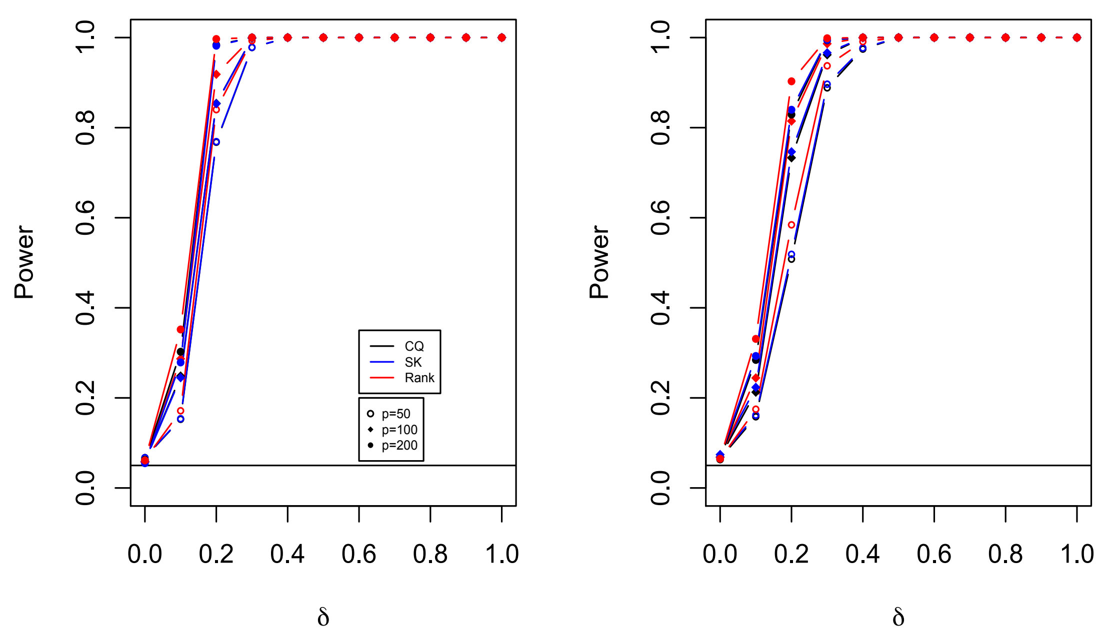
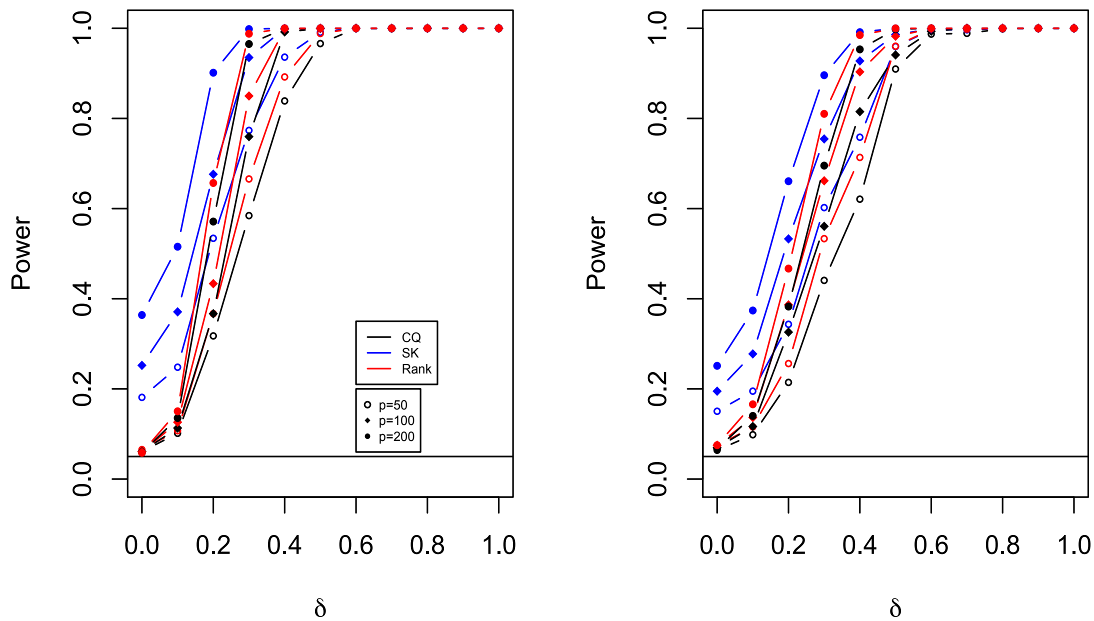
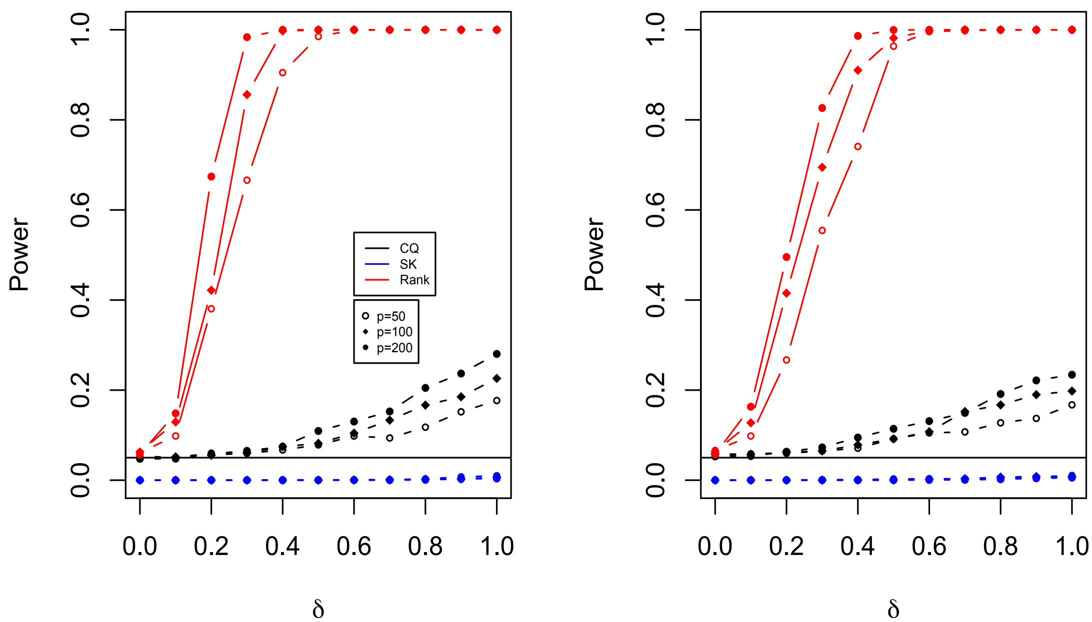


Figure 5.2: Power comparison of the test for the locations when $a = 2$, for $p = 50, 100$ or 200, and $n_1 = 100$ and $n_2 = 150$. Data are generated from multivariate contaminate normal distribution with $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, η_i and α_i for $i = 1, 2$. In both plots, $\boldsymbol{\mu}_1 = \mathbf{0}_p$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \ldots, \mu_{2p})^\top$ where μ_{2k} are iid Uniform$(0, \delta)$; $\eta_1 = 5$ and $\eta_2 = 3$; $\alpha_1 = 0.5$ and $\alpha_2 = 0.1$. In the left panel, $\boldsymbol{\Sigma}_1 = (0.5^{|j-j_1|})$ and $\boldsymbol{\Sigma}_2 = (0.1^{|j-j_1|})$ are used. In the right panel, $\boldsymbol{\Sigma}_1 = (0.5|j - j_1|^{-1/2})$ and $\boldsymbol{\Sigma}_2 = (0.1|j - j_1|^{-1/2})$ are used.



It is clear from Figure 5.1 that the performances of all the three methods (CQ, SK and rank-based) are comparably well for the lighter tail (multivariate t with df ≥ 6) distribution, but rank-based method has a slight edge. For the contaminated-normal distribution (Figure 5.2), SK shows a liberal tendency (see also Table 5.3), but the other two perform well. Here, rank-method has a more pronounced edge over CQ. For Cauchy distributions, the rank-based method which does not require existence of any moments of the population shows an overwhelming power advantage over the other two methods. For all the three distributions, the faster decaying covariances in structure $l = 2$ yield higher power compared to the slower decaying ones in structure $l = 3$. Furthermore, for the alternatives considered in this simulation, larger values of p lead to higher powers than a smaller value of p.

Figure 5.3: Power comparison of the test for the locations when $a = 2$, for $p = 50, 100$ or 200, and $n_1 = 100$ and $n_2 = 150$. Data are generated from multivariate Cauchy distribution with $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, for $i = 1, 2$. In both plots, $\boldsymbol{\mu}_1 = \mathbf{0}_p$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \ldots, \mu_{2p})^\top$ where μ_{2k} are iid Uniform$(0, \delta)$. In the left panel, $\boldsymbol{\Sigma}_1 = (0.5^{|j-j_1|})$ and $\boldsymbol{\Sigma}_2 = (0.1^{|j-j_1|})$ are used. In the right panel, $\boldsymbol{\Sigma}_1 = (0.5|j - j_1|^{-1/2})$ and $\boldsymbol{\Sigma}_2 = (0.1|j - j_1|^{-1/2})$ are used.



5.6 Real Data Application

The Electroencephalograph (EEG) data[1] found at the University of California-Irvine Machine Learning Repository was from a large study to examine EEG correlates of genetic predisposition to alcoholism. Sixty-four electrodes were used to measure Event-Related Potentials (ERP) recorded 256 times for one second. Each channel (electrode) has name identifying the location of the electrode on the scalp. The names are made up of a letter identifying the anatomical location of the placement of the electrode (F–frontal lobe, T–temporal lobe, P–parietal lobe and O–occipital lobe) and a number identifying the hemisphere of the brain (odd number – the left hemisphere and even number – the right hemisphere and letter z (zero) is used for

[1]Web Address: https://archive.ics.uci.edu/ml/datasets/EEG%2BDatabase

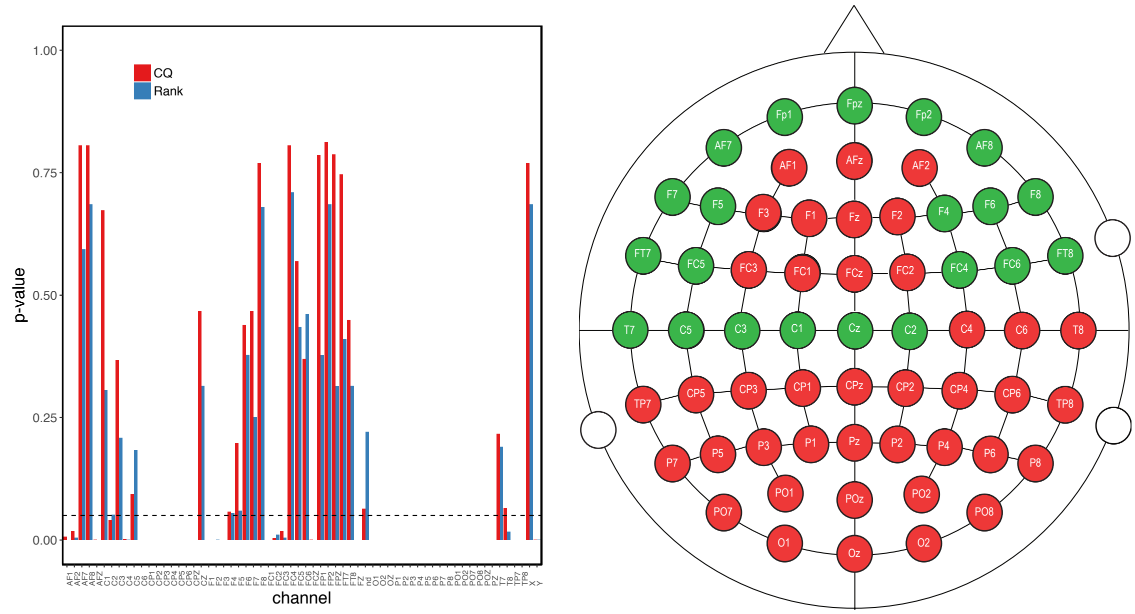
the mid-line). The exception to this naming rule is that, due to their placement and depending on the individual, the “C” electrodes can exhibit/represent EEG activity more typical of Frontal, Temporal, and some Parietal-Occipital activity.

ERP reading from an electrode indicates the level of electrical activity (in volts) in the region of the brain where the electrode is placed. There are two groups of subjects in the study: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from a picture set. For a more detailed account of the EEG data, see Harrar and Kong (2016). In this Chapter, we analyze the data only for the single stimulus (S1) exposure using CQ and the rank method.

FDR adjusted p-values for channel-by-channel results of CQ test and rank-based method are displayed in Figure 5.4. In the left panel, bar plot of the FDR adjusted p-values are shown. The horizontal reference line (black dashed line) marks $\alpha = 0.05$ level of significance. From panel (a), we note that the rank-based method declares the brain activity of one more channel to be significantly different compared to the CQ method. Considering the power advantage the rank-method demonstrated in the simulation study, its results are more reliable and trustworthy. The minor disagreement aside, the locations where differences are detected by rank-method are displayed in panel (b) to put the results in perspective. The picture depicts the scalp of a human viewed from the top, the triangle marking the nose. The locations of the electrodes are indicated by bubbles. The color of the bubbles indicates whether the brain activity pattern for that channel is significantly dissimilar (red) or not significantly different (green).

Interestingly, the results show a markedly-distinct patch of significant difference in brain activity in the central part of the frontal lobe of the brain. This section of the frontal lobe is responsible for cognitive function, emotion control, self awareness, judgement and talking – activities known to be affected by alcohol at least temporarily. No significance difference was found in the outer peripheral channels of the frontal lobe. Significant difference occurs only on some the C channels. No significant difference was detected on the C channels that are expected to show frontal-type activity

Figure 5.4: Channel-by-Channel results for EEG data for testing equality in brain activity between alcoholic and control subjects. (a) Bar plots of FDR adjusted p-values for CQ and rank-based methods (b) Locations (on the scalp) of significant results for rank-based method. Green (Red) means that the the difference between the two groups is statistically significant (insignificant).



(a) CQ and Rank methods

(b) Rank method

except on channels C4 and C6. However, there is always significant difference on channels where parental or occipital type activity is expected. With the exception of T7, the other three temporal lobe channels (T8, TP7 and TP8) are showing significant difference. The activity levels in all the parietal or occipital lobes channels are significantly different between the two groups. These two lobes largely control temperature, taste, touch, movement and vision functions – functions that are likely to sustain effects from alcohol use. In summary, except in the peripheral areas of the frontal lobe, alcohol use is associated with change in the electrical activity of the brain.

5.7 Discussion and Conclusion

A fully nonparametric high-dimensional rank-based method for comparison of treatments or populations is developed. No assumption is made on the distribution of the population except that the dependence between the variables are required to satisfy some mild conditions. The assumptions, for example, hold for linear process type dependence or dependence that decays polynomially fast. The numerical results have unequivocally shown that when data has heavy tails to preclude existence of moments, then the rank-method has a superior power. From data analysis perspective, the application of this method would be a safe strategy when data is in ordinal scale or exhibits outliers. From theoretical stand point, when none of the moments can be assumed to exist, formulation of hypothesis in terms of mean vectors does not make much sense. This Chapter formulates hypothesis in terms of the nonparametric measure of effect, which is always well defined whether the moments or densities exist or not.

The theory is worked out in detail for the two-group (two-sample) situation and the extension to the multi-group case is outlined. The detail derivation in the later case essentially follows along the same lines. The formal extension to a factorial structure is not difficult to envision. The details, with the necessary assumptions, need to be carefully examined. We defer this topic for a future investigation.

5.8 Appendix: Proofs

For notation simplification, denote $\boldsymbol{X}_{1j}$ as $\boldsymbol{X}_j$ and $\boldsymbol{X}_{2j}$ as $\boldsymbol{X}_{n_1+j}$. Let

$$g_1(X_{ik}) = H(X_{ik}) - \omega_{ik}, \quad g_2(X_{ik}) = \widehat{H}(X_{ik}) - H(X_{ik}),$$

and

$$G(X_{ik}, X_{i_1k_1}) = \frac{1}{N}\big\{c(X_{ik}, X_{i_1k_1}) - F_{i_1k_1}(X_{ik})\big\}.$$

It is obvious that

$$g_2(X_{ik}) = \sum_{i_1=1}^{n}\sum_{k_1=1}^{p} G(X_{ik}, X_{i_1k_1}),$$

g_1 and g_2 are bounded by 1 and G is bounded by $1/N$. The mean of $g_1(X_{ik})$ is 0 and that of $G(X_{ik}, X_{i_1k_1})$ can be found by considering the following three cases:

(i) If $i \neq i_1$, then

$$\begin{aligned}
&\mathrm{E}\{G(X_{ik}, X_{i_1k_1})\} = \frac{1}{N}\mathrm{E}\Big[\mathrm{E}\big\{c(X_{ik}, X_{i_1k_1}) - F_{i_1k_1}(X_{ik})|X_{ik}\big\}\Big] \\
&= \frac{1}{N}\mathrm{E}\{\frac{1}{2}P(X_{i_1k_1} = X_{ik}|X_{ik}) + P(X_{i_1k_1} < X_{ik}|X_{ik}) - F_{i_1k_1}(X_{ik})\big\} \\
&= \frac{1}{N}\mathrm{E}\{F_{i_1k_1}(X_{ik}) - F_{i_1k_1}(X_{ik})\} = 0
\end{aligned}$$

(ii) If $i = i_1$ and $k = k_1$, then $G(X_{ik}, X_{ik}) = \frac{1}{2} - F_{ik}(X_{ik})$, which has mean 0.

(iii) If $i = i_1$ but $k \neq k_1$, then $G(X_{ik}, X_{ik_1})$ is bounded by $\frac{1}{N}$, so is its mean.

Proof of Lemma 5.2.2. Using a decomposition similar to Bathke and Lankowski (2005) (seel also, Wang and Akritas, 2010b),

$$\begin{aligned}
T_N &= \mathrm{Vec}(\widehat{\boldsymbol{Y}}_\omega)^\top(\boldsymbol{I}_p \otimes \boldsymbol{C})\mathrm{Vec}(\widehat{\boldsymbol{Y}}_\omega) \\
&= \mathrm{Vec}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y} + \boldsymbol{Y} - \boldsymbol{\omega})^\top(\boldsymbol{I}_p \otimes \boldsymbol{C})\mathrm{Vec}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y} + \boldsymbol{Y} - \boldsymbol{\omega}) \\
&= V_N + 2\mathrm{Vec}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top(\boldsymbol{I}_p \otimes \boldsymbol{C})\mathrm{Vec}(\boldsymbol{Y} - \boldsymbol{\omega}) + \mathrm{Vec}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top(\boldsymbol{I}_p \otimes \boldsymbol{C})\mathrm{Vec}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y}) \\
&= V_N + 2\mathrm{tr}\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\boldsymbol{Y} - \boldsymbol{\omega})\big\} + \mathrm{tr}\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})\big\},
\end{aligned}$$

we have

$$\mathrm{E}\{(T_N - V_N)^2\} \leq 8\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\boldsymbol{Y} - \boldsymbol{\omega})\big\}\Big] + 2\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})\big\}\Big].$$

Therefore, it will be sufficient to show that

$$\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\boldsymbol{Y} - \boldsymbol{\omega})\big\}\Big] = O(D_{\boldsymbol{C}}^2/n^2) + O(S_{\boldsymbol{C}}/n), \tag{5.6}$$

and

$$\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})\big\}\Big] = O(D_{\boldsymbol{C}}^2/n^2) + O(S_{\boldsymbol{C}}/n). \tag{5.7}$$

To prove (5.6), observe that

$$\begin{aligned}
&\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^\top\boldsymbol{C}(\boldsymbol{Y} - \boldsymbol{\omega})\big\}\Big] = \mathrm{E}\Big[\big\{\sum_{i\neq j}^n c_{ij}\sum_{k=1}^p g_1(X_{ik})g_2(X_{jk})\big\}^2\Big] \\
&\leq \sum_{i\neq j, i_1\neq j_1}^n |c_{ij}|\ |c_{i_1j_1}|\mathrm{E}\Big\{\sum_{k=1}^p g_1(X_{ik})g_2(X_{jk})\sum_{k_1=1}^p g_1(X_{i_1k_1})g_2(X_{j_1k_1})\Big\} \\
&= \sum_{i\neq j, i_1\neq j_1}^n |c_{ij}|\ |c_{i_1j_1}|\Big[\sum_{k,k_1,k_2,k_3}^p\sum_{j_2,j_3}^n \mathrm{E}\Big\{g_1(X_{ik})g_1(X_{i_1k_1})G(X_{jk}, X_{j_2k_2})G(X_{j_1k_1}, X_{j_3k_3})\Big\}\Big].
\end{aligned}$$

Note that the summation of the expectations is zero if the number of different indices in the set $\{i, i_1, j, j_1, j_2, j_3\}$ is five or six. We consider several cases to evaluate the term in the square bracket,

$$\sum_{k,k_1,k_2,k_3}^{p} \sum_{j_2,j_3} \mathrm{E}\Big\{g_1(X_{ik})g_1(X_{i_1k_1})G(X_{jk}, X_{j_2k_2})G(X_{j_1k_1}, X_{j_3k_3})\Big\}.$$

Case 1: If j_2, j_3 are both equal to one of the indices i, i_1, j or j_1 and i, i_1, j, j_1 are all different. The summation of expectations vanishes except when $\{j_2 = i, j_3 = i_1\}$ or $\{j_2 = i_1, j_3 = i\}$. For each of these situations, the summation is $O(1/n^2)$ because

$$\begin{aligned}
&\sum_{k,k_1,k_2,k_3}^{p} \mathrm{E}\Big\{g_1(X_{ik})g_1(X_{i_1k_1})G(X_{jk}, X_{ik_2})G(X_{j_1k_1}, X_{i_1k_3})\Big\} \\
&= \sum_{k,k_1,k_2,k_3}^{p} \mathrm{E}\Big[\mathrm{E}\Big\{g_1(X_{ik})G(X_{jk}, X_{ik_2})g_1(X_{i_1k_1})G(X_{j_1k_1}, X_{i_1k_3})|X_{jk}, X_{j_1k_1}\Big\}\Big] \\
&\leq 16 \sum_{k,k_1,k_2,k_3}^{p} \alpha_{|k_2-k|}\alpha_{|k_3-k_1|}\frac{1}{N^2} = O(p^2/N^2) = O(1/n^2),
\end{aligned}$$

where the last inequality follows from Lemma 2 of Billingsley (2012, Section 27).

Case 2: If j_2, j_3 are both equal to one of the indices i, i_1, j or j_1 and there are three different numbers in $\{i, i_1, j, j_1\}$. We breakdown this case into two sub-cases: $i = i_1$ or $i \neq i_1$. In any these cases, we can prove that the summation of expectations (with their respective coefficients $c_{ij}c_{ij_1}$, $c_{ij}c_{i_1i}$ or $c_{ii_1}c_{i_1j}$) is $O(p^3/N^2) = O(1/n)$.

Case 3: If j_2, j_3 are both equal to one of the indices i, i_1, j or j_1 and there are two different numbers in $\{i, i_1, j, j_1\}$. Itt must be that $i = i_1$ and $j = j_1$, or $i = j_1$ and $j = i_1$ because $i \neq j$ and $i_1 \neq j_1$ ($c_{ij} = 0$ and $c_{i_1j_1} = 0$). There are four different possible values for j_2 and j_3: $\{j_2 = j_3 = i\}$, $\{j_2 = j_3 = j\}$, $\{j_2 = i, j_3 = j\}$, or $\{j_2 = j, j_3 = i\}$. For each combination, the summation of expectations (with their corresponding coefficients c_{ij}^2 or $c_{ij}c_{ji}$) is $O(p^3/N^2) = O(1/n)$. To see this, we only prove here the case when $i = j_1, i_1 = j, j_2 = j, j_3 = i$ and k, k_1, k_2, k_3 are all different. The proofs of the other cases follow along the same lines. Indeed,

$$\sum_{k\neq k_1\neq k_2\neq k_3}^{p} g_1(X_{ik})G(X_{ik_1},X_{ik_3})g_1(X_{jk_1})G(X_{jk},X_{jk_2})$$
$$=\sum_{k\neq k_1\neq k_2\neq k_3}^{p} \mathrm{E}\Big\{g_1(X_{ik})G(X_{ik_1},X_{ik_3})\Big\}\mathrm{E}\Big\{g_1(X_{jk_1})G(X_{jk},X_{jk_2})\Big\}$$
$$\leq \sum_{k\neq k_1\neq k_2\neq k_3}^{p} \frac{24^2}{N^2}\alpha_{|k_1-k|}=O(p^3/N^2),$$

while the last inequality follows because

$$\mathrm{E}\Big\{g_1(X_{ik})G(X_{ik_1},X_{ik_3})\Big\}=\frac{1}{N}\mathrm{E}\Big[\mathrm{E}\Big\{g_1(X_{ik})G_2(X_{ik_1},X_{ik_3})|X_{ik},X_{ik_1}\Big\}\Big]$$
$$=\frac{1}{N}\mathrm{E}\Big[g_1(X_{ik})\Big\{F_{X_{ik_3}|X_{ik},X_{ik_1}}(X_{ik_1})-F_{ik_3}(X_{ik_1})\Big\}\Big]\leq\frac{24}{N}\alpha^{1/2}_{|k_1-k|}$$

by Lemma 3 of Billingsley (2012, Section 27).

Case 4: When j_2 equals one of the indices i, i_1, j, j_1, but j_3 is different from all of them, then $X_{j_3k_3}$ is independent of the others, so the summation of the expectation vanishes since

$$\mathrm{E}\{G(X_{j_1k_1},X_{j_3k_3})\}=0,\ \text{for } j_3\neq j_1.$$

Case 5: When both the indices j_2 and j_3 are different from i, i_1, j and j_1, the expectation vanishes again, except when $j_2 = j_3$. In the later case, since $g_1(X_{ik})$ has mean 0, we need to look at the cases $\{i = i_1\}$ or $\{i \neq i_1,\ i = j_1,\ i_1 = j\}$. In both of these cases, we have the summation of the expectations to be $O(1/n)$. To see this, if $i = i_1$,

$$\sum_{k,k_1,k_2,k_3}^{p} \mathrm{E}\Big\{g_1(X_{ik})g_1(X_{ik_1})\sum_{j_2\notin\{i,j,j_1\}} G(X_{jk},X_{j_2k_2})G(X_{j_1k_1},X_{j_2k_3})\Big\}$$
$$\leq 4\sum_{k,k_1}^{p}\alpha_{|k_1-k|}\sum_{j_2\notin\{i,j,j_1\}}\sum_{k_2,k_3}^{p}\frac{4}{N^2}\alpha_{|k_3-k_2|}=O(np^2/N^2)=O(1/n).$$

On the other hand, if $i \neq i_1$, $i = j_1$, $i_1 = j$, then $j \neq j_1$. Therefore, for $k_2 \neq k_3$,

$$\sum_{k,k_1,k_2,k_3}^{p} \mathrm{E}\Big\{g_1(X_{j_1k})g_1(X_{jk_1})\sum_{j_2\notin\{j,j_1\}} G(X_{jk},X_{j_2k_2})G(X_{j_1k_1},X_{j_2k_3})\Big\}$$
$$\leq\sum_{k,k_1,k_2,k_3}^{p}\sum_{j_2\notin\{j,j_1\}}\frac{4}{N^2}\alpha_{|k_3-k_2|}\mathrm{E}\Big\{g_1(X_{j_1k})g_1(X_{jk_1})\Big\}=0,$$

and for $k_2 = k_3$,

$$\sum_{k,k_1,k_2}^{p} \mathrm{E}\Big\{g_1(X_{j_1k})g_1(X_{jk_1}) \sum_{j_2\notin\{j,j_1\}} G(X_{jk}, X_{j_2k_2})G(X_{j_1k_1}, X_{j_2k_2})\Big\}$$

$$= \sum_{k,k_1,k_2}^{p} \sum_{j_2\notin\{j,j_1\}} \mathrm{E}\Big[\mathrm{E}\Big\{g_1(X_{j_1k})G(X_{j_1k_1}, X_{j_2k_2})|X_{j_2k_2}\Big\}$$

$$\cdot\, \mathrm{E}\Big\{g_1(X_{jk_1})G(X_{jk}, X_{j_2k_2})|X_{j_2k_2}\Big\}\Big]$$

$$\le \sum_{k,k_1,k_2}^{p} \sum_{j_2\notin\{j,j_1\}} 64(1+2/N)^2\alpha_{|k_1-k|} = O(np^2/N^2) = O(1/n).$$

Combining the five cases completes the proof of (5.6)

For the proof of (5.7), we first prove that $\mathrm{E}\{g_2(X_{ik})^4\} = O(1/N^2)$. Note that

$$\mathrm{E}\{g_2(X_{ik})^4\}$$

$$= \sum_{i_1,i_2,i_3,i_4}^{n} \sum_{k_1,k_2,k_3,k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_2k_2})G(X_{ik}, X_{i_3k_3})G(X_{ik}, X_{i_4k_4})\Big\}$$

$$= \Big[\sum_{k_1,k_2,k_3,k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{ik_1})G(X_{ik}, X_{ik_2})G(X_{ik}, X_{ik_3})G(X_{ik}, X_{ik_4})\Big\}\Big]$$

$$+ \Big[\sum_{i_1\neq i}^{n} \sum_{k_1,k_2,k_3,k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_1k_3})G(X_{ik}, X_{i_1k_4})$$

$$+ 4G(X_{ik}, X_{ik_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_1k_3})G(X_{ik}, X_{i_1k_4})$$

$$+ 6G(X_{ik}, X_{ik_1})G(X_{ik}, X_{ik_2})G(X_{ik}, X_{i_1k_3})G(X_{ik}, X_{i_1k_4})\Big\}\Big]$$

$$+ \Big[3 \sum_{i_1\neq i_2\neq i}^{n} \sum_{k_1,k_2,k_3,k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_2k_3})G(X_{ik}, X_{i_2k_4})\Big\}\Big]$$

$$= [A] + [B_1 + B_2 + B_3] + [C].$$

The first summation A is at most $O(p^4/N^4) = O(1/n^4) = O(1/N^2)$ since $n/p \to \eta \in (0, \infty)$.

The second summation $B_1 + B_2 + B_3$ is $O(np^3/N^4) = O(1/N^2)$ if the number of different elements in set $\{k_1, k_2, k_3, k_4\}$ is at most three. If all k_1, k_2, k_3, k_4 are different, then

$$
\begin{aligned}
B_1 &= \sum_{i_1 \neq i}^{n} \sum_{k_1,k_2,k_3,k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_1k_3})G(X_{ik}, X_{i_1k_4})\Big\} \\
&= 24 \sum_{i_1 \neq i}^{n} \sum_{k_1<k_2<k_3<k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_1k_3})G(X_{ik}, X_{i_1k_4})\Big\} \\
&\leq 96 \sum_{i_1 \neq i}^{n} \sum_{k_1<k_2<k_3<k_4}^{p} \min\{\alpha_{k_2-k_1}, \alpha_{k_4-k_3}\} \frac{1}{N^4} = O(np^3/N^4) = O(1/N^2),
\end{aligned}
$$

$$
B_2 = O\Big\{ \sum_{i_1 \neq i}^{n} \sum_{k_1 \neq k_2<k_3<k_4}^{p} \min\{\alpha_{k_3-k_2}, \alpha_{k_4-k_3}\} \frac{1}{N^4} \Big\} = O(np^3/N^4) = O(1/N^2),
$$

and

$$
B_3 = O\Big\{ \sum_{i_1 \neq i}^{n} \sum_{k_1 \neq k_2 \neq k_3<k_4}^{p} \alpha_{k_4-k_3} \frac{1}{N^4} \Big\} = O(np^3/N^4) = O(1/N^2).
$$

The last summation C is $O(n^2p^2/N^4) = O(1/N^2)$, if the number of different elements in set $\{k_1, k_2, k_3, k_4\}$ is at most two. If the number is three, without loss of generality, we can assume $k_1 = k_2$ and

$$
\begin{aligned}
C &= \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1 \neq k_3 \neq k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_2k_3})G(X_{ik}, X_{i_2k_4})\Big\} \\
&\leq \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1 \neq k_3 \neq k_4}^{p} \mathrm{E}\Big[\mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_2k_3})G(X_{ik}, X_{i_2k_4}) | X_{ik}\Big\}\Big] \\
&\leq 4 \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1 \neq k_3 \neq k_4}^{p} \alpha_{|k_4-k_3|} \frac{1}{N^4} = O(n^2p^2/N^4) = O(1/N^2).
\end{aligned}
$$

If all $\{k_1, k_2, k_3, k_4\}$ are different, then

$$
\begin{aligned}
C &= \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1 \neq k_2 \neq k_3 \neq k_4}^{p} \mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2})G(X_{ik}, X_{i_2k_3})G(X_{ik}, X_{i_2k_4})\Big\} \\
&\leq 4 \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1<k_2 \neq k_3<k_4}^{p} \mathrm{E}\Big[\mathrm{E}\Big\{G(X_{ik}, X_{i_1k_1})G(X_{ik}, X_{i_1k_2}) \\
&\cdot G(X_{ik}, X_{i_2k_3})G(X_{ik}, X_{i_2k_4}) | X_{ik}\Big\}\Big] \\
&\leq 64 \sum_{i_1 \neq i_2 \neq i}^{n} \sum_{k_1<k_2 \neq k_3<k_4}^{p} \alpha_{k_2-k_1}\alpha_{k_4-k_3} \frac{1}{N^4} = O(n^2p^2/N^4) = O(1/N^2).
\end{aligned}
$$

Finally, the result (5.7) can be easily shown by Cauchy-Schwarz inequality as follows:

$$\mathrm{E}\Big[\mathrm{tr}^2\big\{(\widehat{\boldsymbol{Y}}-\boldsymbol{Y})^\top \boldsymbol{C}(\widehat{\boldsymbol{Y}}-\boldsymbol{Y})\big\}\Big] = \mathrm{E}\Big[\big\{\sum_{i\neq j}^{n} c_{ij}\sum_{k=1}^{p} g_2(X_{ik})g_2(X_{jk})\big\}^2\Big]$$

$$\leq \sum_{i\neq j, i_1\neq j_1}^{n} |c_{ij}|\,|c_{i_1j_1}| \sum_{k,k_1=1}^{p}\Big\{\mathrm{E}\{g_2(X_{ik})^4\}\mathrm{E}\{g_2(X_{jk})^4\}\mathrm{E}\{g_2(X_{i_1k_1})^4\}\mathrm{E}\{g_2(X_{j_1k_1})^4\}\Big\}^{1/4}$$

$$= O(D_{\boldsymbol{C}}^2 p^2/N^2) = O(D_{\boldsymbol{C}}^2/n^2).$$

□

Proof of Theorem 5.3.1. Since

$$T_n(\widehat{\boldsymbol{Y}}^c) - T_n(\boldsymbol{Y}^c) = 2\mathrm{tr}\big\{\boldsymbol{H}(\widehat{\boldsymbol{Y}}) - \boldsymbol{G}(\widehat{\boldsymbol{Y}})\big\} - \big\{(\boldsymbol{H}(\boldsymbol{Y}) - \boldsymbol{G}(\boldsymbol{Y})\big\},$$

where $\boldsymbol{G}$ and $\boldsymbol{H}$ are given in equation (5.2). By Lemma 5.2.2, we have

$$\mathrm{E}\{T_n(\widehat{\boldsymbol{Y}}^c) - T_n(\boldsymbol{Y}^c)\}^2 = O(D_{\boldsymbol{C}}^2/n^2) + O(S_{\boldsymbol{C}}/n),$$

where

$$\boldsymbol{C} = 2\left[\left(\bigoplus_{i=1}^{2}\frac{1}{n_i}\mathbf{1}_{n_i}\right)\boldsymbol{P}_2\left(\bigoplus_{i=1}^{2}\frac{1}{n_i}\mathbf{1}_{n_i}^\top\right)\right] - \left[\bigoplus_{i=1}^{2}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}\right].$$

It is easy to calculate that $D_{\boldsymbol{C}} = 6$ and $S_{\boldsymbol{C}} = \frac{1}{n_1-1} + \frac{1}{n_2-1} + \frac{1}{n_1n_2}$. Therefore,

$$\frac{1}{\sigma_n^2}\{O(D_{\boldsymbol{C}}^2/n^2) + O(S_{\boldsymbol{C}}/n)\} = O(\frac{1}{n^2\sigma_n^2}) = o(1),$$

under assumption $D5$ and $D6$. That finishes the proof of

$$\frac{1}{\sigma_n}\{T_n(\widehat{\boldsymbol{Y}}^c) - T_n(\boldsymbol{Y}^c)\} = o_p(1).$$

□

Chapter 6 Summary

In this dissertation, new high-dimensional methods for profile analysis of mean vectors of repeated measures were introduced. The tests allow the covariance to be equal or unequal. The methods have favorable numerical performance especially when the dimension is large. A more general and flexible test statistic was proposed for a high-dimensional factorial design setting that can be used to make comparisons among cell means (including profile analysis). We also derived a second-order accurate asymptotic null distribution and upper quantiles of it. Simulation results clearly demonstrated the gain improvement from the second-order asymptotic expansions compared to the first-order (limiting distribution) approximation. The methods work well under rather general covariance structures.

By dropping the normality assumption, high-dimensional inferential procedures were proposed and studied in the parametric (mean-base) as well as non-parametric paradigms. The high-dimensional methods of testing equality of mean vectors under non-normality were closely investigated. We relaxed the commonly imposed dependence conditions and broaden the scope of the applicability of the results. The theory is worked out in detail for the two-group situation and the extension to the multi-group was shown to follow along the same lines. The results can also be formally extended to multivariate factorial designs. In fully-nonparametric approach, no assumption is made on the distribution of the population except that the dependence between the variables are required to satisfy some mild conditions. The methods are rank-based and can be applied for variables that are binary, ordered categorical, skewed and heavy tailed. The numerical results have clearly shown that when data comes from distribution with tails too thick for moments to exist, the rank-method has a superior power.

Bibliography

Ahmad, M. R. (2014). A u-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and behrens-fisher setting. *Annals of the Institute of Statistical Mathematics*, 66(1):33–61.

Akritas, M. G. and Arnold, S. F. (2000). Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical Association*, 95(449):212–226.

Akritas, M. G., Arnold, S. F., and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92(437):258–265.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 3rd edition.

Andrews, D. W. K. (1991). Heteroskedastic and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Bathke, A. C. and Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference*, 138(3):588 – 610.

Bathke, A. C., Harrar, S. W., Wang, H., Zhang, K., and Hans-Peter Piepho, H. P. (2010). Series of randomized complete block experiments with non-normal data. *Computational Statistics and Data Analysis*, 54(7):1840–1857.

Bathke, A. C. and Lankowski, D. (2005). Rank procedures for a large number of treatments. *Journal of Statistical Planning and Inference*, 133(2):223–238.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300.

Billingsley, P. (2012). *Probability and Measure.* Wiley Series in Probability and Statistics. Wiley, 3rd edition.

Bock, R. (1963). *Multivariate analysis of variance of repeated measurements. (Problems in measuring change. Editor C. W. Harris).* University of Wisconsin Press, Madison, Wisconsin.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2):290–302.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2(1):107–144.

Brunner, E., Bathke, A. C., and Placzek, M. (2012). Estimation of box's for low- and high-dimensional repeated measures designs with unequal covariance matrices. *Biometrical Journal*, 54(3):301–316.

Brunner, E., Dette, H., and Munk, A. (1997). Box-type approximations in non-parametric factorial designs. *Journal of the American Statistical Association*, 92(440):1494–1502.

Brunner, E., Munzel, U., and Puri, M. L. (1999). Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70(2):286–317.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372.

Cai, T. T. and Xia, Y. (2014). High-dimensional sparse manova. *Journal of Multivariate Analysis*, 131:174–196.

Chen, S. X. and Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Chen, S. X., Zhang, L. X., and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Chi, Y. Y., Gribbin, M., Lamers, Y., Gregory, J. F., and Muller, K. E. (2012). Global hypothesis testing for high-dimensional repeated measures outcomes. *Statistics in Medicine*, 31(8):724–742.

Collier, R. O., Baker, F. B., Mandeville, G. K., and Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika*, 32(3):339–353.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29(4):995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16(1):41–50.

Dong, K., Pang, H., Tong, T., and Gentond, M. G. (2017). Shrinkage-based diagonal hotelling's tests for high-dimensional small sample size data. *Journal of Multivariate Analysis*, 143:127–142.

Fang, K. and Zhang, Y. (1990). *Generalized multivariate analysis*. Springer-Verlag, Berlin; New York.

Feng, L. and Sun, F. (2015). A note on high-dimensional two-sample test. *Statistics & Probability Letters*, 105:29–36.

Feng, L., Zou, C., and Wang, Z. (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *Journal of the American Statistical Association*, 111(514):721–735.

Feng, L., Zou, C., Wang, Z., and Zhu, L. (2015). Two-sample behrens-fisher problem for high-dimensional data. *Statistica Sinica*, 25:1297–1312.

Fujikoshi, Y. (1973). Asymptotic formulas for the distributions of three statistics for multivariate linear hypothesis. *Annals of the Institute of Statistical Mathematics*, 25(1):423–437.

Fujikoshi, Y., Himeno, T., and Wakakii, H. (2004). Asymptotic results of a high dimensional manova test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society*, 34(1):19–26.

Geisser, S. and Greenhouse, S. W. (1958). An extension of box's results on the use of the \$f\$ distribution in multivariate analysis. *The Annals of Mathematical Statistics*, 29(3):885–891.

Ghosh, A. K. and Biswas, M. (2016). Distribution-free high-dimensional two-sample tests based on discriminating hyperplanes. *TEST*, 25(3):525–547.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V., and Lahiri, S. N. (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association*, 110(510):837–849.

Hall, P. and Heyde, C. (1980). *Martingale limit theory and its application*. Probability and mathematical statistics. Academic Press.

Hannan, E. (1970). *Multiple Time Series*. Wiley Series in Probability and Statistics. Wiley.

Happ, M., Harrar, S. W., and Bathke, A. C. (2015). Inference for low- and high-dimensional multi-group repeated measures designs with unequal covariance matrices. *Biometrical Journal*, 58(4):810–830.

Harrar, S. W. (2009). Asymptotics for tests on mean profiles, additional information and dimensionality under non-normality. *Journal of Statistical Planning and Inference*, 139(8):2685–2705.

Harrar, S. W. and Bathke, A. C. (2008). Nonparametric methods for unbalanced multivariate data and many factor levels. *Journal of Multivariate Analysis*, 99(8):1635–1664.

Harrar, S. W. and Kong, X. (2016). High-dimensional multivariate repeated measures analysis with unequal covariance matrices. *Journal of Multivariate Analysis*, 145:1–21.

Harrar, S. W. and Xu, J. (2014). Analyzing mean profiles of nonnormal populations. *Communications in Statistics: Theory & Methods*, 43(17):3553–3573.

Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of cornish-fisher type. *The Annals of Mathematical Statistics*, 39(4):1264–1273.

Himeno, T. and Yamada, T. (2014). Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *Journal of Multivariate Analysis*, 130:27–44.

Hu, J., Bai, Z., Wang, C., and Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics*, 69(2):365–387.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika*, 43(2):161–175.

Huynh, H. and Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact f-distributions. *Journal of the American Statistical Association*, 65(332):1582–1589.

Huynh, H. and Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1(1):69–82.

Hyodo, M. (2017). Tests for the parallelism and flatness hypotheses of multi-group profile analysis for high-dimensional elliptical populations. *Journal of Multivariate Analysis*, 162:82–92.

Hyodo, M., Takahashi, S., and Nishiyama, T. (2014). Multiple comparisons among mean vectors when the dimension is larger than the total sample size. *Communications in Statistics - Simulation and Computation*, 43(10):2283–2306.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education, Inc., Upper Saddle River, New Jersey, 6th edition.

Katayama, S., Kano, Y., and Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 116:410–421.

Kong, X. and Harrar, S. W. (2017). Accurate inference for repeated measures in high dimensions. *submitted*, pages 1–21.

Kong, X. and Harrar, S. W. (2018a). High-dimensional manova under weak conditions. *preprint*, pages 1–27.

Kong, X. and Harrar, S. W. (2018b). High-dimensional rank-based inference. *preprint*, pages 1–24.

Lecoutre, B. (1991). A correction for the formula approximate test in repeated measures designs with two or more independent groups. *Journal of Educational and Behavioral Statistics*, 16(4):371–372.

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2):908–940.

Maruyama, Y. (2007). Asymptotic expansions of the null distributions of some test statistics for profile analysis in general conditions. *Journal of Statistical Planning and Inference*, 137(2):506–526.

Mathai, A. M., Provost, S. B., and Hayakawa, T. (1995). *Bilinear Forms and Zonal Polynomials*. Springer, New York.

Maxwell, S. E. and Arvey, R. D. (1982). Small sample profile analysis with many variables. *Psychological Bulletin*, 92(3):778–785.

Okamoto, N., Miura, N., and Seo, T. (2006). On the distributions of some test statistics for profile analysis in elliptical populations. *American Journal of Mathematics and Management Science*, 26(1–2):1–31.

Pauly, M., Ellenberger, D., and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics*, 49(6):1243–1261.

Rencher, A. C. and Christensen, W. F. (2012). *Methods of multivariate analysis.* John Wiley & Sons, United States, 3rd edition.

Schott, J. (2007a). Some high-dimensional tests for a one-way manova. *Journal of Multivariate Analysis*, 98(9):1825–1839.

Schott, J. (2007b). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Journal Computational Statistics & Data Analysis*, 51(12):6535–6542.

Shiryaev, A. (2015). *Probability.* Springer New York, 3rd edition.

Srivastava, M., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, 114:349–358.

Srivastava, M. and Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, 115:204–216.

Srivastava, M. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101:1319–1329.

Srivastava, M. S. (1987). Profile analysis of several groups. *Communications in Statistics: Theory & Methods*, 16(3):909–926.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, 35(2):251–272.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Stoloff, P. H. (1970). Correcting for heterogeneity of covariance for repeated measures designs of the analysis of variance. *Educational and Psychological Measurement*, 30(4):909–924.

Takahashi, S. and Shutoh, N. (2016). Tests for parallelism and flatness hypotheses of two mean vectors in high-dimensional settings. *Journal of Statistical Computation and Simulation*, 86(6):1150–1165.

Tonda, T. and Fujikoshi, Y. (2004). Asymptotic expansion of the null distribution of lr statistic for multivariate linear hypothesis when the dimension is large. *Communications in Statistics: Theory & Methods*, 33(5):1205–1220.

Wang, H. and Akritas, M. G. (2010a). Inference from heteroscedastic functional data. *Journal of Nonparametric Statistics*, 22(2):149–168.

Wang, H. and Akritas, M. G. (2010b). Rank test for heteroscedastic functional data. *Journal of Multivariate Analysis*, 101(8):1791–1805.

Wang, H., Higgins, J., and Blasi, D. (2010). Distribution-free tests for no effect of treatment in heteroscedastic functional data under both weak and long range dependence. *Statistics & Probability Letters*, 80(5-6):390–402.

Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):1658–1669.

Watamori, Y. (1990). On the moments of traces of wishart and inverted wishart matrices. *South African Statistical Journal*, 24(2):153–176.

Xu, G., Lin, L., Wei, P., and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, 103(3):609–624.

Yamada, T. and Srivastava, M. S. (2012). A test for multivariate analysis of variance in high-dimension. *Communication in Statistics: Theory & Methods*, 41(13-14):2602–2615.

Yang, X. M., Yang, X. Q., and Teo, K. L. (2001). A matrix trace inequality. *Journal of Mathematical Analysis and Applications*, 263(1):327–331.

Yaskov, P. (2015). Variance inequalities for quadratic forms with applications. *Mathematical Methods of Statistics*, 24(4):309–319.

Zhang, C., Bai, Z., Hu, J., and Wang, C. (2018). Multi-sample test for high-dimensional covariance matrices. *Communications in Statistics - Theory and Methods*, 48(13):3161–3177.

Zhou, B., Guo, J., and Zhang, J. (2017). High-dimensional general linear hypothesis testing under heteroscedasticity. *Journal of Statistical Planning and Inference*, 188:36–54.

Vita

Xiaoli Kong

Education

- M.S. in Statistics, University of Kentucky
- Ph.D. in Mathematics, Xiamen University
- M.S. in Mathematics, Nankai University
- B.S in Mathematics, Hebei Normal University

Selected Publications

- K. Mathu-Muju, **X. Kong**, C. Brancato, H. Bush, Utilization of community health workers in Canada's Children's Oral Health Initiative for indigenous communities. Community Dentistry and Oral Epidemiology, 46, 185-193, 2018.
- S.W. Harrar and **X. Kong**, High-dimensional multivariate repeated measures analysis with unequal covariance matrices. Journal of Multivariate Analysis, 145, 1-21, 2016.
- **X. Kong**, C. Bai, and D. Meng, On real simple left-symmetric algebras. Communications in Algebra, 40, 1641–1668, 2012.
- **X. Kong**, H. Chen, and C. Bai, Classification of graded left-symmetric algebraic structures on Witt and Virasoro algebras. International Journal of Mathematics, 22, 201–222, 2011.
- **X. Kong** and S. Tan, Irreducible representations for the baby Tits-Kantor-Koecher algebra. Communications in Algebra, 38, 2177–2183, 2010.
- **X. Kong** and C. Bai, Left-symmetric superalgebraic structures on the super-Virasoro algebras. Pacific Journal of Mathematics, 235, 43–55, 2008.