2021

Novel Nonparametric Testing Approaches for Multivariate Growth Curve Data: Finite-Sample, Resampling and Rank-Based Methods

Ting Zeng
University of Kentucky, zengtingbeidy@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0002-1123-7228
Digital Object Identifier: https://doi.org/10.13023/etd.2021.261

Ting Zeng, Student

Dr. Solomon W. Harrar, Major Professor

Dr. Katherine L. Thompson, Director of Graduate Studies

Novel Nonparametric Testing Approaches for Multivariate Growth Curve Data: Finite-Sample, Resampling and Rank-Based Methods

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
Ting Zeng

Lexington, Kentucky

Director: Dr. Solomon W. Harrar, Professor of Statistics

Lexington, Kentucky

2021

https://orcid.org/0000-0002-1123-7228

ABSTRACT OF DISSERTATION

Novel Nonparametric Testing Approaches for Multivariate Growth Curve Data: Finite-Sample, Resampling and Rank-Based Methods

Multivariate growth curve data naturally arise in various fields, for example, biomedical science, public health, agriculture, social science and so on. For data of this type, the classical approach is to conduct multivariate analysis of variance (MANOVA) based on Wilks' Lambda and other multivariate statistics, which require the assumptions of multivariate normality and homogeneity of within-cell covariance matrices. However, data being analyzed nowadays show marked departure from multivariate normal distribution and homoscedasticity. In this dissertation, we investigate nonparametric testing approaches for multivariate growth curve data from three aspects, i.e., finite-sample, resampling and rank-based methods.

The first project proposes an approximate finite-sample test using modified sums of squares matrices to make them insensitive to the heterogeneity in MANOVA. The modification corrects the associated quadratic forms of the two sums of squares for the effect of heterogeneity. The distribution of the proposed test statistic is invariant to the original data distribution. The proposed approximation method can be used in various experimental designs, for example, factorial design and crossover design. Under various simulation settings, the proposed method outperforms the classical Doubly Multivariate Model and Multivariate Mixed Model, especially for unbalanced sample sizes. The applications of the proposed method are illustrated with ophthalmology data in a factorial design and in a 2×2 crossover design.

In the semiparametric situation, parametric and nonparametric bootstraps are known to have satisfactory finite-sample performance in general factorial designs. In this regard, the second project provides resampling-based tests for multivariate growth curve data. Such tests are useful in situations where data are not necessarily exchangeable under the null hypothesis of interest and with small sample sizes. Simulation studies are conducted to evaluate the finite-sample performance of the proposed test procedures under various practical scenarios. Data from an optometry study are used to illustrate the benefits of the nonparametric methods proposed.

For multivariate growth curve data which are measured in ordered categorical scales, the usual mean- and covariance-based inferences are not appropriate anymore. The third project deals with general nonparametric methods for multivariate growth curve data in factorial designs. Treatment effects are characterized in terms of functionals of distribution functions with the sole assumption of nondegenerate marginal distributions. This model accommodates binary, discrete, ordered categorical, and continuous data in a unified manner. Hypotheses are formulated in terms of meaningful nonparametric measures of treatment effects. In this project, the Wald-type statistic is proposed and its asymptotic properties are investigated. In addition, the ANOVA-type statistic and the modified Wilks' Lambda statistic under the nonparametric framework are also presented. The theory can be used to construct confidence intervals for the nonparametric treatment effects. Simulation studies are conducted to show the finite-sample performance of the proposed methods in comparison with other parametric and nonparametric methods. Data from a study of infantile nystagmus syndrome (INS) are analyzed to illustrate the application of the proposed methods.

KEYWORDS: Heteroscedasticity, Nonnormality, GMANOVA, Nonparametric Analysis, Wilks' Lambda, Relative Effects

Ting Zeng

July 13, 2021

Date

Novel Nonparametric Testing Approaches for Multivariate Growth Curve Data:
Finite-Sample, Resampling and Rank-Based Methods

By

Ting Zeng

Solomon W. Harrar
Director of Dissertation

Katherine L. Thompson
Director of Graduate Studies

July 13, 2021
Date

ACKNOWLEDGMENTS

First and foremost, I wish to express my sincere appreciation to my advisor Dr. Solomon W. Harrar for his invaluable advice and continuous support. This work would not have been possible without his wise guidance and persistent help. The lessons he have taught me about how to question, how to explore and how to search for a deeper answer will stay with me forever.

I am also extremely grateful to my dissertation committee, Dr. Arnold J. Stromberg, Dr. William S. Griffith, Dr. Derek S. Young, Dr. Chi Wang and Dr. Hana Khamfroush, for their insightful comments and helpful suggestions. I would also like to thank Dr. Katherine L. Thompson for her treasured support during my graduate study at the University of Kentucky.

Finally, I would like to thank my friends and colleagues, Yue Cui, Yan Xu and Jing Wei for a cherished time spent together in the office and in social settings. My appreciation also goes out to my family and friends for their tremendous understanding and encouragement over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Background

Multivariate growth curve data appear in a variety of fields, for example, biomedical science, public health, agriculture, social science and so on. For this type of data, several related variables are observed at different occasions or under different conditions for each experimental or observational unit. Often, the units belong to different treatment groups. Typically, interest lies in mean-based inferences where dependencies among the variables and repeated measurements are taken into account. In the context of general factorial design, research questions are whether there is any group effect, time (condition) effect or group-time interaction effect.

The classical parametric and semiparametric procedures proposed are, for example, Bock (1975); Boik (1988); Naik and Rao (2001) and Rencher (2001). However, more assumptions naturally arise when the parametric and semiparametric procedures are applied, for example, normality and homoscedasticity of the data, which are, in fact, difficult to attain in practice (see Xu and Cui, 2008; Suo et al., 2013). A commonly used method of jointly modeling repeated measures data on multiple dependent variables is known as the Doubly Multivariate Model (DMM) as introduced in Bock (1975) and Boik (1988). In this approach, the multivariate repeated measurements from the same experimental unit are stacked to form a vector of measurements of that experimental unit. The hypothesis of interest is then formulated as a general linear hypothesis and the standard multivariate tests are applied. An alternative approach for modeling multivariate repeated measures data is the Multivariate Mixed Model (MMM), where the covariance matrix for the estimated contrast of interest is required to satisfy multivariate sphericity condition as noted in Boik (1988) and Pavur (1987). Such parametric MANOVA inferential procedures are generally robust against mild violation of normality. It is also well known that MANOVA tests are robust to heteroscedasticity in balanced designs. However, in unbalanced heteroscedastic designs, the type-I error rate of MANOVA tests may be significantly impaired (Rencher, 2001).

Without assuming multivariate normality or homogeneity of the covariance matrices, the multivariate Wald-type test statistic (WTS) is asymptotically exact. However, WTS is known to suffer from slow convergence and, hence, may not be satisfactory in finite-sample. Large sample sizes are then required in order to maintain the preassigned type-I error rate. Several improvements were done to tackle the small-sample issue under MANOVA settings (Pesarin, 2001; Pesarin and Salmaso, 2010; Good, 2005, for example), where permutation tests are utilized when the null distribution of the test statistic is invariant under the corresponding randomization group. The permutation idea was further modified for use in situations where the exchangeability under the null hypothesis does not hold by Neuhaus (1993); Janssen and Pauls (2003); Janssen (2005); Omelka and Pauly (2012); Chung and Romano (2013); Pauly et al. (2015); Friedrich et al. (2017). In addition, Konietschke et al. (2015) applied a bootstrap technique to improve the small sample behavior of the WTS for general heteroscedastic factorial design. However, to the best of our knowledge, there does not exist any work done for analyzing multivariate repeated measures data without assuming multivariate normality and homogeneity conditions.

Furthermore, parametric procedures are limited to evaluating continuous data. When count data, discrete data and ordinal data or ordered categorical data are collected, parametric procedures will no longer be appropriate. Thus, a more convenient, robust and reliable method that can handle both metric and nonmetric data is in substantial demand. Nonparametric rank-based methods have been developed for the last few decades to meet this need. In factorial designs, a rank-based test for repeated measures data was proposed by Brunner and Neumann (1982), which is asymptotically distribution-free. This work was later further generalized for multivariate designs (see Thompson, 1990, 1991; Brunner and Denker, 1994). However, they all rely on the assumption of absolutely continuous distribution functions, which is a quite unrealistic assumption for practical applications. Also, hypotheses were originally formulated in terms of marginal distributions in nonparametric tests, for example, Akritas and Arnold (1994); Akritas and Brunner (1997); Brunner et al. (1999); Fan and Zhang (2014, 2017), but it is difficult to interpret the corresponding alternatives. Additionally, with such hypotheses, test procedures cannot be used to construct confidence intervals for the effect size measures.

1.2 Overview of the Dissertation

This dissertation aims to resolve three distinct but closely related problems for growth curve data. The first one is regarding to a robust test without assuming multivariate normality and homoscedasticity, which is ideal when sample size is moderately large. The second problem investigated pertains to resampling-based tests for finite-sample not requiring data exchangeability, multivariate normality and homoscedasticity. Both problems pertain to metric data. The last problem studied deals with nonparametric rank-based inference for both metric and nonmetric data.

The dissertation consists of five chapters. In Chapter 2, without the assumptions of multivariate normality and homoscedasticity, a modified Wilks' Lambda test statistic is proposed to mitigate the effects of potential heteroscedasticity. The null distribution of the proposed test statistic is invariant to the original data distribution. To show the broad scope of application, the detailed calculations are also illustrated for two commonly arising designs.

Chapter 3 introduces Wald-type resampling-based test statistics with finite-sample, where permutation, bootstrap and hybrid permutation-bootstrap procedures are investigated. Due to studentization, the permutation procedure is asymptotically exact despite the time dependencies. The bootstrap procedures also result in asymptotically valid tests in the nonparametric setting. As in Chapter 2, multivariate normality and homoscedasticity are not assumed.

In Chapter 4, general nonparametric rank-based tests are proposed. Treatment effects are characterized as functionals of distribution functions with the sole assumption of nondegenerate marginal distributions. Such nonparametric treatment effects can be used to quantify the magnitude of effects of interest. Hence, hypotheses are formulated in terms of meaningful nonparametric measures of treatment effects. Three

test statistics are considered, the Wald-type statistic, the ANOVA-type statistic and the modified Wilks' Lambda F-approximation under the nonparametric framework. In addition, the asymptotic properties of the nonparametric treatment effects are investigated and are used to construct confidence intervals.

Proofs and mathematical details are included in appendices at the end of each of Chapters 2 to 4. Simulation studies for both type-I error rate and power are also presented in each of these chapters to numerically evaluate the performance of the methods. To illustrate the applications of the proposed methods, analyses of ophthalmology data from a study of infantile nystagmus syndrome (INS) and/or from a study of intraocular pressure are included. Discussion, remarks and possible directions for future research are summarized in Chapter 5.

Chapter 2 Robust Tests for Trials with Multiple Endpoints

2.1 Introduction

Multivariate growth curve data appear in a variety of fields, for example, biomedical science, public health, agriculture, social science, etc. For this type of data, several related variables are observed at different occasions or under different conditions for each experimental or observational unit. Often, the units belong to different treatment groups. Typically, interest lies in mean-based inferences where dependencies among the variables and repeated measurements are taken into account. In the context of general factorial design, research questions are whether there is any group effect, time (condition) effect or group-time interaction effect. For example, for the mandible data in Timm (1980, Table 7.2), the researchers investigated the relative effectiveness of two orthopedic adjustments (treatments) of the mandible with nine subjects in each treatment group. Three dependent variables together reflecting the position and angle of mandibles were measured at three different time points. The objective of the study was to test whether the mean responses of the three mandible variables change over time (time effect), whether the changes in mandible over time are the same across two treatment groups (group effect), and whether the change evolved over time similarly in two treatment groups (group-time interaction effect).

A commonly used way of jointly modeling repeated measures data on multiple dependent variables is known as the Doubly Multivariate Model (DMM) as introduced in Bock (1975) and Boik (1988). In this approach, the multivariate repeated measurements from the same experimental unit are stacked to form a vector of measurements of that experimental unit. The hypothesis of interest is then formulated as a general linear hypothesis and the standard multivariate tests are applied. An alternative approach for modeling multivariate repeated measures data is the Multivariate Mixed Model (MMM), where the covariance matrix for the estimated contrast of interest is required to satisfy multivariate sphericity condition as noted in Boik (1988) and

Pavur (1987). MMM contains fixed treatment, time and interaction effects as well as random block effect for time. The random time effects are used to model the covariance over time in a way similar to the univariate mixed model. Similarly, the hypothesis of interest can be expressed as a general linear hypothesis.

Although DMM and MMM seem very promising approaches in analyzing multivariate repeated measures data, they both assume multivariate normality and constant error covariance matrix across the treatment groups. For example, Naik and Rao (2001) proposed various tests for the three hypotheses in multivariate repeated measures design assuming independence across experimental units and multivariate normality for the matrix of observations from each experimental unit. However, in practical applications, such as in genomics, the assumption of normality is difficult to justify under multivariate settings, as mentioned in Xu and Cui (2008) and Suo et al. (2013). As with general multivariate data, another challenge in analyzing multivariate repeated measures data is the heteroscedasticity issue, which could arise due to the intrinsic difference among the treatment groups (populations) in the dependence structure between the variables and also between the time points.

The parametric MANOVA inferential procedures including DMM and MMM are generally robust against mild violation of normality. It is also well known that MANOVA tests are robust to heteroscedasticity in balanced designs. However, in unbalanced designs, the type-I error rate of MANOVA tests may be significantly impaired. Generally, the effect of cell covariance and sample size difference depend on how the covariances and sample sizes are associated. When larger covariances are associated with larger sample sizes, the test is conservative, and when the covariances and sample sizes matching are inversely related, the test is liberal as noted by Rencher (2001).

It is known that the Wald-type test is asymptotically valid without assuming multivariate normality or homogeneity of covariance matrices. However, large sample sizes are required in order to maintain the preassigned type-I error rate. For example, both Konietschke et al. (2015) and Pauly et al. (2015) utilized Wald-type statistic. Several resampling-based improvements were proposed to tackle the small-sample issue of the Wald-type test under MANOVA settings, for example, Pesarin (2001); Pesarin and Salmaso (2010); Good (2005); Omelka and Pauly (2012); Chung and Romano (2013); Konietschke et al. (2015); Pauly et al. (2015) and Friedrich et al. (2017).

The present chapter aims to develop a robust test for multivariate repeated measures (growth curve) data. To the best of our knowledge, there is no prior work for analyzing multivariate repeated measures data without assuming multivariate normality and homoscedasticity. We modify the Wilks' Lambda test statistic to mitigate the effects of heteroscedasticity and develop an approximation for the null distribution of the modified statistic. To show the broad scope of application, we illustrate the detailed calculations for two commonly arising designs.

This chapter is organized as follows. In Section 2.2, statistical model and hypotheses are presented together with two concrete examples. The robust test is described in detail in Section 2.3. The finite-sample behavior of the proposed robust test along with the classical DMM and MMM are investigated in extensive simulations in Section 2.4. The type-I error rate (size) simulations and power curves are also examined in Section 2.4. The applications of the robust test are illustrated in Section 2.5 with multivariate repeated measures of idiopathic infantile nystagmus syndrome (INS) data and intraocular pressure data. We conclude the chapter with some discussions and remarks in Section 2.6. All relevant technical details are given in Appendix (Section 2.7).

2.2 Models and Hypotheses

Suppose a p-dimensional random vector is observed from each of the study (observational or experimental) subjects (units) at each of the t occasions. Here, the occasions may represent time points or other within-subject factor levels as it happens, for example, in split-plot or crossover designs. We assume that n_i replications are available from the ith group, where $i = 1, \ldots, a$. The groups may represent between-subject treatment groups or other forms of natural groups such as gender or region. Let $\boldsymbol{X}_{ijk}$ be observations from kth study unit in the ith group and at jth occasion. Therefore, a $p \times t$ matrix of dependent observation is made from each subject, where the structure of the matrix is important for the analysis. Note in the context of general factorial design, we consider the multivariate model

$$\boldsymbol{X}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\epsilon}_{ijk}, \tag{2.1}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, t$, and $k = 1, \ldots, n_i$. The index i is for group (treatment), j is for time (repeated measure) and k is for experimental unit (subject). For fixed i and j, the random errors $\boldsymbol{\epsilon}_{ij1}, \ldots, \boldsymbol{\epsilon}_{ijn_i}$ are independent and identically distributed p-dimensional random vectors satisfying

$$\mathrm{E}(\boldsymbol{\epsilon}_{ijk}) = \mathbf{0} \quad \text{and} \quad \mathrm{Cov}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{\Sigma}_{i,jj} > 0,$$

for $i = 1, \ldots, a$ and $j = 1, \ldots, t$. Furthermore, $\mathrm{Cov}(\boldsymbol{\epsilon}_{ijk}, \boldsymbol{\epsilon}_{ij'k}) = \boldsymbol{\Sigma}_{i,jj'}$ for $j \neq j'$.

Stacking the group mean vectors into one vector, let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_a^\top)^\top$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}^\top, \ldots, \boldsymbol{\mu}_{it}^\top)^\top$ and $\boldsymbol{\mu}_{ij} = (\mu_{ij}^{(1)}, \ldots, \mu_{ij}^{(p)})^\top$. Throughout the paper we will use the following notations. The d-dimensional identity matrix is denoted by $\boldsymbol{I}_d$ and a $d \times d$ matrix with all 1s as its components is denoted by $\boldsymbol{J}_d = \mathbf{1}_d \mathbf{1}_d^\top$, where $\mathbf{1}_d = (1, \ldots, 1)_{d\times 1}^\top$. We further denote the centering matrix by $\boldsymbol{P}_d = \boldsymbol{I}_d - \frac{1}{d}\boldsymbol{J}_d$. The operators $\bigoplus$ and $\bigotimes$ represent the Kronecker sum and product, respectively (Schott, 2016, Chap. 8).

Based on the type of effect we are testing, denoted by ϕ, we define the contrast matrices $\boldsymbol{D}_\phi$ and $\boldsymbol{C}_\phi$ which target within-subject and between-subject factor effects, respectively. To conduct a hypothesis test, let $\boldsymbol{H}_\phi = \boldsymbol{C}_\phi^\top \otimes \boldsymbol{D}_\phi \otimes \boldsymbol{I}_p$ be an appropriate contrast matrix. Generally, it is more convenient to use the unique projection matrix $\boldsymbol{T}_\phi = \boldsymbol{H}_\phi^\top (\boldsymbol{H}_\phi \boldsymbol{H}_\phi^\top)^- \boldsymbol{H}_\phi$, where $(\boldsymbol{H}_\phi \boldsymbol{H}_\phi^\top)^-$ is some generalized inverse of $\boldsymbol{H}_\phi \boldsymbol{H}_\phi^\top$, to formulate our hypotheses. It is easy to show that $\boldsymbol{T}_\phi \boldsymbol{\mu} = \mathbf{0}$ if and only if $\boldsymbol{H}_\phi \boldsymbol{\mu} = \mathbf{0}$. Therefore, the unique, symmetric and idempotent contrast matrix $\boldsymbol{T}_\phi$ is equivalent to $\boldsymbol{H}_\phi$ for testing $\boldsymbol{H}_\phi \boldsymbol{\mu} = \mathbf{0}$. To illustrate these notations we give two examples.

Example 1. *Assume a design in which there are one within-subject factor (repeated measures) with t levels and one between-subject factor with a levels. The three common hypotheses for this design are presence of group, occasion (time), and group-occasion* *(group-time) interaction effects. In the notations above, the contrast matrices ($\boldsymbol{D}_\phi$ and $\boldsymbol{C}_\phi$) corresponding to these three hypotheses are $\boldsymbol{D}_{\mathrm{G}} = \mathbf{1}_t^\top/t$, $\boldsymbol{C}_{\mathrm{G}} = \boldsymbol{P}_a/(a-1)$, $\boldsymbol{D}_{\mathrm{T}} = (\boldsymbol{I}_{t-1}, -\mathbf{1}_{t-1})$, $\boldsymbol{C}_{\mathrm{T}} = \boldsymbol{J}_a/a$, and $\boldsymbol{D}_{\mathrm{GT}} = (\boldsymbol{I}_{t-1}, -\mathbf{1}_{t-1})$, $\boldsymbol{C}_{\mathrm{GT}} = \boldsymbol{P}_a/(a-1)$, respectively.*

To be more concrete, for the mandible data (Timm, 1980, Table 7.2), there are $p = 3$ dependent variables and $t = 3$ occasions are involved with two treatment groups $a = 2$ and equal sample sizes $n_1 = n_2 = 9$ per group. The goal of the study was to test whether the treatments induced differential changes over time on the mandibles. Hence, it is of interest to check if there was any treatment (group) effect (G), time effect (T), and treatment (group) by time interaction effect (GT). The contrast matrices are $\boldsymbol{H}_{\mathrm{G}} = \boldsymbol{P}_2 \otimes \frac{1}{3}\mathbf{1}_3^\top \otimes \boldsymbol{I}_3$, $\boldsymbol{H}_{\mathrm{T}} = \frac{1}{2}\mathbf{1}_2^\top \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$ and $\boldsymbol{H}_{\mathrm{GT}} = \boldsymbol{P}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$, respectively, and the corresponding unique projection matrices are $\boldsymbol{T}_{\mathrm{G}} = \boldsymbol{P}_2 \otimes \frac{1}{3}\boldsymbol{J}_3 \otimes \boldsymbol{I}_3$, $\boldsymbol{T}_{\mathrm{T}} = \frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$ and $\boldsymbol{T}_{\mathrm{GT}} = \boldsymbol{P}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$, respectively.

The setup as described before may give the impression that the paper 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.

Example 2. *For an illustration of a more elaborate design, we consider the crossover trial investigated in Li et al. (2020). The trial aims at comparison of two active treatments A and B in a four-period ($b = 4$) crossover design where the active treatments are separate by two 4-week of placebo treatment (P) periods to eliminate carryover effect. Four sequences of the treatments (APBP, PAPB, BPAP, PBPA) are applied to four different groups of participants ($a = 4$) where $n_1 = 37$, $n_2 = 34$, $n_3 = 36$ and $n_4 = 33$, respectively. The response variables are diastolic and systolic blood pressures ($p = 2$). At each period, seven measurements ($t = 7$) of the responses were observed 5 minutes apart. The seven occasions are nested within each of the four periods. Stacking the vector of cell means into a single column vector, let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_4^\top)^\top$,*

$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}^\top, \ldots, \boldsymbol{\mu}_{i4}^\top)^\top$, $\boldsymbol{\mu}_{ij} = (\boldsymbol{\mu}_{ij1}^\top, \ldots, \boldsymbol{\mu}_{ij7}^\top)^\top$ *and* $\boldsymbol{\mu}_{ijk}$ *is a 2-dimensional vector of means for the diastolic and systolic blood pressures. Hypotheses of interest may now be formulated by suitably choosing* $\boldsymbol{C}_\phi$ *and* $\boldsymbol{D}_\phi$ *matrices. For example, in this special design, the hypothesis of no treatment-by-period interaction can be formulated as the intersection of two hypotheses. The first one tackles if the effect of treatment A changes over period,* $\boldsymbol{C}_\phi = \boldsymbol{I}_2 \otimes (1, -1)^\top$ *and* $\boldsymbol{D}_\phi = (\boldsymbol{I}_2 \otimes (1, -1)) \otimes \boldsymbol{1}_7^\top$*. The other hypothesis addresses the same question but for treatment B,* $\boldsymbol{C}_\phi = (\boldsymbol{J}_2 - \boldsymbol{I}_2, -\boldsymbol{I}_2)^\top$ *and* $\boldsymbol{D}_\phi = (0, 1, 1, 0) \otimes \boldsymbol{1}_7^\top$*. The matrices targeting the hypothesis of no time effect within each period and for each sequence are* $\boldsymbol{C}_\phi = \boldsymbol{I}_4$ *and* $\boldsymbol{D}_\phi = \boldsymbol{I}_4 \otimes (\boldsymbol{I}_6, -\boldsymbol{1}_6)$*. In the construction of* $\boldsymbol{D}_\phi$ *and* $\boldsymbol{C}_\phi$*, the order of the matrices in the Kronecker products reflect the arrangement of the mean vectors in* $\boldsymbol{\mu}$*. It should be noted that the occasions nested within each period need not be balanced across periods.*

2.3 Wilks' Lambda Test

Test Statistic

The classical approach for inference in the multivariate growth curve data is based on Wilks' Lambda and other multivariate statistics (see Timm, 2002, p. 294; Rencher, 2001, p. 215; Johnson et al., 2002, p. 398). The classical approaches assume normality and homogeneity of within-cell (group) covariance matrices (Bray and Maxwell, 1985). They perform well if these assumptions are satisfied. Often, data being analyzed nowadays show marked departure from multivariate normal distribution or assessing multivariate normality for them is difficult. Additionally, the equality of covariance matrices does not always hold.

It is well known that the parametric MANOVA inferential procedures are generally robust against mild nonnormality, e.g. O'Brien et al. (1982) and Timm (2002, p. 303). When sample sizes are balanced, those tests are also robust to unequal cell covariances, but when sample sizes are unbalanced the type-I error rate of the MANOVA tests may be substantially affected. If the larger variances and covariances are associated with the larger samples, the type-I error rate tends to be small leading to conservative test results. On the other hand, if the larger variances and covariances are associated with the smaller samples, the type-I error rate is inflated and the test results become liberal (Rencher, 2001). An alternative test which is fairly robust to both nonnormality and heterogeneity assumptions is the Wald-type test statistic (WTS). However, WTS is known to suffer from slow convergence and, hence, may not be satisfactory in finite-samples. The effect of heterogeneity of covariance matrices in MANOVA can be mitigated by modifying the sums of squares and cross product matrices (Harrar and Bathke, 2012b; Zhang and Liu, 2013).

In this section, we propose an approximate finite-sample test for the multivariate growth curve (repeated measures) designs based on the modified Wilks' Lamda statistic. The modification corrects the associated quadratic forms of the hypothesis and error sums of squares and cross product matrices to make them insensitive to inequality of covariances and yet they have equal expectation under the null hypothesis.

Analogous to Harrar and Bathke (2012b), the distributions of the adjusted hypothesis and adjusted error matrices can be approximated by Wishart distributions with their corresponding degrees of freedom obtained by matching the expected values and total variances. Here, normality is less of an issue, because the asymptotic distributions of the test statistics do not depend on the distribution of the data and this property is expected to hold in moderate samples, especially after adjustments for finiteness and heterogeneity of covariance matrices.

The most popular MANOVA test statistics are Wilks' Lambda, Lawley-Hotelling Trace, Pillai's Trace, and Roy's Largest Root (Anderson, 2003). Among the four, Wilks' Lambda plays the dominant role because it corresponds to the likelihood ratio statistic and has the well-known χ^2 and F approximations (Rencher, 2001).

Let $\boldsymbol{X}_{ik} = (\boldsymbol{X}_{i1k}^\top, \ldots, \boldsymbol{X}_{itk}^\top)^\top$, and define

$$\boldsymbol{Z}_{ik}^{(\phi)} = (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p)\boldsymbol{X}_{ik},$$

and $r = \text{rank}(\boldsymbol{D}_\phi)$, i.e. $\boldsymbol{D}_\phi$ is full row rank with dimension $r \times t$. In Example 1, r is either 1 or $(t-1)$ depending on the effect we are testing. Let $\boldsymbol{V}_i = \operatorname{Cov}(\boldsymbol{X}_{ik})$,

$$\boldsymbol{\Sigma}_i^{(\phi)} = \operatorname{Cov}\left(\boldsymbol{Z}_{ik}^{(\phi)}\right) = \left(\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p\right) \boldsymbol{V}_i \left(\boldsymbol{D}_\phi^\top \otimes \boldsymbol{I}_p\right) \quad \text{and} \quad \boldsymbol{\Sigma}_\phi = \sum_{i=1}^{a} \boldsymbol{C}_{\phi,ii} \cdot \frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}.$$

We further define the adjusted hypothesis matrix $\boldsymbol{H}^{(\phi)}$ and error matrix $\boldsymbol{G}^{(\phi)}$ as

$$\boldsymbol{H}^{(\phi)} = \overline{\boldsymbol{Z}}^{(\phi)} \boldsymbol{C}_\phi \overline{\boldsymbol{Z}}^{(\phi)\top} \quad \text{and} \tag{2.2}$$

$$\boldsymbol{G}^{(\phi)} = \sum_{i=1}^{a} \frac{\boldsymbol{C}_{\phi,ii}}{n_i(n_i-1)} \sum_{k=1}^{n_i} \left(\boldsymbol{Z}_{ik}^{(\phi)} - \overline{\boldsymbol{Z}}_i^{(\phi)}\right) \left(\boldsymbol{Z}_{ik}^{(\phi)} - \overline{\boldsymbol{Z}}_i^{(\phi)}\right)^\top, \tag{2.3}$$

where $\overline{\boldsymbol{Z}}^{(\phi)} = \left(\overline{\boldsymbol{Z}}_1^{(\phi)}, \ldots, \overline{\boldsymbol{Z}}_a^{(\phi)}\right)$ and $\overline{\boldsymbol{Z}}_i^{(\phi)} = \frac{1}{n_i} \sum_{k=1}^{n_i} \boldsymbol{Z}_{ik}^{(\phi)}$. Following the idea in Harrar and Bathke (2012b), we approximate the distributions of the adjusted hypothesis and error matrices using Wishart distributions,

$$f_{\boldsymbol{H}^{(\phi)}} \cdot \boldsymbol{H}^{(\phi)} \underset{H_0^{(\phi)}}{\overset{approx}{\sim}} W_{rp}\left(f_{\boldsymbol{H}^{(\phi)}}, \boldsymbol{\Sigma}_\phi\right) \quad \text{and} \tag{2.4}$$

$$f_{\boldsymbol{G}^{(\phi)}} \cdot \boldsymbol{G}^{(\phi)} \overset{approx}{\sim} W_{rp}\left(f_{\boldsymbol{G}^{(\phi)}}, \boldsymbol{\Sigma}_\phi\right). \tag{2.5}$$

By matching the expected values and total variances of the adjusted hypothesis and error matrices with that of their approximate Wishart distributions, we solve for the degrees of freedom $f_{\boldsymbol{H}^{(\phi)}}$ and $f_{\boldsymbol{G}^{(\phi)}}$. The final results of the degrees of freedom are

$$f_{\boldsymbol{H}^{(\phi)}} = \frac{\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi^2\right) + \left[\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi\right)\right]^2}{\sum_{i=1}^{a} \sum_{i'=1}^{a} \left(\boldsymbol{C}_{\phi,ii'}\right)^2 \frac{1}{n_i n_{i'}} \left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)} \boldsymbol{\Sigma}_{i'}^{(\phi)}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}_{i'}^{(\phi)}\right)\right]} \quad \text{and} \tag{2.6}$$

$$f_{\boldsymbol{G}^{(\phi)}} = \frac{\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi^2\right) + \left[\operatorname{tr}(\boldsymbol{\Sigma}_\phi)\right]^2}{\sum_{i=1}^{a} \frac{\boldsymbol{C}_{\phi,ii}}{n_i^2(n_i-1)} \left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right) + \left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2\right\}}. \tag{2.7}$$

The mathematical details are presented in the Appendix (Section 2.7).

We propose the modified Wilks' Lambda test statistic as

$$U = U_{rp, f_{\boldsymbol{H}^{(\phi)}}, f_{\boldsymbol{G}^{(\phi)}}} = \frac{|f_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)}|}{|f_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)} + f_{\boldsymbol{H}^{(\phi)}} \boldsymbol{H}^{(\phi)}|}.$$

Using (2.4) and (2.5), the corresponding Rao's F-approximation (Rencher, 2001) for the null distribution would be

$$F = \frac{1 - U^{1/s}}{U^{1/s}} \cdot \frac{\mathrm{df}_2}{\mathrm{df}_1}, \tag{2.8}$$

where

$$s = \begin{cases} \sqrt{\dfrac{(rp)^2 f^2_{\boldsymbol{H}^{(\phi)}} - 4}{(rp)^2 + f^2_{\boldsymbol{H}^{(\phi)}} - 5}}; & \text{if } (rp)^2 + f^2_{\boldsymbol{H}^{(\phi)}} - 5 > 0 \\ 1; & \text{if } (rp)^2 + f^2_{\boldsymbol{H}^{(\phi)}} - 5 \leq 0 \end{cases},$$

$\mathrm{df}_1 = rp \cdot f_{\boldsymbol{H}^{(\phi)}}$, $\mathrm{df}_2 = \omega s - \frac{1}{2}(rp \cdot f_{\boldsymbol{H}^{(\phi)}} - 2)$ and $\omega = f_{\boldsymbol{G}^{(\phi)}} - \frac{1}{2}(rp - f_{\boldsymbol{H}^{(\phi)}} + 1)$. The test will reject the null hypothesis if the F statistic is greater than $F_{1-\alpha,\mathrm{df}_1,\mathrm{df}_2}$, where α is the significance level.

When the sample size is large relative to the dimension, which is the total number of variables, i.e. $p \times t$ variables in the multivariate growth curve data model context with p variables measured at t different occasions, the distribution of the test statistic F in (2.8) can be approximated by a χ^2 distribution. In general, this approximation is less accurate than the Rao's F-approximation.

Affine Invariance

An important property in multivariate analysis is the inferential procedure to be invariant to translation, scaling and rotation of the coordinate system. More generally, the transformation $\boldsymbol{Y}_{ik} = (\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{X}_{ik} + \boldsymbol{b}$ for any $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are $p \times p$ and $t \times t$, respectively, nonsingular matrices and $\boldsymbol{b}$ is an tp-dimensional vector, should not alter the inference. This property is known as *affine invariance*. For transformation that makes sense with multivariate growth curve data, we let $\boldsymbol{A}$ be any nonsingular $p \times p$ matrix, $\boldsymbol{B} = \boldsymbol{I}_t$ and $\boldsymbol{b} = \boldsymbol{1}_t \otimes \boldsymbol{c}$, where $\boldsymbol{c}$ is an p-dimensional vector. Obviously, the degrees of freedom in (2.6) and (2.7) and, therefore, the test statistic U are not affine invariant.

To alleviate this problem, note that the approximations in (2.4) and (2.5) imply

$$f^*_{\boldsymbol{H}^{(\phi)}} \cdot \boldsymbol{H}^{(\phi)}_* \underset{H_0^{(\phi)}}{\overset{approx}{\sim}} W_{rp}\left(f^*_{\boldsymbol{H}^{(\phi)}}, \boldsymbol{I}_{rp}\right) \quad \text{and} \quad f^*_{\boldsymbol{G}^{(\phi)}} \cdot \boldsymbol{G}^{(\phi)}_* \overset{approx}{\sim} W_{rp}\left(f^*_{\boldsymbol{G}^{(\phi)}}, \boldsymbol{I}_{rp}\right), \quad (2.9)$$

where $\boldsymbol{H}^{(\phi)}_* = \boldsymbol{\Sigma}_\phi^{-1/2}\boldsymbol{H}^{(\phi)}\boldsymbol{\Sigma}_\phi^{-1/2}$ and $\boldsymbol{G}^{(\phi)}_* = \boldsymbol{\Sigma}_\phi^{-1/2}\boldsymbol{G}^{(\phi)}\boldsymbol{\Sigma}_\phi^{-1/2}$. These redefinitions of the hypothesis and error sums of squares and cross products matrices change the degrees of freedom in (2.6) and (2.7) to

$$f^*_{\boldsymbol{H}^{(\phi)}} = \frac{rp(1+rp)}{\sum_{i=1}^a \sum_{i'=1}^a \left(\boldsymbol{C}_{\phi,ii'}\right)^2 \frac{1}{n_i n_{i'}} \left[\mathrm{tr}\left(\boldsymbol{A}_i \boldsymbol{A}_{i'}\right) + \mathrm{tr}\left(\boldsymbol{A}_i\right)\mathrm{tr}\left(\boldsymbol{A}_{i'}\right)\right]} \quad (2.10)$$

and

$$f^*_{\boldsymbol{G}^{(\phi)}} = \frac{rp(1+rp)}{\sum_{i=1}^{a} \left(\boldsymbol{C}_{\phi,ii}\right)^2 \frac{1}{n_i^2(n_i-1)} \left\{\operatorname{tr}\left(\boldsymbol{A}_i^2\right) + \left[\operatorname{tr}\left(\boldsymbol{A}_i\right)\right]^2\right\}}, \tag{2.11}$$

respectively, where $\boldsymbol{A}_i = \boldsymbol{\Sigma}_\phi^{-1/2}\boldsymbol{\Sigma}_i^{(\phi)}\boldsymbol{\Sigma}_\phi^{-1/2}$. These degrees of freedom are affine invariant. For $a = 1$, $f^*_{\boldsymbol{H}^{(\phi)}} = 1$, $f^*_{\boldsymbol{G}^{(\phi)}} = n_1 - 1$, and $\boldsymbol{\Sigma}_\phi = \frac{1}{n_1}\boldsymbol{\Sigma}_1^{(\phi)}$. Therefore, the modified version of the approximation reduces to the usual DMM analysis given in Boik (1988). In fact, when $a = 1$, the modified version of the Wilks' Lambda approximation is exactly the same as both DMM and MMM in Boik (1988). When $a = 2$, MMM results in the same test statistic as DMM due to the unique projection matrix $\boldsymbol{C}_\phi$ and generalized inverse being used in MMM method. Our numerical investigations and other researches, for example, Krishnamoorthy and Yu (2004) and Gamage and Mathew (2008) have shown that the results of the affine invariant versions are more stable and consistently better than the unmodified versions. We will use the revised versions in the simulation studies in the next section.

2.4 Simulation Studies

Simulation Design

In this section, we numerically evaluate the proposed test along with the two classical methods for multivariate repeated measures data, namely DMM and MMM, under different simulation settings and with different experimental designs. To investigate the finite-sample performance of the proposed test, we examine its empirical type-I error rates and compare them with the type-I error rates produced by DMM and MMM. For a comprehensive evaluation of the performances, we conduct our simulations under various settings of data distribution, sample size, covariance structure, and factor effects. The proposed test will hereinafter be referred to as WLF. The specific objectives of our simulations are to:

(i) investigate the effect of sample size, including balanced and unbalanced cases,

(ii) analyze the effect of data distribution,

(iii) examine the effect of covariance structure, and

(iv) study the effect of the hypothesis of interest,

on the performance of the tests. All simulations are conducted in **R** version 3.6.0 with $5,000$ simulation runs. The nominal type-I error rate is $\alpha = 0.05$. We consider multivariate growth curve model in the context of factorial designs as well as crossover designs. Specifically, for factorial designs we set the number of groups to be $a = 2$ with sample sizes in the two groups denoted by $\boldsymbol{n} = (n_1, n_2)$. Both balanced and unbalanced cases (increasing sizes and decreasing sizes) are explored. For balanced designs we take $\boldsymbol{n} \in \{(20, 20), (30, 30)\}$ and for unbalanced designs we set $\boldsymbol{n} \in \{(20, 35), (30, 45)\}$.

In addition, we set the dimension (number of response variables) to be $p = 2$ and the number of time points (occasions) to be $t = 2$. In such factorial designs, we are interested in whether there is any Time effect, Group effect and Group $\times$ Time interaction effect. The corresponding contrast matrices for these hypotheses are introduced in Example 1.

For crossover designs, four-period crossover designs (e.g. Li et al., 2020) are investigated in our simulations. There are four periods ($t = 4$). Two different active treatments and two identical placebos are applied in each period separately. The two active treatments are designed to be one period apart and within that one period placebo is applied to avoid carryover effects. There are four different sequences of applications for treatments and placebo ($a = 4$). The number of replications per sequence is denoted by $\boldsymbol{n} = (n_1, n_2, n_3, n_4)$. For balanced designs we use $\boldsymbol{n} \in \{(25, 25, 25, 25), (35, 35, 35, 35)\}$ and for unbalanced designs we take $\boldsymbol{n} \in \{(10, 20, 30, 40), (20, 30, 40, 50)\}$. The number of the response variables is set to be $p = 3$.

In the four-period crossover designs, Period effect, Sequence effect, Treatment effect, Sequence $\times$ Period effect, and Treatment $\times$ Period effect are usually of special interest to medical researchers. To test those effects, the corresponding contrast matrices ($\boldsymbol{D}_\phi$ and $\boldsymbol{C}_\phi$) are $\boldsymbol{D}_{\mathrm{P}} = (\boldsymbol{I}_{t-1}, -\boldsymbol{1}_{t-1})$ and $\boldsymbol{C}_{\mathrm{P}} = \boldsymbol{J}_a/a$ for Period effect,

$\boldsymbol{D}_{\mathrm{S}} = \mathbf{1}_t^\top/t$ and $\boldsymbol{C}_{\mathrm{S}} = \boldsymbol{P}_a/(a-1)$ for Sequence effect, $\boldsymbol{D}_{\mathrm{Tr}} = (1,-1,-1,1)$ and $\boldsymbol{C}_{\mathrm{Tr}} = H_{\mathrm{Tr}}^\top \left(H_{\mathrm{Tr}} H_{\mathrm{Tr}}^\top\right)^- H_{\mathrm{Tr}} = (2\boldsymbol{I}_2 - \boldsymbol{J}_2) \otimes (2\boldsymbol{I}_2 - \boldsymbol{J}_2)/4$, where $H_{\mathrm{Tr}} = (1,-1,-1,1)$, for Treatment effect, $\boldsymbol{D}_{\mathrm{SP}} = (\boldsymbol{I}_{t-1}, -\mathbf{1}_{t-1})$ and $\boldsymbol{C}_{\mathrm{SP}} = \boldsymbol{P}_a/(a-1)$ for Sequence$\times$Period effect, $\boldsymbol{D}_{\mathrm{TrP1}} = \boldsymbol{I}_2 \otimes \mathbf{1}_2^\top$, $\boldsymbol{D}_{\mathrm{TrP2}} = (\boldsymbol{I}_2, \boldsymbol{J}_2 - \boldsymbol{I}_2)$, $\boldsymbol{C}_{\mathrm{TrP1}} = H_{\mathrm{TrP1}}^\top \left(H_{\mathrm{TrP1}} H_{\mathrm{TrP1}}^\top\right)^- H_{\mathrm{TrP1}} = \boldsymbol{I}_2 \otimes (2\boldsymbol{I}_2 - \boldsymbol{J}_2)/2$, where $H_{\mathrm{TrP1}} = \boldsymbol{I}_2 \otimes (1,-1)$, and $\boldsymbol{C}_{\mathrm{TrP2}} = H_{\mathrm{TrP2}}^\top \left(H_{\mathrm{TrP2}} H_{\mathrm{TrP2}}^\top\right)^- H_{\mathrm{TrP2}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 \\ 0 & -1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, where $H_{\mathrm{TrP2}} = (0,1,-1,0)$, for Treatment $\times$ Period effect.

For covariance structure, homoscedastic and heteroscedastic covariances with compound symmetric and autoregressive structures are considered. Partitioning the covariance for the i^{th} group (sequence) as $\boldsymbol{\Sigma}_i = (\boldsymbol{\Sigma}_{i,jk})$, the two covariance structures considered are below.

(i) Compound symmetry:

$$\boldsymbol{\Sigma}_{i,jk} = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}_p & k = j = 1, \ldots, t \\ \boldsymbol{\Sigma}_{i,jk} = \rho_i \boldsymbol{J}_p & k \neq j \text{ and } k, j = 1, \ldots, t \end{cases},$$

where for the factorial designs we take $\boldsymbol{\rho} = (0.2, 0.2)$ as the equal covariances setting and $\boldsymbol{\rho} = (0.2, 0.7)$ as the unequal covariances setting, and for the crossover designs we consider $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$ as the equal covariances setting and $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$ as the unequal covariances setting.

(ii) Autoregressive structure AR(1):

$$\boldsymbol{\Sigma}_{i,jk} = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}_p & k = j = 1, \ldots, t \\ \boldsymbol{\Sigma}_{i,jk} = \rho_i^{|k-j|} \boldsymbol{J}_p & k \neq j \text{ and } k = 1, \ldots, t \end{cases},$$

where for the factorial designs we let $\boldsymbol{\rho} = (0.2, 0.2)$ be the equal covariances setting and $\boldsymbol{\rho} = (0.2, 0.7)$ be the unequal covariances setting, and for the crossover designs we make $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$ the equal covariances setting and $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$ the unequal covariances setting.

The compound symmetric and the first-order autoregressive covariance structures will hereinafter be referred to as CS and AR, respectively. The equal covariances settings have the same correlation value in different groups or sequences, representing homoscedasticity. However, the unequal covariances settings have different correlation values in different groups or sequences, serving as heteroscedastic cases. In the heteroscedastic cases, we investigate the performance of the tests by considering the settings where the unequal group (sequence) covariances are associated with either increasing or decreasing sample sizes.

Simulation Results

Factorial Design

Tables 2.1–2.2 show empirical type-I error rates of factorial designs for testing Time, Group and Group $\times$ Time effects with equal covariances. Table 2.1 shows the cases in balanced designs, but Table 2.2 shows the cases in unbalanced designs. Note that for settings in our factorial designs, i.e. $a = 2$, DMM and MMM are identical. Therefore, we only present comparisons with DMM. When covariances are equal in two groups, the type-I error rates are maintained well for both WLF and DMM with normal, $t(5)$ and $\chi^2(5)$ distributions and the results are comparable. When data are highly skewed from lognormal distribution, the type-I error rates produced by three methods are still comparable but a little more conservative than that of the normal, $t(5)$ and $\chi^2(5)$ cases.

Tables 2.3–2.5 present simulation results of factorial designs testing Time, Group and Group $\times$ Time effects with unequal covariances. When sample sizes are equal as shown in Table 2.3, the results of all three methods are roughly the same. When the data are from normal, $t(5)$ and $\chi^2(5)$ distributions, the type-I error rates are almost exact. When data are from highly skewed lognormal distribution, all three methods are still comparable but a little more conservative. However, Tables 2.4 and 2.5 show that the proposed test WLF performs very well with almost exact type-I error rates in unbalanced heteroscedastic designs, for both positive and negative pairing situations, especially for normal, $t(5)$ and $\chi^2(5)$ distributions and for testing all effects (Time, Group and Group×Time). The classical method DMM and MMM, on the other hand, lead to very liberal type-I error rates for positive pairing cases (Table 2.4) and very conservative type-I error rates for negative pairing cases (Table 2.5), especially for testing Time and Group × Time effects. Generally, when data come from lognormal distribution, WLF is still the best one among all three tests producing type-I error rates that are fairly close to 5% compared with the other methods.

Four-Period Crossover Design

Tables 2.6–2.7 present the test results when covariances in the four sequences are equal. Table 2.6 shows the cases in balanced designs. Table 2.7, on the other hand, shows the cases in unbalanced designs. Generally, when data come from normal, $t(5)$ and $\chi^2(5)$ distributions with equal covariances, the type-I error rates of WLF and DMM are almost exact. However, the classical MMM test performs well only when testing the main effects of Period and Treatment. When testing Treatment × Period effect, the test results of MMM tend to be more liberal. When data are lognormal with equal covariances, all of the three methods are still comparable for testing Period and Treatment effects.

Simulation results in the crossover designs with unequal covariances are shown in Tables 2.8–2.10. Like in the factorial designs, WLF has a clear advantage over DMM and MMM, especially when sample sizes are unequal. Table 2.8 shows type-I error rates in balanced designs with unequal covariances. All of the three methods perform comparably well in most cases, especially with normal, $t(5)$ and $\chi^2(5)$ distributions. When sample sizes are unequal in Tables 2.9 and 2.10, WLF maintains the preassigned type-I error rates well to the 0.05 level for normal, $t(5)$ and $\chi^2(5)$ data and for testing all effects (Period, Treatment and Treatment × Period). However, under the same settings, DMM and MMM tend to have very liberal type-I error rates when sample sizes increase (positive pairing) as shown in Table 2.9, but tend to be very conservative when sample sizes decrease (negative pairing) as shown in Table 2.10. It can further be seen in Tables 2.9 and 2.10 that for lognormal data and for testing Period and

Treatment effects, WLF is still the best among the three methods and it keeps type-I error rates closer to the nominal level. Note that when testing Treatment effect, DMM and MMM lead to the same numerical results. The reason is that when testing Treatment effect, the test statistics in the two methods reduce to the same quantity.

Table 2.1: Type-I error rate (×100) of WLF, DMM and MMM for equal covariances with equal sample sizes in factorial designs, $a = 2$, $p = 2$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Time			Group			Group×Time		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	(20,20)	5.0	5.1	5.1	5.2	5.2	5.2	4.9	5.0	5.0
		(30,30)	5.0	5.0	5.0	5.2	5.2	5.2	5.0	5.0	5.0
	AR	(20,20)	4.5	4.6	4.6	4.7	4.8	4.8	5.0	5.0	5.0
		(30,30)	4.5	4.6	4.6	5.5	5.6	5.6	5.3	5.3	5.3
$t(5)$	CS	(20,20)	4.2	4.2	4.2	4.7	4.8	4.8	4.7	4.8	4.8
		(30,30)	4.8	4.8	4.8	4.9	4.9	4.9	5.0	5.0	5.0
	AR	(20,20)	4.8	4.9	4.9	4.7	4.8	4.8	4.4	4.4	4.4
		(30,30)	4.7	4.7	4.7	4.7	4.7	4.7	4.9	4.9	4.9
$\chi^2(5)$	CS	(20,20)	5.0	5.2	5.2	4.9	5.1	5.1	4.8	4.9	4.9
		(30,30)	5.1	5.2	5.2	4.3	4.4	4.4	4.7	4.9	4.9
	AR	(20,20)	5.0	5.1	5.1	4.6	4.7	4.7	4.7	4.8	4.8
		(30,30)	4.9	5.0	5.0	5.2	5.2	5.2	4.6	4.6	4.6
Lognormal	CS	(20,20)	3.6	3.7	3.7	3.4	3.7	3.7	3.6	3.7	3.7
		(30,30)	4.0	4.1	4.1	4.1	4.3	4.3	4.2	4.3	4.3
	AR	(20,20)	3.7	3.9	3.9	3.5	3.8	3.8	3.8	4.0	4.0
		(30,30)	4.0	4.2	4.2	4.4	4.6	4.6	3.9	4.1	4.1

Power Studies

Power and Effect Size

We will investigate the empirical power of the three methods to detect a fixed alternative in factorial designs. For power simulations of the crossover designs, the results are similar to that of the factorial designs and, hence, we only show the cases in factorial designs for illustration purpose. It is widely known that with a fixed sample size and a fixed effect size, a test procedure tends to have a larger (smaller) power when the type-I error rate is larger (smaller) (Cohen, 2013). In order to show the

Table 2.2: Type-I error rate ($\times 100$) of WLF, DMM and MMM for equal covariances with unequal sample sizes in factorial designs, $a = 2$, $p = 2$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Time			Group			Group×Time		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	(35,20)	4.8	4.9	4.9	5.1	5.4	5.4	4.2	4.2	4.2
		(45,30)	4.8	4.8	4.8	5.1	5.2	5.2	5.3	5.3	5.3
	AR	(35,20)	4.8	4.6	4.6	5.0	4.9	4.9	4.9	5.0	5.0
		(45,30)	4.6	4.7	4.7	4.6	4.6	4.6	5.5	5.6	5.6
$t(5)$	CS	(35,20)	4.5	4.6	4.6	4.7	4.8	4.8	4.6	4.5	4.5
		(45,30)	4.9	5.0	5.0	4.3	4.4	4.4	4.7	4.7	4.7
	AR	(35,20)	4.4	4.6	4.6	5.1	5.2	5.2	5.0	5.4	5.4
		(45,30)	4.9	5.1	5.1	4.9	4.9	4.9	4.4	4.5	4.5
$\chi^2\,(5)$	CS	(35,20)	5.0	5.1	5.1	4.9	4.8	4.8	4.8	4.9	4.9
		(45,30)	5.0	5.1	5.1	4.9	4.7	4.7	4.6	4.6	4.6
	AR	(35,20)	5.5	5.5	5.5	4.8	4.8	4.8	4.4	4.5	4.5
		(45,30)	4.8	4.7	4.7	4.9	4.9	4.9	4.8	4.9	4.9
Lognormal	CS	(35,20)	3.8	4.5	4.5	4.5	4.2	4.2	3.8	4.4	4.4
		(45,30)	4.3	4.7	4.7	4.4	4.5	4.5	4.2	4.3	4.3
	AR	(35,20)	4.1	4.1	4.1	4.5	4.2	4.2	3.9	4.1	4.1
		(45,30)	3.9	4.1	4.1	4.1	4.2	4.2	4.5	4.6	4.6

power advantage of the proposed WLF test, we present the worst-case scenarios of the WLF power simulations, i.e., in positive pairing unbalanced heteroscedastic designs. Figure 2.1 shows the power curves for testing Group effect of all the three tests in factorial design with different data distributions. The number of variables is $p = 2$ and the number of time points (occasions) is $t = 3$. There are two groups ($a = 2$) with sample sizes $\boldsymbol{n} = (20, 35)$. The covariance structure used is compound symmetry with correlation values $\boldsymbol{\rho} = (0.2, 0.7)$. The bottom row of Figure 2.1 presents the power curves of all the three tests along different effect sizes of the Group factor when the effect of the Time factor is 0. With such settings, we have seen from the previous section that, in general, the type-I error rates of WLF are nearly exact for all data distributions (normal, $t(5)$, $\chi^2\,(5)$ and lognormal). However, DMM and MMM are much more liberal under the same conditions. The power advantage of WLF can be readily seen when we increase the effect size of the Group factor just by a little, where the power of WLF catches up with that of DMM and MMM. Moreover, WLF

Table 2.3: Type-I error rate ($\times 100$) of WLF, DMM and MMM for unequal covariances with equal sample sizes in factorial designs, $a = 2$, $p = 2$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	n	Time			Group			Group×Time		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	(20,20)	5.7	6.0	6.0	4.6	4.7	4.7	5.4	5.7	5.7
		(30,30)	5.0	5.1	5.1	4.7	4.7	4.7	5.4	5.7	5.7
	AR	(20,20)	4.6	5.0	5.0	4.4	4.5	4.5	4.8	5.1	5.1
		(30,30)	5.1	5.3	5.3	5.3	5.4	5.4	4.8	4.9	4.9
$t(5)$	CS	(20,20)	4.7	5.0	5.0	4.6	4.7	4.7	4.8	5.0	5.0
		(30,30)	4.9	5.3	5.3	5.0	5.1	5.1	4.9	5.0	5.0
	AR	(20,20)	4.9	5.3	5.3	4.8	5.0	5.0	5.0	5.4	5.4
		(30,30)	4.3	4.4	4.4	5.0	5.1	5.1	4.8	4.9	4.9
$\chi^2(5)$	CS	(20,20)	5.1	5.3	5.3	4.7	4.9	4.9	5.1	5.5	5.5
		(30,30)	5.4	5.5	5.5	4.5	4.7	4.7	5.0	5.3	5.3
	AR	(20,20)	4.8	5.3	5.3	4.8	5.0	5.0	4.9	5.2	5.2
		(30,30)	4.8	5.1	5.1	5.0	5.1	5.1	4.4	4.7	4.7
Lognormal	CS	(20,20)	3.9	4.1	4.1	3.9	4.2	4.2	4.0	4.2	4.2
		(30,30)	4.1	4.3	4.3	4.7	4.9	4.9	3.6	3.8	3.8
	AR	(20,20)	3.7	4.1	4.1	3.9	4.2	4.2	3.9	4.2	4.2
		(30,30)	4.1	4.4	4.4	3.8	3.9	3.9	4.0	4.2	4.2

keeps its leading position until all the three tests reach the plateau at 1. The top row exhibits that along different effect sizes of the Group factor, the mean power at all values of effect of the Time factor. It is obvious that the power curves on the bottom row and the mean power curves on the top row are almost identical. Likewise, WLF has the overall advantage in terms of power.

Power and Sample Size

In this section, we examine the empirical power of all three methods to detect a fixed alternative in factorial designs at different sample size levels. For the same reason mentioned above, we investigate the power performance of all the three methods against sample size in positive pairing unbalanced heteroscedastic designs. Again, the number of variables is $p = 2$. There are three time points, $t = 3$. Two groups ($a = 2$) of data are simulated with sample sizes $\boldsymbol{n}^{(1)} = (8, 14)$, $\boldsymbol{n}^{(2)} = (12, 21)$, $\boldsymbol{n}^{(3)} = (16, 28)$, $\boldsymbol{n}^{(4)} = (20, 35)$, $\boldsymbol{n}^{(5)} = (24, 42)$ and $\boldsymbol{n}^{(6)} = (28, 49)$. The covariance

Table 2.4: Type-I error rate (×100) of WLF, DMM and MMM for unequal covariances with increasing sample sizes in factorial designs, $a = 2$, $p = 2$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Time			Group			Group×Time		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	(20,35)	5.2	10.3	10.3	4.7	6.2	6.2	5.5	10.7	10.7
		(30,45)	5.1	8.5	8.5	4.8	5.7	5.7	4.8	8.2	8.2
	AR	(20,35)	5.4	10.3	10.3	4.6	5.9	5.9	4.5	9.3	9.3
		(30,45)	4.8	8.0	8.0	5.1	5.8	5.8	4.8	8.1	8.1
$t(5)$	CS	(20,35)	4.9	9.9	9.9	5.6	7.1	7.1	5.4	10.0	10.0
		(30,45)	5.4	8.7	8.7	5.3	6.0	6.0	5.0	8.2	8.2
	AR	(20,35)	4.5	9.2	9.2	4.6	5.3	5.3	5.1	10.1	10.1
		(30,45)	4.8	8.0	8.0	4.7	5.6	5.6	4.9	8.1	8.1
$\chi^2(5)$	CS	(20,35)	4.6	9.1	9.1	5.3	6.2	6.2	4.6	9.4	9.4
		(30,45)	4.6	7.9	7.9	5.2	6.0	6.0	4.8	8.5	8.5
	AR	(20,35)	4.7	9.6	9.6	4.7	5.9	5.9	4.6	9.5	9.5
		(30,45)	5.5	8.4	8.4	4.7	5.8	5.8	5.3	8.4	8.4
Lognormal	CS	(20,35)	3.3	7.5	7.5	4.9	5.2	5.2	3.6	7.8	7.8
		(30,45)	3.7	6.7	6.7	4.1	4.8	4.8	4.2	6.6	6.6
	AR	(20,35)	4.2	8.5	8.5	4.4	5.8	5.8	3.9	8.3	8.3
		(30,45)	3.6	6.8	6.8	4.7	5.1	5.1	4.3	6.8	6.8

structures for the two groups are compound symmetric with $\boldsymbol{\rho} = (0.2, 0.7)$. Four data distributions are investigated (normal, $t(5)$, $\chi^2(5)$ and lognormal). The effect of the Time factor is set to be 0 throughout our simulations investigating the relationship between power and sample size.

Figure 2.2 shows the power curves along the sample sizes (increasing sample size pairs) for testing Group effect at three different effect sizes of Group factor $\delta_{\text{Group}} = 0, 0.2$ and 0.3. For all data distributions, the type-I error rates of WLF are almost exact as shown in the bottom row ($\delta_{\text{Group}} = 0$), but DMM and MMM are much more liberal. In the second row with $\delta_{\text{Group}} = 0.2$, the power of WLF appears to catch up with that of both DMM and MMM as sample sizes increase for all four data distributions. In the top row ($\delta_{\text{Group}} = 0.3$), the power of WLF exceeds that of DMM and MMM and it keeps the leading position until it plateaus at 1. It is generally true that a smaller (larger) type-I error rate results in a smaller (larger) power (Cohen, 2013). However, the transition from the bottom row to the top row indicates that the

Table 2.5: Type-I error rate ($\times 100$) of WLF, DMM and MMM for unequal covariances with decreasing sample sizes in factorial designs, $a = 2$, $p = 2$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Time			Group			Group×Time		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	(35,20)	5.1	2.4	2.4	5.3	5.4	5.4	4.6	2.1	2.1
		(45,30)	4.9	3.1	3.1	4.6	4.5	4.5	4.5	2.5	2.5
	AR	(35,20)	4.7	2.1	2.1	4.7	4.4	4.4	5.4	2.5	2.5
		(45,30)	4.8	2.7	2.7	5.0	5.0	5.0	4.9	2.8	2.8
$t(5)$	CS	(35,20)	4.5	1.9	1.9	4.6	4.7	4.7	4.9	2.2	2.2
		(45,30)	4.8	2.7	2.7	4.8	5.0	5.0	4.8	2.9	2.9
	AR	(35,20)	4.7	2.4	2.4	4.7	4.4	4.4	4.7	2.5	2.5
		(45,30)	5.4	3.2	3.2	4.4	4.5	4.5	5.2	2.9	2.9
$\chi^2(5)$	CS	(35,20)	5.0	2.3	2.3	5.0	4.9	4.9	5.3	2.6	2.6
		(45,30)	5.3	3.2	3.2	5.1	5.1	5.1	4.9	2.8	2.8
	AR	(35,20)	4.9	2.1	2.1	4.8	5.0	5.0	4.6	2.2	2.2
		(45,30)	4.5	2.3	2.3	4.6	4.3	4.3	5.2	3.2	3.2
Lognormal	CS	(35,20)	4.2	2.4	2.4	5.2	5.1	5.1	4.1	2.3	2.3
		(45,30)	4.0	2.4	2.4	4.9	4.4	4.4	3.6	2.2	2.2
	AR	(35,20)	4.0	2.0	2.0	4.9	4.5	4.5	4.0	2.3	2.3
		(45,30)	4.4	2.4	2.4	4.6	4.3	4.3	4.2	2.7	2.7

proposed WLF method not only achieves the best type-I error rate but also quickly catches up in power and further outperforms the other two competing methods. In other words, the best performance of WLF in terms of type-I error rate does not come at the expense of reduced power.

2.5 Application

To stimulate readers' interest and to illustrate how our methodology can be applied in real life situations, we demonstrate the method using one example for factorial design (Idiopathic Infantile Nystagmus Syndrome data) and another example for crossover design (Intraocular Pressure data). In both examples, the data are nonnormal multivariate repeated measures with small sample.

Table 2.6: Type-I error rate ($\times 100$) of WLF, DMM and MMM for equal covariances with equal sample sizes in crossover designs, $a = 4$, $p = 3$, $t = 4$, CS: $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$, AR: $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$, $\boldsymbol{n}^{(1)} = (25, 25, 25, 25)$, $\boldsymbol{n}^{(2)} = (35, 35, 35, 35)$, $\boldsymbol{n}^{(3)} = (10, 20, 30, 40)$, $\boldsymbol{n}^{(4)} = (20, 30, 40, 50)$, $\boldsymbol{n}^{(5)} = (40, 30, 20, 10)$, $\boldsymbol{n}^{(6)} = (50, 40, 30, 20)$.

Dist	Cov	**n**	Period			Treatment			Treatment×Period		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	$\boldsymbol{n}^{(1)}$	5.2	5.3	5.4	4.6	4.7	4.7	4.6	4.6	6.3
		$\boldsymbol{n}^{(2)}$	4.7	4.8	4.6	5.5	5.5	5.5	4.4	4.7	5.9
	AR	$\boldsymbol{n}^{(1)}$	4.9	5.0	5.5	4.9	4.9	4.9	4.4	4.7	5.5
		$\boldsymbol{n}^{(2)}$	5.1	5.1	6.0	4.7	4.7	4.7	4.9	4.9	5.9
$t(5)$	CS	$\boldsymbol{n}^{(1)}$	4.3	4.5	4.9	5.2	5.2	5.2	4.0	4.3	5.8
		$\boldsymbol{n}^{(2)}$	4.8	4.9	4.8	5.6	5.6	5.6	4.6	5.0	6.6
	AR	$\boldsymbol{n}^{(1)}$	4.0	4.2	5.7	4.6	4.7	4.7	4.5	4.7	5.4
		$\boldsymbol{n}^{(2)}$	4.7	4.9	5.7	4.7	4.7	4.7	5.1	5.2	5.1
$\chi^2(5)$	CS	$\boldsymbol{n}^{(1)}$	5.3	5.3	4.9	5.2	5.3	5.3	4.0	4.3	6.3
		$\boldsymbol{n}^{(2)}$	5.1	5.3	4.8	4.7	4.7	4.7	5.0	5.1	6.8
	AR	$\boldsymbol{n}^{(1)}$	5.1	5.2	5.9	5.0	5.0	5.0	4.2	4.5	5.1
		$\boldsymbol{n}^{(2)}$	4.6	4.7	5.6	5.1	5.3	5.3	4.7	5.1	5.9
Lognormal	CS	$\boldsymbol{n}^{(1)}$	5.6	6.2	4.0	4.3	4.5	4.5	3.5	4.1	6.1
		$\boldsymbol{n}^{(2)}$	5.8	6.1	4.0	4.1	4.1	4.1	3.5	3.8	5.8
	AR	$\boldsymbol{n}^{(1)}$	6.3	6.7	5.6	3.9	4.1	4.1	2.7	3.7	4.7
		$\boldsymbol{n}^{(2)}$	6.4	6.7	5.6	4.2	4.3	4.3	3.3	3.9	4.9

Idiopathic Infantile Nystagmus Syndrome

In this section, we introduce an example where multivariate growth curve data on idiopathic infantile nystagmus syndrome (INS) need to be analyzed in the context of general factorial design. In this study (Fadardi et al., 2017), 15 voluntary participants with idiopathic INS were recruited from a referring ophthalmologist. Participants were asked to carry out acuity tasks identifying the direction of horizontal Tumbling-E targets under different mental load settings. For the low mental load setting, participants were given unlimited time to respond. After responding, they were required to view a fixation cross for 100 milliseconds prior to the presence of the next acuity target. For the high mental load setting, participants were given only 0.5 second to view the target and then 300 milliseconds to view a visual noise mask. Participants were required to respond while they were viewing a fixation cross for 1

Table 2.7: Type-I error rate (×100) of WLF, DMM and MMM for equal covariances with unequal sample sizes in crossover designs, $a = 4$, $p = 3$, $t = 4$, CS: $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$, AR: $\boldsymbol{\rho} = (0.3, 0.3, 0.3, 0.3)$, $\boldsymbol{n}^{(1)} = (25, 25, 25, 25)$, $\boldsymbol{n}^{(2)} = (35, 35, 35, 35)$, $\boldsymbol{n}^{(3)} = (10, 20, 30, 40)$, $\boldsymbol{n}^{(4)} = (20, 30, 40, 50)$, $\boldsymbol{n}^{(5)} = (40, 30, 20, 10)$, $\boldsymbol{n}^{(6)} = (50, 40, 30, 20)$.

Dist	Cov	**n**	Period			Treatment			Treatment×Period		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	$\boldsymbol{n}^{(5)}$	5.4	4.9	5.1	4.8	5.1	5.1	5.1	4.4	6.4
		$\boldsymbol{n}^{(6)}$	4.8	5.0	4.9	4.8	4.6	4.6	5.2	4.9	6.5
	AR	$\boldsymbol{n}^{(5)}$	5.3	4.9	6.0	5.3	5.0	5.0	5.6	4.7	5.9
		$\boldsymbol{n}^{(6)}$	5.1	5.0	5.6	4.9	5.1	5.1	4.7	4.4	5.2
$t(5)$	CS	$\boldsymbol{n}^{(5)}$	4.6	4.6	5.0	5.2	4.9	4.9	4.4	4.6	6.4
		$\boldsymbol{n}^{(6)}$	4.6	4.9	4.6	4.8	5.1	5.1	4.1	4.2	5.8
	AR	$\boldsymbol{n}^{(5)}$	4.9	4.6	5.5	4.9	5.2	5.2	5.1	4.9	5.1
		$\boldsymbol{n}^{(6)}$	5.3	5.1	6.0	4.7	4.8	4.8	4.9	4.8	5.7
$\chi^2(5)$	CS	$\boldsymbol{n}^{(5)}$	5.3	5.0	5.0	4.8	4.8	4.8	5.0	5.1	6.4
		$\boldsymbol{n}^{(6)}$	5.3	5.4	4.6	5.1	5.2	5.2	5.1	4.7	6.0
	AR	$\boldsymbol{n}^{(5)}$	5.2	5.0	5.4	4.8	5.3	5.3	5.3	5.3	5.6
		$\boldsymbol{n}^{(6)}$	5.5	4.9	5.8	5.4	5.0	5.0	5.2	5.0	5.9
Lognormal	CS	$\boldsymbol{n}^{(5)}$	5.8	6.7	6.2	4.7	5.6	5.6	3.8	4.7	7.5
		$\boldsymbol{n}^{(6)}$	6.1	6.4	4.9	4.5	4.8	4.8	3.5	4.6	6.5
	AR	$\boldsymbol{n}^{(5)}$	5.8	6.6	6.7	4.4	5.0	5.0	3.5	4.8	6.2
		$\boldsymbol{n}^{(6)}$	6.3	6.2	5.4	4.5	4.9	4.9	3.3	4.5	5.1

second. In addition, participants were also asked to conduct mental arithmetic (continuously subtracting 7 from a number randomly selected between 100 and 120 and given by the examiner during the task) simultaneously with the acuity task. Both the low and the high mental load effects were evaluated at two gaze positions (null position and away position). Eventually, the size and contrast of the target at which participants' task performance plateaued were recorded. The main objective of the study is to investigate whether there is any main effect of mental load (M), main effect of gaze position (P), and interaction effect between the mental load and gaze position (MP).

Among all 15 participants with idiopathic infantile nystagmus syndrome, 11 of them finished the task with no missing data. To test the interaction effect mentioned above in the context of our method, we need the key parameters. Since all

Table 2.8: Type-I error rate (×100) of WLF, DMM and MMM for unequal covariances with equal sample sizes in crossover designs, $a = 4$, $p = 3$, $t = 4$, CS: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, AR: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, $\boldsymbol{n}^{(1)} = (25, 25, 25, 25)$, $\boldsymbol{n}^{(2)} = (35, 35, 35, 35)$, $\boldsymbol{n}^{(3)} = (10, 20, 30, 40)$, $\boldsymbol{n}^{(4)} = (20, 30, 40, 50)$, $\boldsymbol{n}^{(5)} = (40, 30, 20, 10)$, $\boldsymbol{n}^{(6)} = (50, 40, 30, 20)$.

Dist	Cov	**n**	Period			Treatment			Treatment×Period		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	$\boldsymbol{n}^{(1)}$	4.8	4.9	5.2	4.5	4.5	4.5	5.1	5.1	7.0
		$\boldsymbol{n}^{(2)}$	4.8	4.9	4.9	4.6	4.7	4.7	4.9	5.1	6.3
	AR	$\boldsymbol{n}^{(1)}$	5.1	5.4	6.2	5.1	5.2	5.2	4.9	5.3	6.6
		$\boldsymbol{n}^{(2)}$	5.2	5.5	6.4	5.2	5.2	5.2	5.3	5.5	6.0
$t(5)$	CS	$\boldsymbol{n}^{(1)}$	4.7	4.8	5.2	5.6	5.7	5.7	4.6	5.3	6.7
		$\boldsymbol{n}^{(2)}$	4.6	4.8	4.4	5.2	5.4	5.4	4.9	5.1	6.7
	AR	$\boldsymbol{n}^{(1)}$	4.5	4.6	6.1	4.7	4.8	4.8	4.6	4.9	6.5
		$\boldsymbol{n}^{(2)}$	5.0	5.2	6.3	5.0	5.0	5.0	4.3	4.6	5.8
$\chi^2(5)$	CS	$\boldsymbol{n}^{(1)}$	4.8	5.2	4.8	4.6	4.6	4.6	4.8	5.1	6.9
		$\boldsymbol{n}^{(2)}$	5.3	5.5	5.0	5.1	5.2	5.2	4.8	4.9	6.7
	AR	$\boldsymbol{n}^{(1)}$	4.3	4.6	5.6	4.8	4.9	4.9	4.2	4.5	5.3
		$\boldsymbol{n}^{(2)}$	4.9	5.1	5.8	5.0	5.1	5.1	4.6	4.8	5.1
Lognormal	CS	$\boldsymbol{n}^{(1)}$	5.9	6.5	4.4	4.3	4.5	4.5	3.2	4.5	6.8
		$\boldsymbol{n}^{(2)}$	5.7	5.9	4.6	4.2	4.3	4.3	3.9	4.8	7.0
	AR	$\boldsymbol{n}^{(1)}$	6.3	6.8	5.4	4.5	4.6	4.6	3.1	4.0	5.3
		$\boldsymbol{n}^{(2)}$	6.1	6.4	5.7	4.5	4.6	4.6	3.5	4.2	5.2

participants have the disease, there is only one group, i.e., $a = 1$. There are two response variables measured each time, target size and contrast ($p = 2$). There are four repeated measures ($t = 4$) representing four different occasions, low mental load at null position, low mental load at away position, high mental load at null position, and high mental load at away position. Since $a = 1$, $\boldsymbol{C}_\phi$ is always 1, where $\phi = \{\mathrm{M}, \mathrm{P}, \mathrm{MP}\}$. The corresponding contrast matrix for testing the interaction effect between the mental load and gaze position is $\boldsymbol{D}_{\mathrm{MP}} = (1, -1, -1, 1)$. The contrast matrices for testing main effects of mental load and gaze position are $\boldsymbol{D}_{\mathrm{M}} = (1, 1, -1, -1)$ and $\boldsymbol{D}_{\mathrm{P}} = (1, -1, 1, -1)$, respectively.

Of the 11 participants, 3 are female and 8 are male. It might also be interesting to investigate whether there are other interaction effects, for example, gender × mental (GM), gender × position (GP), and gender × mental × position (GMP). In this case,

Table 2.9: Type-I error rate ($\times 100$) of WLF, DMM and MMM for unequal covariances with increasing sample sizes in crossover designs (positive pairing), $a = 4$, $p = 3$, $t = 4$, CS: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, AR: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, $\boldsymbol{n}^{(1)} = (25, 25, 25, 25)$, $\boldsymbol{n}^{(2)} = (35, 35, 35, 35)$, $\boldsymbol{n}^{(3)} = (10, 20, 30, 40)$, $\boldsymbol{n}^{(4)} = (20, 30, 40, 50)$, $\boldsymbol{n}^{(5)} = (40, 30, 20, 10)$, $\boldsymbol{n}^{(6)} = (50, 40, 30, 20)$.

Dist	Cov	n	Period			Treatment			Treatment×Period		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	$\boldsymbol{n}^{(3)}$	5.8	12.7	13.2	4.3	9.0	9.0	5.2	7.2	7.8
		$\boldsymbol{n}^{(4)}$	4.7	9.4	9.8	5.5	8.1	8.1	4.8	7.1	8.2
	AR	$\boldsymbol{n}^{(3)}$	5.7	13.1	10.9	5.0	9.2	9.2	5.2	7.7	8.4
		$\boldsymbol{n}^{(4)}$	5.3	10.6	9.7	5.2	7.8	7.8	5.4	6.5	7.0
$t(5)$	CS	$\boldsymbol{n}^{(3)}$	5.4	11.9	12.2	4.8	9.4	9.4	5.1	8.0	8.7
		$\boldsymbol{n}^{(4)}$	4.5	9.4	9.4	4.5	7.2	7.2	4.6	7.0	7.9
	AR	$\boldsymbol{n}^{(3)}$	5.8	14.0	12.5	5.4	9.4	9.4	4.7	7.6	7.9
		$\boldsymbol{n}^{(4)}$	4.9	11.1	10.1	4.6	7.4	7.4	4.9	6.4	7.2
$\chi^2(5)$	CS	$\boldsymbol{n}^{(3)}$	5.8	12.3	12.0	4.7	9.1	9.1	4.9	7.8	8.8
		$\boldsymbol{n}^{(4)}$	4.6	8.7	8.8	4.8	7.5	7.5	4.5	7.0	8.1
	AR	$\boldsymbol{n}^{(3)}$	6.3	14.4	12.0	4.9	9.3	9.3	4.9	7.7	8.2
		$\boldsymbol{n}^{(4)}$	5.3	10.4	9.0	5.1	7.6	7.6	5.3	7.0	7.3
Lognormal	CS	$\boldsymbol{n}^{(3)}$	6.0	13.0	12.2	4.0	9.5	9.5	3.2	6.3	8.2
		$\boldsymbol{n}^{(4)}$	6.0	10.6	9.1	4.5	7.4	7.4	3.8	6.4	7.5
	AR	$\boldsymbol{n}^{(3)}$	6.8	14.6	11.7	4.4	9.5	9.5	3.6	6.4	7.7
		$\boldsymbol{n}^{(4)}$	6.6	11.7	9.7	4.3	6.9	6.9	4.0	5.9	6.9

we have $a = 2$ for two groups, i.e., female group and male group, $p = 2$ for two response variables and $t = 4$ for the four occasions. Accordingly, the contrast matrix $\boldsymbol{C}_\phi$ is always $\boldsymbol{P}_2$, where $\phi = \{\text{GM}, \text{GP}, \text{GMP}\}$ for testing all three interaction effects. However, $\boldsymbol{D}_\phi$ varies with effect being tested and they are $\boldsymbol{D}_{\text{GM}} = (1, -1) \otimes \frac{1}{2}\mathbf{1}_2^\top$, $\boldsymbol{D}_{\text{GP}} = \frac{1}{2}\mathbf{1}_2^\top \otimes (1, -1)$, and $\boldsymbol{D}_{\text{GMP}} = (1, -1) \otimes (1, -1)$.

Before conducting hypotheses tests, we perform a preliminary check on the data. The marginal distributions of the variables, size and contrast, are both highly right-skewed and clearly they are nonnormal. Further, the covariance matrices of the female and male groups are not equal. As we have seen in Section 2.4, the proposed WLF method performs the best when data are nonnormal with unequal sample sizes and unequal covariances. Hence, the WLF method is the most reliable method for analyzing the optometry data.

Table 2.10: Type-I error rate (×100) of WLF, DMM and MMM for unequal covariances with decreasing sample sizes in crossover designs (negative pairing), $a = 4$, $p = 3$, $t = 4$, CS: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, AR: $\boldsymbol{\rho} = (0.3, 0.4, 0.5, 0.6)$, $\boldsymbol{n}^{(1)} = (25, 25, 25, 25)$, $\boldsymbol{n}^{(2)} = (35, 35, 35, 35)$, $\boldsymbol{n}^{(3)} = (10, 20, 30, 40)$, $\boldsymbol{n}^{(4)} = (20, 30, 40, 50)$, $\boldsymbol{n}^{(5)} = (40, 30, 20, 10)$, $\boldsymbol{n}^{(6)} = (50, 40, 30, 20)$.

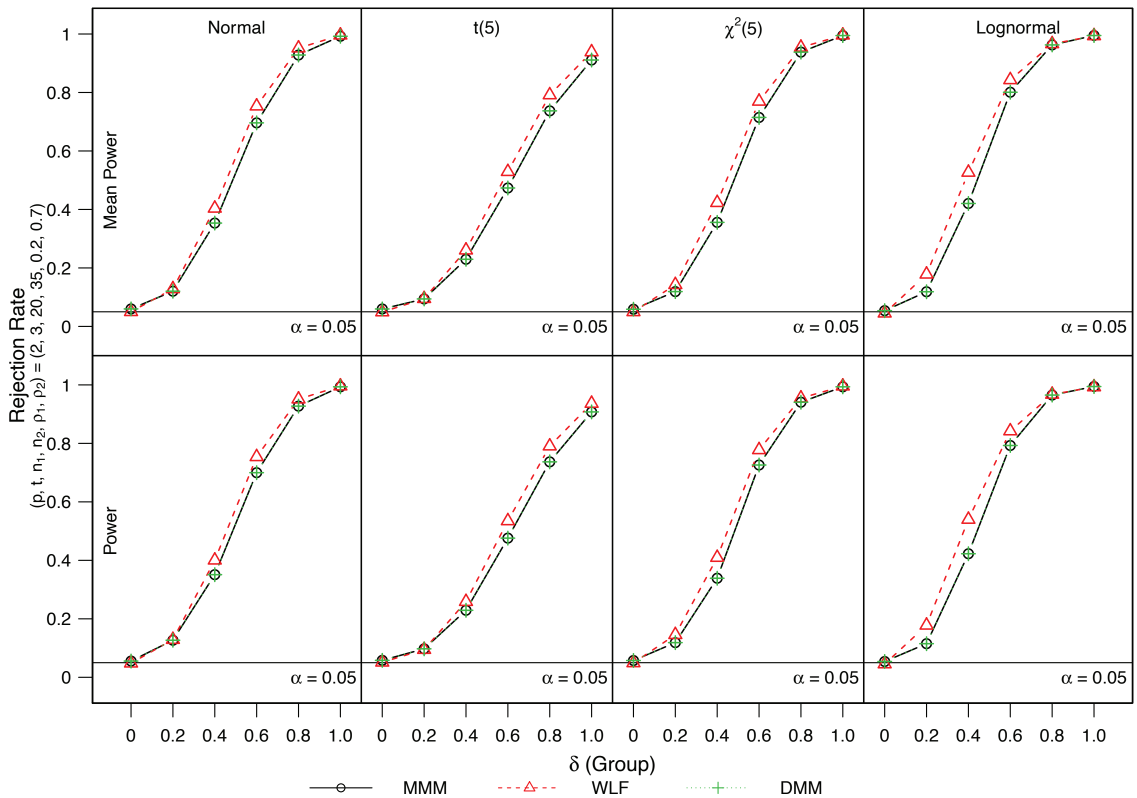
Dist	Cov	**n**	Period			Treatment			Treatment×Period		
			WLF	DMM	MMM	WLF	DMM	MMM	WLF	DMM	MMM
Normal	CS	$\boldsymbol{n}^{(5)}$	4.7	1.2	1.3	5.1	2.3	2.3	5.3	3.7	6.7
		$\boldsymbol{n}^{(6)}$	5.1	2.6	2.5	4.6	3.0	3.0	5.2	3.9	7.0
	AR	$\boldsymbol{n}^{(5)}$	4.7	1.3	2.7	5.1	2.5	2.5	5.8	4.0	5.7
		$\boldsymbol{n}^{(6)}$	5.4	2.5	4.2	5.2	3.2	3.2	4.9	4.1	5.2
$t(5)$	CS	$\boldsymbol{n}^{(5)}$	4.5	1.3	1.3	4.6	2.2	2.2	4.2	3.4	6.6
		$\boldsymbol{n}^{(6)}$	5.2	2.6	2.4	4.7	2.7	2.7	4.6	3.6	6.5
	AR	$\boldsymbol{n}^{(5)}$	4.4	1.4	3.0	4.8	2.6	2.6	5.0	3.9	5.5
		$\boldsymbol{n}^{(6)}$	4.8	2.2	4.2	4.8	2.7	2.7	4.5	3.8	5.7
$\chi^2(5)$	CS	$\boldsymbol{n}^{(5)}$	5.2	1.8	1.4	5.0	2.5	2.5	4.8	3.5	6.5
		$\boldsymbol{n}^{(6)}$	5.6	2.9	2.9	4.9	3.3	3.3	4.7	3.7	6.2
	AR	$\boldsymbol{n}^{(5)}$	5.2	1.9	3.0	4.7	2.2	2.2	4.9	3.7	6.0
		$\boldsymbol{n}^{(6)}$	5.5	2.9	4.3	5.4	3.4	3.4	4.5	4.0	5.7
Lognormal	CS	$\boldsymbol{n}^{(5)}$	5.9	2.9	1.9	3.9	2.1	2.1	3.2	2.9	6.3
		$\boldsymbol{n}^{(6)}$	6.4	3.5	2.3	5.0	3.2	3.2	3.8	3.5	6.2
	AR	$\boldsymbol{n}^{(5)}$	6.3	3.0	3.9	4.2	2.6	2.6	3.8	3.7	6.3
		$\boldsymbol{n}^{(6)}$	6.5	3.5	4.1	4.3	2.6	2.6	3.8	3.4	5.6

As shown in Table 2.11, the p-values of the main effect of gaze position and all interaction effects are greater than 0.05, leading to insignificant test results. It is worth mentioning that the p-values are considerably large for testing the interaction effects containing the gender effect, which might due to the small sample sizes in both female and male groups. Also, the number of levels in gender is only two, which is too small to satisfy the asymptotic framework mentioned in Harrar and Bathke (2012b). The only significant effect is the main effect of mental load.

Table 2.11: Analysis of the idiopathic infantile nystagmus syndrome (INS) data using WLF.

Effect	WLF	Effect	WLF
Mental Load	0.0239	Gender × Mental	0.3553
Gaze Position	0.1131	Gender × Position	0.8713
Mental × Position	0.0858	Gender × Mental × Position	0.6422

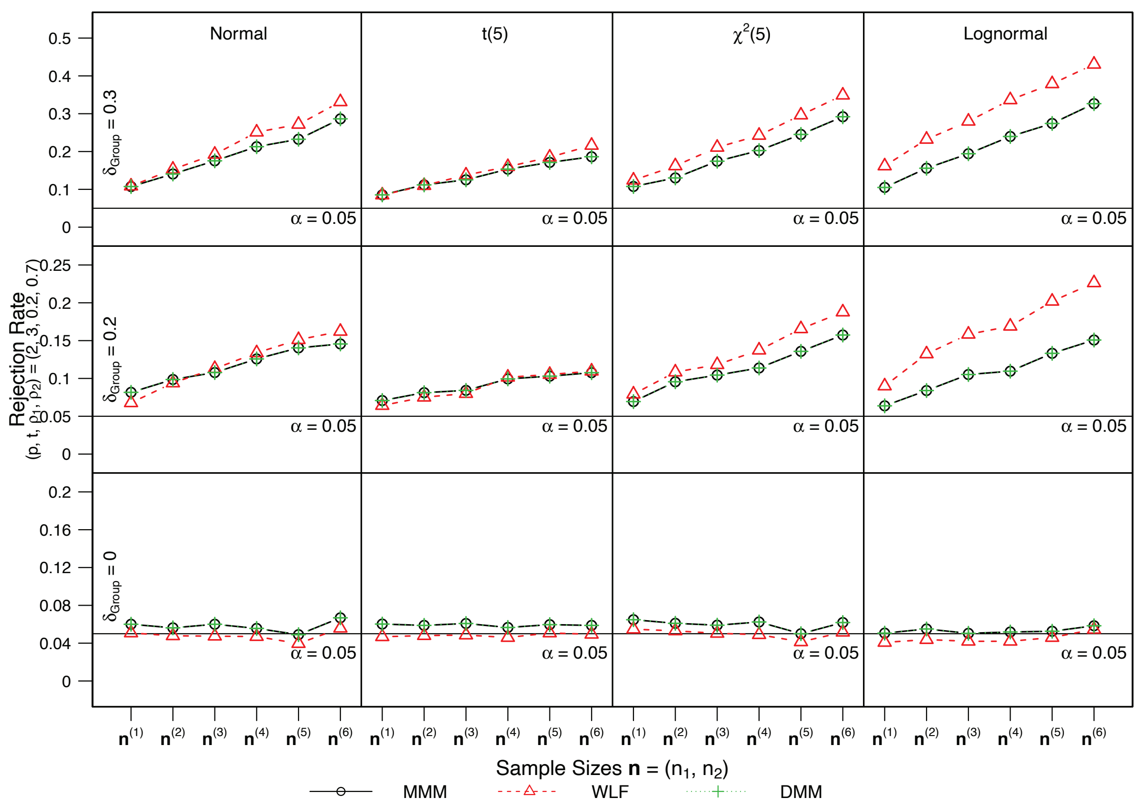
Figure 2.1: Power curves of detecting fixed alternative of Group effect in factorial designs with $\alpha = 0.05$, $a = 2$, $p = 2$, $t = 3$, $\boldsymbol{n} = (20, 35)$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$. The columns are for different distributions. The bottom row presents the power curves along different effect sizes of the Group factor when effect size of the Time factor is 0. The top row shows, along effect sizes of the Group factor, the mean power at all values of effect size of the Time factor.



Intraocular Pressure

In this section, we analyze intraocular pressure data. This data set contains multivariate repeated measures data in a 2×2 crossover design (Grender and Johnson, 1994). The data were recorded from an experiment to evaluate presence of collagen bits in a 1:8 concentration of Optipranolol suspended in Murocel. The bivariate response, intraocular pressure, both in the right eye and left eye was measured at baseline (pre-treatment measurement) and at 1, 8 and 24 hours following the topical application of treatment. There are two sequences, BITS/NO BITS and NO BITS/BITS.

Figure 2.2: Power curves of detecting fixed alternative of Group effect in factorial designs with $\alpha = 0.05$, $a = 2$, $p = 2$, $t = 3$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$. In the panel plot, columns are for different distributions and rows are for different effect sizes of the Group factor, denoted by δ_{Group}. In fact, the bottom row shows type-I error rates for testing Group effect. The effect of the Time factor is set to be 0 throughout the simulations investigating the relationship between power and sample size. The sample sizes are $\boldsymbol{n}^{(1)} = (8, 14)$, $\boldsymbol{n}^{(2)} = (12, 21)$, $\boldsymbol{n}^{(3)} = (16, 28)$, $\boldsymbol{n}^{(4)} = (20, 35)$, $\boldsymbol{n}^{(5)} = (24, 42)$ and $\boldsymbol{n}^{(6)} = (28, 49)$.



Seven subjects were given the treatments in the order of BITS/NO BITS ($n_1 = 7$) and six subjects were given the treatments in the reverse order of NO BITS/BITS ($n_2 = 6$), both with a seven-day washout period between drug treatments. The goal of the study was to analyze difference in intraocular pressure between baseline and the three post-treatment measurements.

The data set contains two sequences ($a = 2$). There are two periods and within each period there are there time points, which makes the total number of time points $2 \times 3 = 6$. At each time point, two intraocular pressure variables (in the right eye and left eye) were measured ($p = 2$). Five different null hypotheses of interest about main and interaction effects can be tested by choosing appropriate contrast matrices. For illustration purpose, we test effects including Time effect within each Period for each Sequence, Time effect, Period effect, Treatment effect and Treatment $\times$ Period interaction effect. Assuming there is no treatment carryover effect, we test the hypothesis $H_0^{(\phi)} : \boldsymbol{H}_\phi \boldsymbol{\mu} = \boldsymbol{0}$, where $\boldsymbol{H}_\phi = \boldsymbol{C}_\phi^\top \otimes \boldsymbol{D}_\phi \otimes \boldsymbol{I}_p$. The corresponding contrast matrices are $\boldsymbol{D}_{\mathrm{T(P)}} = \boldsymbol{I}_2 \otimes (\boldsymbol{I}_2, -\boldsymbol{1}_2)$ and $\boldsymbol{C}_{\mathrm{T(P)}} = \boldsymbol{I}_2$ for testing Time effect within each Period, $\boldsymbol{D}_{\mathrm{T}} = \boldsymbol{1}_2^\top \otimes (\boldsymbol{I}_2, -\boldsymbol{1}_2)$ and $\boldsymbol{C}_{\mathrm{T}} = \boldsymbol{J}_2/2$ for testing Time effect, $\boldsymbol{D}_{\mathrm{P}} = \boldsymbol{I}_2 \otimes \boldsymbol{1}_3^\top$ and $\boldsymbol{C}_{\mathrm{P}} = \boldsymbol{J}_2/2$ for testing Period effect, $\boldsymbol{D}_{\mathrm{Tr}} = (1, -1) \otimes \boldsymbol{1}_3^\top$ and $\boldsymbol{C}_{\mathrm{Tr}} = \boldsymbol{P}_2$ for testing Treatment effect, and $\boldsymbol{D}_{\mathrm{TrP}} = \boldsymbol{1}_6^\top/6$ and $\boldsymbol{C}_{\mathrm{TrP}} = \boldsymbol{P}_2$ for testing Treatment $\times$ Period effect.

Again, we perform a preliminary check on the Intraocular Pressure data. The marginal distributions of the two response variables, right eye pressure and left eye pressure are both right-skewed. Moreover, the covariance matrices of the two sequences are unequal and the sample sizes are unbalanced. The simulation results in Section 2.4 indicate that WLF method is the best for use when data are nonnormal in unbalanced heteroscedastic designs.

Table 2.12 shows the p-values for testing the aforementioned effects using WLF method. At the significance level of 0.05, there is no Time effect within each Period for each Sequence and there is no significant difference between the two treatments (BITS and NO BITS) in lowering eye pressure. Both Time and Period have significant influences on the reduction of intraocular pressure. However, the interaction effect of

Treatment $\times$ Period is not significant.

Table 2.12: Analysis of the difference in intraocular pressure (both right eye and left eye) across three time points (1, 8, 24 hours) following treatment using WLF.

Effect	WLF	Effect	WLF
Time (Period)	0.466	Period	<0.001
Treatment	0.147	Time	0.008
Treatment× Period	0.269		

2.6 Discussion and Conclusion

Multivariate growth curve data or data with repeated measures on multiple outcome variables have become more and more common in a variety of fields. There are some effective methods developed for analyzing repeated measures, but with the assumptions of multivariate normality and homoscedasticity. However, these assumptions are difficult to justify and attain in reality. In this chapter, we have considered inference methods for generalized MANOVA designs without assuming multivariate normality and homogeneous covariance matrices.

Compared with the classical methods, the proposed method has the advantage that it is applicable for a wide range of designs in a unified way, by specifying appropriate contrast matrices. For illustration of the methodology, general factorial design and crossover design have been considered. The proposed approximate finite-sample test uses adjusted sums of square matrices to mitigate the influence of potential heterogeneity in data. The proposed methodology is applicable to data with any distribution and with finite second moment.

Our simulation studies showed that when covariance matrices are equal across groups, the proposed method performs roughly the same as the classical methods. However, when heteroscedasticity presents, the proposed method outperforms both competing tests, especially when sample sizes are unequal across groups for both positive and negative pairings. We further investigated test performance in terms of power. Compared with the classical methods, the proposed method achieved the most accurate type-I error rate. Moreover, such superiority in type-I error rate does not come at the expense of reduced power.

In general, with multivariate growth curve data, the proposed method has an edge over classical methods for growth curve (also known as generalized MANOVA) models with heterogeneity of covariance matrices. Such advantage is more pronounced in unbalanced heteroscedastic designs. On the contrary, the classical methods exhibit unstable behavior in such designs. Due to the robustness to nonnormality and heteroscedasticity, we recommend using the proposed test in real applications. The proposed method may not be appropriate when count data, discrete data and ordered categorical data are collected. We plan to investigate the idea for applications with nonmetric data in future research.

2.7 Appendix

Wishart Approximation

Suppose $\boldsymbol{Y}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ are independent for $i = 1, \ldots, n, \boldsymbol{\Sigma}_i > 0$. Define, $\boldsymbol{Q} = \boldsymbol{Y}\boldsymbol{C}\boldsymbol{Y}^\top$, where $\boldsymbol{C} = (c_{ij})$ is an $n \times n$ symmetric nonnegative definite matrix and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$ is a $p \times n$ matrix. The idea is to approximate the distribution of $\boldsymbol{Q}$ by a p-dimensional central Wishart distribution with degrees of freedom f and mean $f\boldsymbol{\Psi}$, which is denoted by $W_p(f, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} > \boldsymbol{0}$. The quantities f and $\boldsymbol{\Psi}$ are to be approximated by matching the means and the total variances of $\boldsymbol{Q}$ and $W_p(f, \boldsymbol{\Psi})$. Note that the total variance is the trace of the variance-covariance matrix. Given a random matrix $\boldsymbol{W} \sim W_p(f, \boldsymbol{\Psi})$, we have $\mathrm{E}(\boldsymbol{W}) = f\boldsymbol{\Psi}$ and $\mathrm{Var}(\boldsymbol{W}) = f(\boldsymbol{I}_{p^2} + K_{p,p})(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi})$.

By Lemma 1 in Harrar and Bathke (2012b), the mean and variance of $\boldsymbol{Q}$ can be calculated.

Lemma 2.7.1. *Let* $\boldsymbol{A} = (a_{ij})$ *and* $\boldsymbol{B} = (b_{ij})$ *be* $n \times n$ *matrices. Then,* $\mathrm{E}(\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^\top) = \sum_{i=1}^n a_{ii}\boldsymbol{\Sigma}_i$ *and* $\mathrm{Cov}\left(\mathrm{Vec}(\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^\top), \mathrm{Vec}(\boldsymbol{Y}\boldsymbol{B}\boldsymbol{Y}^\top)\right) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ij}(\boldsymbol{I}_{p^2}+K_{p,p})(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_j) + \sum_{i=1}^n a_{ii}b_{ii}K_4(\boldsymbol{Y}_i)$, *where* $K_4(\boldsymbol{Y}_i) = \mathrm{E}(\mathrm{Vec}(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top)\mathrm{Vec}(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top)^\top) - (\boldsymbol{I}_{p^2}+K_{p,p})(\boldsymbol{\Sigma}_i \otimes$

$\boldsymbol{\Sigma}_i)-\mathrm{Vec}(\boldsymbol{\Sigma}_i)\mathrm{Vec}(\boldsymbol{\Sigma}_i)^\top$. *Note that under normality,* $K_4(\boldsymbol{Y}_i)=0$. *(Harrar and Bathke, 2012b)*

Calculations for $\mathbf{H}^{(\phi)}$

Equation (2.2), $\boldsymbol{H}^{(\phi)}=\overline{\boldsymbol{Z}}^{(\phi)}\boldsymbol{C}_\phi\overline{\boldsymbol{Z}}^{(\phi)^\top}$, is a quadratic form. $\overline{\boldsymbol{Z}}^{(\phi)}=\left(\overline{\boldsymbol{Z}}_1^{(\phi)},\ldots,\overline{\boldsymbol{Z}}_a^{(\phi)}\right)$, where $\overline{\boldsymbol{Z}}_i^{(\phi)}=\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{Z}_{ik}^{(\phi)}$, is a $(rp)\times a$ matrix, where $r=\mathrm{rank}(\boldsymbol{D}_\phi)$ and it depends on the effect being tested.

By multivariate central limit theorem, $\sqrt{n_i}\left(\overline{\boldsymbol{Z}}_i^{(\phi)}-\tilde{\boldsymbol{\mu}}_i^{(\phi)}\right)\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{\Sigma}_i^{(\phi)})$. We use $\tilde{\boldsymbol{\mu}}_i^{(\phi)}$ to denote the theoretical mean vectors of $\boldsymbol{Z}_{i1}^{(\phi)}, i=1,2,\ldots,a$. Without loss of generality, we assume $\tilde{\boldsymbol{\mu}}_i^{(\phi)}=\mathbf{0}, i=1,2,\ldots,a$. Under this assumption, $\mathrm{E}(\overline{\boldsymbol{Z}}_i^{(\phi)})=\mathbf{0}$ and $\mathrm{Cov}(\overline{\boldsymbol{Z}}_i^{(\phi)})=\boldsymbol{\Sigma}_i^{(\phi)}/n_i$. By Lemma 2.7.1 above,

$$\mathrm{E}(\boldsymbol{H}^{(\phi)})=\mathrm{E}\left(\overline{\boldsymbol{Z}}^{(\phi)}\boldsymbol{C}_\phi\overline{\boldsymbol{Z}}^{(\phi)^\top}\right)=\sum_{i=1}^{a}\boldsymbol{C}_{\phi,ii}\cdot\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}=\boldsymbol{\Sigma}_\phi. \tag{2.12}$$

Again, by Lemma 2.7.1, we calculate the covariance matrix of $\boldsymbol{H}^{(\phi)}$. Note that under normality $K_4(\overline{\boldsymbol{Z}}_i^{(\phi)})=0$. Further,

$$\begin{aligned}\mathrm{Cov}\left(\mathrm{Vec}\left(\boldsymbol{H}^{(\phi)}\right)\right)&=\mathrm{Cov}\left(\mathrm{Vec}\left(\overline{\boldsymbol{Z}}^{(\phi)}\boldsymbol{C}_\phi\overline{\boldsymbol{Z}}^{(\phi)^\top}\right),\mathrm{Vec}\left(\overline{\boldsymbol{Z}}^{(\phi)}\boldsymbol{C}_\phi\overline{\boldsymbol{Z}}^{(\phi)^\top}\right)\right)\\&=\sum_{i=1}^{a}\sum_{j=1}^{a}\left(\boldsymbol{C}_{\phi,ij}\right)^2\left(\boldsymbol{I}_{(rp)^2}+K_{rp,rp}\right)\left(\frac{1}{n_i}\boldsymbol{\Sigma}_i^{(\phi)}\otimes\frac{1}{n_j}\boldsymbol{\Sigma}_j^{(\phi)}\right).\end{aligned} \tag{2.13}$$

To approximate the hypothesis matrix $\boldsymbol{H}^{(\phi)}$, we use a Wishart distribution. Let $\boldsymbol{W}_{\boldsymbol{H}^{(\phi)}}\sim W_{rp}(f_{\boldsymbol{H}^{(\phi)}},\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}})$, where $r=\mathrm{rank}(\boldsymbol{D}_\phi)$ and $\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}>\mathbf{0}$. It is well known that (see Magnus and Neudecker, 1979)

$$\mathrm{E}(\boldsymbol{W}_{\boldsymbol{H}^{(\phi)}})=f_{\boldsymbol{H}^{(\phi)}}\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}\ \text{and} \tag{2.14}$$

$$\mathrm{Var}(\boldsymbol{W}_{\boldsymbol{H}^{(\phi)}})=f_{\boldsymbol{H}^{(\phi)}}(\boldsymbol{I}_{(rp)^2}+K_{rp,rp})(\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}\otimes\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}). \tag{2.15}$$

Setting (2.12) = (2.14) and $\mathrm{tr}\left\{(2.13)\right\}=\mathrm{tr}\left\{(2.15)\right\}$ we get

$$\begin{aligned}&\sum_{i=1}^{a}\boldsymbol{C}_{\phi,ii}\cdot\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}=f_{\boldsymbol{H}^{(\phi)}}\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}},\ \text{and}\\&\mathrm{tr}\left\{f_{\boldsymbol{H}^{(\phi)}}(\boldsymbol{I}_{(rp)^2}+K_{rp,rp})(\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}\otimes\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}})\right\}=\\&\mathrm{tr}\left\{\sum_{i=1}^{a}\sum_{j=1}^{a}\left(\boldsymbol{C}_{\phi,ij}\right)^2\left(\boldsymbol{I}_{(rp)^2}+K_{rp,rp}\right)\left(\frac{1}{n_i}\boldsymbol{\Sigma}_i^{(\phi)}\otimes\frac{1}{n_j}\boldsymbol{\Sigma}_j^{(\phi)}\right)\right\}.\end{aligned} \tag{2.16}$$

Hence, $f_{\boldsymbol{H}^{(\phi)}}$ and $\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}}$ can be solved. By Harrar and Bathke (2012b,a) , we get the solutions to equations in (2.16),

$$
\begin{aligned}
f_{\boldsymbol{H}^{(\phi)}} &= \frac{\operatorname{tr}\left\{\left(\sum_{i=1}^{a}\boldsymbol{C}_{\phi,ii}\cdot\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}\right)^2\right\}+\left\{\operatorname{tr}\left(\sum_{i=1}^{a}\boldsymbol{C}_{\phi,ii}\cdot\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}\right)\right\}^2}{\sum_{i=1}^{a}\sum_{j=1}^{a}\left(\boldsymbol{C}_{\phi,ij}\right)^2\left\{\operatorname{tr}\left(\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}\frac{\boldsymbol{\Sigma}_j^{(\phi)}}{n_j}\right)+\operatorname{tr}\left(\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}\right)\operatorname{tr}\left(\frac{\boldsymbol{\Sigma}_j^{(\phi)}}{n_j}\right)\right\}} \\
&= \frac{\operatorname{tr}\left\{\boldsymbol{\Sigma}_\phi^2\right\}+\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi\right)\right\}^2}{\sum_{i=1}^{a}\sum_{j=1}^{a}\left(\boldsymbol{C}_{\phi,ij}\right)^2\frac{1}{n_in_j}\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\boldsymbol{\Sigma}_j^{(\phi)}\right)+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_j^{(\phi)}\right)\right\}}.
\end{aligned}
$$

and

$$
\boldsymbol{\Psi}_{\boldsymbol{H}^{(\phi)}} = \frac{1}{f_{\boldsymbol{H}^{(\phi)}}}\boldsymbol{\Sigma}_\phi = \frac{\sum_{i=1}^{a}\sum_{j=1}^{a}\left(\boldsymbol{C}_{\phi,ij}\right)^2\frac{1}{n_in_j}\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\boldsymbol{\Sigma}_j^{(\phi)}\right)+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_j^{(\phi)}\right)\right\}}{\operatorname{tr}\left\{\boldsymbol{\Sigma}_\phi^2\right\}+\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi\right)\right\}^2}\cdot\boldsymbol{\Sigma}_\phi.
$$

Calculations for $\mathbf{G}^{(\phi)}$

By equation (2.3), $\boldsymbol{G}^{(\phi)} = \sum_{i=1}^{a}\frac{\boldsymbol{C}_{\phi,ii}}{n_i(n_i-1)}\sum_{k=1}^{n_i}\left(\boldsymbol{Z}_{ik}^{(\phi)}-\overline{\boldsymbol{Z}}_i^{(\phi)}\right)\left(\boldsymbol{Z}_{ik}^{(\phi)}-\overline{\boldsymbol{Z}}_i^{(\phi)}\right)^\top$. For fixed group i and let $\boldsymbol{F}_i = \left(\boldsymbol{Z}_{i1}^{(\phi)},\boldsymbol{Z}_{i2}^{(\phi)},\ldots,\boldsymbol{Z}_{in_i}^{(\phi)}\right)$. Then $\boldsymbol{G}^{(\phi)}$ can be written as

$$
\boldsymbol{G}^{(\phi)} = \sum_{i=1}^{a}\frac{\boldsymbol{C}_{\phi,ii}}{n_i\left(n_i-1\right)}\left[\boldsymbol{F}_i\left(\boldsymbol{I}_{n_i}-\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\boldsymbol{F}_i^\top\right].
$$

By the definition of $\boldsymbol{Z}_{ij}^{(\phi)}$, $\boldsymbol{Z}_{i1}^{(\phi)},\boldsymbol{Z}_{i2}^{(\phi)},\ldots,\boldsymbol{Z}_{in_i}^{(\phi)}$ are *i.i.d.* with theoretical mean $\tilde{\boldsymbol{\mu}}_i^{(\phi)}$ and covariance matrix $\boldsymbol{\Sigma}_i^{(\phi)}$. Also, we assume normality of the vectors $\boldsymbol{Z}_{i1}^{(\phi)},\boldsymbol{Z}_{i2}^{(\phi)},\ldots,\boldsymbol{Z}_{in_i}^{(\phi)}$ in $\boldsymbol{F}_i$ since the asymptotic distribution of the test statistic does not depend on the distribution of the data.

By Lemma 2.7.1,

$$
\begin{aligned}
\operatorname{E}\left(\boldsymbol{G}^{(\phi)}\right) &= \sum_{i=1}^{a}\frac{\boldsymbol{C}_{\phi,ii}}{n_i\left(n_i-1\right)}\operatorname{E}\left[\boldsymbol{F}_i\left(\boldsymbol{I}_{n_i}-\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\boldsymbol{F}_i^\top\right] \\
&= \sum_{i=1}^{a}\frac{\boldsymbol{C}_{\phi,ii}}{n_i\left(n_i-1\right)}\sum_{k=1}^{n_i}\left(1-\frac{1}{n_i}\right)\boldsymbol{\Sigma}_i^{(\phi)} \\
&= \sum_{i=1}^{a}\boldsymbol{C}_{\phi,ii}\cdot\frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i} = \boldsymbol{\Sigma}_\phi.
\end{aligned}
$$

Next, we calculate the covariance matrix of $\boldsymbol{G}^{(\phi)}$. For convenience, we use matrix $\boldsymbol{M} = (m_{uv})$, where $u,v = 1,\ldots,n_i$, to denote $\left(\boldsymbol{I}_{n_i}-\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)$. Note that under normality, $K_4\left(\boldsymbol{Z}_{iu}^{(\phi)}\right) = 0$ by Lemma 2.7.1.

$$\begin{aligned}\operatorname{Cov}\left(\operatorname{Vec}\left(\boldsymbol{G}^{\phi)}\right)\right) &= \sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2} \operatorname{Cov}\left[\operatorname{Vec}\left(\boldsymbol{F}_i\left(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\boldsymbol{F}_i^\top\right)\right] \\ &= \sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2} \left\{\sum_{u=1}^{n_i}\sum_{v=1}^{n_i} m_{uv}^2\left(\boldsymbol{I}_{(rp)^2} + K_{rp,rp}\right)\left(\boldsymbol{\Sigma}_i^{(\phi)} \otimes \boldsymbol{\Sigma}_i^{(\phi)}\right)\right\}.\end{aligned}$$

To approximate the adjusted error matrix $\boldsymbol{G}^{(\phi)}$, we use a Wishart distribution $\boldsymbol{W}_{\boldsymbol{G}^{(\phi)}} \sim W_{rp}(f_{\boldsymbol{G}^{(\phi)}}, \boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}})$, where $r = \operatorname{rank}(\boldsymbol{D}_\phi)$ and $\boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}} > 0$. Matching the means and total variances of $\boldsymbol{G}^{(\phi)}$ and $\boldsymbol{W}_{\boldsymbol{G}^{(\phi)}}$, the equations are

$$\sum_{i=1}^{a} \boldsymbol{C}_{\phi,ii} \cdot \frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i} = f_{\boldsymbol{G}^{(\phi)}} \boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}} \quad \text{and} \tag{2.17}$$

$$\begin{aligned}&\operatorname{tr}\left\{f_{\boldsymbol{G}^{(\phi)}}(\boldsymbol{I}_{(rp)^2} + K_{rp,rp})(\boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}} \otimes \boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}})\right\} = \\ &\operatorname{tr}\left\{\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2}\left[\sum_{u=1}^{n_i}\sum_{v=1}^{n_i} m_{uv}^2\left(\boldsymbol{I}_{(rp)^2} + K_{rp,rp}\right)\left(\boldsymbol{\Sigma}_i^{(\phi)} \otimes \boldsymbol{\Sigma}_i^{(\phi)}\right)\right]\right\}.\end{aligned} \tag{2.18}$$

Rewriting equation (2.17), we get $\boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}} = \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\boldsymbol{\Sigma}_\phi$. Using this, we solve for $f_{\boldsymbol{G}^{(\phi)}}$ in equation (2.18). For the left-hand side of equation (2.18), we have

$$\begin{aligned}&\operatorname{tr}\left\{f_{\boldsymbol{G}^{(\phi)}}(\boldsymbol{I}_{(rp)^2} + K_{rp,rp})(\frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\boldsymbol{\Sigma}_\phi \otimes \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\boldsymbol{\Sigma}_\phi)\right\} \\ &= \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\operatorname{tr}\left\{(\boldsymbol{I}_{(rp)^2} + K_{rp,rp})(\boldsymbol{\Sigma}_\phi \otimes \boldsymbol{\Sigma}_\phi)\right\} \\ &= \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\operatorname{tr}\left\{\boldsymbol{\Sigma}_\phi \otimes \boldsymbol{\Sigma}_\phi + K_{rp,rp}(\boldsymbol{\Sigma}_\phi \otimes \boldsymbol{\Sigma}_\phi)\right\} \\ &= \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi \otimes \boldsymbol{\Sigma}_\phi\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_\phi\boldsymbol{\Sigma}_\phi\right)\right] \\ &= \frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\left\{\left[\operatorname{tr}(\boldsymbol{\Sigma}_\phi)\right]^2 + \operatorname{tr}\left(\boldsymbol{\Sigma}_\phi^2\right)\right\},\end{aligned} \tag{2.19}$$

where the third equality is by the property of commutation matrix (Magnus and Neudecker, 1979), along the same line as in (2.19). For the right-hand side of equation

(2.18),

$$
\begin{aligned}
&\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2} \operatorname{tr}\left\{\sum_{u=1}^{n_i}\sum_{v=1}^{n_i} m_{uv}^2\left(\boldsymbol{I}_{(rp)^2}+K_{rp,rp}\right)\left(\boldsymbol{\Sigma}_i^{(\phi)}\otimes\boldsymbol{\Sigma}_i^{(\phi)}\right)\right\} \\
&=\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2}\sum_{u=1}^{n_i}\sum_{v=1}^{n_i} m_{uv}^2\left\{\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)\right\} \\
&=\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2}\left\{\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)\right\}\sum_{u=1}^{n_i}\sum_{v=1}^{n_i} m_{uv}^2 \\
&=\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)^2}\left\{\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)\right\}\cdot\left\{n_i\left(1-\frac{1}{n_i}\right)^2\right. \\
&\quad\left.+\left(n_i^2-n_i\right)\left(-\frac{1}{n_i}\right)^2\right\} \\
&=\sum_{i=1}^{a} \frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)}\left\{\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)\right\}.
\end{aligned}
$$

Equation (2.18) can be written as

$$
\frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\left\{[\operatorname{tr}(\boldsymbol{\Sigma}_\phi)]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_\phi^2\right)\right\}=\sum_{i=1}^{a}\frac{\left(\boldsymbol{C}_{\phi,ii}\right)^2}{n_i^2\left(n_i-1\right)}\left\{\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2+\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)\right\}.
$$

Solving for $f_{\boldsymbol{G}^{(\phi)}}$, we get

$$
f_{\boldsymbol{G}^{(\phi)}}=\frac{\operatorname{tr}\left\{\boldsymbol{\Sigma}_\phi^2\right\}+\left\{\operatorname{tr}(\boldsymbol{\Sigma}_\phi)\right\}^2}{\sum_{i=1}^{a}\left(\boldsymbol{C}_{\phi,ii}\right)^2\frac{1}{n_i^2(n_i-1)}\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)+\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2\right\}}
$$

and

$$
\boldsymbol{\Psi}_{\boldsymbol{G}^{(\phi)}}=\frac{1}{f_{\boldsymbol{G}^{(\phi)}}}\boldsymbol{\Sigma}_\phi=\frac{\sum_{i=1}^{a}\left(\boldsymbol{C}_{\phi,ii}\right)^2\frac{1}{n_i^2(n_i-1)}\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)2}\right)+\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_i^{(\phi)}\right)\right]^2\right\}}{\operatorname{tr}\left\{\boldsymbol{\Sigma}_\phi^2\right\}+\left\{\operatorname{tr}(\boldsymbol{\Sigma}_\phi)\right\}^2}\cdot\boldsymbol{\Sigma}_\phi.
$$

Affine Invariance

With the original data $\boldsymbol{X}_{ik}$, where $i=1,\ldots,a$, $k=1,\ldots,n_i$, we define $\overline{\boldsymbol{X}}_i=\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{X}_{ik}$ and $\overline{\boldsymbol{X}}=\left(\overline{\boldsymbol{X}}_1,\ldots,\overline{\boldsymbol{X}}_a\right)$. Now, with the transformed data $\boldsymbol{Y}_{ik}=(\boldsymbol{I}_t\otimes\boldsymbol{A})\boldsymbol{X}_{ik}+\boldsymbol{1}_t\otimes\boldsymbol{c}$, where $\boldsymbol{A}$ is any nonsingular $p\times p$ matrix and $\boldsymbol{c}$ is a p-dimensional vector, we want to show the degrees of freedom in (2.10) and (2.11) and the U statistic are affine invariant.

Analogously, for the transformed data, we have $\boldsymbol{\Sigma}_{i,\boldsymbol{Y}}^{(\phi)} = (\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \boldsymbol{V}_i (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top$ and $\boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}} = \sum_{i=1}^a \boldsymbol{C}_{\phi,ii} \boldsymbol{\Sigma}_{i,\boldsymbol{Y}}^{(\phi)}/n_i = (\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \left(\sum_{i=1}^a \frac{\boldsymbol{C}_{\phi,ii}}{n_i} \boldsymbol{V}_i\right) (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top$. The adjusted hypothesis and error matrices are

$$\boldsymbol{H}_{\boldsymbol{Y}}^{(\phi)} = (\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \overline{\boldsymbol{X}} \boldsymbol{C}_\phi \overline{\boldsymbol{X}}^\top (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top \quad \text{and}$$

$$\boldsymbol{G}_{\boldsymbol{Y}}^{(\phi)} = \sum_{i=1}^a \frac{\boldsymbol{C}_{\phi,ii}}{n_i (n_i - 1)} \sum_{k=1}^{n_i} (\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i\right) \left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i\right)^\top (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top,$$

respectively. Note that for testing Time and Group×Time effects, $\boldsymbol{D}_\phi \boldsymbol{1}_t = \boldsymbol{0}$ and for testing Group and Group×Time effects, $\boldsymbol{1}_a^\top \otimes [(\boldsymbol{D}_\phi \boldsymbol{1}_t) \otimes \boldsymbol{c}] \boldsymbol{C}_\phi = \boldsymbol{0}$.

The approximation in (2.9) implies

$$f^*_{\boldsymbol{H}^{(\phi)},\boldsymbol{Y}} \cdot \boldsymbol{H}_{*,\boldsymbol{Y}}^{(\phi)} \underset{H_0^{(\phi)}}{\overset{approx}{\sim}} W_{rp}\left(f^*_{\boldsymbol{H}^{(\phi)},\boldsymbol{Y}}, \boldsymbol{I}_{rp}\right) \quad \text{and} \quad f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}} \cdot \boldsymbol{G}_{*,\boldsymbol{Y}}^{(\phi)} \overset{approx}{\sim} W_{rp}\left(f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}}, \boldsymbol{I}_{rp}\right),$$

where $\boldsymbol{H}_{*,\boldsymbol{Y}}^{(\phi)} = \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2} \boldsymbol{H}_{\boldsymbol{Y}}^{(\phi)} \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2}$ and $\boldsymbol{G}_{*,\boldsymbol{Y}}^{(\phi)} = \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2} \boldsymbol{G}_{\boldsymbol{Y}}^{(\phi)} \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2}$. From (2.9), (2.10) and (2.11), it is easy to see the degrees of freedom for the transformed data are

$$f^*_{\boldsymbol{H}^{(\phi)},\boldsymbol{Y}} = \frac{rp\,(1 + rp)}{\sum_{i=1}^a \sum_{i'=1}^a (\boldsymbol{C}_{\phi,ii'})^2 \frac{1}{n_i n_{i'}} \left\{\mathrm{tr}\,(\boldsymbol{A}_{i,\boldsymbol{Y}} \boldsymbol{A}_{i',\boldsymbol{Y}}) + \mathrm{tr}\,(\boldsymbol{A}_{i,\boldsymbol{Y}})\, \mathrm{tr}\,(\boldsymbol{A}_{i',\boldsymbol{Y}})\right\}}$$

and

$$f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}} = \frac{rp\,(1 + rp)}{\sum_{i=1}^a (\boldsymbol{C}_{\phi,ii})^2 \frac{1}{n_i^2 (n_i - 1)} \left\{\mathrm{tr}\left(\boldsymbol{A}_{i,\boldsymbol{Y}}^2\right) + [\mathrm{tr}\,(\boldsymbol{A}_{i,\boldsymbol{Y}})]^2\right\}},$$

where $\boldsymbol{A}_{i,\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2} \boldsymbol{\Sigma}_{i,\boldsymbol{Y}}^{(\phi)} \boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1/2}$. Note that $\boldsymbol{D}_\phi \otimes \boldsymbol{A} = (\boldsymbol{I}_r \otimes \boldsymbol{A})(\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p)$ and, therefore,

$$\begin{aligned}
\mathrm{tr}\,(\boldsymbol{A}_{i,\boldsymbol{Y}}) &= \mathrm{tr}\left(\boldsymbol{\Sigma}_{\phi,\boldsymbol{Y}}^{-1} \boldsymbol{\Sigma}_{i,\boldsymbol{Y}}^{(\phi)}\right) \\
&= \mathrm{tr}\left\{\left[(\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \left(\sum_{i=1}^a \frac{\boldsymbol{C}_{\phi,ii}}{n_i} \boldsymbol{V}_i\right) (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top\right]^{-1} (\boldsymbol{D}_\phi \otimes \boldsymbol{A}) \boldsymbol{V}_i (\boldsymbol{D}_\phi \otimes \boldsymbol{A})^\top\right\} \\
&= \mathrm{tr}\left\{\left[(\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p) \left(\sum_{i=1}^a \frac{\boldsymbol{C}_{\phi,ii}}{n_i} \boldsymbol{V}_i\right) (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p)^\top\right]^{-1} (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p) \boldsymbol{V}_i (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p)^\top\right\} \\
&= \mathrm{tr}\,(\boldsymbol{A}_i).
\end{aligned}$$

Similarly, $\mathrm{tr}\,(\boldsymbol{A}_{i,\boldsymbol{Y}} \boldsymbol{A}_{i',\boldsymbol{Y}}) = \mathrm{tr}\,(\boldsymbol{A}_i \boldsymbol{A}_{i'})$. Therefore, $f^*_{\boldsymbol{H}^{(\phi)},\boldsymbol{Y}} = f^*_{\boldsymbol{H}^{(\phi)}}$ and $f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}} = f^*_{\boldsymbol{G}^{(\phi)}}$ and the degrees of freedom in (2.10) and (2.11) are affine invariant. Because

$\boldsymbol{D}_\phi \otimes \boldsymbol{A} = (\boldsymbol{I}_r \otimes \boldsymbol{A})(\boldsymbol{D}_\phi \otimes \boldsymbol{I}_p)$, it is easy to see that $|f^*_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)}_*| = |f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}} \boldsymbol{G}^{(\phi)}_{*,\boldsymbol{Y}}|$ and $|f^*_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)}_* + f^*_{\boldsymbol{H}^{(\phi)}} \boldsymbol{H}^{(\phi)}_*| = |f^*_{\boldsymbol{G}^{(\phi)},\boldsymbol{Y}} \boldsymbol{G}^{(\phi)}_{*,\boldsymbol{Y}} + f^*_{\boldsymbol{H}^{(\phi)},\boldsymbol{Y}} \boldsymbol{H}^{(\phi)}_{*,\boldsymbol{Y}}|$. Therefore, the Wilks' Lambda statistic U is affine invariant as well.

Chapter 3 Resampling-based Tests for Multivariate Growth Curve Data

3.1 Introduction

Robust tests were proposed in Zeng and Harrar (2021b); Harrar and Bathke (2012b); Zhang and Liu (2013) without assuming multivariate normality or homogeneity of covariance matrices. Such tests applied modified MANOVA, where Wishart distributions are used to approximate the hypothesis and the error matrices by matching their corresponding expected values and total variances. These robust tests are ideal only when sample sizes are moderately large. When sample sizes are small, test results tend to be liberal. Moreover, the robust tests are only asymptotic approximations and they do not lead to asymptotically exact tests.

Without assuming multivariate normality and homogeneity of the covariance matrices the multivariate Wald-type test statistic (WTS) is asymptotically exact. However, WTS is known to suffer from slow convergence and, hence, may not be satisfactory in finite-sample. Large sample sizes are then required in order to maintain the preassigned type-I error rate. Several improvements were done to tackle the small-sample issue under MANOVA settings (Pesarin, 2001; Pesarin and Salmaso, 2010; Good, 2005, for example), where permutation tests are utilized when the null distribution of the test statistic is invariant under the corresponding randomization group. The permutation idea was further modified for use in situations where the exchangeability under the null hypothesis does not hold by researchers (see Neuhaus, 1993; Janssen and Pauls, 2003; Janssen, 2005; Omelka and Pauly, 2012; Chung and Romano, 2013; Pauly et al., 2015; Friedrich et al., 2017). In addition, Konietschke et al. (2015) proposed another technique, bootstrap, to improve the small sample behavior of the WTS for general heteroscedastic factorial designs. However, so far, to the best of our knowledge, there is no existing work done for analyzing multivariate repeated measures data without assuming multivariate normality and homoscedasticity.

The present chapter aims to develop resampling-based test statistics, where permutation, bootstrap and hybrid permutation-bootstrap procedures are investigated. Under such resampling schemes, no assumption of multivariate normality or homoscedasticity is required. The permutation scheme extends the ideas in Pauly et al. (2015) and Friedrich et al. (2017) of permuting all longitudinal data with only one response variable in general split-plot designs. The bootstrap scheme extends the idea in Konietschke et al. (2015) of using a parametric bootstrap to analyze multivariate repeated measures data in nonparametric settings. The small sample properties are investigated via simulation studies, which show that our resampling approaches are much more accurate than the classical WTS and some other comparable tests in most cases.

This chapter is organized as follows. In Section 3.2, statistical models and hypotheses are introduced. The Wald-type test statistics, resampling procedures and relevant mathematical results are presented in Section 3.3. The finite-sample behavior of the proposed resampling-based WTS tests, classical Wald's asymptotic test, Doubly Multivariate Model (DMM) and Multivariate Mixed Model (MMM) are investigated in extensive simulation studies in Section 3.4. Along with the type-I error rate (size) simulations, we also investigate power performance of the tests in Section 3.4. The application of the proposed resampling-based WTS tests is illustrated in Section 3.5 with an optometry data set. We conclude the chapter with some discussions and remarks in Section 3.6. All relevant proofs and technical details are given in the Appendix (Section 3.7).

3.2 Models and Hypotheses

Throughout this chapter we will use the following notations. The d-dimensional identity matrix is denoted by $\boldsymbol{I}_d$ and a $d \times d$ matrix with all 1s as its components is denoted by $\boldsymbol{J}_d = \mathbf{1}_d \mathbf{1}_d^\top$, where $\mathbf{1}_d = (1, \dots, 1)_{d\times 1}^\top$. We further denote the centering matrix by $\boldsymbol{P}_d = \boldsymbol{I}_d - \frac{1}{d}\boldsymbol{J}_d$. The operators $\bigoplus$ and $\bigotimes$ represent the Kronecker sum and product, respectively (Schott, 2016, Chap. 8).

In the context of general factorial design, consider the multivariate model

$$\boldsymbol{X}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\epsilon}_{ijk}, \tag{3.1}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, t$, and $k = 1, \ldots, n_i$. The index i is for group (treatment), j is for time (occasion), and k is for experimental unit (subject). For each experimental unit and at each time point, a p-variate observation is made. For fixed i and j, the random errors $\boldsymbol{\epsilon}_{ij1}, \ldots, \boldsymbol{\epsilon}_{ijn_i}$ are independent and identically distributed p-dimensional random vectors satisfying

$$\mathrm{E}(\boldsymbol{\epsilon}_{ijk}) = \mathbf{0} \quad \text{and} \quad \mathrm{Cov}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{\Sigma}_{i,jj} > 0. \tag{3.2}$$

Stacking the group mean vectors into one vector, define $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_a^\top)^\top$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}^\top, \ldots, \boldsymbol{\mu}_{it}^\top)^\top$ and $\boldsymbol{\mu}_{ij} = (\mu_{ij}^{(1)}, \ldots, \mu_{ij}^{(p)})^\top$. Note that we allow different covariance matrices $\boldsymbol{\Sigma}_i$ and different sample sizes n_i in different groups . The distribution of the error terms can also be different across groups. The total sample size is denoted by $N = \sum_{i=1}^a n_i$ and $\tilde{N} = tN$ is the total number of p-variate observations. The following proportional divergence of the sample sizes is assumed in order to derive asymptotic results.

$$A_1: \quad \frac{n_i}{N} \to \kappa_i > 0 \text{ as } N \to \infty, \quad i = 1, \ldots, a. \tag{3.3}$$

$$A_2: \quad \sup_{i,j} \mathrm{E}(||\boldsymbol{\epsilon}_{ij1}||^4) < \infty, \quad i = 1, \ldots, a; j = 1, \ldots, t. \tag{3.4}$$

Let $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_a^\top)^\top$, where $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}^\top, \ldots, \boldsymbol{X}_{in_i}^\top)^\top$, $\boldsymbol{X}_{ik} = (\boldsymbol{X}_{i1k}^\top, \ldots, \boldsymbol{X}_{itk}^\top)^\top$ and $\boldsymbol{X}_{ijk} = (X_{ijk}^{(1)}, \ldots, X_{ijk}^{(p)})^\top$. We denote the overall sample mean vector by $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_{1\cdot}^\top, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^\top)^\top$, where $\overline{\boldsymbol{X}}_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{X}_{ik}$, and the covariance of $\sqrt{N}\overline{\boldsymbol{X}}$ by

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\sqrt{N}\overline{\boldsymbol{X}}) = \mathrm{diag}(\frac{N}{n_i}\boldsymbol{\Sigma}_i : 1 \leq i \leq a),$$

a block diagonal matrix, where diagonal blocks are the covariances of $\sqrt{N}\overline{\boldsymbol{X}}_{i\cdot}$. Similarly, we denote the empirical covariance by

$$\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\frac{N}{n_i}\widehat{\boldsymbol{\Sigma}}_i : 1 \leq i \leq a), \tag{3.5}$$

where

$$\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot})^\top.$$

Note that the estimator $\widehat{\boldsymbol{\Sigma}}$ is strongly consistent under the asymptotic setting A_1 in (3.3) because $\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}$ converges almost surely to the identity matrix $\boldsymbol{I}$.

To conduct a hypothesis test, let $\boldsymbol{H}$ be an appropriate contrast matrix. Generally, it is more convenient to use the unique projection matrix $\boldsymbol{T} = \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^-\boldsymbol{H}$, where $(\boldsymbol{H}\boldsymbol{H}^\top)^-$ is some generalized inverse of $\boldsymbol{H}\boldsymbol{H}^\top$, to formulate our hypotheses. It is easy to show that $\boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$ if and only if $\boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$. Therefore, the unique, symmetric and idempotent contrast matrix $\boldsymbol{T}$ is equivalent to $\boldsymbol{H}$ for testing $\boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$. For the mandible data in Timm (1980, Table 7.2), there are three dependent variables ($p = 3$), three time points ($t = 3$) and two treatment groups ($a = 2$) with equal sample sizes $n_1 = n_2 = 9$ in two groups. The goal of the study was to test whether the treatments induced differential changes over time on the mandibles. Hence, it is of interest to check if there was any treatment (group) effect (G), time effect (T), and treatment (group) by time interaction effect (GT). The contrast matrices are $\boldsymbol{H}_G = \boldsymbol{P}_2 \otimes \frac{1}{3}\boldsymbol{1}_3^\top \otimes \boldsymbol{I}_3$, $\boldsymbol{H}_T = \frac{1}{2}\boldsymbol{1}_2^\top \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$ and $\boldsymbol{H}_{GT} = \boldsymbol{P}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$, respectively. The corresponding unique projection matrices are $\boldsymbol{T}_G = \boldsymbol{P}_2 \otimes \frac{1}{3}\boldsymbol{J}_3 \otimes \boldsymbol{I}_3$, $\boldsymbol{T}_T = \frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$, and $\boldsymbol{T}_{GT} = \boldsymbol{P}_2 \otimes \boldsymbol{P}_3 \otimes \boldsymbol{I}_3$, respectively.

3.3 Test Statistics and Distributions

Wald's Asymptotic Test

One classical approach to conducting a multivariate hypothesis test is to use the Wald-type statistic (WTS). For testing $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, the WTS is defined by

$$Q_N(\boldsymbol{T}) = N\overline{\boldsymbol{X}}^\top \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}, \tag{3.6}$$

where $(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T})^+$ denotes the pseudo inverse or Moore-Penrose inverse of $\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T}$. Under the null hypothesis, the statistic $Q_N(\boldsymbol{T})$ has an asymptotic chi-square distribution, which is formally stated in the following theorem.

Theorem 3.3.1. *Assume the model* (3.1) *and* (3.2)*. Under the null hypothesis* $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \mathbf{0}$*, and under the conditions* A_1 *and* A_2*,* Q_N *in* (3.6) *has, asymptotically as* $N \to \infty$*, a central* $\chi^2_{f_{\boldsymbol{T}}}$ *distribution with degrees of freedom* $f_{\boldsymbol{T}} = \text{rank}(\boldsymbol{T})$*.*

The corresponding test is given by $\varphi_{\text{WTS}} = \mathbb{I}\{Q_N(\boldsymbol{T}) > \chi^2_{f_{\boldsymbol{T}},1-\alpha}\}$, where $\chi^2_{f_{\boldsymbol{T}},1-\alpha}$ denotes the $(1-\alpha)$-quantile of the $\chi^2_{f_{\boldsymbol{T}}}$ distribution. The test is an asymptotic level α test and is consistent for general fixed alternatives $\boldsymbol{T}\boldsymbol{\mu} \neq \mathbf{0}$. The proof of Theorem 3.3.1 is given in the Appendix (Section 3.7).

It is well known that the Wald-type statistic has slow convergence due to the fat-tailed limiting distribution, chi-square distribution. Therefore, in order to achieve a satisfactory approximation from the chi-square asymptotic distribution, large sample sizes are needed. In the next section, we will develop simple resampling approximations to the null distribution to overcome the large sample size requirement. We will then prove that the procedures are actually asymptotically valid in general factorial designs where no assumptions of equal covariances, sample sizes and error distributions are required.

Two-stage Resampling Method

In this section, we delve into the resampling procedures with permutation, bootstrap and hybrid permutation-bootstrap to propose test procedures for the multivariate repeated measures data. We also demonstrate the validity of the test procedures by deriving resampling Central Limit Theorem.

Permutation over Time

Let $\boldsymbol{X}^\pi$ denote a fixed but arbitrary permutation of all $\tilde{N}$ p-variate observations of $\boldsymbol{X}$ for the total of N experimental or observational units,

$$\boldsymbol{X}^\pi = \pi(\boldsymbol{X}_{111}^\top, \ldots, \boldsymbol{X}_{an_at}^\top)^\top = ({\boldsymbol{X}_{111}^\pi}^\top, \ldots, {\boldsymbol{X}_{an_at}^\pi}^\top)^\top.$$

The notation $\boldsymbol{X}_{ijk}^\pi$ denotes the p-variate vector of the kth subject in the ith group at the jth time point from the permuted vector $\boldsymbol{X}^\pi$. In other words, the permutation process permutes the individual $\boldsymbol{X}_{ijk}$ vectors from $\boldsymbol{X}$ and results in a new vector $\boldsymbol{X}^{\pi}$, which is composed of the same individual $\boldsymbol{X}_{ijk}$ vectors but reordered. Analogously, we use $\overline{\boldsymbol{X}}^{\pi}$ to denote the overall sample mean vector of the permuted vector $\boldsymbol{X}^{\pi}$, and $\widehat{\boldsymbol{\Sigma}}^{\pi} = \bigoplus_{i=1}^{a} \frac{N}{n_i} \widehat{\boldsymbol{\Sigma}}_i^{\pi}$ to denote the empirical covariance matrix of $\sqrt{N}\overline{\boldsymbol{X}}^{\pi}$,where $\widehat{\boldsymbol{\Sigma}}_i^{\pi} = \frac{1}{n_i-1} \sum_{k=1}^{n_i} (\boldsymbol{X}_{ik}^{\pi} - \overline{\boldsymbol{X}}_{i\cdot}^{\pi})(\boldsymbol{X}_{ik}^{\pi} - \overline{\boldsymbol{X}}_{i\cdot}^{\pi})^{\top}$.

As Good (2005) noted, with classical permutation tests, one has to ensure exchangeablitity to provide meaningful results. For example, if a set of observations $\boldsymbol{X}_{ijk}$, $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n_i$ are i.i.d., then they are exchangeable, which means that their joint distribution will still be the same for any relabeling.

In our context, if the p-variate observations are exchangeable under the null hypothesis $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, $\boldsymbol{X}$ and $\boldsymbol{X}^{\pi}$ would have the same distribution. However, with general factorial design, it is not the case. For example, in Huang et al. (2006) the general two- or higher-way layouts observations are not exchangeable. The studentization methodology (see Chung and Romano, 2013; Janssen, 1997; Neuhaus, 1993; Janssen and Pauls, 2003; Janssen, 2005; Omelka and Pauly, 2012; Pauly et al., 2015; Neubert and Brunner, 2007) can be applied to the permuted overall sample mean vector $\sqrt{N}\overline{\boldsymbol{X}}^{\pi}$ to overcome the heteroscedasticity problem. Projecting the permuted overall sample mean vector onto the hypothesis space, we obtain the Wald-type statistic for the permuted vector in an analogous manner as in (3.6),

$$Q_N^{\pi}(\boldsymbol{T}) = N\, \overline{\boldsymbol{X}}^{\pi\,\top} \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}^{\pi}\boldsymbol{T})^{+}\boldsymbol{T}\overline{\boldsymbol{X}}^{\pi}, \tag{3.7}$$

where $\overline{\boldsymbol{X}}^{\pi} = (\overline{\boldsymbol{X}}_{1\cdot}^{\pi\,\top}, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^{\pi\,\top})^{\top}$ and $\overline{\boldsymbol{X}}_{i\cdot}^{\pi} = \frac{1}{n_i} \sum_{k=1}^{n_i} \boldsymbol{X}_{ik}^{\pi}$ for $1 \leq i \leq a$.

The validity of the permutation procedure is demonstrated in the following theorem where the conditional distribution of the permutation version of the Wald-type statistic $Q_N^{\pi}(\boldsymbol{T})$ is ensured to always approximate the distribution of $Q_N(\boldsymbol{T})$ under the null hypothesis. In particular, the weak convergence of Q_N^{π} conditioned on the data $\boldsymbol{X}$ is estimated. This is a desirable property, which justifies that the resampling method works and is often hard to come by.

Theorem 3.3.2. *Assume the model described in* (3.1) *and* (3.2). *Under the null hypothesis* $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, *and the conditions* A_1 *and* A_2, *conditional on the observed* *data* $\boldsymbol{X}$*, the studentized permutation distribution of* Q_N^π *in* (3.7) *weakly converges to the central* $\chi^2_{f_{\boldsymbol{T}}}$ *distribution in probability, where* $f_{\boldsymbol{T}} = rank(\boldsymbol{T})$.

The corresponding permutation test is given by $\varphi_{\text{WTPS}} = \mathbb{I}\{Q_N(\boldsymbol{T}) > c_{1-\alpha}\}$, where $c_{1-\alpha}$ is the conditional $(1-\alpha)$-quantile of the permutation distribution of Q_N^π given the data. The proof of Theorem 3.3.2 essentially involves showing the Prohorov distance between the conditional distribution of $Q_N^\pi(\boldsymbol{T})$ given $\boldsymbol{X}$ and the null distribution of $Q_N(\boldsymbol{T})$ converges to zero in probability as $N \to \infty$. This conditional Central Limit Theorem holds under both the null and the alternative hypotheses. More precisely, for any underlying parameter $\boldsymbol{\mu} \in \mathbb{R}^{atp}$ and the null value $\boldsymbol{\mu}_0$ satisfying $\boldsymbol{T}\boldsymbol{\mu}_0 = \boldsymbol{0}$,

$$\sup_{x\in\mathbb{R}} |P_{\boldsymbol{\mu}}\left(Q_N^\pi(\boldsymbol{T}) \le x|\boldsymbol{X}\right) - P_{\boldsymbol{\mu}_0}\left(Q_N(\boldsymbol{T}) \le x\right)| \to 0$$

in probability, where $P_{\boldsymbol{\mu}}\left(Q_N^\pi(\boldsymbol{T}) \le x|\boldsymbol{X}\right)$ denotes the conditional distribution function of $Q_N^\pi(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}$ is the true parameter and $P_{\boldsymbol{\mu}_0}\left(Q_N(\boldsymbol{T}) \le x\right)$ denotes the unconditional distribution function of $Q_N(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}_0$ is the true parameter. This convergence guarantees that the level α critical value obtained from the resampling distribution always converges to the $\chi^2_{f_{\boldsymbol{T}},1-\alpha}$, which is the $(1-\alpha)$-quantile of the asymptotic distribution of the Wald-type statistic.

In addition, the permutation test, asymptotically, keeps the preassigned level α under the null hypothesis. Moreover, it is consistent for any fixed alternative point $\boldsymbol{\mu}$ for which $\boldsymbol{T}\boldsymbol{\mu} \neq \boldsymbol{0}$. Thus, asymptotically, the permutation test has power 1 for any fixed alternative. The numerical studies in Section 3.4 conclusively confirm this theoretical result in finite-samples. Furthermore, when the observed data $\boldsymbol{X}$ is exchangeable under the null hypothesis, the joint distribution is invariant under all rearrangements of the observations. That is, $\boldsymbol{X}$ and $\boldsymbol{X}^\pi$ have the same distribution resulting in the distribution of the test statistic Q_N^π in (3.7) asymptotically follows central $\chi^2_{f_{\boldsymbol{T}}}$ distribution. In other words, the permutation test is asymptotically exact. The proof of Theorem 3.3.2 is given in the Appendix (Section 3.7).

Bootstrap Sampling the Subjects

So far, we described a permutation resampling technique to deal with the repeated measures data despite the time dependencies. Notice that we have p variables measured each time for every experimental unit. To deal with the multivariate aspect of growth curve data , we use bootstrap resampling technique. To better capture the covariance structure of the original data and to get a more accurate finite sample approximation, we employ parametric bootstrap. As we reviewed in Section 3.1, for existing mean-based inference methods on multivariate data, there are few that do not assume either multivariate normality or homoscedasticity. For example, Xu and Cui (2008) does not assume multivariate normality, but it is median-based MANOVA and requires equal covariances across groups. When covariance matrices are heterogeneous, Vallejo et al. (2001) compares a couple of multivariate test procedures assuming multivariate normality. Without the assumptions of multivariate normality and homogeneity of covariance matrices, parametric bootstrap has long been a prevalent resampling technique and has been applied in the context of one-way and two-way factorial designs (see Zhang and Liu, 2013; Xu, 2015; Zhang, 2012; Krishnamoorthy and Lu, 2010; Xu et al., 2013).

In general, parametric bootstrap is utilized for parametric models. However, we show below a parametric bootstrap which results in an asymptotically valid procedure in our nonparametric settings. This parametric bootstrap arises from the use of multivariate normality with the rationale that it can better mimic the covariance structure of the original observations and further lead to a more accurate finite-sample approximation. Given the p-variate observations at t time points, we generate n_i samples from $N(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_i)$, for $i = 1, \ldots, a$,

$$\boldsymbol{X}_{i1}^*, \ldots, \boldsymbol{X}_{in_i}^* \overset{\text{i.i.d}}{\sim} N(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_i).$$

With tp-variate observations $\boldsymbol{X}_{11}^*, \ldots, \boldsymbol{X}_{an_a}^*$, the bootstrap version of the Wald-type statistic can be calculated in the way analogous to Q_N in (3.6),

$$Q_N^*(\boldsymbol{T}) = N\, \overline{\boldsymbol{X}}^{*\top} \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}^*\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}^*, \tag{3.8}$$

where $\overline{\boldsymbol{X}}^* = (\overline{\boldsymbol{X}}_{1\cdot}^{*\top}, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^{*\top})^\top$, $\overline{\boldsymbol{X}}_{i\cdot}^* = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{X}_{ik}^*$ and $\widehat{\boldsymbol{\Sigma}}^* = \bigoplus_{i=1}^{a} \frac{N}{n_i}\widehat{\boldsymbol{\Sigma}}_i^*$ and $\widehat{\boldsymbol{\Sigma}}_i^* = \frac{1}{n_i-1}\sum_{k=1}^{n_i}(\boldsymbol{X}_{ik}^* - \overline{\boldsymbol{X}}_{i\cdot}^*)(\boldsymbol{X}_{ik}^* - \overline{\boldsymbol{X}}_{i\cdot}^*)^\top$. The following theorem confirms that the parametric bootstrap version of the Wald-type statistic $Q_N^*(\boldsymbol{T})$ has the same asymptotic properties as the permutation version of the Wald-type statistic $Q_N^\pi(\boldsymbol{T})$. Specifically, under the null hypothesis, conditioned on the observed data $\boldsymbol{X}$, the bootstrap distribution of $Q_N^*(\boldsymbol{T})$ always approximates the distribution of the original Wald-type statistic $Q_N(\boldsymbol{T})$.

Theorem 3.3.3. *Assume the model described in* (3.1) *and* (3.2)*. Under the null hypothesis $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, and the conditions A_1 and A_2, conditional on the observed data $\boldsymbol{X}$, the bootstrap distribution of Q_N^* in* (3.8) *weakly converges to the central $\chi^2_{f_{\boldsymbol{T}}}$ distribution in probability, where $f_{\boldsymbol{T}} = rank(\boldsymbol{T})$.*

The corresponding bootstrap test is given by $\varphi_{\text{WTBS}} = \mathbb{I}\{Q_N(\boldsymbol{T}) > c^*_{1-\alpha}\}$, where $c^*_{1-\alpha}$ is the conditional $(1-\alpha)$-quantile of the bootstrap distribution of Q_N^* given the data. As in the permutation test, the conditional Central Limit Theorem remains true under both the null and alternative hypothesis. More precisely, $\forall \boldsymbol{\mu} \in \mathbb{R}^{atp}$ and the null value $\boldsymbol{\mu}_0$ satisfying $\boldsymbol{T}\boldsymbol{\mu}_0 = \boldsymbol{0}$,

$$\sup_{x\in\mathbb{R}} \left| P_{\boldsymbol{\mu}}\left(Q_N^*(\boldsymbol{T}) \le x | \boldsymbol{X}\right) - P_{\boldsymbol{\mu}_0}\left(Q_N(\boldsymbol{T}) \le x\right)\right| \to 0$$

in probability, where $P_{\boldsymbol{\mu}}\left(Q_N^*(\boldsymbol{T}) \le x | \boldsymbol{X}\right)$ denotes the conditional distribution function of $Q_N^*(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}$ is the true parameter and $P_{\boldsymbol{\mu}_0}\left(Q_N(\boldsymbol{T}) \le x\right)$ denotes the unconditional distribution function of $Q_N(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}_0$ is the true parameter. Accordingly, the conditional distribution of the parametric bootstrap version of the Wald-type statistic $Q_N^*(\boldsymbol{T})$ approximates the distribution of $Q_N(\boldsymbol{T})$ under the null hypothesis. Further, the level α critical value obtained from the bootstrap distribution of $Q_N^*(\boldsymbol{T})$ converges to the $\chi^2_{f_{\boldsymbol{T}},1-\alpha}$, which is the $(1-\alpha)$-quantile of the asymptotic distribution of the Wald-type statistic. In other words, the parametric bootstrap test and the permutation test share the same asymptotic properties. The proof of Theorem 3.3.3 is given in the Appendix (Section 3.7).

Another well known bootstrap is called wild bootstrap, which was originally developed by Wu (1986). It was designed to analyze data with heteroscedasticity. The idea of wild bootstrap is similar to that of the residual bootstrap, where the response variables are resampled based on the residuals. Specifically, the residuals are randomly multiplied by a random variable. For this random variable, Davidson and Flachaire (2008) proposed the Rademacher distribution, which is a discrete probability distribution. A Rademacher random variable has 50% probability of taking -1 and 50% probability of taking 1. The tp-variate wild bootstrap samples are

$$\boldsymbol{X}_{ik}^{\dagger} = \overline{\boldsymbol{X}}_{i\cdot} + s_{ik}^{\dagger}\left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}\right),$$

where $i = 1, \ldots, a$, $k = 1, \ldots, n_i$ and $s_{ik}^{\dagger}$ follows Rademacher distribution, i.e.,

$$s_{ik}^{\dagger} = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}.$$

Analogously, the wild bootstrap version of the Wald-type statistic can be calculated and is denoted by $Q_N^{\dagger}$,

$$Q_N^{\dagger}(\boldsymbol{T}) = N\, \overline{\boldsymbol{X}}^{\dagger\,\top} \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}^{\dagger}\boldsymbol{T})^{+}\boldsymbol{T}\overline{\boldsymbol{X}}^{\dagger}, \tag{3.9}$$

where $\overline{\boldsymbol{X}}^{\dagger} = (\overline{\boldsymbol{X}}_{1\cdot}^{\dagger\,\top}, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^{\dagger\,\top})^{\top}$, $\overline{\boldsymbol{X}}_{i\cdot}^{\dagger} = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{X}_{ik}^{\dagger}$ and $\widehat{\boldsymbol{\Sigma}}^{\dagger} = \bigoplus_{i=1}^{a} \frac{N}{n_i}\widehat{\boldsymbol{\Sigma}}_i^{\dagger}$ and $\widehat{\boldsymbol{\Sigma}}_i^{\dagger} = \frac{1}{n_i-1}\sum_{k=1}^{n_i}(\boldsymbol{X}_{ik}^{\dagger} - \overline{\boldsymbol{X}}_{i\cdot}^{\dagger})(\boldsymbol{X}_{ik}^{\dagger} - \overline{\boldsymbol{X}}_{i\cdot}^{\dagger})^{\top}$. The asymptotic properties of $Q_N^{\dagger}$ are formally stated in the following theorem.

Theorem 3.3.4. *Assume the model described in* (3.1) *and* (3.2). *Under the null hypothesis* $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, *and the conditions* A_1 *and* A_2, *conditional on the observed data* $\boldsymbol{X}$, *the wild bootstrap distribution of* $Q_N^{\dagger}$ *in* (3.9) *weakly converges to the central* $\chi^2_{f_{\boldsymbol{T}}}$ *distribution in probability, where* $f_{\boldsymbol{T}} = rank(\boldsymbol{T})$.

The corresponding wild bootstrap test is $\varphi_{\text{WTWBS}} = \mathbb{I}\{Q_N(\boldsymbol{T}) > c_{1-\alpha}^{\dagger}\}$, where $c_{1-\alpha}^{\dagger}$ is the conditional $(1-\alpha)$-quantile of the wild bootstrap distribution of $Q_N^{\dagger}$ given the data. Moreover, the following result holds for wild bootstrap procedure as well. $\forall \boldsymbol{\mu} \in \mathbb{R}^{atp}$ and the null value $\boldsymbol{\mu}_0$ satisfying $\boldsymbol{T}\boldsymbol{\mu}_0 = \boldsymbol{0}$,

$$\sup_{x\in\mathbb{R}} |P_{\boldsymbol{\mu}}(Q_N^{\dagger}(\boldsymbol{T}) \leq x|\boldsymbol{X}) - P_{\boldsymbol{\mu}_0}(Q_N(\boldsymbol{T}) \leq x)| \to 0$$

in probability, where $P_{\boldsymbol{\mu}}(Q_N^{\dagger}(\boldsymbol{T}) \leq x|\boldsymbol{X})$ denotes the conditional distribution function of $Q_N^{\dagger}(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}$ is the true parameter and $P_{\boldsymbol{\mu}_0}\left(Q_N(\boldsymbol{T}) \leq x\right)$ denotes the unconditional distribution function of $Q_N(\boldsymbol{T})$ under the assumption that $\boldsymbol{\mu}_0$ is the true parameter. The proof of Theorem 3.3.4 is analogous to that of Theorem 3.3.3 and it is briefly given in the Appendix (Section 3.7).

Hybrid Permutation–Bootstrap Resampling

With the validity of separate permutation test and parametric bootstrap test established, we propose a Wald-type hybrid permutation-bootstrap procedure. With the original stacked data $\boldsymbol{X}$, keeping each p-variate observation $\boldsymbol{X}_{ijk}$ intact, we permute or relabel the observations. Next, we conduct parametric bootstrap for each group using sample mean and sample covariance of the permuted vector $\boldsymbol{X}_i^{\pi}$, specifically, generating multivariate normal data for each group. We denote the permuted and then bootstrapped data by $\boldsymbol{X}^{\star}$. Analogously, the permutation-bootstrap version of the Wald-type statistic can be calculate,

$$Q_N^{\star}(\boldsymbol{T}) = N\, \overline{\boldsymbol{X}}^{\star\top} \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}^{\star}\boldsymbol{T})^{+}\boldsymbol{T}\overline{\boldsymbol{X}}^{\star}, \tag{3.10}$$

where $\overline{\boldsymbol{X}}^{\star} = (\overline{\boldsymbol{X}}_{1\cdot}^{\star\top}, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^{\star\top})^{\top}$, $\overline{\boldsymbol{X}}_{i\cdot}^{\star} = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{X}_{ik}^{\star}$ and $\widehat{\boldsymbol{\Sigma}}^{\star} = \bigoplus_{i=1}^{a} \frac{N}{n_i}\widehat{\boldsymbol{\Sigma}}_i^{\star}$ and $\widehat{\boldsymbol{\Sigma}}_i^{\star} = \frac{1}{n_i-1}\sum_{k=1}^{n_i}(\boldsymbol{X}_{ik}^{\star} - \overline{\boldsymbol{X}}_{i\cdot}^{\star})(\boldsymbol{X}_{ik}^{\star} - \overline{\boldsymbol{X}}_{i\cdot}^{\star})^{\top}$. We repeat such bootstrap process for so many times, say 100, to get a distribution of $Q_N^{\star}$.

Further, we repeat the entire permutation-bootstrap process for 100 times, which creates 100 of the distributions of $Q_N^{\star}$. Combining $100 \times 100 = 10,000$ of the $Q_N^{\star}$ values, we obtain a final distribution of $Q_N^{\star}$. The corresponding hybrid permutation-bootstrap test is given by $\varphi_{\text{WTHS}} = \mathbb{I}\{Q_N(\boldsymbol{T}) > c_{1-\alpha}^{\star}\}$, where $c_{1-\alpha}^{\star}$ is the conditional $(1-\alpha)$-quantile of the permutation-bootstrap combined distribution of $Q_N^{\star}$ given the data. Moreover, we calculate the p-value by comparing the original WTS Q_N with $10,000$ of the $Q_N^{\star}$ values. The step by step numerical algorithm is given below:

Step1: Given $\boldsymbol{X}$, calculate the original Wald-type statistic $Q_N(\boldsymbol{T})$ for appropriate choice of the projection matrix.

Step 2: Randomly permute $\boldsymbol{X}$ with each $p \times 1$ vector intact. In other words, permute the p-variate observations as a whole.

Step 3: Using group sample means and group sample covariance matrices of the permuted data from step 2, conduct $\mathbb{B}$ parametric bootstraps.

Step 4: Repeat steps 2-3 a large number of, say $\mathbb{P}$, times.

Step 5: Calculate Wald-type statistic $Q_N^{\star}(\boldsymbol{T})$ for each of the $\mathbb{P} \times \mathbb{B}$ data sets from step 4, denoted by $Q_{N(1)}^{\star}, \ldots, Q_{N(L)}^{\star}$, where $L = \mathbb{P} \times \mathbb{B}$

Step 6: Compute the p-value by

$$p\text{-value} = \frac{1}{\mathbb{P} \times \mathbb{B}} \sum_{l=1}^{\mathbb{P} \times \mathbb{B}} \mathbb{I}\{Q_N(\boldsymbol{T}) \leq Q_{N(l)}^{\star}(\boldsymbol{T})\}.$$

3.4 Simulation Studies

Simulation Design

In this section we numerically evaluate the testing schemes developed in Section 3.3 using several specific designs. To study the finite-sample behavior of the resampling-based tests, i.e., the Wald-type permutation test (PT), parametric bootstrap test (BT), wild bootstrap (WBT) and the hybrid permutation-bootstrap test (PBT), we investigate the empirical type-I error rate and power for detecting fixed alternatives via simulations. To make the investigation more comprehensive, we also include the classical multivariate tests, DMM, MMM and the Wald's asymptotic tests (AT) as competing procedures. We compare the performance of the aforementioned seven test procedures in terms of their ability to control the preassigned type-I error rate and to detect fixed alternatives under different settings. These tests will hereinafter be referred to as their corresponding abbreviations. We conduct our simulations under various designs, including multiple settings for sample size, data distribution, covariance structure and effect tested. Specifically, we have the following objectives for our simulation studies.

1. Investigate the effect of covariance structure on the performance of the tests.

2. Analyze the effect of distribution on the performance of the tests.

3. Examine the effect of hypothesis type on the performance of the tests.

4. Study the effect of sample size on the performance of the tests.

All simulations are conducted in **R** version 3.6.2 with 5000 of simulations for one set of test results and the nominal type-I error rate is $\alpha = 0.05$. Within each simulation, 100 permutations and 100 bootstraps are carried out for PBT. The number of repetitions is set to be 1000 for PT, BT and WBT. Generally, the larger number of repetitions and simulation runs is used, the better simulation results will be. However, it also means that more computational resources are required. We choose these numbers because not only they are large enough to confirm the validity of our methodology, but also they are considerably efficient in terms of computing time.

In our simulations, we simulate multivariate growth curve data in the context of general factorial design. To be concrete, we set the number of groups to be $a = 2$ with group sample sizes denoted by $\boldsymbol{n} = (n_1, n_2)$. We adequately evaluate both balanced and unbalanced cases. For balanced settings, we consider $\boldsymbol{n} = \{(20, 20), (30, 30)\}$. For unbalanced cases, we investigate $\boldsymbol{n} = \{(20, 30), (30, 40)\}$.

In addition, the computational complexity of multivariate data analysis can grow exponentially as the dimension increases even just by one. Therefore, to illustrate the performance of our methods in a more succinct and efficient way, we set the dimension of the observations to be $p = 4$, and the number of repeated measures to be $t = 2$. We are interested in whether there is any Group (G), Time (T) and Group×Time (GT) effects. The corresponding unique projection matrices for testing these hypotheses are $\boldsymbol{T}_G = \boldsymbol{P}_a \otimes \frac{1}{t}\boldsymbol{J}_t \otimes \boldsymbol{I}_p$, $\boldsymbol{T}_T = \frac{1}{a}\boldsymbol{J}_a \otimes \boldsymbol{P}_t \otimes \boldsymbol{I}_p$ and $\boldsymbol{T}_{GT} = \boldsymbol{P}_a \otimes \boldsymbol{P}_t \otimes \boldsymbol{I}_p$, respectively.

For covariance structure, we consider both homoscedastic and heteroscedastic designs. The covariance structures investigated are compound symmetry and the first order autoregressive structure. With the partitioned matrix $\boldsymbol{\Sigma}_i = (\boldsymbol{\Sigma}_{i,jk})$ representation, two covariance structures to be considered are below.

(i) Compound symmetry

$$\boldsymbol{\Sigma}_i = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}_p & k = 1,\dots,t \\ \boldsymbol{\Sigma}_{i,kj} = \rho_i \boldsymbol{J}_p & k \neq j \text{ and } k,j = 1,\dots,t \end{cases},$$

where we take $\boldsymbol{\rho} = (0.2, 0.2)$ as the homoscedastic setting and $\boldsymbol{\rho} = (0.2, 0.7)$ as the heteroscedastic setting.

(ii) Autoregressive structure AR(1)

$$\boldsymbol{\Sigma}_i = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}_p & k = 1,\dots,t \\ \boldsymbol{\Sigma}_{i,kj} = \rho_i^{|k-j|} \boldsymbol{J}_p & j \neq k \text{ and } k = 1,\dots,t \end{cases},$$

where we consider $\boldsymbol{\rho} = (0.2, 0.2)$ as the equal covariances setting and $\boldsymbol{\rho} = (0.2, 0.7)$ as the unequal covariances setting.

The compound symmetric and the autoregressive covariance structures will hereinafter be referred to as CS and AR, respectively. The equal covariances settings have the same correlation value in different groups, representing homoscedasticity. However, the unequal covariances settings have different correlation values in different groups, serving as heteroscedastic cases. Heteroscedastic designs with $\boldsymbol{\rho} = (0.2, 0.7)$ are considered as positive pairings if n_2 is larger than n_1, but are considered as negative pairings if n_2 is smaller than n_1.

Data are generated according to the model

$$\boldsymbol{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\epsilon}_{ik}, \quad i = 1,\dots,a; k = 1,\dots n_i, \tag{3.11}$$

where $\boldsymbol{\Sigma}_i^{1/2}$ is the square root of $\boldsymbol{\Sigma}_i$. To generate independent and identically distributed random vectors $\boldsymbol{\epsilon}_{ik} = (\boldsymbol{\epsilon}_{i1k}^\top, \dots, \boldsymbol{\epsilon}_{itk}^\top)^\top$, where $\boldsymbol{\epsilon}_{ijk} = (\epsilon_{ijk}^{(1)}, \dots, \epsilon_{ijk}^{(p)})^\top$, we generate each component from the same standardized distribution by

$$\epsilon_{ijk}^{(s)} = \frac{X_{ijk}^{(s)} - \mathrm{E}(X_{ijk}^{(s)})}{\sqrt{\mathrm{Var}(X_{ijk}^{(s)})}}, \quad s = 1,\dots,p,$$

where $X_{ijk}^{(s)}$ can be normal, $t(5)$, $\chi^2(5)$ and lognormal random variables.

Simulation Results

The simulation results for testing Time and Group $\times$ Time effects are presented here for illustration purpose. The simulation results for testing Group effect are shown in the Appendix (Section 3.7).

Table 3.1–3.2 present simulation results for homoscedastic settings with balanced and unbalanced sample sizes, respectively. It can be seen from these two tables that for all four data distributions (normal, $t(5)$, $\chi^2(5)$ and lognormal) and for testing all three effects (Time, Group, and Group$\times$Time effects), AT tends to highly over-reject the null hypothesis in all scenarios, especially when sample sizes are smaller, i.e., $\boldsymbol{n} = (20, 20)$ in Table 3.1 and $\boldsymbol{n} = (20, 30)$ in Table 3.2. However, under homoscedastic settings, PBT, BT, DMM and MMM produce type-I error rates that are almost exact for normal, $t(5)$ and $\chi^2(5)$ distributions and for testing all three effects. When the data are lognormal, these four tests are conservative. Remarkably, under the same settings, WBT and PT perform very well in terms of controlling the nominal type-I error rate in all scenarios.

Next, we investigate the test performance under heteroscedastic settings in Tables 3.3, 3.4 and 3.5. Likewise, test results of AT are extremely liberal in all scenarios of heteroscedastic settings. In some cases the type-I error rates are larger than 10%. Under balanced heteroscedastic settings in Table 3.3, PBT, BT, DMM and MMM perform well for normal, $t(5)$ and $\chi^2(5)$ distributions, but they lead to conservative test results for lognormal data. However, WBT and PT maintain the type-I error rates to the nominal level in all cases.

Table 3.4 shows test results of positive pairing cases where the larger (smaller) covariance is associated with the larger (smaller) sample size. Under such settings, test results of AT, DMM and MMM are very liberal in general. Although test results of PBT and PT are slightly liberal for the smaller sample size set $\boldsymbol{n} = (20, 30)$, they perform well for the larger sample size set $\boldsymbol{n} = (30, 40)$. Additionally, BT keeps the test results to the nominal level, especially for normal, $t(5)$ and $\chi^2(5)$ distributions. WBT performs well in all positive pairing cases, for all four distributions and for all three effects.

Under negative pairing settings, as shown in Table 3.5, DMM and MMM are extremely conservative. However, PBT, PT, BT and WBT perform very well in general. PBT and BT are slightly conservative for testing Time and Group×Time effects with highly skewed lognormal data. It's worth noting that test results of WBT and PT are almost exact in all scenarios.

Power Studies

Power versus Effect Size

We also investigate empirical power of the aforementioned test procedures to detect a fixed alternative. To assure that all competing methods control the type-I error rate under the null hypothesis, we restrict our power simulations to balanced homoscedastic designs. Specifically, we set the number of groups to be $a = 2$ with sample sizes $\boldsymbol{n} = (30, 30)$. There are four response variables ($p = 4$) and three time points ($t = 3$). The covariance structure is compound symmetric with $\boldsymbol{\rho} = (0.2, 0.2)$.

Let δ vary from 0 to 0.6 with 0.1 increment. For testing Time effect, we set $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \delta(1, 2, \ldots, t)^\top \otimes \mathbf{1}_p$. For testing Group effect, we let $\boldsymbol{\mu}_1 = \mathbf{0}_{tp\times 1}$ and $\boldsymbol{\mu}_2 = \delta \mathbf{1}_t \otimes \mathbf{1}_p$. For testing Group×Time, we have $\boldsymbol{\mu}_1 = \mathbf{0}_{tp\times 1}$ and $\boldsymbol{\mu}_2 = \delta(1, 2, \ldots, t)^\top \otimes \mathbf{1}_p$.

Figure 3.1 shows the empirical power curves for detecting Time, Group and Group×Time effects of all seven test procedures along increasing δ value, with data from normal, $t(5)$, $\chi^2(5)$ and lognormal distributions. It can be seen that, in general, AT has the most liberal test results at $\delta = 0$, which leads to the largest power among all methods, especially for testing Time and Group×Time effects. However, due to its extreme liberality, the AT test is not recommended for use in practice. We include the AT method for illustration purpose only. For comparison of the test performance in terms of power, we only compare the remaining six methods, i.e., DMM, MMM, WBT, PBT, PT and BT.

For testing Group effect in the bottom row of Figure 3.1, all six methods achieve almost identical power along different δ values for normal, $t(5)$ and $\chi^2(5)$ distributions.

Table 3.1: Type-I error rate ($\times 100$) of DMM, MMM, AT, WBT, PBT, PT and BT for balanced homoscedastic factorial designs for testing Time and Group×Time effects, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	4.4	4.4	8.9	4.5	4.4	4.6	4.4
		(30,30)	4.6	4.6	7.2	4.9	4.6	4.9	4.7
	AR	(20,20)	4.7	4.7	9.0	4.9	4.6	4.9	4.9
		(30,30)	4.7	4.7	7.7	4.7	4.5	4.7	4.7
$t(5)$	CS	(20,20)	4.9	4.9	8.8	5.1	4.7	4.9	4.8
		(30,30)	4.9	4.9	7.7	5.3	4.8	5.0	4.8
	AR	(20,20)	4.6	4.6	8.3	4.7	4.5	4.7	4.5
		(30,30)	4.9	4.9	7.3	5.0	4.7	5.1	4.9
$\chi^2(5)$	CS	(20,20)	4.8	4.8	9.5	5.1	4.5	5.0	4.8
		(30,30)	4.8	4.8	7.1	4.7	4.7	5.1	4.8
	AR	(20,20)	5.0	5.0	9.4	5.2	4.9	5.3	4.9
		(30,30)	4.7	4.7	7.1	4.9	4.6	4.9	4.5
Lognormal	CS	(20,20)	3.7	3.7	7.5	4.8	3.5	4.8	3.5
		(30,30)	3.9	3.9	6.6	5.0	3.8	5.0	3.9
	AR	(20,20)	3.5	3.5	7.3	4.5	3.3	4.6	3.4
		(30,30)	3.2	3.2	6.1	4.2	3.2	4.4	3.3
Dist	Cov	**n**	Group×Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	5.6	5.6	9.9	5.7	5.6	5.6	5.5
		(30,30)	4.7	4.7	7.6	4.9	4.7	4.8	4.9
	AR	(20,20)	4.8	4.8	8.8	4.9	4.8	4.8	4.7
		(30,30)	4.9	4.9	7.4	5.1	4.8	5.0	5.0
$t(5)$	CS	(20,20)	4.9	4.9	8.7	5.2	4.8	5.3	4.9
		(30,30)	5.1	5.1	7.5	5.1	4.9	5.3	5.1
	AR	(20,20)	4.8	4.8	8.7	5.2	4.8	5.1	4.9
		(30,30)	5.0	5.0	7.4	5.3	4.9	5.3	5.0
$\chi^2(5)$	CS	(20,20)	4.4	4.4	8.8	4.7	4.3	4.8	4.4
		(30,30)	4.9	4.9	7.5	5.1	4.9	5.3	5.1
	AR	(20,20)	4.6	4.6	8.4	4.9	4.4	4.8	4.7
		(30,30)	4.8	4.8	7.2	4.9	4.8	4.9	4.9
Lognormal	CS	(20,20)	3.5	3.5	7.7	5.0	3.4	4.9	3.6
		(30,30)	3.9	3.9	6.1	4.6	3.8	5.0	4.0
	AR	(20,20)	4.0	4.0	7.8	5.0	3.7	5.0	3.7
		(30,30)	3.7	3.7	6.1	4.8	3.5	4.7	3.7

Table 3.2: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced homoscedastic factorial designs for testing Time and Group×Time effects, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	4.7	4.7	8.4	4.7	4.4	4.6	4.7
		(30,40)	4.8	4.8	7.5	5.1	4.9	4.9	4.8
	AR	(20,30)	5.0	5.0	9.0	5.3	5.2	5.2	5.2
		(30,40)	5.4	5.4	7.8	5.4	5.3	5.5	5.5
$t(5)$	CS	(20,30)	4.4	4.4	8.5	4.5	4.2	4.7	4.5
		(30,40)	4.6	4.6	7.1	4.9	4.6	4.8	4.6
	AR	(20,30)	4.8	4.8	8.6	4.6	4.6	5.0	4.7
		(30,40)	4.9	4.9	7.1	5.1	4.9	5.0	4.9
$\chi^2(5)$	CS	(20,30)	5.0	5.0	8.3	5.0	4.8	4.9	4.9
		(30,40)	5.0	5.0	7.8	5.4	5.0	5.3	5.0
	AR	(20,30)	4.8	4.8	8.8	5.2	4.9	5.2	5.1
		(30,40)	4.7	4.7	7.0	5.0	4.7	5.0	4.8
Lognormal	CS	(20,30)	3.5	3.5	6.8	4.7	3.5	4.6	3.6
		(30,40)	4.2	4.2	6.1	5.1	3.9	4.9	4.2
	AR	(20,30)	3.7	3.7	7.3	4.9	3.5	5.1	3.7
		(30,40)	4.1	4.1	6.5	5.1	4.0	4.9	4.1
Dist	Cov	**n**	Group×Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	4.8	4.8	9.0	4.8	5.0	5.0	5.1
		(30,40)	4.6	4.6	7.2	4.8	4.7	4.7	5.0
	AR	(20,30)	4.8	4.8	8.3	5.0	4.8	4.9	4.9
		(30,40)	5.0	5.0	7.2	5.0	4.8	5.2	5.1
$t(5)$	CS	(20,30)	4.9	4.9	8.4	5.2	5.1	5.3	5.0
		(30,40)	4.8	4.8	7.1	5.1	4.7	4.9	4.8
	AR	(20,30)	5.1	5.1	8.6	5.2	4.9	5.2	4.8
		(30,40)	4.8	4.8	7.2	4.9	4.7	5.0	4.9
$\chi^2(5)$	CS	(20,30)	4.8	4.8	8.5	5.5	5.0	5.3	5.2
		(30,40)	4.8	4.8	7.2	5.1	4.8	5.0	4.8
	AR	(20,30)	5.1	5.1	8.5	5.2	4.9	5.1	5.1
		(30,40)	5.0	5.0	7.5	5.1	4.9	4.8	5.2
Lognormal	CS	(20,30)	3.9	3.9	6.9	4.9	3.5	4.9	3.8
		(30,40)	3.6	3.6	5.4	4.5	3.7	4.4	3.6
	AR	(20,30)	3.9	3.9	7.0	4.9	3.6	5.0	3.6
		(30,40)	3.5	3.5	5.3	4.1	3.4	4.5	3.5

Table 3.3: Type-I error rate ($\times 100$) of DMM, MMM, AT, WBT, PBT, PT and BT for balanced heteroscedastic factorial designs for testing Time and Group×Time effects, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	5.3	5.3	9.5	4.9	5.2	5.6	4.8
		(30,30)	5.2	5.2	7.7	4.9	5.1	5.3	4.9
	AR	(20,20)	5.5	5.5	9.9	4.8	5.2	5.6	4.9
		(30,30)	5.5	5.5	8.1	5.1	5.3	5.6	4.9
$t(5)$	CS	(20,20)	5.7	5.7	9.6	5.4	5.5	6.0	5.2
		(30,30)	5.6	5.6	8.2	5.4	5.4	5.8	5.1
	AR	(20,20)	5.0	5.0	9.8	4.8	5.0	5.5	4.6
		(30,30)	5.5	5.5	8.2	5.3	5.4	5.7	5.1
$\chi^2(5)$	CS	(20,20)	4.8	4.8	8.8	4.5	4.6	5.1	4.4
		(30,30)	5.5	5.5	8.0	5.3	5.4	5.7	5.2
	AR	(20,20)	5.4	5.4	9.9	5.1	5.2	5.7	5.0
		(30,30)	5.3	5.3	7.6	5.2	5.2	5.6	5.0
Lognormal	CS	(20,20)	4.1	4.1	8.2	4.8	3.7	5.5	3.6
		(30,30)	4.2	4.2	6.5	4.7	4.1	5.3	4.0
	AR	(20,20)	4.3	4.3	8.2	5.2	4.2	5.6	3.9
		(30,30)	3.9	3.9	6.5	4.6	3.8	5.0	3.6
Dist	Cov	**n**	Group×Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	5.6	5.6	9.6	5.0	5.5	5.7	5.2
		(30,30)	5.8	5.8	8.4	5.4	5.8	5.9	5.5
	AR	(20,20)	5.9	5.9	10.6	5.3	5.8	6.0	5.3
		(30,30)	5.3	5.3	7.7	4.9	5.2	5.3	5.0
$t(5)$	CS	(20,20)	5.2	5.2	9.5	4.7	5.0	5.4	4.8
		(30,30)	5.4	5.4	7.4	5.4	5.3	5.6	5.1
	AR	(20,20)	5.1	5.1	8.8	4.9	4.9	5.5	4.7
		(30,30)	5.7	5.7	8.4	5.4	5.6	6.1	5.2
$\chi^2(5)$	CS	(20,20)	4.9	4.9	9.5	4.5	4.8	5.3	4.3
		(30,30)	4.6	4.6	7.0	4.2	4.5	4.8	4.1
	AR	(20,20)	6.2	6.2	10.4	5.7	6.1	6.5	5.4
		(30,30)	5.1	5.1	7.8	4.9	5.1	5.4	4.9
Lognormal	CS	(20,20)	4.0	4.0	7.5	4.6	3.7	4.8	3.6
		(30,30)	4.3	4.3	6.6	4.8	4.1	5.1	4.1
	AR	(20,20)	4.1	4.1	8.1	5.0	3.9	5.5	3.7
		(30,30)	4.3	4.3	6.8	5.0	4.2	5.4	4.0

Table 3.4: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced (increasing sizes) heteroscedastic factorial designs for testing Time and Group×Time effects, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	10.2	10.2	10.6	5.4	6.5	6.5	5.6
		(30,40)	8.1	8.1	7.9	4.8	5.4	5.7	4.8
	AR	(20,30)	10.0	10.0	10.4	5.6	6.9	7.0	5.8
		(30,40)	9.2	9.2	9.0	5.4	6.1	6.4	5.4
$t(5)$	CS	(20,30)	9.3	9.3	10.0	5.0	5.7	6.3	4.9
		(30,40)	7.4	7.4	7.5	4.8	5.3	5.4	4.7
	AR	(20,30)	9.4	9.4	10.0	4.9	5.9	6.2	4.7
		(30,40)	8.4	8.4	8.4	5.3	5.6	6.1	5.0
$\chi^2(5)$	CS	(20,30)	9.4	9.4	9.8	5.3	6.2	6.4	5.2
		(30,40)	7.7	7.7	7.7	4.7	5.3	5.4	4.6
	AR	(20,30)	9.8	9.8	10.3	5.1	6.1	6.7	5.2
		(30,40)	8.2	8.2	8.1	5.3	5.7	6.1	5.3
Lognormal	CS	(20,30)	7.9	7.9	8.4	5.1	4.4	6.1	3.7
		(30,40)	6.7	6.7	6.8	5.1	4.3	5.5	4.1
	AR	(20,30)	7.8	7.8	8.3	5.1	4.5	6.1	3.9
		(30,40)	7.6	7.6	7.6	5.8	5.0	6.4	4.6

Dist	Cov	**n**	Group×Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	9.7	9.7	10.5	5.0	6.4	6.6	5.2
		(30,40)	8.7	8.7	8.6	5.5	6.2	6.2	5.5
	AR	(20,30)	9.3	9.3	9.8	5.3	6.4	6.6	5.6
		(30,40)	8.0	8.0	8.1	5.1	5.6	5.9	5.1
$t(5)$	CS	(20,30)	9.9	9.9	10.3	5.4	6.0	6.6	5.1
		(30,40)	8.0	8.0	8.2	5.5	5.8	6.0	5.2
	AR	(20,30)	8.9	8.9	9.2	4.7	5.4	5.8	4.5
		(30,40)	8.4	8.4	8.4	5.2	5.7	6.1	5.1
$\chi^2(5)$	CS	(20,30)	9.7	9.7	10.1	4.9	5.6	6.0	4.8
		(30,40)	8.1	8.1	8.0	5.0	5.3	5.6	4.8
	AR	(20,30)	9.4	9.4	9.7	5.1	5.8	6.1	4.9
		(30,40)	8.4	8.4	8.3	5.0	5.6	6.2	5.2
Lognormal	CS	(20,30)	7.6	7.6	8.6	4.9	4.4	6.0	3.9
		(30,40)	7.1	7.1	7.2	5.4	4.8	5.9	4.3
	AR	(20,30)	7.1	7.1	8.0	4.6	4.2	5.8	3.5
		(30,40)	6.7	6.7	6.8	5.0	4.1	5.4	3.8

Table 3.5: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced (decreasing sizes) heteroscedastic factorial designs for testing Time and Group×Time effects, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(30,20)	2.5	2.5	8.5	5.2	4.8	4.8	5.4
		(40,30)	4.0	4.0	7.9	5.6	5.6	5.7	5.6
	AR	(30,20)	3.1	3.1	8.3	5.4	5.0	5.1	5.3
		(40,30)	3.2	3.2	7.5	5.2	4.8	4.9	5.2
$t(5)$	CS	(30,20)	2.3	2.3	7.7	4.8	4.1	4.6	4.4
		(40,30)	3.0	3.0	7.0	5.0	4.7	5.1	4.9
	AR	(30,20)	2.4	2.4	7.4	4.7	4.0	4.3	4.2
		(40,30)	3.1	3.1	6.9	5.2	4.6	4.9	4.9
$\chi^2(5)$	CS	(30,20)	2.6	2.6	8.1	4.9	4.2	4.4	4.6
		(40,30)	2.8	2.8	6.8	4.8	4.5	4.7	4.8
	AR	(30,20)	2.5	2.5	7.4	4.7	3.9	4.3	4.4
		(40,30)	3.0	3.0	6.8	4.8	4.5	4.7	4.5
Lognormal	CS	(30,20)	2.0	2.0	6.5	4.7	3.3	4.4	3.5
		(40,30)	2.5	2.5	6.3	5.1	3.9	5.1	3.9
	AR	(30,20)	2.2	2.2	7.3	5.1	3.6	5.0	3.8
		(40,30)	2.7	2.7	5.9	5.0	4.0	4.9	4.2
Dist	Cov	**n**	Group×Time						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(30,20)	2.9	2.9	8.1	5.1	4.8	4.9	5.1
		(40,30)	3.6	3.6	7.9	5.6	5.4	5.6	5.6
	AR	(30,20)	2.8	2.8	8.1	4.8	4.4	4.6	4.9
		(40,30)	3.0	3.0	6.9	4.8	4.5	4.7	4.6
$t(5)$	CS	(30,20)	2.5	2.5	8.2	5.1	4.4	4.8	4.8
		(40,30)	2.9	2.9	7.0	4.9	4.6	4.9	4.8
	AR	(30,20)	2.1	2.1	7.1	4.2	3.9	4.1	4.2
		(40,30)	3.3	3.3	7.4	5.3	4.9	5.3	5.2
$\chi^2(5)$	CS	(30,20)	2.8	2.8	8.3	5.0	4.3	4.7	4.6
		(40,30)	3.2	3.2	7.3	5.2	5.0	5.1	5.1
	AR	(30,20)	2.6	2.6	8.2	5.0	4.3	4.5	4.6
		(40,30)	3.3	3.3	7.6	5.5	5.3	5.5	5.5
Lognormal	CS	(30,20)	1.9	1.9	6.4	4.7	3.3	4.5	3.4
		(40,30)	2.5	2.5	5.9	4.9	3.5	4.6	3.7
	AR	(30,20)	1.9	1.9	6.3	4.4	3.1	4.3	3.5
		(40,30)	2.4	2.4	6.0	4.8	3.8	4.8	3.8

However, when data come from lognormal distribution, WBT and PT are almost identical and have larger power than the other four tests, which makes those two tests the best ones in terms of power among all six competing methods.

MMM has larger power than the other five tests, with data from normal, $t(5)$ and $\chi^2(5)$ distributions, and for testing Time and Group×Time effects, as shown in the middle and top rows of Figure 3.1, respectively. However, it is well known that MMM performs well in practice only if multivariate normality and sphericity are satisfied. Further, as the type-I error rate simulations indicate, MMM is highly unstable. It produces extremely liberal test results with positive pairing cases and extremely conservative test results with negative pairing cases. Hence, MMM is not recommended even if it has larger power here in the balanced homoscedastic designs. On the other hand, WBT and PT achieve slightly larger power than the rest three tests (DMM, PBT and BT) with data from normal, $t(5)$ and $\chi^2(5)$ distributions, for testing Time and Group×Time effects.

With lognormal data for testing Group and Group×Time effects, WBT and PT are almost identical and have an clear advantage in power over the other four tests. Nevertheless, with lognormal data for testing Time effect, WBT is the best in power among the six competing methods.

Power versus Sample Size

Next, we investigate how power changes along increasing sample sizes. In order to have a fair comparison of the competing tests, we restrict our simulations to balanced homoscedastic designs for testing Group effect. As shown in Figure 3.1, all six methods, DMM, MMM, WBT, PBT, PT and BT tend to have the same power for different δ values and for testing Group effect. Therefore, it is reasonable to see how power changes along increasing sample sizes under such settings. If there exists any patterns of the power curves, it is most likely due to the increasing sample sizes. The simulation settings are $a = 2$, $p = 4$, $t = 3$ and compound symmetric covariance structure with $\boldsymbol{\rho} = (0.2, 0.2)$. The sample sizes are $\boldsymbol{n}^{(1)} = (15, 15)$, $\boldsymbol{n}^{(2)} = (20, 20)$, $\boldsymbol{n}^{(3)} = (25, 25)$, $\boldsymbol{n}^{(4)} = (30, 30)$ and $\boldsymbol{n}^{(5)} = (35, 35)$.

Figure 3.2 shows the power curves of all seven methods along increasing sample sizes. The bottom, middle and top rows are for power curves at δ being 0, 0.2 and 0.4, respectively. In fact, the bottom row shows the empirical type-I error rates at $\delta = 0$ of all seven methods with data from normal, $t(5)$, $\chi^2(5)$ and lognormal distributions. It is obvious that AT is extremely liberal producing type-I error rate larger than 10% for $\boldsymbol{n}^{(1)} = (15, 15)$. As we mentioned above, AT is not recommended for use due to its extreme liberality. In the following part, we only compare the remaining six methods. When data are from normal, $t(5)$ and $\chi^2(5)$ distributions, all six methods, DMM, MMM, WBT, PBT, PT and BT, are almost exact with type-I error rates along the $\alpha = 0.05$ horizontal line. When data are from lognormal distribution, WBT and PT are still almost exact, whereas the other four test, DMM, MMM, PBT and BT are very conservative.

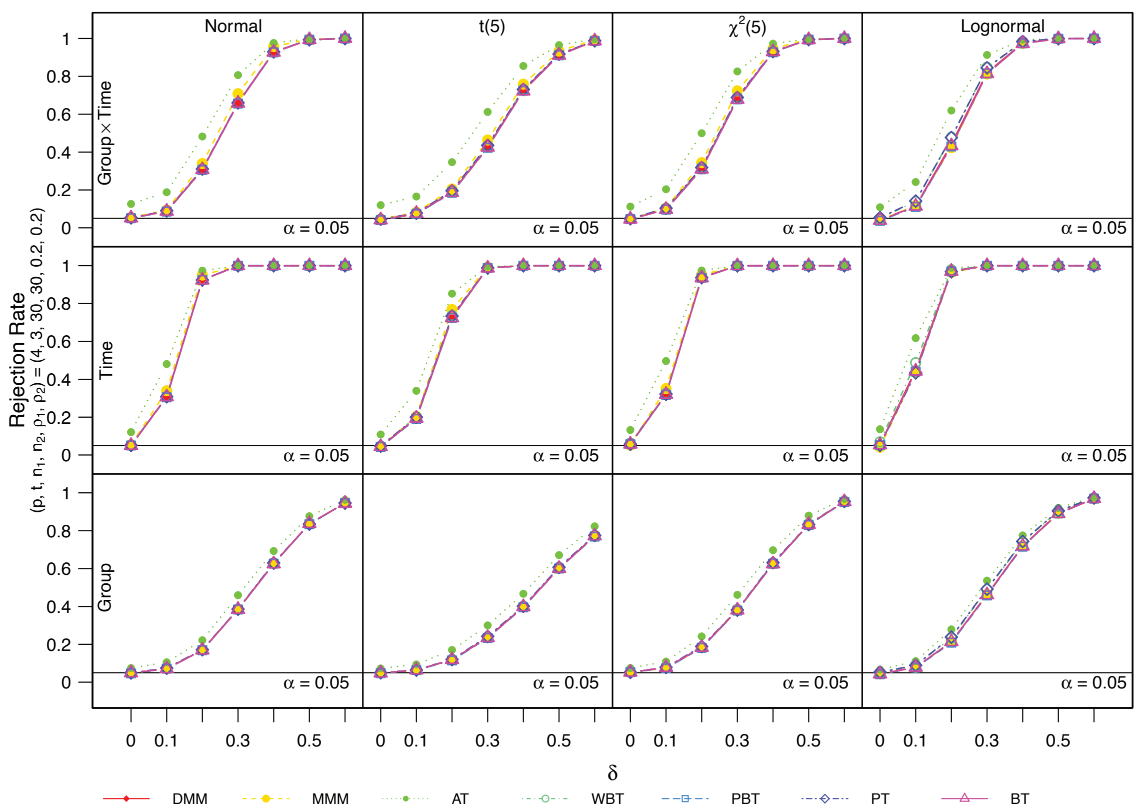
In the middle row for $\delta = 0.2$ and the top row for $\delta = 0.4$, WBT and PT have power slightly larger than that of the remaining four tests with data from normal, $t(5)$ and $\chi^2(5)$ distributions. However, the power superiority of WBT and PT are more obvious when data come from highly skewed lognormal distribution.

With both type-I error rate and power considered, the simulations above indicate an absolute advantage of WBT and PT especially when data are from highly skewed lognormal distribution. Further, PBT and BT perform well in terms of both type-I error rate and power when data are from normal, $t(5)$ and $\chi^2(5)$ distributions. All four tests, WBT, PT, PBT and BT, have very good performance under both homoscedastic and heteroscedastic designs with both balanced and unbalanced sample sizes. However, DMM and MMM perform well only under homoscedastic and balanced heteroscedastic designs.

3.5 Application

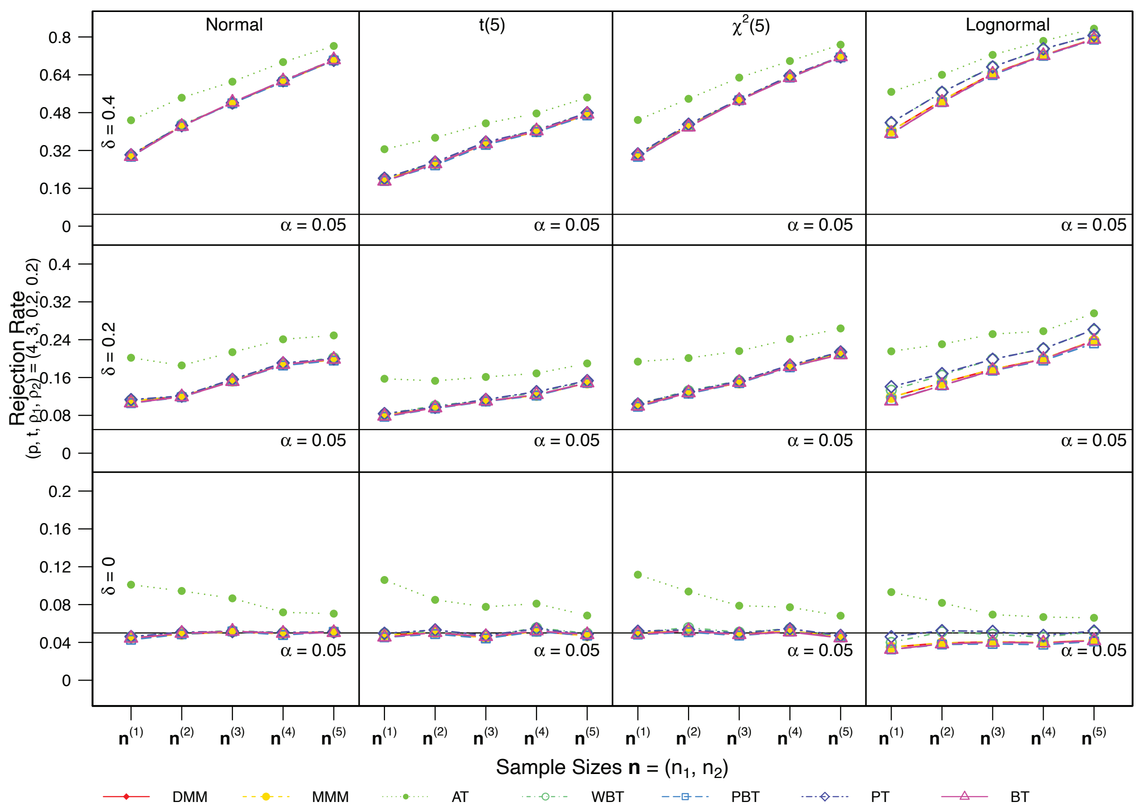
To illustrate how our methodology can be applied and to stimulate readers' interest, we introduce an example where multivariate growth curve data on idiopathic infantile nystagmus syndrome (INS) need to be analyzed in the context of general factorial design. In this study (Fadardi et al., 2017), 15 voluntary participants with idiopathic

Figure 3.1: Power curves of detecting fixed alternative of Group, Time and Group×Time effects in factorial designs on the bottom, middle and top rows, respectively, with $\alpha = 0.05$, $a = 2$, $p = 4$, $t = 3$, $\boldsymbol{n} = (30, 30)$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$. The columns are for different distributions.



INS were recruited from a referring ophthalmologist. Participants were asked to carry out acuity tasks identifying the direction of horizontal Tumbling-E targets under different mental load settings. For the low mental load setting, participants were given unlimited time to respond. After responding, they were required to view a fixation cross for 100 milliseconds prior to the presence of the next acuity target. For the high mental load setting, participants were given only 0.5 second to view the target and then 300 milliseconds to view a visual noise mask. Participants were required to respond while they were viewing a fixation cross for 1 second. In addition, participants were also asked to conduct mental arithmetic (continuously subtracting 7 from a number randomly selected between 100 and 120 and given by the examiner during

Figure 3.2: Power curves of detecting fixed alternative of Group effect with $\alpha = 0.05$, $a = 2$, $p = 4$, $t = 3$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$. The effect size δ is set to be 0, 0.2 and 0.4 on the bottom, middle and top rows, respectively. The columns are for different distributions. The sample sizes are $\boldsymbol{n}^{(1)} = (15, 15)$, $\boldsymbol{n}^{(2)} = (20, 20)$, $\boldsymbol{n}^{(3)} = (25, 25)$, $\boldsymbol{n}^{(4)} = (30, 30)$ and $\boldsymbol{n}^{(5)} = (35, 35)$.



the task) simultaneously with the acuity task. Both the low and the high mental load effects were evaluated at two gaze positions (null position and away position). Eventually, the size and contrast of the target at which participants' task performance plateaued were recorded. The main objective of the study is to investigate whether there is any main effect of mental load (M), main effect of gaze position (P), and interaction effect between the mental load and gaze position (MP).

Among all 15 participants with idiopathic infantile nystagmus syndrome, 11 of them finished the task with no missing data. To test the interaction effect mentioned above in the context of our method, we need the key parameters. Since all participants have the disease, there is only one group ($a = 1$). There are two response variables measured each time, target size and contrast ($p = 2$). There are four repeated measures ($t = 4$) representing four different occasions, low mental load at null position, low mental load at away position, high mental load at null position, and high mental load at away position. The corresponding contrast matrix for testing the interaction effect between the mental load and gaze position is $\boldsymbol{H}_{\text{MP}} = (1, -1, -1, 1) \otimes \boldsymbol{I}_p$. The contrast matrices for testing the main effects of mental load and gaze position are $\boldsymbol{H}_{\text{M}} = (1, 1, -1, -1) \otimes \boldsymbol{I}_p$ and $\boldsymbol{H}_{\text{P}} = (1, -1, 1, -1) \otimes \boldsymbol{I}_p$, respectively.

As shown in Table 3.6, the p-values of AT are smaller than $\alpha = 0.05$, leading to significant results. However, due to the small sample size of 11, test results of the asymptotic test are not trustworthy. The test results of DMM and MMM are the same. This is because when $a = 1$, DMM and MMM are theoretically identical. Except for AT, the p-values of the other six methods (DMM, MMM, WBT, PBT, PT and BT) for testing the main effect of gaze position and the interaction effect of mental$\times$position are larger than $\alpha = 0.05$, indicating that there are no such effects. Moreover, these test results agree with the findings of Fadardi et al. (2017) who conducted univariate analysis and adjusted for multiplicity in lieu of a multivariate analysis.

It is worth mentioning that, except for AT, the remaining six methods lead to different decisions for testing the main effect of mental load. The p-values of WBT, PBT and BT show that we fail to reject the null hypothesis whereas the p-values of DMM, MMM and PT reveal that the null hypothesis should be rejected. Methods based on bootstrap, i.e., WBT, PBT and BT, rely on the assumption of positive definite covariance matrix in data. Given the fact that the covariance matrix of the optometry data is almost singular, these three tests are not trustworthy. Also, it is well known that both DMM and MMM assume multivariate normality in order to have valid test results. However, the marginal distributions of the optometry data are highly skewed, which makes it less ideal to use those two methods. Further, although the optometry data have almost singular covariance matrix, by permuting the p-variate observations, the covariance of the permuted data is no longer singular.

The permutation procedure corrects the covariance matrix and makes it more ideal for the permutation version of the Wald-type statistic. Hence, the most reasonable and reliable method is PT, which means that there is a significant main effect of mental load.

To support the claims above, we further conduct sensitivity analysis. It can be seen from Figure 3.3 that there is a clear difference between high (red) and low (blue) mental load. Based on (3.11), we simulate data using the empirical mean and covariance matrix of the optometry data with error terms from normal, $t(5)$, $\chi^2(5)$ and lognormal distributions. The empirical power of PT is always the largest for detecting the mental load effect, whereas WBT and BT have zero power and PBT has very small power. It further confirms that PT can detect the difference between low and high mental load even though the covariance matrix is almost singular. This is also the reason why the p-values for the three bootstrap based methods (WBT, PBT and BT) are generally larger than that of PT.

Of the 11 participants, 3 are female and 8 are male. It might also be interesting to investigate whether there are other interaction effects, for example, gender × mental (GM), gender × position (GP), and gender × mental × position (GMP). In this case, we have two groups ($a = 2$), i.e., female group and male group. Accordingly, the contrast matrices are $\boldsymbol{H}_{\text{GM}} = \boldsymbol{P}_a \otimes \boldsymbol{P}_2 \otimes \frac{1}{2}\mathbf{1}_2^\top \otimes \boldsymbol{I}_p$, $\boldsymbol{H}_{\text{GP}} = \boldsymbol{P}_a \otimes \frac{1}{2}\mathbf{1}_2^\top \otimes \boldsymbol{P}_2 \otimes \boldsymbol{I}_p$, and $\boldsymbol{H}_{\text{GMP}} = \boldsymbol{P}_a \otimes \boldsymbol{P}_2 \otimes \boldsymbol{P}_2 \otimes \boldsymbol{I}_p$.

Test results are shown in Table 3.7. The p-values of all methods for testing all three interaction effects are greater than 0.05, leading to insignificant test results. It means that there are no gender×mental, gender×position and gender×mental×position effects.

Table 3.6: Analysis of the idiopathic infantile nystagmus syndrome (INS) data using resampling-based methods.

Effect	DMM	MMM	AT	WBT	PBT	PT	BT
Mental Load	0.024	0.024	0.002	0.582	0.130	0.025	0.636
Gaze Position	0.113	0.113	0.044	0.622	0.352	0.105	0.616
Mental×Position	0.086	0.086	0.027	0.589	0.305	0.109	0.650

Figure 3.3: Plot of optometry data of the 11 subjects with two variables, size and contrast. Subjects were measured under both low and high mental loads with null and away gaze positions.

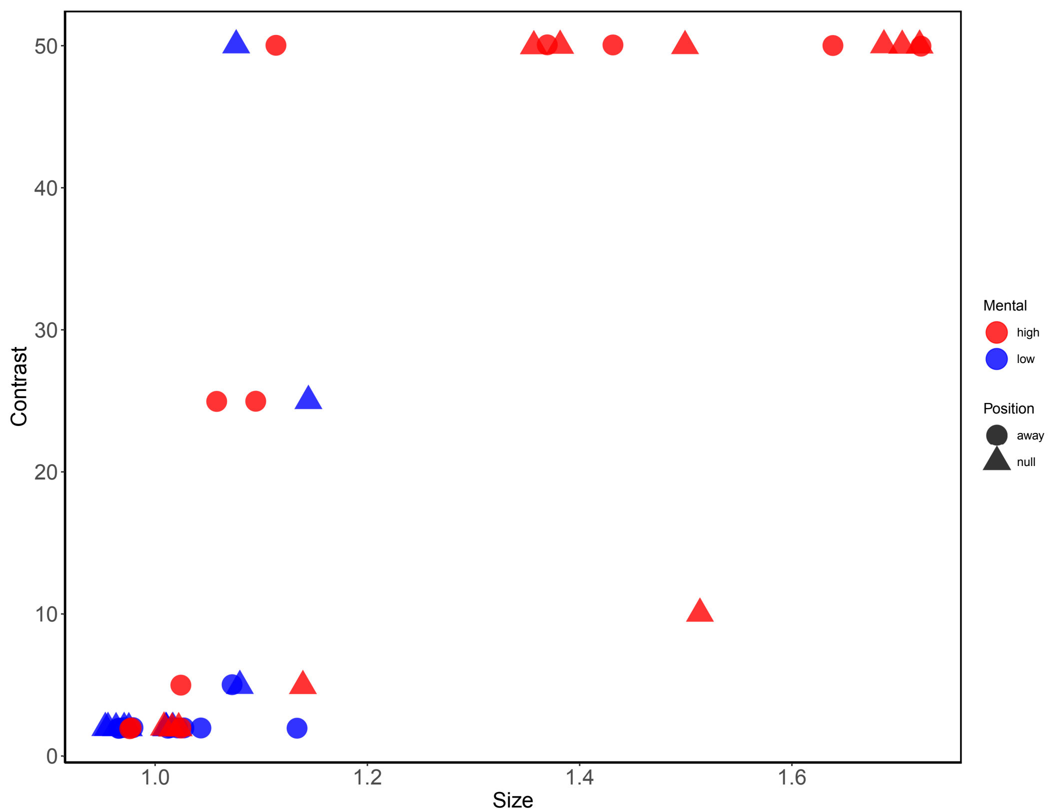


Table 3.7: Analysis of the idiopathic infantile nystagmus syndrome (INS) data for two genders using resampling-based methods.

Effect	DMM	MMM	AT	WBT	PBT	PT	BT
Gender×Mental	0.410	0.410	0.136	0.692	0.586	0.325	0.693
Gender×Position	0.738	0.738	0.794	0.800	0.930	0.854	0.903
Gender×Mental×Position	0.214	0.214	0.333	0.686	0.727	0.481	0.787

3.6 Discussion and Conclusion

Multivariate growth curve data, which are essentially repeated measures of multivariate data, have become more and more common in various areas. There are some effective methods developed for analyzing such repeated measures data, but with the assumptions of multivariate normality and homoscedasticity. However, these assumptions are difficult to justify and attain in reality.

In this chapter, we generalized the idea of permutation in Pauly et al. (2015);

Friedrich et al. (2017) from independent univariate repeated measures factorial designs to multivariate repeated measures factorial designs and we also, in another sense, generalized the concept of bootstrap in Konietschke et al. (2015) from multivariate factorial designs to multivariate repeated measures factorial designs. Due to the use of WTS, resampling tests, i.e., PT, BT and WBT are asymptotically exact and valid without the assumptions of multivariate normality and homoscedasticity. We rigorously proved in Theorem 3.3.2 that the studentized permutation distribution of the WTS always approximates the null distribution of WTS. Moreover, we proved in Theorems 3.3.3 and 3.3.4 that the parametric bootstrap and wild bootstrap distributions of WTS always approximate the null distribution of WTS as well.

In practical applications, WBT and PT are highly recommended due to their favorable performance in terms of controlling type-I error rate and having power advantage over the other competing tests. The data analyses in Section 3.5 further demonstrate that PT is far more superior when multivariate data have almost singular covariance matrix. Moreover, PBT and BT are mostly good tests when data are not highly skewed. On the other hand, AT cannot be used in small samples and DMM and MMM are not recommended for unbalanced heteroscedastic designs.

Resampling procedures, including permutation and bootstrap, are generally computationally expensive. However, the proposed resampling-based tests have high accuracy and reasonable efficiency. For example, on a MacBook Pro Intel Core i7 processor with 2.2 GHz speed and 16GB RAM, the processing time to run all seven methods (DMM, MMM, AT, WBT, PBT, PT and BT) in a factorial design (with $\chi^2(5)$ data, $a = 2$, $p = 4$, $t = 2$, sample sizes $\boldsymbol{n} = (30, 30)$, compound symmetric covariance structure with $\boldsymbol{\rho} = (0.2, 0.2)$ for testing Time effect) is only about 2.6 seconds.

The current chapter proposed Wald-type resampling-based methods without assuming multivariate normality or homoscedasticity. For the permutation test, studentization is used to correct the covariance of the permuted data leading to an asymptotically exact test despite the time dependencies. Further, the parametric and wild bootstrap tests are also shown to be asymptotically exact in the nonparametric settings. Although the resampling-based tests have satisfactory performance, they are only appropriate for metric data. We plan to investigate nonparametric methods that can accommodate both metric and nonmetric data.

3.7 Appendix

Proof of Theorem 3.3.1

Proof. Within each group, i.e., for fixed i, $\boldsymbol{X}_{ik}, k = 1, \ldots, n_i$ are independent and identically distributed random vectors. By Multivariate Central Limit Theorem,

$$\sqrt{n_i}(\overline{\boldsymbol{X}}_{i\cdot} - \boldsymbol{\mu}_i) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_i).$$

Under the asymptotic framework A_1, we have the following convergence for $\overline{\boldsymbol{X}}$,

$$\sqrt{N}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \text{diag}(\frac{N}{n_i}\boldsymbol{\Sigma}_i : 1 \leq i \leq a)$. Hence, under the null hypothesis $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \mathbf{0}$,

$$\sqrt{N}\boldsymbol{T}\overline{\boldsymbol{X}} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}\right).$$

By Theorem 9.2.2 in Rao and Mitra (1971, p. 173), the quadratic form $\tilde{Q}_N(\boldsymbol{T}) = N\overline{\boldsymbol{X}}^\top \boldsymbol{T}\left(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}\right)^+ \boldsymbol{T}\overline{\boldsymbol{X}}$ has, asymptotically, $\chi^2_{f_{\boldsymbol{T}}}$, where $f_{\boldsymbol{T}} = \text{rank}(\boldsymbol{T})$. It is easy to see that $\widehat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$. By Continuous Mapping Theorem, $\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T} \xrightarrow{p} \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}$. Further, by Theorem 4.2 in Rakočević (1997), the continuity of Moore-Penrose inverse holds. Therefore, by the Continuous Mapping Theorem again, $(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T})^+ \xrightarrow{p} (\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T})^+$. Replacing $(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T})^+$ with $(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T})^+$, we have $Q_N(\boldsymbol{T}) - \tilde{Q}_N(\boldsymbol{T}) \xrightarrow{p} 0$. Finally, by the Multivariate Slutsky's Theorem $Q_N(\boldsymbol{T})$ has, asymptotically, $\chi^2_{f_{\boldsymbol{T}}}$ distribution with $f_{\boldsymbol{T}} = \text{rank}(\boldsymbol{T})$. □

Proof of Theorem 3.3.2

The result follows by applying Propositions 3.7.1–3.7.3 together with the Continuous Mapping Theorem. For convenience, we define

$$\boldsymbol{Z} = \left(\boldsymbol{Z}_{N,1}^\top, \boldsymbol{Z}_{N,2}^\top, \ldots, \boldsymbol{Z}_{N,\tilde{N}}^\top\right)^\top = \left(\boldsymbol{X}_{111}^\top, \boldsymbol{X}_{112}^\top, \ldots, \boldsymbol{X}_{11n_1}^\top, \boldsymbol{X}_{121}^\top, \ldots, \boldsymbol{X}_{atn_a}^\top\right)^\top,$$

where $\boldsymbol{Z}_{N,l}$ is a $p\times 1$ vector with $l = 1,\ldots,\tilde{N} = Nt$, for the pooled sample. The average of the $p\times 1$ vectors is denoted by $\overline{\boldsymbol{Z}}_{N,\cdot} = 1/\tilde{N}\sum_{l=1}^{\tilde{N}} \boldsymbol{Z}_{N,l} = 1/\tilde{N}\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i} \boldsymbol{X}_{ijk} = \overline{\boldsymbol{X}}_{\ldots}$.

Proposition 3.7.1. *Under the assumptions of Theorem 3.3.2 and under the null hypothesis* $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, $\sqrt{\tilde{N}}(\overline{\boldsymbol{X}}^{\pi} - \boldsymbol{1}_a \otimes \boldsymbol{1}_t \otimes \overline{\boldsymbol{X}}_{\ldots})$ *given the observed data* $\boldsymbol{X}$ *weakly converges to a multivariate normal distribution with mean vector* $\boldsymbol{0}$ *and covariance matrix* $\mathrm{diag}(\kappa_i^{-1},\ldots,\kappa_a^{-1}) \otimes \boldsymbol{I}_t \otimes \boldsymbol{\Gamma}$.

Proof. Let $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top,\ldots,\boldsymbol{\lambda}_{at}^\top)^\top$ and $\boldsymbol{\lambda}_i$ be a p-dimesional vector. For any $\boldsymbol{\lambda} \in \mathbb{R}^{apt}$,

$$\sqrt{\tilde{N}}\boldsymbol{\lambda}^\top(\overline{\boldsymbol{X}}^{\pi} - \boldsymbol{1}_a \otimes \boldsymbol{1}_t \otimes \overline{\boldsymbol{X}}_{\ldots}) = \sqrt{N}\sum_{s=1}^{\tilde{N}} c_{N,s}\boldsymbol{\lambda}_s^{*\top}(\boldsymbol{Z}_s - \overline{\boldsymbol{Z}}_{N,\cdot}),$$

where

$$c_{N,s} = \sum_{i=1}^{T} \boldsymbol{1}\{M_{i-1}+1 \le s \le M_i\}\frac{\boldsymbol{1}}{\sqrt{\tilde{N}}} \text{ and } \boldsymbol{\lambda}_s^{*} = \sum_{i=1}^{T} \boldsymbol{1}\{M_{i-1}+1 \le s \le M_i\}\frac{\boldsymbol{\lambda}_i}{\sqrt{\tilde{N}}}.$$

For a vector analog of Theorem 4.1 in Pauly et al. (2011), we check

$$\max_{1\le s\le \tilde{N}} |c_{N,s} - \overline{c}| \xrightarrow{p} 0, \tag{3.12}$$

$$\sum_{s=1}^{\tilde{N}}(c_{N,s} - \overline{c})^2 \xrightarrow{p} \sigma^2 = \sum_{i=1}^{T} b_i^{-1} - T^2, \tag{3.13}$$

$$\sqrt{\tilde{N}}(c_{N,\pi(1)} - \overline{c}) \xrightarrow{d} W \text{ where } E(W) = 0 \text{ and } \mathrm{Var}(W) = \sigma^2, \tag{3.14}$$

$$\frac{1}{\sqrt{\tilde{N}}}\max_{1\le s\le \tilde{N}} |\boldsymbol{\lambda}_s^{*\top}(\boldsymbol{Z}_{N,s} - \overline{\boldsymbol{Z}}_{N,\cdot})| \xrightarrow{p} 0 \tag{3.15}$$

and

$$\frac{1}{\tilde{N}}\sum_{s=1}^{\tilde{N}} \boldsymbol{\lambda}_s^{*\top}(\boldsymbol{Z}_{N,s} - \overline{\boldsymbol{Z}}_{N,\cdot})(\boldsymbol{Z}_{N,s} - \overline{\boldsymbol{Z}}_{N,\cdot})^\top\boldsymbol{\lambda}_s^{*} \xrightarrow{p} \boldsymbol{\lambda}^\top\boldsymbol{\Gamma}\boldsymbol{\lambda}, \tag{3.16}$$

where $b_i = \lim_{\min(n_i)\to\infty}\tilde{N}/n_i$ for $i = 1,\ldots,a$ due to the assumption A_1 in (3.3), $T = at$, $1 < t < \infty$, $1 < p < \infty$ and

$$\boldsymbol{\Gamma} = \sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\left(\boldsymbol{\Sigma}_{ij} + \boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^\top\right) - \left(\sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}\right)\left(\sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}\right)^\top.$$

The validity of (3.12)–(3.15) follows in steps similar to Friedrich et al. (2017) and (3.16) follows by using Proposition 3.7.2. Therefore, given the data $\boldsymbol{X}$,

$$\sqrt{\tilde{N}}\boldsymbol{\lambda}^\top(\overline{\boldsymbol{X}}^\pi - \mathbf{1}_a \otimes \mathbf{1}_t \otimes \overline{\boldsymbol{X}}_{\dots}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\lambda}^\top(\mathrm{diag}(\kappa_i^{-1}, \dots, \kappa_a^{-1}) \otimes \boldsymbol{I}_t \otimes \boldsymbol{\Gamma})\boldsymbol{\lambda}).$$

Finally the desired result follows by subsequential arguments. For the technical details see Pauly et al. (2011) and Friedrich et al. (2017). □

Proposition 3.7.2. *Assume the model* (3.1) *and* (3.2)*. Under the assumptions A_1 and A_2 in* (3.3) *and* (3.4)*, respectively,*

$$\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)^\top \xrightarrow{p} \boldsymbol{\Gamma}.$$

Proof. Notice that

$$\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)^\top = \frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\boldsymbol{Z}_{N,l}^\top - \overline{\boldsymbol{Z}}_{N,\cdot}\overline{\boldsymbol{Z}}_{N,\cdot}^\top$$

Now

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\boldsymbol{Z}_{N,l}^\top\right) &= \frac{1}{\tilde{N}}\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i}\mathrm{E}\left(\boldsymbol{X}_{ijk}\boldsymbol{X}_{ijk}^\top\right) \\
&= \sum_{i=1}^{a}\frac{n_i}{\tilde{N}}\sum_{j=1}^{t}\left(\boldsymbol{\Sigma}_{ij} + \boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^\top\right) \\
&\longrightarrow \sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\left(\boldsymbol{\Sigma}_{ij} + \boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^\top\right),
\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{E}\left(\overline{\boldsymbol{Z}}_{N,\cdot}\overline{\boldsymbol{Z}}_{N,\cdot}^{\top}\right) &= \mathrm{E}\left[\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right)\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}^{\top}\right)\right] \\
&= \mathrm{Cov}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right) + \mathrm{E}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right)\mathrm{E}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}^{\top}\right) \\
&= \mathrm{Cov}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right) + \mathrm{E}\left(\frac{1}{\tilde{N}}\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}\right)\mathrm{E}\left(\frac{1}{\tilde{N}}\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}^{\top}\right) \\
&= \mathrm{Cov}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right) + \left(\sum_{i=1}^{a}\frac{n_i}{\tilde{N}}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}\right)\left(\sum_{i=1}^{a}\frac{n_i}{\tilde{N}}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}^{\top}\right) \\
&\longrightarrow \left(\sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}\right)\left(\sum_{i=1}^{a}\frac{1}{b_i}\sum_{j=1}^{t}\boldsymbol{\mu}_{ij}^{\top}\right).
\end{aligned}
$$

Note that the elements of the covariance matrix $\mathrm{Cov}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\right)$ are $\mathcal{O}\left(\frac{1}{N}\right)$ due to independence and the assumption A_2 in (3.4).

Next, we investigate the covariance matrix. Note that $\overline{\boldsymbol{X}}_{\dots} = \frac{1}{\tilde{N}}\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}$. By independence and the assumption A_2 in (3.4), we have

$$
\begin{aligned}
&\mathrm{Cov}\left[\mathrm{Vec}\left(\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\boldsymbol{Z}_{N,l}\boldsymbol{Z}_{N,l}^{\top} - \overline{\boldsymbol{Z}}_{N,\cdot}\overline{\boldsymbol{Z}}_{N,\cdot}^{\top}\right)\right] \\
&= \frac{1}{\tilde{N}^2}\mathrm{Cov}\left[\mathrm{Vec}\left(\sum_{i=1}^{a}\sum_{j=1}^{t}\sum_{k=1}^{n_i}\left(\boldsymbol{X}_{ijk}\boldsymbol{X}_{ijk}^{\top} - \overline{\boldsymbol{X}}_{\dots}\overline{\boldsymbol{X}}_{\dots}^{\top}\right)\right)\right] \\
&= \frac{1}{\tilde{N}^2}\sum_{i=1}^{a}\sum_{k=1}^{n_i}\mathrm{Cov}\left[\sum_{j=1}^{t}\mathrm{Vec}\left(\boldsymbol{X}_{ijk}\boldsymbol{X}_{ijk}^{\top} - \overline{\boldsymbol{X}}_{\dots}\overline{\boldsymbol{X}}_{\dots}^{\top}\right)\right] \\
&\longrightarrow \boldsymbol{0}_{p^2\times p^2}.
\end{aligned}
$$

Note that the elements in the covariance matrix are $\mathcal{O}\left(\frac{1}{N}\right)$ and, hence, are zeros as $N\to\infty$. Therefore, by Chebyshev's inequality,

$$
\frac{1}{\tilde{N}}\sum_{l=1}^{\tilde{N}}\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)\left(\boldsymbol{Z}_{N,l} - \overline{\boldsymbol{Z}}_{N,\cdot}\right)^{\top} \xrightarrow{p} \boldsymbol{\Gamma}.
$$

$\square$

Proposition 3.7.3. *Assume the model* (3.1) *and* (3.2). *Under the assumptions* A_1 *and* A_2 *in* (3.3) *and* (3.4), *respectively,*

$$\widehat{\boldsymbol{\Sigma}}^{\pi} = \bigoplus_{i=1}^{a} \frac{N}{n_i} \widehat{\boldsymbol{\Sigma}}_i^{\pi} \xrightarrow{p} \operatorname{diag}\left(\kappa_i^{-1}, \ldots, \kappa_a^{-1}\right) \otimes \boldsymbol{I}_t \otimes \boldsymbol{\Gamma},$$

where $\widehat{\boldsymbol{\Sigma}}_i^{\pi} \xrightarrow{p} \boldsymbol{I}_t \otimes \boldsymbol{\Gamma}$.

Proof. Let $r, s \in \{1, \ldots, t\}$ be indices for time point and $u, v \in \{1, \ldots, p\}$ be indices for variable. An element of $\widehat{\boldsymbol{\Sigma}}_i^{\pi}$ is denoted as $(\widehat{\boldsymbol{\Sigma}}_i^{\pi})_{ru,sv}$. To complete the proof, it suffices to show that

$$\frac{n_i - 1}{n_i} \left(\widehat{\boldsymbol{\Sigma}}_i^{\pi}\right)_{ru,sv} \xrightarrow{p} \begin{cases} 0 & r \neq s \\ \sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \left(\sigma_{ij}^{(u,v)} + \mu_{ij}^{(u)} \mu_{ij}^{(v)}\right) - & \\ \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(u)}\right) \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(v)}\right) & r = s \end{cases}.$$

Notice that

$$\frac{n_i - 1}{n_i} \left(\widehat{\boldsymbol{\Sigma}}_i^{\pi}\right)_{ru,sv} = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{irk}^{(u)\pi} X_{isk}^{(v)\pi} - \bar{X}_{ir\cdot}^{(u)\pi} \bar{X}_{is\cdot}^{(v)\pi}.$$

Then

$$\frac{1}{n_i} \sum_{k=1}^{n_i} X_{irk}^{(u)\pi} X_{isk}^{(v)\pi} \xrightarrow{p} \begin{cases} \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(u)}\right) \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(v)}\right) & r \neq s \\ \sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \left(\sigma_{ij}^{(u,v)} + \mu_{ij}^{(u)} \mu_{ij}^{(v)}\right) & r = s \end{cases}.$$

Further,

$$\bar{X}_{ir\cdot}^{(u)\pi} \bar{X}_{is\cdot}^{(v)\pi} \xrightarrow{p} \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(u)}\right) \left(\sum_{i=1}^{a} \frac{1}{b_i} \sum_{j=1}^{t} \mu_{ij}^{(v)}\right).$$

The proofs of these claims are analogous to Friedrich et al. (2017). □

Proof of Theorem 3.3.3

Proof. Following the same idea of Konietschke et al. (2015), it is easy to see that $\widehat{\boldsymbol{\Sigma}}_i$ converges almost surely to $\boldsymbol{\Sigma}_i$ for $i = 1, \ldots, a$. Also, by Multivariate Lyapunov's Central Limit Theorem, WLLN for triangular array and Multivariate Slutsky Theorem, the asymptotic multivariate normality can be shown. The remainder of the proof is analogous to the proof of Theorem 3.3.1.

□

Proof of Theorem 3.3.4

Proof. For fixed i and k, $\mathrm{E}(s_{ik}^{\dagger}) = 0$ and $\mathrm{Var}(s_{ik}^{\dagger}) = 1$. It is easy to see that $\mathrm{E}(\boldsymbol{X}_{ik}^{\dagger}) = \mathrm{E}(\overline{\boldsymbol{X}}_{i\cdot}) = \boldsymbol{\mu}_i$ and $\mathrm{Cov}(\boldsymbol{X}_{ik}^{\dagger}|\boldsymbol{X}) = \left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}\right)\left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}\right)^{\top}$. Further,

$$\mathrm{Cov}\left(\sum_{k=1}^{n_i} \boldsymbol{X}_{ik}^{\dagger}|\boldsymbol{X}\right) = \sum_{k=1}^{n_i} \left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}\right)\left(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}\right)^{\top} = (n_i - 1)\widehat{\boldsymbol{\Sigma}}_i.$$

The remainder of the proof is analogous to the proof of Theorem 3.3.3. □

Tables for Testing Group Effect

Table 3.8: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for balanced homoscedastic factorial designs for testing Group effect, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Group						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	4.5	4.5	9.0	4.6	4.4	4.7	4.6
		(30,30)	4.3	4.3	6.7	4.4	4.2	4.4	4.4
	AR	(20,20)	5.2	5.2	9.6	5.4	5.2	5.3	5.3
		(30,30)	5.2	5.2	7.3	5.0	5.1	5.1	5.2
$t(5)$	CS	(20,20)	4.6	4.6	8.3	4.8	4.5	4.7	4.4
		(30,30)	4.8	4.8	7.7	5.1	4.6	5.1	4.9
	AR	(20,20)	4.3	4.3	8.6	4.5	4.3	4.6	4.3
		(30,30)	5.3	5.3	7.9	5.5	5.3	5.6	5.4
$\chi^2(5)$	CS	(20,20)	4.9	4.9	8.6	5.2	4.8	5.2	4.9
		(30,30)	4.8	4.8	7.4	5.0	4.8	5.1	4.9
	AR	(20,20)	4.9	4.9	9.5	5.0	4.7	5.1	4.7
		(30,30)	4.9	4.9	7.6	5.1	4.9	5.2	5.0
Lognormal	CS	(20,20)	4.1	4.1	8.1	5.2	3.7	5.4	3.9
		(30,30)	4.2	4.2	6.5	5.0	4.0	5.3	3.9
	AR	(20,20)	3.4	3.4	7.6	4.2	3.2	4.8	3.2
		(30,30)	4.3	4.3	6.8	5.2	4.1	5.4	4.1

Table 3.9: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced homoscedastic factorial designs for testing Group effect, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Group						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	5.9	5.9	9.5	6.2	5.8	6.1	5.7
		(30,40)	5.5	5.5	7.8	5.6	5.5	5.7	5.6
	AR	(20,30)	5.1	5.1	9.0	5.2	5.3	5.3	5.3
		(30,40)	4.9	4.9	7.2	5.0	4.8	5.0	5.0
$t(5)$	CS	(20,30)	4.9	4.9	8.4	5.0	4.6	4.8	4.7
		(30,40)	5.0	5.0	7.3	5.3	5.1	5.4	5.3
	AR	(20,30)	4.7	4.7	8.5	5.0	4.9	5.0	4.8
		(30,40)	4.9	4.9	7.5	5.2	4.9	5.2	5.1
$\chi^2(5)$	CS	(20,30)	4.8	4.8	8.6	5.0	4.9	5.0	5.0
		(30,40)	4.8	4.8	7.2	5.3	5.0	5.1	4.9
	AR	(20,30)	4.8	4.8	8.7	5.0	4.7	5.0	4.9
		(30,40)	4.6	4.6	6.8	4.6	4.4	4.8	4.7
Lognormal	CS	(20,30)	4.5	4.5	8.3	5.7	4.5	5.4	4.6
		(30,40)	3.7	3.7	6.0	4.5	3.8	4.7	3.9
	AR	(20,30)	4.2	4.2	8.4	6.0	4.5	5.7	4.7
		(30,40)	4.2	4.2	6.3	5.0	4.0	5.0	4.2

Table 3.10: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for balanced heteroscedastic factorial designs for testing Group effect, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,20)	4.9	4.9	8.9	4.8	4.9	5.1	4.8
		(30,30)	4.7	4.7	7.2	4.6	4.7	5.0	4.6
	AR	(20,20)	4.9	4.9	9.3	4.6	4.9	5.0	4.6
		(30,30)	4.6	4.6	7.6	4.5	4.6	4.7	4.5
$t(5)$	CS	(20,20)	5.5	5.5	9.5	5.6	5.4	5.9	5.3
		(30,30)	5.5	5.5	7.8	5.3	5.5	5.8	5.2
	AR	(20,20)	5.1	5.1	9.7	5.2	5.0	5.3	4.8
		(30,30)	4.8	4.8	7.7	5.0	4.8	5.3	4.7
$\chi^2(5)$	CS	(20,20)	5.8	5.8	9.9	5.7	5.8	6.0	5.4
		(30,30)	5.1	5.1	7.4	5.2	5.0	5.1	5.0
	AR	(20,20)	4.8	4.8	9.4	4.6	4.6	5.1	4.6
		(30,30)	5.4	5.4	7.8	5.3	5.4	5.6	5.3
Lognormal	CS	(20,20)	4.1	4.1	7.8	4.9	3.8	5.3	3.8
		(30,30)	4.7	4.7	7.1	5.5	4.5	5.6	4.5
	AR	(20,20)	4.5	4.5	9.3	5.4	4.1	5.7	4.0
		(30,30)	3.6	3.6	6.4	4.7	3.5	4.7	3.7

Table 3.11: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced (increasing sizes) heteroscedastic factorial designs for testing Group effect, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(20,30)	7.5	7.5	8.9	5.1	5.6	5.9	5.0
		(30,40)	6.1	6.1	7.2	4.8	4.9	5.0	4.7
	AR	(20,30)	7.2	7.2	9.2	5.2	5.6	5.7	5.2
		(30,40)	7.1	7.1	8.3	5.3	5.6	5.5	5.3
$t(5)$	CS	(20,30)	7.6	7.6	9.4	5.2	5.4	5.7	5.1
		(30,40)	6.0	6.0	7.3	4.7	4.6	4.9	4.5
	AR	(20,30)	7.1	7.1	8.7	5.1	5.4	5.8	4.9
		(30,40)	6.9	6.9	8.0	5.4	5.5	5.7	5.1
$\chi^2(5)$	CS	(20,30)	7.5	7.5	9.7	5.5	6.0	6.2	5.4
		(30,40)	6.4	6.4	7.6	4.9	5.1	5.2	4.9
	AR	(20,30)	7.7	7.7	10.1	5.1	5.6	5.7	5.3
		(30,40)	7.1	7.1	8.1	5.5	5.8	5.8	5.4
Lognormal	CS	(20,30)	6.8	6.8	9.4	5.9	4.9	6.4	4.8
		(30,40)	6.5	6.5	8.1	6.3	5.4	6.5	5.2
	AR	(20,30)	6.0	6.0	8.3	5.5	4.5	5.5	4.4
		(30,40)	6.0	6.0	7.3	5.8	5.1	6.1	5.0

Table 3.12: Type-I error rate (×100) of DMM, MMM, AT, WBT, PBT, PT and BT for unbalanced (decreasing sizes) heteroscedastic factorial designs for testing Group effect, $a = 2$, $p = 4$, $t = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group						
			DMM	MMM	AT	WBT	PBT	PT	BT
Normal	CS	(30,20)	4.2	4.2	8.8	5.7	5.3	5.3	5.6
		(40,30)	4.3	4.3	7.6	5.2	5.3	5.2	5.3
	AR	(30,20)	3.7	3.7	8.4	5.0	4.3	4.5	4.8
		(40,30)	4.6	4.6	7.5	5.7	5.4	5.7	5.7
$t(5)$	CS	(30,20)	4.4	4.4	8.9	5.6	5.2	5.5	5.4
		(40,30)	4.4	4.4	7.7	5.6	5.2	5.5	5.4
	AR	(30,20)	4.1	4.1	8.7	5.7	5.0	5.2	5.3
		(40,30)	3.9	3.9	6.8	5.0	4.8	5.1	4.8
$\chi^2(5)$	CS	(30,20)	4.2	4.2	8.7	5.7	5.3	5.5	5.6
		(40,30)	4.3	4.3	7.3	5.6	5.2	5.3	5.2
	AR	(30,20)	3.8	3.8	8.4	5.0	4.4	4.7	4.7
		(40,30)	4.3	4.3	7.5	5.2	4.9	5.1	5.0
Lognormal	CS	(30,20)	3.4	3.4	7.1	4.9	3.7	4.7	4.1
		(40,30)	4.3	4.3	7.0	5.7	4.8	5.5	4.9
	AR	(30,20)	3.7	3.7	8.0	5.9	4.5	5.5	4.6
		(40,30)	4.1	4.1	7.1	5.8	4.7	5.8	4.6

Chapter 4 Nonparametric Effect Measures in Multivariate Growth Curve Data

4.1 Introduction

Multivariate growth data appear in a great variety of disciplines, for example, biomedical science, public health, agriculture, social science, etc. For this type of data, several related variables are observed at different time points or occasions for each experimental unit or observational unit. Typically, mean- and covariance-based inferences are of interest to most scholars and researchers due to their interpretability. The classical parametric and semiparametric procedures developed are, for example, Bock (1975), Boik (1988), Naik and Rao (2001) and Rencher (2001). However, with such procedures, normality and homoscedasticity are required, which are difficult to attain in practice (see Xu and Cui, 2008; Suo et al., 2013). Furthermore, these methods are limited to analyzing continuous data. When binary, discrete and ordered categorical data are collected, parametric procedures are no longer appropriate. Thus, a more convenient, robust and reliable method that can accommodate both metric and non-metric data is in substantial demand. Under such circumstances, methods based on nonparametric relative treatment effects become promising solutions. Such nonparametric treatment effects are functionals of distribution functions and can be used to quantify the magnitude of effects of interest (see Brunner and Munzel, 2000; Brunner et al., 2002). Empirical distribution functions are used to estimate these treatment effects, which makes them rank-based methods in nature.

Nonparametric rank-based methods have been developed for the last few decades. In factorial designs, a asymptotically distribution-free rank-based test for repeated measures data was proposed by Brunner and Neumann (1982). It was later generalized for multivariate designs as well (Thompson, 1990, 1991; Brunner and Denker, 1994). However, they rely on the assumption of absolutely continuous distribution functions. Based on the idea of Munzel (1999), the normalized version of distribution function was used to derive asymptotic results for stratified two-sample designs in mixed models (Brunner et al., 1995).

Assuming continuity, hypotheses were originally formulated in terms of marginal distributions in nonparametric tests (Akritas and Arnold, 1994). Dropping the continuity assumption, Akritas and Brunner (1997) and Brunner et al. (1999) further advanced the formulation of hypotheses. The advantage of formulating hypotheses in terms of distribution functions is that it embraces models with both metric and nonmetric data. However, with such hypotheses, it is difficult to interpret the corresponding alternatives and test procedures cannot be used to construct confidence intervals for the effect size measure. Hence, test procedures with hypotheses formulated in terms of the nonparametric treatment effect are more appropriate (Konietschke et al., 2012a; Brunner et al., 2017).

It is our intention to propose nonparametric test procedures in multivariate repeated measures model for data types including discrete, ordered categorical and continuous data in a unified manner. In the meanwhile, the unweighted treatment effects measures along with the corresponding confidence intervals are derived to quantify the magnitude of effects of interest. In order to do so, asymptotic normality of the estimators and consistency of the covariance matrix estimator are established through nonparametric rank-based theories.

This chapter is organized as follows. In Section 4.2, statistical models and hypotheses are introduced and the nonparametric treatment effects and the estimators are defined. The asymptotic properties of the estimators are investigated in Section 4.3. In addition, the consistent estimators for the unknown covariance matrix in multivariate repeated measures model are derived in Section 4.3. In Section 4.4, we propose test statistics in the context of factorial designs and study their asymptotic properties. We also conduct extensive simulations under various settings in Section 4.5. In Section 4.6, we illustrate the application of our proposed methods with an optometry data set. We conclude the chapter with some discussions and remarks in Section 4.7. All relevant proofs and technical details are given in the Appendix (Section 4.8).

4.2 Models and Hypotheses

Assume nonparametric model

$$X_{ijk}^{(s)} \sim F_{ij}^{(s)}, \quad i = 1, \ldots, a; j = 1, \ldots, b; s = 1, \ldots, c; k = 1, \ldots, n_i,$$

where i is the treatment group index, j is the time (occasion) index, s is the variable index and k is the subject index. With nonparametric model where normalized version of distribution function is used, the assumption of continuous distributions can be dropped (Kruskal et al., 1952; Ruymgaart, 1980). It provides a unified treatment of ties for discrete, ordered categorical and continuous data. A formal definition of the normalized distribution function is $F(x) = \frac{1}{2}\left[F^+(x) + F^-(x)\right]$, where $F^-(x) = P(X < x)$ is the left continuous distribution function and $F^+(x) = P(X \leqslant x)$ is the right continuous distribution function (Brunner et al., 2002).

In this set-up, to specify the treatment effects, we use nonparametric relative treatment effects between the distribution of group i at time point (occasion) j and the distribution of group l at time point (occasion) m for the variable s,

$$w_{lm,ij}^{(s)} = P\left(X_{lm1}^{(s)} < X_{ij1}^{(s)}\right) + \frac{1}{2} P\left(X_{lm1}^{(s)} = X_{ij1}^{(s)}\right) = \int F_{lm}^{(s)} \mathrm{d} F_{ij}^{(s)},$$

where $i, l = 1, \ldots, a; j, m = 1, \ldots, b; s = 1, \ldots, c$. We further define the unweighted mean distribution function by

$$G_s(x) = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} F_{ij}^{(s)}(x). \tag{4.1}$$

The relative treatment effect of distribution of group i at time point j for variable s with respect to all distributions for variable s is denoted as follows by comparing each marginal distribution function with the unweighted mean distribution function,

$$p_{ij}^{(s)} = \int G_s \mathrm{d} F_{ij}^{(s)}, \tag{4.2}$$

which, by definition, takes value in $[\frac{1}{2ab}, 1 - \frac{1}{2ab}]$ and can also be expressed as $p_{ij}^{(s)} = P(Z_s < X_{ij1}^{(s)}) + \frac{1}{2} P(Z_s = X_{ij1}^{(s)})$, where $Z_s \sim G_s$ and Z_s is independent of $X_{ij1}^{(s)}$. When $p_{ij}^{(s)} < \frac{1}{2}$, it means that observations from the distribution $F_{ij}^{(s)}$ tend to have smaller values than those from the unweighted mean distribution G_s. To be specific, for example, $p_{ij}^{(s)} = 0.43$ can be interpreted as that the probability of a randomly chosen observation from mean distribution of variable s resulting in a smaller value than a randomly chosen observation from the distribution of group i at time point (occasion) j of variable s is 43%. Further, if $p_{uv}^{(s)} = 0.65$, then $p_{ij}^{(s)} < p_{uv}^{(s)}$. It means that the observations from group i at time point (occasion) j tend to result in smaller values than those from group u at time point (occasion) v.

Unlike the weighted mean distribution H in Brunner et al. (2002), the use of unweighted mean distribution removes the dependence of test results on sample size allocation. To formulate the hypotheses on relative effects, we define the vectors of nonparametric treatment effects,

$$\boldsymbol{p} = \left(\boldsymbol{p}_1^\top, \ldots, \boldsymbol{p}_a^\top\right)^\top, \quad \boldsymbol{p}_i = \left(\boldsymbol{p}_{i1}^\top, \ldots, \boldsymbol{p}_{ib}^\top\right)^\top, \text{ and } \boldsymbol{p}_{ij} = \left(p_{ij}^{(1)}, \ldots, p_{ij}^{(c)}\right)^\top, \qquad (4.3)$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$. Let $\boldsymbol{H}$ denote an appropriate contrast matrix and a linear hypothesis can be expressed as

$$H_0^p : \boldsymbol{Hp} = \boldsymbol{0},$$

which is sometimes referred to as generalized or nonparametric Behrens-Fisher situation. (see Fligner and Policello, 1981; Brunner et al., 1995; Brunner and Munzel, 2000).

To estimate the relative effects, we replace the distribution functions $G_s(x)$ and $F_{ij}^{(s)}(x)$ with the corresponding empirical distribution functions. We define the empirical distribution function of $F_{ij}^{(s)}$ by $\widehat{F}_{ij}^{(s)}(x) = \frac{1}{n_i}\sum_{k=1}^{n_i} c(x - X_{ijk}^{(s)})$, where $c(u)$ is the normalized version of the counting function and it takes the value of $0, \frac{1}{2}, 1$, if $u < 0, u = 0$ or $u > 1$, respectively. By using the normalized version of the empirical distribution function, we can accommodate discrete, ordered categorical and continuous data in a unified form. The estimator of relative effect is denoted by

$$\widehat{p}_{ij}^{(s)} = \int \widehat{G}_s \mathrm{d}\widehat{F}_{ij}^{(s)} = \frac{1}{ab}\sum_{l=1}^{a}\sum_{m=1}^{b}\int \widehat{F}_{lm}^{(s)} \mathrm{d}\widehat{F}_{ij}^{(s)} = \frac{1}{ab}\sum_{l=1}^{a}\sum_{m=1}^{b} \widehat{w}_{lm,ij}^{(s)},$$

where

$$\widehat{w}_{lm,ij}^{(s)} = \frac{1}{n_l}\left(\bar{R}_{ij\cdot}^{(s)}(lm, ij) - \frac{n_i+1}{2}\right), \quad \bar{R}_{ij\cdot}^{(s)}(lm, ij) = \frac{1}{n_i}\sum_{k=1}^{n_i} R_{ijk}^{(s)}(lm, ij),$$

where $R_{ijk}^{(s)}(lm,ij)$ denotes the mid-rank of $X_{ijk}^{(s)}$ among all $n_i + n_l$ observations within the two samples $X_{lm1}^{(s)},\ldots,X_{lmn_l}^{(s)}$ and $X_{ij1}^{(s)},\ldots,X_{ijn_i}^{(s)}$. The proof of this follows the same idea in Brunner et al. (2017). Define $\widehat{\boldsymbol{p}}$ in an analogous way as in (4.3), $\widehat{\boldsymbol{p}} = (\widehat{\boldsymbol{p}}_1^\top,\ldots,\widehat{\boldsymbol{p}}_a^\top)^\top$, where $\widehat{\boldsymbol{p}}_i = (\widehat{\boldsymbol{p}}_{i1}^\top,\ldots,\widehat{\boldsymbol{p}}_{ib}^\top)^\top$ and $\widehat{\boldsymbol{p}}_{ij} = (\widehat{p}_{ij}^{(1)},\ldots,\widehat{p}_{ij}^{(c)})^\top$. It is well known that $\widehat{w}_{lm,ij}^{(s)}$ is an unbiased and L_2 consistent estimator of $w_{lm,ij}^{(s)}$ for $i,l = 1,\ldots,a; j,m = 1,\ldots,b; s = 1,\ldots,c$, which has been proved in Brunner and Munzel (2000) and Brunner and Puri (2001). Given $\widehat{p}_{ij}^{(s)}$ are linear combinations of $\widehat{w}_{lm,ij}^{(s)}$, $\widehat{p}_{ij}^{(s)}$ retain the same properties of being an unbiased and L_2 consistent estimator of $p_{ij}^{(s)}$.

Let $\boldsymbol{w}_{lm,ij} = (w_{lm,ij}^{(1)},\ldots,w_{lm,ij}^{(c)})^\top = (\int F_{lm}^{(1)}\mathrm{d}F_{ij}^{(1)},\ldots,\int F_{lm}^{(c)}\mathrm{d}F_{ij}^{(c)})^\top$, we further write $\boldsymbol{w}$ analogously as follows,

$$\boldsymbol{w} = \left(\boldsymbol{w}_{11}^\top,\ldots,\boldsymbol{w}_{1b}^\top,\ldots,\boldsymbol{w}_{a1}^\top,\ldots,\boldsymbol{w}_{ab}^\top\right)^\top \quad \text{and}$$
$$\boldsymbol{w}_{ij} = \left(\boldsymbol{w}_{11,ij}^\top,\ldots,\boldsymbol{w}_{1b,ij}^\top,\ldots,\boldsymbol{w}_{a1,ij}^\top,\ldots,\boldsymbol{w}_{ab,ij}^\top\right)^\top,$$

where $i = 1,\ldots a; j = 1,\ldots,b$. $\boldsymbol{w}$ is a vector of dimension $(ab)^2c \times 1$. The empirical version $\widehat{\boldsymbol{w}}$ is defined in the similar way but with empirical distributions, i.e., $\widehat{\boldsymbol{w}}_{lm,ij} = (\widehat{w}_{lm,ij}^{(1)},\ldots,\widehat{w}_{lm,ij}^{(c)})^\top = (\int \widehat{F}_{lm}^{(1)}\mathrm{d}\widehat{F}_{ij}^{(1)},\ldots,\int \widehat{F}_{lm}^{(c)}\mathrm{d}\widehat{F}_{ij}^{(c)})^\top$. Further, let $\boldsymbol{E} = \boldsymbol{I}_{ab} \otimes \left(\frac{1}{ab}\boldsymbol{1}_{ab}^\top\right) \otimes \boldsymbol{I}_c$. The vector $\boldsymbol{p}$ and its estimator $\widehat{\boldsymbol{p}}$ can be written as

$$\boldsymbol{p} = \boldsymbol{E}\boldsymbol{w} \quad \text{and} \tag{4.4}$$
$$\widehat{\boldsymbol{p}} = \boldsymbol{E}\widehat{\boldsymbol{w}}. \tag{4.5}$$

The asymptotic covariance matrix $\boldsymbol{V}$ of $\sqrt{N}(\widehat{\boldsymbol{p}}-\boldsymbol{p})$ can be expressed as $\boldsymbol{V} = \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^\top$, where $\boldsymbol{S}$ is the asymptotic covariance matrix of $\sqrt{N}\widehat{\boldsymbol{w}}$ and it is of dimension $(ab)^2c \times (ab)^2c$. To derive the elements of $\boldsymbol{S}$, we partition $\boldsymbol{S}$ into blocks with dimension $abc \times abc$ and we denote them by $\boldsymbol{S}_{ij,i'j'}$, where $i,i' = 1,\ldots,a; j,j' = 1,\ldots,b$. There are ab of such blocks.

Throughout the paper we will use the following notations. The d-dimensional identity matrix is denoted by $\boldsymbol{I}_d$ and a $d \times d$ matrix with all 1s as its elements is denoted by $\boldsymbol{J}_d = \boldsymbol{1}_d\boldsymbol{1}_d^\top$, where $\boldsymbol{1}_d = (1,\ldots,1)_{d\times 1}^\top$. We further denote the centering matrix by $\boldsymbol{P}_d = \boldsymbol{I}_d - \frac{1}{d}\boldsymbol{J}_d$. The operators $\bigoplus$ and $\bigotimes$ represent the Kronecker sum and product, respectively (Schott, 2016, Chap. 8).

4.3 Asymptotic Results

Asymptotic Distribution

To derive the asymptotic distribution of $\boldsymbol{t}_N = \sqrt{N}(\widehat{\boldsymbol{p}} - \boldsymbol{p})$, we assume

$$\min_{1\leqslant i\leqslant a}(n_i) \to \infty \text{ such that } N/n_i \leqslant N_0 < \infty \text{ for all } i = 1, \ldots, a, \tag{4.6}$$

where $N = \sum_{i=1}^{a} n_i$. Follow the same idea in Brunner and Munzel (2000), we have the following asymptotic equivalence for $\sqrt{N}(\widehat{w}^{(s)}_{i'j',ij} - w^{(s)}_{i'j',ij}) = \sqrt{N}(\int \widehat{F}^{(s)}_{i'j'} \mathrm{d}\widehat{F}^{(s)}_{ij} - \int F^{(s)}_{i'j'} \mathrm{d}F^{(s)}_{ij})$, which is the cornerstone of showing more complicated asymptotic results of $\boldsymbol{t}_N$. The asymptotic equivalence theorem in the current setting is restated as follows.

Theorem 4.3.1. *(Asymptotic Equivalence) If* $\min(n_{i'}, n_i) \to \infty$*, then* $t^{(s)}_N(i'j', ij) = \sqrt{N}(\widehat{w}^{(s)}_{i'j',ij} - w^{(s)}_{i'j',ij})$ *is asymptotically equivalent to*

$$U^{(s)}_N(i'j', ij) = \sqrt{N}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\left[F^{(s)}_{i'j'}\left(X^{(s)}_{ijk}\right) - w^{(s)}_{i'j',ij}\right] - \frac{1}{n_{i'}}\sum_{k=1}^{n_{i'}}\left[F^{(s)}_{ij}\left(X^{(s)}_{i'j'k}\right) - w^{(s)}_{ij,i'j'}\right]\right\}, \tag{4.7}$$

for all $i, i' = 1, \ldots, a$*,* $j, j' = 1, \ldots, b$*, and* $s = 1, \ldots, c$*.*

For the proof of Theorem 4.3.1, we refer to Brunner and Munzel (2000). To simplify the notation, we denote equation (4.7) by $\sqrt{N} Z^{(s)}_{i'j',ij}$. Further, we construct the following vector,

$$\boldsymbol{Z} = \left(\boldsymbol{Z}^\top_{11}, \ldots, \boldsymbol{Z}^\top_{1b}, \ldots, \boldsymbol{Z}^\top_{a1}, \ldots, \boldsymbol{Z}^\top_{ab}\right)^\top, \tag{4.8}$$

where

$$\boldsymbol{Z}_{ij} = \left(\boldsymbol{Z}^\top_{11,ij}, \ldots, \boldsymbol{Z}^\top_{1b,ij}, \ldots, \boldsymbol{Z}^\top_{a1,ij}, \ldots, \boldsymbol{Z}^\top_{ab,ij}\right)^\top \text{ and } \boldsymbol{Z}_{i'j',ij} = \left(Z^{(1)}_{i'j',ij}, \ldots, Z^{(c)}_{i'j',ij}\right)^\top. \tag{4.9}$$

Note that $Z_{ij,ij} = 0$, $Z_{i'j',ij} = -Z_{ij,i'j'}$ and $\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{0}$. By asymptotic equivalence in Theorem 4.3.1, $\sqrt{N}\boldsymbol{E}(\widehat{\boldsymbol{w}} - \boldsymbol{w})$ and $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$ are asymptotically equivalent and by equations (4.4) and (4.5), $\sqrt{N}(\widehat{\boldsymbol{p}} - \boldsymbol{p})$ and $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$ are asymptotically equivalent. Therefore, the asymptotic distribution of $\sqrt{N}(\widehat{\boldsymbol{p}} - \boldsymbol{p})$ is the same as that of $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$, which implies that they have the same asymptotic covariance matrix.

Theorem 4.3.2. *(Asymptotic Multivariate Normality) Under the assumptions of Theorem 4.3.1,* $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$ *has an asymptotic multivariate normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{V} = \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^{\top}$*. Further,*

$$\sqrt{N}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{V}\right),$$

where $\boldsymbol{S}$ *is given in Section 4.3.*

The proof of Theorem 4.3.2 is given in the Appendix (Section 4.8).

Asymptotic Covariance

Now, we derive the asymptotic covariance matrix $\boldsymbol{S}$ of $\sqrt{N}\boldsymbol{Z}$. From equations (4.8) and (4.9), $\boldsymbol{Z}$ is partitioned into ab vectors, $\boldsymbol{Z}_{ij}$ for $i = 1, \ldots, a$ and $j = 1, \ldots, b$. Analogously, $\boldsymbol{S}$ is partitioned into ab blocks, $\boldsymbol{S}_{ij,i'j'}$ for $i, i' = 1, \ldots, a$ and $j, j' = 1, \ldots, b$. On the main diagonal blocks, $\boldsymbol{S}_{ij,ij} = \mathrm{Cov}(\sqrt{N}\boldsymbol{Z}_{ij})$. Otherwise, $\boldsymbol{S}_{ij,i'j'} = \mathrm{Cov}(\sqrt{N}\boldsymbol{Z}_{ij}, \sqrt{N}\boldsymbol{Z}_{i'j'})$, where $(i,j) \neq (i',j')$.

Next, we show the cases of elements of the covariance matrix blocks $\boldsymbol{S}_{ij,ij}$ and $\boldsymbol{S}_{ij,i'j'}$. Let $s_{ij}^{(s,s')}(lm, l'm')$ denote the elements of $\boldsymbol{S}_{ij,ij}$ and $s_{ij,i'j'}^{(s,s')}(lm, l'm')$ denote the elements of $\boldsymbol{S}_{ij,i'j'}$ for $(i,j) \neq (i',j')$. For brevity, we consider $c = 2$ and the general case follows in an obvious manner. For presentational convenience, we define

$$\gamma_{ij,ij'}^{(s,s')}(lm, l'm') = \frac{N}{n_i} E\left\{\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right) - w_{lm,ij}^{(s)}\right]\left[F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right) - w_{l'm',ij'}^{(s')}\right]\right\}. \quad (4.10)$$

On the main diagonal blocks of the covariance matrix $\boldsymbol{S}_{ij,ij}$, there are two different cases (equations (4.11) and (4.12)). There are five different cases for the off-diagonal blocks (equations (4.13), (4.14), (4.15), (4.16), and (4.17)). All possible results of $s_{ij}^{(s,s')}(lm, l'm')$ are listed below.

1. $l = l', m = m', l \neq i$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(lm, lm) + \gamma_{lm,lm}^{(s,s)}(ij, ij) & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,,ij}^{(s,s')}(lm, lm) + \gamma_{lm,lm}^{(s,s')}(ij, ij) & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.11)$$

2. $l = l', m = m', l = i, m \neq j$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(im,im) + \gamma_{im,im}^{(s,s)}(ij,ij) - & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ 2\gamma_{ij,im}^{(s,s)}(im,ij) & \\ \gamma_{ij,ij}^{(s,s')}(im,im) - \gamma_{ij,im}^{(s,s')}(im,ij) - & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \\ \gamma_{im,ij}^{(s,s')}(ij,im) + \gamma_{im,im}^{(s,s')}(ij,ij) & \end{cases}, \quad (4.12)$$

3. $l = l', m \neq m', l = i, m \neq j, m' \neq j$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(im,im') - \gamma_{im,ij}^{(s,s)}(ij,im') - & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,im'}^{(s,s)}(im,ij) + \gamma_{im,im'}^{(s,s)}(ij,ij) & \\ \gamma_{ij,ij}^{(s,s')}(im,im') - \gamma_{ij,im'}^{(s,s')}(im,ij) - & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \\ \gamma_{im,ij}^{(s,s')}(ij,im') + \gamma_{im,im'}^{(s,s')}(ij,ij) & \end{cases}, \quad (4.13)$$

4. $l = l', m \neq m', l \neq i$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(lm,lm') + \gamma_{lm,lm'}^{(s,s)}(ij,ij) & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij}^{(s,s')}(lm,lm') + \gamma_{lm,lm'}^{(s,s')}(ij,ij) & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.14)$$

5. $l \neq l', l \neq i, l' \neq i$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(lm,l'm') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij}^{(s,s')}(lm,l'm') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.15)$$

6. $l \neq l', l = i, l' \neq i, m \neq j$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(im,l'm') - \gamma_{im,ij}^{(s,s)}(ij,l'm') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij}^{(s,s')}(im,l'm') - \gamma_{im,ij}^{(s,s')}(ij,l'm') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.16)$$

7. $l \neq l', l \neq i, l' = i, m' \neq j$

$$\begin{cases} \gamma_{ij,ij}^{(s,s)}(lm,im') - \gamma_{ij,im'}^{(s,s)}(lm,ij) & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij}^{(s,s')}(lm,im') - \gamma_{ij,im'}^{(s,s')}(lm,ij) & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.17)$$

8. otherwise 0.

For the elements $s_{ij,i'j'}^{(s,s')}(lm, l'm')$ of the covariance matrix block $\boldsymbol{S}_{ij,i'j'}$, there are two main cases, $i = i', j \neq j'$ and $i \neq i'$. There are five subcases (equations (4.18), (4.19), (4.20), (4.21), and (4.22)) for $i = i'$ and $j \neq j'$ and six subcases (equations (4.23), (4.24), (4.25), (4.26), (4.27) and (4.28)) for $i \neq i'$.

1. $i = i', j \neq j', l = l', l = i, m \neq j, m' \neq j'$

$$\begin{cases} \gamma_{ij,ij'}^{(s,s)}(im, im') - \gamma_{ij,im'}^{(s,s)}(im, ij') - & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{im,ij'}^{(s,s)}(ij, im') + \gamma_{im,im'}^{(s,s)}(ij, ij') & \\ \gamma_{ij,ij'}^{(s,s')}(im, im') - \gamma_{ij,im'}^{(s,s')}(im, ij') - & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \\ \gamma_{im,ij'}^{(s,s')}(ij, im') + \gamma_{im,im'}^{(s,s')}(ij, ij') & \end{cases}, \quad (4.18)$$

2. $i = i', j \neq j', l = l', l \neq i$

$$\begin{cases} \gamma_{ij,ij'}^{(s,s)}(lm, lm') + \gamma_{lm,lm'}^{(s,s)}(ij, ij') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij'}^{(s,s')}(lm, lm') + \gamma_{lm,lm'}^{(s,s')}(ij, ij') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.19)$$

3. $i = i', j \neq j', l \neq l', l = i, l' \neq i, m \neq j$

$$\begin{cases} \gamma_{ij,ij'}^{(s,s)}(im, l'm') - \gamma_{im,ij'}^{(s,s)}(ij, l'm') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij'}^{(s,s')}(im, l'm') - \gamma_{im,ij'}^{(s,s')}(ij, l'm') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.20)$$

4. $i = i', j \neq j', l \neq l', l \neq i, l' = i, m' \neq j'$

$$\begin{cases} \gamma_{ij,ij'}^{(s,s)}(lm, im') - \gamma_{ij,im'}^{(s,s)}(lm, ij') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij'}^{(s,s')}(lm, im') - \gamma_{ij,im'}^{(s,s')}(lm, ij') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.21)$$

5. $i = i', j \neq j', l \neq l', l \neq i, l' \neq i$

$$\begin{cases} \gamma_{ij,ij'}^{(s,s)}(lm, l'm') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma_{ij,ij'}^{(s,s')}(lm, l'm') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \quad (4.22)$$

6. $i \neq i', l = l', l \neq i, l \neq i'$

$$\begin{cases} \gamma^{(s,s)}_{lm,lm'}(ij, i'j') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ \gamma^{(s,s')}_{lm,lm'}(ij, i'j') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.23}$$

7. $i \neq i', l = l', l = i, l \neq i', m \neq j$

$$\begin{cases} -\gamma^{(s,s)}_{ij,im'}(im, i'j') + \gamma^{(s,s)}_{im,im'}(ij, i'j') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ -\gamma^{(s,s')}_{ij,im'}(im, i'j') + \gamma^{(s,s')}_{im,im'}(ij, i'j') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.24}$$

8. $i \neq i', l = l', l \neq i, l = i', m' \neq j'$

$$\begin{cases} -\gamma^{(s,s)}_{i'm,i'j'}(ij, i'm') + \gamma^{(s,s)}_{i'm,i'm'}(ij, i'j') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ -\gamma^{(s,s')}_{i'm,i'j'}(ij, i'm') + \gamma^{(s,s')}_{i'm,i'm'}(ij, i'j') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.25}$$

9. $i \neq i', l \neq l', l = i', l' \neq i$

$$\begin{cases} -\gamma^{(s,s)}_{i'm,i'j'}(ij, l'm') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ -\gamma^{(s,s')}_{i'm,i'j'}(ij, l'm') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.26}$$

10. $i \neq i', l \neq l', l \neq i', l' = i$

$$\begin{cases} -\gamma^{(s,s)}_{ij,im'}(lm, i'j') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ -\gamma^{(s,s')}_{ij,im'}(lm, i'j') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.27}$$

11. $i \neq i', l \neq l', l = i', l' = i$

$$\begin{cases} -\gamma^{(s,s)}_{ij,im'}(i'm, i'j') - \gamma^{(s,s)}_{i'm,i'j'}(ij, im') & \text{if } s = s' = 1 \text{ or } s = s' = 2 \\ -\gamma^{(s,s')}_{ij,im'}(i'm, i'j') - \gamma^{(s,s')}_{i'm,i'j'}(ij, im') & \text{if } s = 1, s' = 2 \text{ or } s = 2, s' = 1 \end{cases}, \tag{4.28}$$

12. otherwise 0.

The explicit derivation of equations from (4.11) to (4.28) is given in the Appendix (Section 4.8).

Estimator of the Asymptotic Covariance

Finally, we obtain the asymptotic covariance matrix of $\sqrt{N}\,(\widehat{\boldsymbol{p}} - \boldsymbol{p})$, denoted by $\boldsymbol{V}$. The unknown quantity in equation (4.10) can be equivalently written as

$$\gamma_{ij,ij'}^{(s,s')}\,(lm, l'm') = \frac{N}{n_i}\text{Cov}\left\{\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right) - w_{lm,ij}^{(s)}\right], \left[F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right) - w_{l'm',ij'}^{(s')}\right]\right\},$$

which can be consistently estimated by replacing the distribution functions with their empirical counterparts. Let $D_{ijk}^{(s)}\,(lm) = \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - \widehat{w}_{lm,ij}^{(s)}$. It has been shown in Brunner et al. (2017) that

$$\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - \widehat{w}_{lm,ij}^{(s)} = \frac{1}{n_l}\left[\left(R_{ijk}^{(s)}(lm, ij) - R_{ijk}^{(s)}(ij)\right) - \left(\bar{R}_{ij\cdot}^{(s)}(lm, ij) - \frac{n_i + 1}{2}\right)\right].$$

Then the estimator of $\gamma_{ij,ij'}^{(s,s')}\,(lm, l'm')$ in equation (4.10) is given by

$$\widehat{\gamma}_{ij,ij'}^{(s,s')}\,(lm, l'm') = \frac{N}{n_i(n_i - 1)}\sum_{k=1}^{n_i} D_{ijk}^{(s)}(lm) \cdot D_{ij'k}^{(s')}\,(l'm')\,. \tag{4.29}$$

Replacing the quantities $\gamma_{ij,ij'}^{(s,s')}\,(lm, l'm')$ with $\widehat{\gamma}_{ij,ij'}^{(s,s')}\,(lm, l'm')$ in equations (4.11) to (4.28), we obtain L_2 consistent estimators $\widehat{s}_{ij}^{(s,s')}\,(lm, l'm')$ and $\widehat{s}_{ij,i'j'}^{(s,s')}\,(lm, l'm')$ for the elements of covariance matrix blocks $\boldsymbol{S}_{ij,ij}$ and $\boldsymbol{S}_{ij,i'j'}$. The resulting estimator of the asymptotic covariance matrix $\boldsymbol{S}$ of $\sqrt{N}\boldsymbol{Z}$ is denoted by $\widehat{\boldsymbol{S}}$. Hence, $\widehat{\boldsymbol{S}}$ is also an L_2 consistent estimator of the unknown quantity $\boldsymbol{S}$. Then an estimator $\widehat{\boldsymbol{V}}$ of the asymptotic covariance matrix $\boldsymbol{V}$ of the statistic $\sqrt{N}\,(\widehat{\boldsymbol{p}} - \boldsymbol{p})$ is

$$\widehat{\boldsymbol{V}} = \boldsymbol{E}\widehat{\boldsymbol{S}}\boldsymbol{E}^\top.$$

Theorem 4.3.3. *(L_2 Consistent Estimator) Under the assumptions of Theorem 4.3.1 and Theorem 4.3.2, $\widehat{\gamma}_{ij,ij'}^{(s,s')}(lm, l'm')$ defined in equation* (4.29) *is an L_2 consistent estimator of the unknown quantity $\gamma_{ij,ij'}^{(s,s')}(lm, l'm')$. Further, the estimators $\widehat{\boldsymbol{S}}$ and $\widehat{\boldsymbol{V}}$ are L_2 consistent estimators of the unknown covariance matrices $\boldsymbol{S}$ and $\boldsymbol{V}$, respectively.*

The L_2 consistency of the estimators $\widehat{\gamma}_{ij,ij'}^{(s,s')}\,(lm, l'm')$ is proved in the Appendix (Section 4.8).

4.4 Test Statistics

Asymptotic Wald-type Statistic (WTS)

The construction of test statistics relies on the asymptotic multivariate normality of $\boldsymbol{t}_N = \sqrt{N}\,(\widehat{\boldsymbol{p}} - \boldsymbol{p})$. Let $\boldsymbol{T} = \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^-\boldsymbol{H}$ be the unique project matrix onto the column space of $\boldsymbol{H}$ and $(\boldsymbol{H}\boldsymbol{H}^\top)^-$ is some generalized inverse of $\boldsymbol{H}\boldsymbol{H}^\top$. It is easy to show that $\boldsymbol{T}\boldsymbol{p} = \boldsymbol{0}$ if and only if $\boldsymbol{H}\boldsymbol{p} = \boldsymbol{0}$. Therefore, the unique, symmetric and idempotent contrast matrix $\boldsymbol{T}$ is equivalent to $\boldsymbol{H}$ for testing $H_0^p : \boldsymbol{H}\boldsymbol{p} = \boldsymbol{0}$. The Wald-type statistic (WTS) is denoted by

$$W_N\,(\boldsymbol{T}) = N\widehat{\boldsymbol{p}}^\top\boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{V}}\boldsymbol{T})^+\boldsymbol{T}\widehat{\boldsymbol{p}}, \tag{4.30}$$

where $(\boldsymbol{T}\widehat{\boldsymbol{V}}\boldsymbol{T})^+$ denotes the pseudo inverse or Moore-Penrose inverse of $\boldsymbol{T}\widehat{\boldsymbol{V}}\boldsymbol{T}$. Under the null hypothesis, the statistic $W_N(\boldsymbol{T})$ has an asymptotic chi-square distribution, which is formally stated in the following theorem.

Theorem 4.4.1. *Under the null hypothesis $H_0^p : \boldsymbol{T}\boldsymbol{p} = \boldsymbol{0}$ and the condition in equation* (4.6)*, the Wald-type statistic W_N in equation* (4.30) *has, asymptotically as $N \to \infty$, a central $\chi^2_{f_{\boldsymbol{T}}}$ distribution with degrees of freedom $f_{\boldsymbol{T}} = \operatorname{rank}(\boldsymbol{T})$.*

The test is given by $\varphi_N = \mathbb{I}\{W_N(\boldsymbol{T}) > \chi^2_{f_{\boldsymbol{T}},1-\alpha}\}$, where $\chi^2_{f_{\boldsymbol{T}},1-\alpha}$ denotes the $(1-\alpha)$-quantile of the $\chi^2_{f_{\boldsymbol{T}}}$ distribution. Under the asymptotic framework in equation (4.6) and the null hypothesis, by Theorem 9.2.2 in Rao and Mitra (1971, p. 173), the quadratic form $W_N\,(\boldsymbol{T})$ has, asymptotically, $\chi^2_{f_{\boldsymbol{T}}}$ distribution, where $f_{\boldsymbol{T}} = \operatorname{rank}(\boldsymbol{T})$. The proof of Theorem 4.4.1 follows the same idea in Zeng and Harrar (2021a).

Wilks' Lambda F-Approximation

In this section, we propose a multivariate test applied to a transformed version of the raw data $X_{ijk}^{(s)}$. Let

$$\widehat{Y}_{ijk}^{(s)} = \widehat{G}_s(X_{ijk}^{(s)}), \tag{4.31}$$

where $\widehat{G}_s(x) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{F}_{ij}^{(s)}(x)$ and $\widehat{F}_{ij}^{(s)}(x) = \frac{1}{n_i}\sum_{k=1}^{n_i} c(x - X_{ijk}^{(s)})$. The asymptotic version of $\widehat{Y}_{ijk}^{(s)}$ denoted by $Y_{ijk}^{(s)} = G_s(X_{ijk}^{(s)})$ are independent for different values of i or k. However, the covariance matrices differ for different values of i. Therefore, it would be reasonable to apply the MANOVA test introduced in Zeng and Harrar (2021b) on $\widehat{Y}_{ijk}^{(s)}$. More specifically, with the observed $c \times 1$ vector, $\boldsymbol{X}_{ijk} = (X_{ijk}^{(1)}, \ldots, X_{ijk}^{(c)})^\top$, we have the corresponding transformed $c \times 1$ vector, $\widehat{\boldsymbol{Y}}_{ijk} = (\widehat{Y}_{ijk}^{(1)}, \ldots, \widehat{Y}_{ijk}^{(c)})^\top$, where $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n_i$. The type of effect being tested is denoted by ϕ. The contrast matrices $\boldsymbol{D}_\phi$ and $\boldsymbol{C}_\phi$ target within-subject and between-subject factor effects, respectively. Let $\boldsymbol{H}_\phi = \boldsymbol{C}_\phi^\top \otimes \boldsymbol{D}_\phi \otimes \boldsymbol{I}_c$ be an appropriate contrast matrix. The unique projection matrix $\boldsymbol{T}_\phi = \boldsymbol{H}_\phi^\top (\boldsymbol{H}_\phi \boldsymbol{H}_\phi^\top)^- \boldsymbol{H}_\phi$ is more convenient and equivalent to $\boldsymbol{H}_\phi$ for testing the null hypothesis $H_0 : \boldsymbol{H}_\phi \boldsymbol{\mu} = \boldsymbol{0}$, where $\boldsymbol{\mu}$ is the overall mean vector of the transformed data $\boldsymbol{Y}$. Note that the mean vector $\boldsymbol{\mu}$ is, in fact, $\boldsymbol{p}$. Thus, the null hypothesis is equivalent to $H_0^p : \boldsymbol{H}_\phi \boldsymbol{p} = \boldsymbol{0}$. For the $bc \times 1$ vector, $\widehat{\boldsymbol{Y}}_{ik} = (\widehat{\boldsymbol{Y}}_{i1k}^\top, \ldots, \widehat{\boldsymbol{Y}}_{ibk}^\top)^\top$, $i = 1, \ldots, a; k = 1, \ldots, n_i$, based on Zeng and Harrar (2021b), we define

$$\boldsymbol{Z}_{ik}^{(\phi)} = (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_c)\, \widehat{\boldsymbol{Y}}_{ik},$$

where $\boldsymbol{D}_\phi$ is an appropriate within-subject contrast matrix and is full row rank with dimension $r \times b$ with $r = \text{rank}(\boldsymbol{D}_\phi)$. We further define the adjusted hypothesis and error matrices $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$,

$$\boldsymbol{H}^{(\phi)} = \overline{\boldsymbol{Z}}^{(\phi)} \boldsymbol{C}_\phi\, \overline{\boldsymbol{Z}}^{(\phi)\top} \quad \text{and}$$
$$\boldsymbol{G}^{(\phi)} = \sum_{i=1}^{a} \frac{\boldsymbol{C}_{\phi,ii}}{n_i(n_i - 1)} \sum_{k=1}^{n_i} \left(\boldsymbol{Z}_{ik}^{(\phi)} - \overline{\boldsymbol{Z}}_i^{(\phi)}\right) \left(\boldsymbol{Z}_{ik}^{(\phi)} - \overline{\boldsymbol{Z}}_i^{(\phi)}\right)^\top,$$

where $\overline{\boldsymbol{Z}}^{(\phi)} = (\overline{\boldsymbol{Z}}_1^{(\phi)}, \ldots \overline{\boldsymbol{Z}}_a^{(\phi)})$, $\overline{\boldsymbol{Z}}_i^{(\phi)} = \frac{1}{n_i} \sum_{k=1}^{n_i} \boldsymbol{Z}_{ik}^{(\phi)}$ and $\boldsymbol{C}_\phi$ is the between-subject contrast matrix with dimension $a \times a$. Let $\boldsymbol{V}_i = \text{Cov}(\widehat{\boldsymbol{Y}}_{ik})$,

$$\boldsymbol{\Sigma}_i^{(\phi)} = \text{Cov}\left(\boldsymbol{Z}_{ik}^{(\phi)}\right) = (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_c)\, \boldsymbol{V}_i\, (\boldsymbol{D}_\phi \otimes \boldsymbol{I}_c)^\top \text{ and } \boldsymbol{\Sigma}_\phi = \sum_{i=1}^{a} \boldsymbol{C}_{\phi,ii} \frac{\boldsymbol{\Sigma}_i^{(\phi)}}{n_i}.$$

Following the idea in Harrar and Bathke (2012a), the distributions of the adjusted hypothesis and error matrices can be approximated by the following Wishart distributions,

$$f^*_{\boldsymbol{H}^{(\phi)}} \cdot \boldsymbol{H}^{(\phi)}_* \underset{H_0^{(\phi)}}{\overset{approx}{\sim}} W_{rc}\left(f^*_{\boldsymbol{H}^{(\phi)}}, \boldsymbol{I}_{rc}\right) \quad \text{and}$$

$$f^*_{\boldsymbol{G}^{(\phi)}} \cdot \boldsymbol{G}^{(\phi)}_* \overset{approx}{\sim} W_{rc}\left(f^*_{\boldsymbol{G}^{(\phi)}}, \boldsymbol{I}_{rc}\right),$$

where $f^*_{\boldsymbol{H}^{(\phi)}}$, $f^*_{\boldsymbol{G}^{(\phi)}}$, $\boldsymbol{H}^{(\phi)}_* = \boldsymbol{\Sigma}_\phi^{-1/2}\boldsymbol{H}^{(\phi)}\boldsymbol{\Sigma}_\phi^{-1/2}$ and $\boldsymbol{G}^{(\phi)}_* = \boldsymbol{\Sigma}_\phi^{-1/2}\boldsymbol{G}^{(\phi)}\boldsymbol{\Sigma}_\phi^{-1/2}$ can be calculated following the results in Zeng and Harrar (2021b). Then we conduct the modified Wilks' Lambda test (Zeng and Harrar, 2021b), where the statistic is defined by

$$U = U_{rc, f^*_{\boldsymbol{H}^{(\phi)}}, f^*_{\boldsymbol{G}^{(\phi)}}} = \frac{|f^*_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)}_*|}{|f^*_{\boldsymbol{G}^{(\phi)}} \boldsymbol{G}^{(\phi)}_* + f^*_{\boldsymbol{H}^{(\phi)}} \boldsymbol{H}^{(\phi)}_*|}.$$

The Rao's F-approximation (Rencher, 2001) for the null distribution is

$$F = \frac{1 - U^{1/q}}{U^{1/q}} \cdot \frac{\mathrm{df}_2}{\mathrm{df}_1}, \tag{4.32}$$

where

$$q = \begin{cases} \sqrt{\dfrac{(rc)^2\, {f^*}^2_{\boldsymbol{H}^{(\phi)}} - 4}{(rc)^2 + {f^*}^2_{\boldsymbol{H}^{(\phi)}} - 5}}; & \text{if } (rc)^2 + {f^*}^2_{\boldsymbol{H}^{(\phi)}} - 5 > 0 \\ 1; & \text{if } (rc)^2 + {f^*}^2_{\boldsymbol{H}^{(\phi)}} - 5 \leqslant 0 \end{cases},$$

$\mathrm{df}_1 = rc \cdot f^*_{\boldsymbol{H}^{(\phi)}}$, $\mathrm{df}_2 = \omega q - \frac{1}{2}(rc \cdot f^*_{\boldsymbol{H}^{(\phi)}} - 2)$, and $\omega = f^*_{\boldsymbol{G}^{(\phi)}} - \frac{1}{2}(rc - f^*_{\boldsymbol{H}^{(\phi)}} + 1)$.

The corresponding test is given by $\tilde{\varphi}_N = \mathbb{I}\left\{F > F_{1-\alpha, \mathrm{df}_1, \mathrm{df}_2}\right\}$, where $F_{1-\alpha, \mathrm{df}_1, \mathrm{df}_2}$ is the $(1-\alpha)$-quantile of the $F_{\mathrm{df}_1, \mathrm{df}_2}$ distribution.

ANOVA-type Statistic F-Approximation (ATS)

We also consider another commonly used test statistic for multivariate repeated measures data, known as the analysis-of-variance (ANOVA) type of statistic, defined by

$$Q_N(\boldsymbol{T}) = \frac{N}{\mathrm{tr}\left(\boldsymbol{T}\widehat{\boldsymbol{V}}\right)} \widehat{\boldsymbol{p}}^\top \boldsymbol{T} \widehat{\boldsymbol{p}},$$

for which we assume $\mathrm{tr}(\boldsymbol{T}\boldsymbol{V}) \neq 0$. This assumption means that the projection of $\widehat{\boldsymbol{p}}$ onto the hypothesis space is almost surely nonconstant and it is a quite weak assumption (Brunner et al., 2017). We approximate the null distribution of $Q_N(\boldsymbol{T})$ using an F distribution from the modified Wilks' Lambda test in equation (4.32). The test is $\widehat{\varphi}_N = \mathbb{I}\left\{Q_N(\boldsymbol{T}) > F_{1-\alpha, \mathrm{df}_1, \mathrm{df}_2}\right\}$, where $F_{1-\alpha, \mathrm{df}_1, \mathrm{df}_2}$ is the $(1-\alpha)$-quantile of the $F_{\mathrm{df}_1, \mathrm{df}_2}$ distribution.

Confidence Intervals

By applying delta method on the multivariate normal limiting distribution of $\boldsymbol{t}_N = \sqrt{N}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right)$, we derive range preserving $(1-\alpha)$ confidence intervals for the nonparametric treatment effects $p_{ij}^{(s)}$,

$$\mathrm{CI}_{g,i,j,s} = g^{-1}\left\{ g\left(\widehat{p}_{ij}^{(s)}\right) \pm \frac{z_{1-\alpha/2}}{\sqrt{N}} \sqrt{\widehat{v}_{ij,ij}^{(s)}} |g'\left(\widehat{p}_{ij}^{(s)}\right)| \right\}, \tag{4.33}$$

where $i = 1, \ldots a; j = 1, \ldots, b; s = 1, \ldots, c$, g is differentiable in p_{ij} with $g'(p_{ij}) \neq 0$ and $\widehat{v}_{ij,ij}^{(s)}$ are diagonal elements of the covariance matrix $\widehat{\boldsymbol{V}}$. A possible choice of g is the logit function (Konietschke et al., 2012b).

4.5 Simulation Studies

Simulation Design

The test procedures derived in the previous sections are investigated and compared in extensive simulation studies in terms of their ability to maintain the preassigned type-I error rate ($\alpha = 0.05$) under the null hypothesis and to detect fixed alternatives under different settings.

To study the finite-sample behavior of the Wald-type test (WTS), modified Wilks' Lambda F-approximation test (WLF) and ANOVA-type test (ATS), we explore extensively with different sample sizes, covariance structures, data distributions and effects tested.

All simulations are conducted in **R** version 3.6.2. The number of simulations used to get the type-I error rates and powers is nsim = 5000. For our simulation designs, we simulate multivariate growth curve data in the context of general factorial designs. To be specific, we set the number of groups to be $a = 2$, with sample sizes $\boldsymbol{n} = (n_1, n_2)$. Both balanced and unbalanced sample sizes are analyzed. For balanced designs, $\boldsymbol{n} \in \{(20, 20), (30, 30)\}$ are investigated. For unbalanced designs, $\boldsymbol{n} \in \{(20, 35), (30, 45)\}$ are examined.

In addition, the computational complexity of multivariate data analyses grows exponentially as the dimension increases. Therefore, to illustrate our methodology in an efficient way, we set the dimension of the observations to be $c = 2$, and the number of repeated measures to be $b = 2$. We are interested in whether there are group, time and group-time interaction effects. The corresponding hypotheses are as follows.

1. The hypothesis of no Group effect is

   $H_0 : \boldsymbol{H}_G \boldsymbol{p} = \boldsymbol{0}$, where $\boldsymbol{H}_G = \boldsymbol{P}_a \otimes \frac{1}{b}\boldsymbol{1}_b^\top \otimes \boldsymbol{I}_c$.

2. The hypothesis of no Time effect is

   $H_0 : \boldsymbol{H}_T \boldsymbol{p} = \boldsymbol{0}$, where $\frac{1}{a}\boldsymbol{1}_a^\top \otimes \boldsymbol{P}_b \otimes \boldsymbol{I}_c$.

3. The hypothesis of no Group $\times$ Time effect is

   $H_0 : \boldsymbol{H}_{GT} \boldsymbol{p} = \boldsymbol{0}$, where $\boldsymbol{H}_{GT} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b \otimes \boldsymbol{I}_c$.

For covariance structure, we consider compound symmetric and autoregressive covariance structures. With these two covariance structures, both homoscedastic and heteroscedastic designs are investigated. Specifically, with the partitioned matrix $\boldsymbol{\Sigma}_i = (\boldsymbol{\Sigma}_{i,jk})$ representation, two covariance structures to be considered are below.

1. Compound Symmetry

$$\boldsymbol{\Sigma}_{i,jk} = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_c + \rho_i \boldsymbol{J}_c & k = j = 1, \dots, b \\ \boldsymbol{\Sigma}_{i,jk} = \rho_i \boldsymbol{J}_c & k \neq j \text{ and } k, j = 1, \dots, b \end{cases},$$

   where we take $\boldsymbol{\rho} = (0.2, 0.2)$ as the homoscedastic setting and $\boldsymbol{\rho} = (0.2, 0.7)$ as the heteroscedastic setting.

2. Autoregressive Structure AR(1)

$$\boldsymbol{\Sigma}_{i,jk} = \begin{cases} \boldsymbol{\Sigma}_{i,kk} = (1-\rho_i)\boldsymbol{I}_c + \rho_i \boldsymbol{J}_c & k = j = 1, \dots, b \\ \boldsymbol{\Sigma}_{i,jk} = \rho_i^{|k-j|} \boldsymbol{J}_c & k \neq j \text{ and } k = j = 1, \dots, b \end{cases},$$

   where we consider $\boldsymbol{\rho} = (0.2, 0.2)$ as the equal covariances setting and $\boldsymbol{\rho} = (0.2, 0.7)$ as the unequal covariances setting.

The compound symmetric and the autoregressive covariance structures will hereinafter be referred to as CS and AR, respectively. In unbalanced heteroscedastic designs, we also consider how the unequal covariance matrices are associated with the unequal sample sizes in two groups. For example, in heteroscedastic designs with correlation values $\boldsymbol{\rho} = (0.2, 0.7)$, we investigate unbalanced sample sizes in both increasing order $\boldsymbol{n} \in \{(20, 35), (30, 45)\}$ and in decreasing order $\boldsymbol{n} = \{(35, 20), (45, 30)\}$, for positive pairing and negative pairing, respectively.

Data are generated according to the model

$$\boldsymbol{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\epsilon}_{ik}, \quad i = 1, \ldots, a; k = 1, \ldots n_i,$$

where $\boldsymbol{\Sigma}_i^{1/2}$ is the square root of $\boldsymbol{\Sigma}_i$. To generate independent and identically distributed random vectors $\boldsymbol{\epsilon}_{ik} = (\boldsymbol{\epsilon}_{i1k}^\top, \ldots, \boldsymbol{\epsilon}_{ibk}^\top)^\top$, where $\boldsymbol{\epsilon}_{ijk} = (\epsilon_{ijk}^{(1)}, \ldots, \epsilon_{ijk}^{(c)})^\top$, we generate each element from the same standardized distribution by

$$\epsilon_{ijk}^{(s)} = \frac{X_{ijk}^{(s)} - \mathrm{E}(X_{ijk}^{(s)})}{\sqrt{\mathrm{Var}(X_{ijk}^{(s)})}}, \quad s = 1, \ldots, c,$$

where $X_{ijk}^{(s)}$ are generated from symmetric distributions (standard normal and $t(5)$ distributions), skewed distributions ($\chi^2(5)$ and standard lognormal distributions) and discrete normal distribution, where data are generated by taking the integer part of the standard normal data.

Simulation Results

For the simulation results, we consider five major scenarios, balanced homoscedastic (Table 4.1), unbalanced homoscedastic (Table 4.2), balanced heteroscedastic (Table 4.3), unbalanced heteroscedastic with positive pairing (Table 4.4) and unbalanced heteroscedastic with negative pairing (Table 4.5).

Table 4.1 shows the empirical type-I error rates in balanced homoscedastic settings. The WTS and WLF methods are roughly the same for testing Group effect. However, under the same settings, ATS is a little conservative when data are from normal, $t(5)$ and $\chi^2(5)$ distributions. When data are from lognormal and discrete normal distributions, the three tests WTS, WLF and ATS are roughly the same for testing the Group effect. For testing Time effect and Group $\times$ Time interaction effect, WLF is the best test maintaining the type-I error rates well to the 0.05 level for all data distributions, while WTS and ATS are much more liberal. Table 4.2 presents the type-I error rates in unbalanced homoscedastic settings. Although the sample sizes are unbalanced, the patterns of the test results are similar to that of the balanced homoscedastic designs shown in Table 4.1.

Next, we comment on the balanced heteroscedastic settings shown in Table 4.3. All three methods WTS, WLF and ATS are comparable for testing Group effect with normal, $t(5)$, $\chi^2(5)$ and discrete normal data. When data come from the highly skewed lognormal distribution for testing Group effect, WTS and WLF still perform well, but ATS produces test results that are more liberal. For testing Time and Group $\times$ Time effects with all data distributions, WLF is the best method leading to test results that are almost exact. However, compared with WLF, WTS and ATS are generally more liberal under the same settings. The liberality of ATS is less severe than that of WTS. Table 4.4 shows the test results for unbalanced heteroscedastic designs with positive pairing which share the same patterns as test results in balanced heteroscedastic designs shown Table 4.3.

Test results for unbalanced heteroscedastic designs with negative pairing are displayed in Table 4.5. In general, WTS and WLF perform well in maintaining the preassigned type-I error rate and are comparable for testing Group effect with all five data distributions, normal, $t(5)$, $\chi^2(5)$, lognormal and discrete normal. However, under the same settings, ATS test results are more liberal. For testing Time and Group $\times$ Time effects, WLF performs the best producing test results that are almost exact, but test results of WTS and ATS are generally more liberal. Like before, the liberality of ATS is less severe than that of WTS.

In summary, the modified Wilks' Lambda test $\tilde{\varphi}_N$ turns out to control the preassigned type-I error rate the best in all settings investigated. Therefore, it is the most recommended method for practical applications in terms of maintaining type-I error rate. The Wald-type test φ_N, on the other hand, performs well when testing

Group effect. Hence, it is highly recommended for testing Group effect. In addition to test results, the nonparametric treatment effects along with the corresponding confidence intervals can be calculated, which are interpretable and are used to quantify the magnitude of effects.

Table 4.1: Type-I error rate ($\times 100$) of WTS, WLF and ATS for balanced homoscedastic designs with $a = 2$, $b = 2$, $c = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Group			Time			Group×Time		
			WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	CS	(20,20)	5.2	5.0	3.6	10.0	5.2	6.9	9.2	5.0	6.4
		(30,30)	4.8	5.1	4.3	8.5	5.2	6.7	8.2	5.0	6.7
	AR	(20,20)	5.5	5.4	4.4	8.9	4.5	6.0	8.8	4.6	6.1
		(30,30)	4.8	4.8	3.9	8.4	4.8	6.5	8.5	5.3	7.0
$t(5)$	CS	(20,20)	5.0	4.7	3.9	9.1	5.0	6.0	9.1	5.1	6.7
		(30,30)	5.0	5.3	4.3	8.0	4.7	5.9	7.8	4.7	6.3
	AR	(20,20)	6.1	5.5	4.6	8.8	4.5	6.0	9.6	5.3	7.0
		(30,30)	4.7	4.8	4.0	8.2	5.0	6.6	8.4	5.2	6.7
$\chi^2(5)$	CS	(20,20)	5.6	5.3	4.5	8.4	4.8	5.9	8.7	4.9	6.1
		(30,30)	5.1	5.2	4.4	8.3	4.9	6.5	8.2	5.0	6.6
	AR	(20,20)	6.1	5.6	4.8	8.9	5.3	6.4	9.2	5.3	6.4
		(30,30)	5.0	5.3	4.1	8.7	5.8	7.2	7.9	5.0	6.2
Lognormal	CS	(20,20)	5.8	5.1	5.1	8.5	5.0	6.0	8.0	4.7	5.7
		(30,30)	5.3	5.3	5.5	7.4	4.1	5.9	8.3	4.9	6.5
	AR	(20,20)	5.3	4.5	5.0	7.8	4.9	5.8	7.9	4.5	6.0
		(30,30)	4.9	4.9	5.2	8.5	5.4	6.7	8.3	4.8	6.5
Disc. Norm	CS	(20,20)	5.7	5.0	4.9	8.7	5.3	6.3	8.7	5.2	6.0
		(30,30)	5.3	5.2	5.1	8.0	4.7	6.0	8.0	4.9	6.3
	AR	(20,20)	5.4	4.6	4.5	9.0	5.6	6.4	9.3	5.2	6.5
		(30,30)	4.7	4.7	4.8	8.5	5.2	6.7	7.9	4.8	6.4

Power Studies

We also investigate the empirical power of the test procedures to detect fixed alternatives. Similar to the type-I error rate simulations above, data are generated from five different distributions. For symmetric distributions, normal and $t\,(5)$ are investigated. For skewed data, $\chi^2\,(5)$ and lognormal are examined. For ordered categorical data, discrete normal is analyzed. To compare the power performance of the three tests, WTS, WLF and ATS, with different data distributions and for detecting dif-

Table 4.2: Type-I error rate ($\times 100$) of WTS, WLF and ATS for unbalanced homoscedastic designs with $a = 2$, $b = 2$, $c = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$, AR: $\boldsymbol{\rho} = (0.2, 0.2)$.

Dist	Cov	**n**	Group			Time			Group×Time		
			WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	CS	(20,35)	4.7	4.5	3.8	8.8	5.1	6.5	8.7	5.3	6.7
		(30,45)	4.9	4.9	4.3	8.8	5.6	7.0	8.2	5.0	6.5
	AR	(20,35)	5.6	5.3	4.2	8.7	4.9	6.0	9.2	5.1	6.4
		(30,45)	4.4	4.6	4.2	8.4	4.8	6.3	8.1	4.9	6.6
$t(5)$	CS	(20,35)	5.0	4.9	4.0	8.6	4.7	6.0	8.5	4.8	6.0
		(30,45)	4.6	4.6	4.3	8.0	5.1	6.4	8.3	5.1	6.6
	AR	(20,35)	5.2	5.1	4.4	9.1	4.8	6.6	8.4	4.7	6.3
		(30,45)	4.9	5.3	4.4	8.5	4.7	6.5	8.5	4.8	6.4
$\chi^2(5)$	CS	(20,35)	5.8	5.6	4.4	8.7	4.9	6.3	8.9	5.2	6.5
		(30,45)	5.1	5.2	4.5	8.2	5.2	6.5	7.6	4.6	6.3
	AR	(20,35)	4.9	4.5	3.6	8.7	5.0	6.2	9.1	5.2	6.6
		(30,45)	4.6	4.9	4.4	8.4	5.1	6.6	7.8	4.7	6.4
Lognormal	CS	(20,35)	5.2	4.8	4.9	8.4	5.2	6.3	8.6	5.4	6.2
		(30,45)	4.9	5.2	5.1	8.3	5.2	6.9	8.1	5.1	6.5
	AR	(20,35)	5.5	5.0	4.7	8.7	5.4	6.4	8.6	5.0	6.3
		(30,45)	5.2	5.4	5.2	8.3	4.9	6.5	8.7	5.5	7.2
Disc. Norm	CS	(20,35)	5.6	5.1	5.0	8.6	5.1	6.5	8.3	4.8	5.9
		(30,45)	4.7	4.9	4.8	7.6	5.1	6.4	8.4	5.1	6.7
	AR	(20,35)	6.1	5.7	5.2	8.0	4.9	6.1	8.8	5.4	6.4
		(30,45)	4.8	4.7	5.1	8.0	5.5	6.9	7.9	4.6	6.4

ferent fixed alternatives, we set two groups ($a = 2$), two time points ($b = 2$) and two response variables ($c = 2$). The sample sizes are balanced $\boldsymbol{n} = (30, 30)$. For covariance structure, compound symmetry with correlation values $\boldsymbol{\rho} = (0.2, 0.2)$ is used. We consider the following three fixed alternatives. For true Group effect, we set $\boldsymbol{\mu}_1 = 0 \times \mathbf{1}_2 \otimes \mathbf{1}_2$ and $\boldsymbol{\mu}_2 = \delta \mathbf{1}_2 \otimes \mathbf{1}_2$. For true Time effect, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \delta\,(1,2)^\top \otimes \mathbf{1}_2$. For true Group×Time effect, $\boldsymbol{\mu}_1 = 0 \times \mathbf{1}_2 \otimes \mathbf{1}_2$ and $\boldsymbol{\mu}_2 = \delta\,(1,2)^\top \otimes \mathbf{1}_2$. In all three cases, $\delta \in \{0, 0.1, 0.2, \ldots, 1.0\}$ takes value from small to large. The simulation results displayed in Table 4.6 are for symmetric distributions, in Table 4.7 are for skewed distributions and in Table 4.8 are for discrete normal data. Generally, the power of all three tests, WTS, WLF and ATS, are comparable but with the tendency that WTS and ATS have slightly larger power than WLF does. The power advantage of WTS and ATS for testing Time and Group×Time effects may be explained by the

Table 4.3: Type-I error rate ($\times 100$) of WTS, WLF and ATS for balanced heteroscedastic designs with $a = 2$, $b = 2$, $c = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group			Time			Group×Time		
			WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	CS	(20,20)	6.2	5.4	5.8	8.3	4.8	5.8	8.7	4.7	5.7
		(30,30)	5.1	5.1	5.6	8.3	5.0	6.2	8.3	5.1	6.6
	AR	(20,20)	5.7	5.1	5.5	8.0	4.5	5.5	9.1	5.0	6.2
		(30,30)	4.6	4.4	5.3	8.1	4.5	5.8	8.1	4.8	6.4
$t(5)$	CS	(20,20)	5.3	4.5	5.2	8.5	4.6	5.6	8.6	5.3	6.0
		(30,30)	4.7	4.4	5.1	7.8	4.8	6.1	7.8	4.8	5.7
	AR	(20,20)	5.4	4.6	5.6	8.6	4.7	5.4	8.2	4.6	5.4
		(30,30)	5.2	4.9	5.8	8.5	5.1	6.2	7.8	4.8	6.1
$\chi^2(5)$	CS	(20,20)	6.2	5.4	6.1	8.5	5.1	5.9	7.8	4.7	5.0
		(30,30)	4.8	4.6	5.1	8.3	5.0	6.1	7.3	4.5	5.8
	AR	(20,20)	5.8	5.0	5.2	8.7	5.4	6.2	9.2	5.3	6.0
		(30,30)	5.5	5.4	6.3	8.1	5.1	6.5	8.2	5.2	6.5
Lognormal	CS	(20,20)	5.4	5.0	6.5	7.8	4.8	5.3	7.3	4.7	4.8
		(30,30)	5.4	5.3	7.3	7.8	5.0	5.9	7.7	5.5	6.3
	AR	(20,20)	5.6	5.0	6.4	8.3	5.3	5.8	7.1	4.7	4.9
		(30,30)	5.1	4.9	6.9	7.6	4.8	5.9	7.5	4.6	5.6
Disc. Norm	CS	(20,20)	6.0	5.6	5.7	8.5	5.3	5.8	9.0	5.4	5.8
		(30,30)	4.9	5.0	5.5	7.7	4.8	6.0	7.9	4.8	6.1
	AR	(20,20)	6.4	5.7	5.6	8.2	5.0	5.4	8.3	4.8	5.7
		(30,30)	5.4	5.7	5.4	7.7	4.8	5.9	7.6	4.7	5.8

liberal behavior in type-I error rates.

4.6 Application

To illustrate how our methodology can be applied and to stimulate readers' interest, we introduce an example where multivariate growth curve data on idiopathic infantile nystagmus syndrome (INS) are analyzed in the context of general factorial design. In this study (Fadardi et al., 2017), 15 voluntary participants with idiopathic INS were recruited from a referring ophthalmologist. Participants were asked to carry out acuity tasks identifying the direction of horizontal Tumbling-E targets under different mental load settings. For the low mental load setting, participants were given unlimited time to respond. After responding, they were required to view a

Table 4.4: Type-I error rate (×100) of WTS, WLF and ATS for unbalanced heteroscedastic designs (positive pairing) with $a = 2$, $b = 2$, $c = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group			Time			Group×Time		
			WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	CS	(20,35)	5.1	4.8	4.9	8.5	4.7	5.6	8.9	4.6	5.8
		(30,45)	5.0	5.0	5.6	8.3	4.8	6.2	8.7	5.1	6.5
	AR	(20,35)	5.9	5.3	5.2	9.2	5.0	5.9	9.5	5.0	6.2
		(30,45)	4.6	4.7	5.1	8.8	5.1	6.4	8.7	4.8	6.2
$t(5)$	CS	(20,35)	4.8	4.8	4.5	9.3	5.3	6.1	9.9	5.3	6.4
		(30,45)	4.8	4.8	5.3	8.4	4.7	6.0	7.8	4.8	6.0
	AR	(20,35)	5.4	4.8	5.2	9.2	5.0	5.9	9.5	5.0	5.7
		(30,45)	4.8	4.7	5.2	8.8	5.2	6.3	8.7	5.1	6.3
$\chi^2(5)$	CS	(20,35)	5.3	4.8	5.2	9.3	5.2	6.0	9.1	5.3	6.2
		(30,45)	5.5	5.4	5.7	7.7	4.4	5.9	8.6	5.2	6.4
	AR	(20,35)	5.3	4.8	5.0	10.0	5.7	6.6	9.3	5.3	5.8
		(30,45)	5.4	5.4	5.7	8.6	5.1	6.2	8.5	5.1	6.5
Lognormal	CS	(20,35)	5.3	4.5	5.8	8.6	4.9	5.3	8.7	5.4	5.7
		(30,45)	5.6	5.4	7.1	8.8	5.5	6.8	8.1	5.0	6.0
	AR	(20,35)	5.6	4.4	6.7	9.1	5.1	6.0	8.4	5.0	5.5
		(30,45)	5.1	4.9	6.7	8.4	5.0	6.1	7.7	5.0	6.1
Disc. Norm	CS	(20,35)	5.5	5.1	5.1	8.4	4.6	5.5	9.0	5.3	5.7
		(30,45)	5.3	5.7	5.1	8.2	5.0	5.9	7.7	4.9	6.1
	AR	(20,35)	5.6	5.4	4.7	9.3	5.5	6.2	8.9	5.2	5.6
		(30,45)	4.8	5.0	5.2	8.4	5.1	6.2	8.3	5.1	6.6

fixation cross for 100 milliseconds prior to the presence of the next acuity target. For the high mental load setting, participants were given only 0.5 second to view the target and then 300 milliseconds to view a visual noise mask. Participants were required to respond while they were viewing a fixation cross for 1 second. In addition, participants were also asked to conduct mental arithmetic (continuously subtracting 7 from a number randomly selected between 100 and 120 and given by the examiner during the task) simultaneously with the acuity task. Both the low and the high mental load effects were evaluated at two gaze positions (null position and away position). Eventually, the size and contrast of the target at which participants' task performance plateaued were recorded. The main objective of the study is to investigate whether there is any main effect of mental load (M), main effect of gaze position (P), and

Table 4.5: Type-I error rate (×100) of WTS, WLF and ATS for unbalanced heteroscedastic designs (negative pairing) with $a = 2$, $b = 2$, $c = 2$, CS: $\boldsymbol{\rho} = (0.2, 0.7)$, AR: $\boldsymbol{\rho} = (0.2, 0.7)$.

Dist	Cov	**n**	Group			Time			Group×Time		
			WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	CS	(35,20)	5.4	4.6	5.8	8.1	5.3	6.5	7.8	5.0	6.1
		(45,30)	4.7	4.8	6.2	7.7	5.3	6.5	7.4	4.7	6.3
	AR	(35,20)	5.3	4.5	5.7	7.3	4.9	5.9	7.6	5.0	5.9
		(45,30)	4.6	4.4	6.1	7.7	5.1	6.4	7.0	4.4	5.8
$t(5)$	CS	(35,20)	5.2	5.0	6.2	7.7	4.9	5.7	8.1	5.5	6.6
		(45,30)	5.4	5.5	6.2	8.0	5.5	6.8	7.3	4.7	6.3
	AR	(35,20)	5.8	5.0	6.5	7.4	4.8	5.7	8.0	5.1	5.9
		(45,30)	4.8	4.6	6.1	8.1	5.3	7.0	6.5	4.2	5.5
$\chi^2(5)$	CS	(35,20)	6.1	5.3	6.8	7.2	4.9	5.8	7.6	5.1	6.1
		(45,30)	5.5	5.4	6.4	7.5	5.4	6.7	6.6	4.5	5.8
	AR	(35,20)	5.7	5.0	6.4	6.9	4.6	5.4	7.1	5.0	5.8
		(45,30)	5.0	4.7	6.3	7.7	5.0	6.4	7.5	5.3	6.3
Lognormal	CS	(35,20)	6.3	5.6	7.9	6.4	4.5	5.2	6.8	4.7	5.3
		(45,30)	5.2	5.1	7.2	7.2	5.3	6.2	7.3	5.2	6.2
	AR	(35,20)	5.9	5.5	7.8	6.8	4.6	5.6	6.9	5.1	5.5
		(45,30)	4.8	4.9	6.9	7.4	5.2	6.3	7.0	5.0	6.1
Disc. Norm	CS	(35,20)	5.6	5.3	6.2	7.5	4.8	5.9	7.8	5.2	6.2
		(45,30)	5.4	5.6	6.3	7.4	5.0	6.2	7.3	5.1	6.3
	AR	(35,20)	6.2	5.8	6.7	7.3	4.7	5.6	7.3	5.0	5.5
		(45,30)	5.2	5.4	6.0	7.4	4.7	5.9	8.0	5.5	6.6

interaction effect between the mental load and gaze position (MP).

Among all 15 participants with idiopathic infantile nystagmus syndrome, 11 of them finished the task with no missing data. To test the interaction effect mentioned above in the context of our method, we need the key parameters. All participants have the disease. Thus, there is only one group ($a = 1$). There are four repeated measures ($b = 4$) representing four different occasions, low mental load at null position, low mental load at away position, high mental load at null position, and high mental load at away position. There are two response variables measured each time, target size and contrast ($c = 2$). For the Wald-type and ANOVA-type tests, the corresponding contrast matrix for testing the interaction effect between the mental load and gaze position is $\boldsymbol{H}_{\text{MP}} = (1, -1, -1, 1) \otimes \boldsymbol{I}_2$. The contrast matrices for testing main effects of

Table 4.6: Power ($\times 100$) at $\alpha = 0.05$ of WTS, WLF and ATS for three fixed alternatives with $a = 2$, $b = 2$, $c = 2$, $\boldsymbol{n} = (30, 30)$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$. Symmetric distributions, normal and $t\,(5)$, are investigated.

Dist	δ	Group			Time			Group×Time		
		WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Normal	0.0	4.7	5.0	3.7	7.9	4.6	6.4	7.9	4.8	6.3
	0.1	7.7	8.0	7.6	15.6	10.2	12.9	10.6	6.4	8.3
	0.2	17.7	17.3	18.8	37.3	28.5	33.5	15.2	9.8	12.2
	0.3	34.5	33.5	37.9	67.8	58.5	63.8	24.9	17.3	21.6
	0.4	55.9	55.0	61.5	89.9	84.5	88.6	36.6	28.1	33.0
	0.5	76.8	75.9	80.5	98.0	96.1	97.7	51.5	42.0	48.0
	0.6	89.7	89.2	92.4	99.7	99.4	99.7	67.5	58.5	64.0
	0.7	97.1	96.7	98.0	100.0	100.0	100.0	79.5	72.2	77.2
	0.8	99.2	99.2	99.5	100.0	100.0	100.0	87.5	82.4	86.0
	0.9	99.9	99.9	99.9	100.0	100.0	100.0	93.7	90.2	92.8
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	97.5	95.6	97.3
$t(5)$	0.0	4.3	4.4	3.6	7.7	4.5	5.9	8.0	4.2	5.9
	0.1	6.6	6.5	6.3	14.7	9.7	12.1	9.0	5.5	7.0
	0.2	13.2	13.0	14.6	30.8	22.9	27.5	13.9	9.1	11.6
	0.3	26.4	25.9	30.2	58.4	49.1	54.2	21.8	15.2	19.2
	0.4	43.5	42.7	48.1	80.7	74.0	78.7	31.2	23.3	27.4
	0.5	62.1	61.0	67.2	93.8	90.4	92.8	42.7	33.4	39.0
	0.6	79.3	78.5	83.8	98.7	97.7	98.5	54.8	45.6	51.6
	0.7	90.7	89.8	93.2	99.9	99.7	99.9	67.5	59.1	64.4
	0.8	96.4	96.1	97.6	100.0	100.0	100.0	78.1	71.3	76.0
	0.9	98.8	98.6	99.2	100.0	100.0	100.0	85.1	79.0	83.0
	1.0	99.7	99.7	99.8	100.0	100.0	100.0	90.6	86.8	89.7

mental load and gaze position are $\boldsymbol{H}_{\mathrm{M}} = (1, 1, -1, -1) \otimes \boldsymbol{I}_2$ and $\boldsymbol{H}_{\mathrm{P}} = (1, -1, 1, -1) \otimes \boldsymbol{I}_2$, respectively. For the modified Wilks' Lambda F-approximation test, since $a = 1$, $\boldsymbol{C}_\phi$ is always 1, where $\phi = \{\mathrm{M}, \mathrm{P}, \mathrm{MP}\}$. However, $\boldsymbol{D}_\phi$ varies with effect tested, $\boldsymbol{D}_{\mathrm{MP}} = (1, -1, -1, 1)$ for testing the interaction effect, $\boldsymbol{D}_{\mathrm{M}} = (1, 1, -1, -1)$ for testing mental load effect and $\boldsymbol{D}_{\mathrm{P}} = (1, -1, 1, -1)$ for testing gaze position effect.

The original study (Fadardi et al., 2017) noted that due to the limitations imposed by the projector resolution, acuity target sizes were limited to 1.8, 1.75, 1.7, 1.65, 1.6, 1.5, 1.4, 1.3, 1.1 and 1 logMAR. Hence, it is reasonable to treat the target size as an order categorical variable. Also, tumbling-E targets were presented on the screen background with contrast of 98, 50, 25, 10, 5, and 2%, which makes the contrast variable an ordered categorical variable as well. In this case, our test procedures, WTS and ATS, which are based on nonparametric treatment effects in equation

Table 4.7: Power ($\times 100$) at $\alpha = 0.05$ of WTS, WLF and ATS for three fixed alternatives with $a = 2$, $b = 2$, $c = 2$, $\boldsymbol{n} = (30, 30)$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$. Skewed distributions, $\chi^2(5)$ and lognormal, are investigated.

Dist	δ	Group			Time			Group×Time		
		WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
$\chi^2(5)$	0.0	5.3	5.5	5.0	7.8	5.0	6.1	8.7	5.4	7.2
	0.1	8.4	8.3	8.7	18.7	12.9	15.8	10.9	6.7	8.6
	0.2	21.6	21.1	24.0	48.7	39.3	44.8	17.6	12.1	14.5
	0.3	41.6	40.3	46.5	80.1	73.5	77.8	30.3	22.8	27.5
	0.4	66.7	65.6	71.8	96.4	94.0	95.8	45.5	36.3	42.0
	0.5	85.5	84.7	89.3	99.7	99.3	99.7	60.7	51.3	57.3
	0.6	96.0	95.6	97.4	100.0	100.0	100.0	76.7	68.7	73.8
	0.7	99.1	99.0	99.5	100.0	100.0	100.0	86.1	80.4	84.6
	0.8	99.9	99.9	100.0	100.0	100.0	100.0	92.3	88.3	91.1
	0.9	100.0	100.0	100.0	100.0	100.0	100.0	96.5	94.0	96.1
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	98.4	97.3	98.3
Lognormal	0.0	4.5	4.7	4.8	7.8	4.7	6.0	8.2	5.1	6.5
	0.1	18.1	16.8	21.6	54.7	46.8	50.8	18.8	13.4	16.3
	0.2	57.1	55.0	65.4	96.7	95.5	96.4	43.8	36.3	40.3
	0.3	89.7	88.6	93.4	100.0	100.0	100.0	69.0	62.3	66.6
	0.4	98.7	98.6	99.2	100.0	100.0	100.0	84.5	79.9	82.7
	0.5	99.9	99.9	100.0	100.0	100.0	100.0	93.5	91.0	92.5
	0.6	100.0	100.0	100.0	100.0	100.0	100.0	97.7	96.5	97.2
	0.7	100.0	100.0	100.0	100.0	100.0	100.0	98.9	98.1	98.7
	0.8	100.0	100.0	100.0	100.0	100.0	100.0	99.6	99.4	99.6
	0.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	99.8
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.9

Table 4.8: Power ($\times 100$) at $\alpha = 0.05$ of WTS, WLF and ATS for three fixed alternatives with $a = 2$, $b = 2$, $c = 2$, $\boldsymbol{n} = (30, 30)$, CS: $\boldsymbol{\rho} = (0.2, 0.2)$. Discrete normal is investigated.

Dist	δ	Group			Time			Group×Time		
		WTS	WLF	ATS	WTS	WLF	ATS	WTS	WLF	ATS
Disc. Norm	0.0	5.1	4.9	4.6	8.3	5.1	6.5	8.5	5.3	6.8
	0.1	17.3	16.4	20.7	46.1	40.3	42.2	13.8	9.3	11.4
	0.2	38.1	34.5	43.4	74.7	72.1	72.0	17.9	13.1	15.3
	0.3	61.3	55.2	65.0	87.8	87.3	86.1	19.8	15.0	16.7
	0.4	76.6	68.7	77.8	91.4	91.5	90.5	26.2	20.7	23.0
	0.5	84.2	75.7	84.4	94.6	94.5	93.2	50.8	44.3	47.2
	0.6	89.1	82.1	89.1	96.7	96.6	96.2	67.6	61.6	64.4
	0.7	91.8	86.8	92.3	98.6	98.3	98.5	77.5	71.8	74.4
	0.8	94.5	91.6	95.4	99.9	99.8	99.8	80.7	76.4	79.0
	0.9	99.1	98.8	99.6	100.0	100.0	100.0	94.5	91.3	93.5
	1.0	99.8	99.8	99.9	100.0	100.0	100.0	98.7	98.0	98.5

(4.2), can be applied here. Moreover, the modified Wilks' Lambda F-approximation test based on the transformed data in equation (4.31) can also be applied in this nonparametric setting.

As shown in Table 4.9, the p-values of WTS, ATS and WLF are less than $\alpha = 0.05$ for testing the main effect of mental load, leading to significant results. However, for testing the main effect of gaze position, the p-values of all three tests are larger than 0.05. Thus, there is no main effect of gaze position. For testing the mental $\times$ position interaction effect, WTS method leads to significant result, whereas WLF and ATS methods produce insignificant results. Based on our simulations in Section 4.5, WLF is the best method in all settings, so it is more reasonable to claim that there is no significant mental $\times$ position interaction effect. It is also worth mentioning that with the total sample size of only 11, the asymptotic Wald-type test is not trustworthy.

Of the 11 participants, 3 are female and 8 are male. It might also be interesting to investigate whether there are other interaction effects, for example, gender $\times$ mental (GM), gender $\times$ position (GP), and gender $\times$ mental $\times$ position (GMP). In this case, there are two groups ($a = 2$), i.e., female group and male group, four occasions ($b = 4$) and two response variables ($c = 2$). Accordingly, the contrast matrices are $\boldsymbol{H}_{\text{GM}} = \boldsymbol{P}_a \otimes \boldsymbol{P}_2 \otimes \frac{1}{2}\mathbf{1}_2^\top \otimes \boldsymbol{I}_c$, $\boldsymbol{H}_{\text{GP}} = \boldsymbol{P}_a \otimes \frac{1}{2}\mathbf{1}_2^\top \otimes \boldsymbol{P}_2 \otimes \boldsymbol{I}_c$, and $\boldsymbol{H}_{\text{GMP}} = \boldsymbol{P}_a \otimes \boldsymbol{P}_2 \otimes \boldsymbol{P}_2 \otimes \boldsymbol{I}_c$. For the modified Wilks' Lambda F-approximation test, $\boldsymbol{C}_\phi$ is always $\boldsymbol{P}_2$ for testing all three interaction effects. However, $\boldsymbol{D}_\phi$ varies with effect tested and they are $\boldsymbol{D}_{\text{GM}} = (1, -1) \otimes \frac{1}{2}\mathbf{1}_2^\top$, $\boldsymbol{D}_{\text{GP}} = \frac{1}{2}\mathbf{1}_2^\top \otimes (1, -1)$, and $\boldsymbol{D}_{\text{GMP}} = (1, -1) \otimes (1, -1)$.

Test results of these two- and three-way interactions are shown in Table 4.10. With all three tests, the p-values for testing the two-way interactions gender $\times$ mental and gender$\times$position are greater than 0.05, leading to insignificant test results. The results of the three methods for testing the three-way interaction gender $\times$ mental $\times$ position are not aligned with each other. Given the group sample sizes of 3 and 8 for the two genders, the test results of WTS and WLF are not trustworthy. However, the test result of ATS is the most reliable one leading to an insignificant result.

In addition, point estimates of the nonparametric treatment effects $p_{ij}^{(s)}$ in equation (4.2) for two variables and for the four different occasions are computed. The

Table 4.9: Analysis of the idiopathic infantile nystagmus syndrome (INS) data using fully nonparametric methods.

Effect	WTS	WLF	ATS
Mental Load	< 0.001	0.00566	0.00346
Gaze Position	0.29863	0.37348	0.22134
Mental $\times$ Position	0.04788	0.06728	0.08743

index i refers to the group ($i = 1$) and the index j refers to the time point (occasion) ($j = 1$, low mental load at null position; $j = 2$, low mental load at away position; $j = 3$, high mental load at null position; $j = 4$, high mental load at away position). Lastly, the index s refers to the variable ($s = 1$, size; $s = 2$, contrast). The two-sided 95% confidence intervals for both with and without range preserving of the nonparametric treatment effect $p_{ij}^{(s)}$ are calculated based on equation (4.33). The estimated nonparametric treatment effects and their corresponding confidence intervals are presented in Table 4.11.

The estimated nonparametric treatment effect $\widehat{p}_{11}^{(2)} = 0.3936$ for the contrast variable with low mental load at null position means that the observations from $F_{11}^{(2)}$ tend to be smaller than those from the mean distribution $G_2 = \frac{1}{4}\sum_{i=1}^{1}\sum_{j=1}^{4} F_{ij}^{(2)}$. More precisely, it can be interpreted as that the probability that a randomly selected observation of variable 2 (contrast), Z_2, from the mean distribution G_2 is smaller than a randomly selected observation $X_{111}^{(2)}$ from $F_{11}^{(2)}$ equals 0.3936. Analogously, the estimated treatment effect $\widehat{p}_{13}^{(1)} = 0.6942$ for the size variable with high mental load at null position means that the observations from $F_{13}^{(1)}$ tend to be larger than those from the mean distribution G_1.

The confidence intervals of both variables do not overlap within each gaze position, which may be interpreted as that the mental load has an effect on the task performance in both gaze positions. This result is aligned with the result of testing the main effect of mental load as shown in Table 4.9. The visual representation of the estimated nonparametric treatment effects along with their corresponding confidence intervals (without range preserving) is shown in Figure 4.1. To further compare the with and without range preserving confidence intervals, the side-by-side plots are shown in Figure 4.2. It can be seen that these two types of confidence intervals are almost identical with the given optometry data.

Table 4.10: Analysis of the idiopathic infantile nystagmus syndrome (INS) data for two genders using fully nonparametric methods.

Effect	WTS	WLF	ATS
Gender × Mental	0.20337	0.32876	0.26494
Gender × Position	0.64313	0.84303	0.61250
Gender × Mental × Position	< 0.001	0.03151	0.26456

Table 4.11: Estimated nonparametric treatment effects and their corresponding 95% confidence intervals (with and without range preserving). The range preserving confidence intervals are denoted by RP.

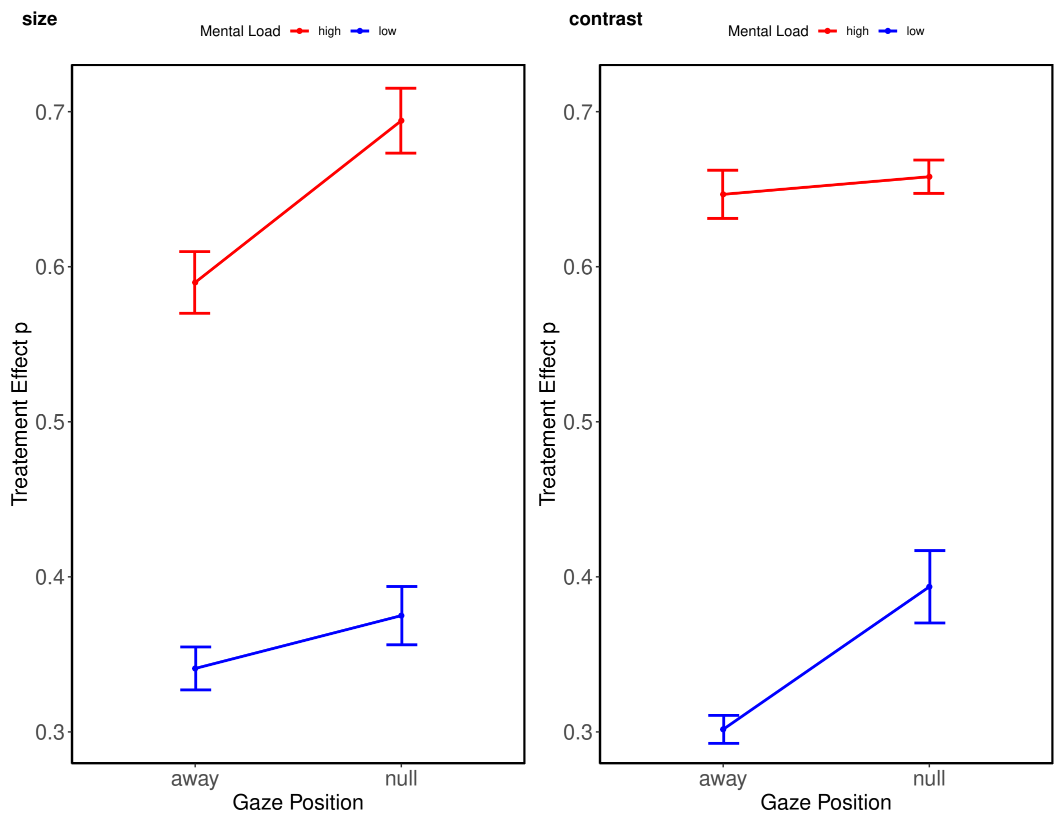
Group	Time	Mental	Position	Variable	$\widehat{\boldsymbol{p}}$	95% L	95% U	RP 95% L	RP 95% U
1	1	low	null	size	0.3750	0.3562	0.3938	0.3563	0.3940
1	1	low	null	contrast	0.3936	0.3702	0.4170	0.3705	0.4172
1	2	low	away	size	0.3409	0.3271	0.3548	0.3272	0.3549
1	2	low	away	contrast	0.3017	0.2926	0.3107	0.2927	0.3108
1	3	high	null	size	0.6942	0.6733	0.7151	0.6729	0.7147
1	3	high	null	contrast	0.6581	0.6473	0.6688	0.6472	0.6688
1	4	high	away	size	0.5899	0.5701	0.6097	0.5699	0.6095
1	4	high	away	contrast	0.6467	0.6311	0.6623	0.6310	0.6621

4.7 Discussion and Conclusion

Multivariate growth curve data occur in various disciplines, which urges us to develop robust and efficient methods for analyzing data of this type. Scholars and researchers have developed some classical parametric and semiparametric procedures to analyze such data but assumptions like multivariate normality and homoscedasticity are usually assumed. Moreover, due to the intrinsic nature of parametric methods, data that can be analyzed are restricted to continuous data.

With the proposed methods, null hypotheses can be formulated in terms of meaningful nonparametric measures of treatment effects. Such nonparametric treatment effects are characterized in terms of functionals of distribution functions with the sole assumption of nondegenerate marginal distributions. For the WTS method, we showed that it is asymptotically exact under the null hypothesis of interest. For

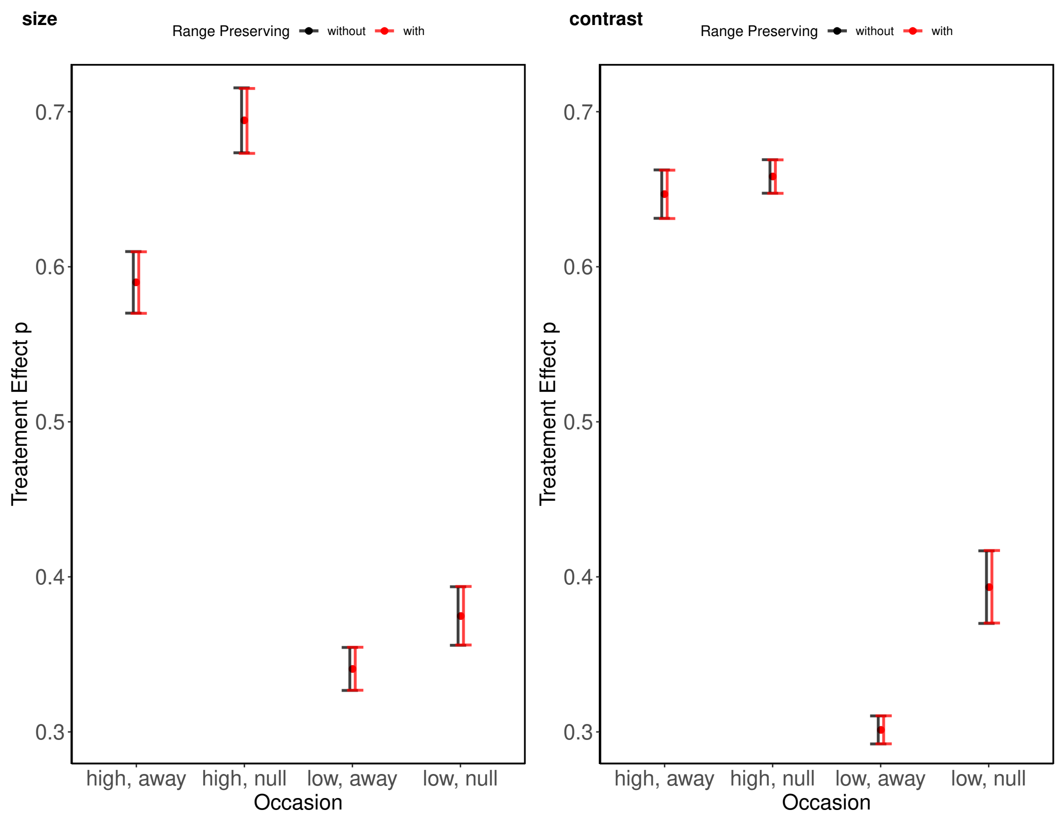
Figure 4.1: Visual representation of the estimated nonparametric treatment effects and their corresponding 95% confidence intervals (without range preserving).



the WLF method, instead of analyzing the original data, we proposed to analyze the transformed data using the modified Wilks' Lambda F-approximation test which was developed in Zeng and Harrar (2021b). For the ATS method, we used an F distribution obtained from the WLF method to approximate the distribution of the ANOVA-type statistic. In addition, we showed asymptotic multivariate normality of the estimated nonparametric treatment effects vector $\widehat{\boldsymbol{p}}$, which leads to the construction of confidence intervals of the treatment effects.

Our simulation studies indicated that WLF method is the most recommended method for testing Time and Group×Time effects in practical applications. However, for testing Group effect, WTS is highly recommended due to its superiority in maintaining the type-I error rate under the null hypotheses and achieving large power for detecting fixed alternatives.

Figure 4.2: Side-by-side plots of estimated nonparametric treatment effects and their corresponding 95% confidence intervals (with and without range preserving).



Compared with the classical parametric test procedures, our nonparametric tests are appropriate for analyzing multivariate growth curve data without assuming normality or homoscedasticity. Also, these methods can deal with discrete, ordered categorical and continuous data in the same way. Some other existing rank-based nonparametric methods were developed but are suitable for univariate data only.

The proposed methods are most appropriate when sample size is large. However, when sample size is small, they may lead to liberal test results. For future research, we plan to investigate other multiple contrast statistics and resampling methods based on nonparametric treatment effects to overcome the small sample issue.

4.8 Appendix

Derivation of the Covariance S in (4.11) **to** (4.28)

We show the derivation of the main diagonal covariance blocks $\boldsymbol{S}_{ij,ij} = \text{Cov}\left(\sqrt{N}\boldsymbol{Z}_{ij}\right)$ and the off-diagonal covariance blocks $\boldsymbol{S}_{ij,i'j'} = \text{Cov}\left(\sqrt{N}\boldsymbol{Z}_{ij}, \sqrt{N}\boldsymbol{Z}_{i'j'}\right)$, where $(ij) \neq (i'j')$. For presentational convenience, we use the following notations,

$$\xi^{(s)}_{lm,ij;k} = F^{(s)}_{lm}\left(X^{(s)}_{ijk}\right) - w^{(s)}_{lm,ij}$$

and

$$\xi^{(s)}_{lm,ij} = \xi^{(s)}_{lm,ij;1} = F^{(s)}_{lm}\left(X^{(s)}_{ij1}\right) - w^{(s)}_{lm,ij}.$$

Note that $\text{E}\left(\xi^{(s)}_{lm,ij}\right) = \text{E}\left(F^{(s)}_{lm}\left(X^{(s)}_{ij1}\right)\right) - w^{(s)}_{lm,ij} = 0$ and that $\text{E}\left(\xi^{(s)}_{lm,ij}\xi^{(s)}_{lm,i'j'}\right) = 0$ when $i \neq i'$ due to independence. In the case of $l = l', m = m', l \neq i$ and $s = s' = 1$, the element $s^{(s,s')}_{ij}(lm, l'm')$ is

$$\begin{aligned}
& s^{(1,1)}_{ij}(lm, lm) \\
&= N\text{Var}\left(Z^{(1)}_{lm,ij}\right) \\
&= N\text{Var}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\left[F^{(1)}_{lm}\left(X^{(1)}_{ijk}\right) - w^{(1)}_{lm,ij}\right] - \frac{1}{n_l}\sum_{k'=1}^{n_l}\left[F^{(1)}_{ij}\left(X^{(1)}_{lmk'}\right) - w^{(1)}_{ij,lm}\right]\right\} \\
&= \frac{N}{n_i}\text{E}\left\{\left[F^{(1)}_{lm}\left(X^{(1)}_{ij1}\right) - w^{(1)}_{lm,ij}\right]^2\right\} + \frac{N}{n_l}\left\{\left[F^{(1)}_{ij}\left(X^{(1)}_{lm1}\right) - w^{(1)}_{ij,lm}\right]^2\right\} \\
&= \frac{N}{n_i}\text{E}\left\{\left[\xi^{(1)}_{lm,ij}\right]^2\right\} + \frac{N}{n_l}\text{E}\left\{\left[\xi^{(1)}_{ij,lm}\right]^2\right\} \\
&= \gamma^{(1,1)}_{ij,ij}(lm, lm) + \gamma^{(1,1)}_{lm,lm}(ij, ij).
\end{aligned}$$

In the case of $i = i', j \neq j', l = l', l = i, m \neq j, m' \neq j'$ and $s = s' = 1$, the element $s^{(s,s')}_{ij,i'j'}(lm, l'm')$ is

$$
\begin{aligned}
& s^{(1,1)}_{ij,ij'}(im, im') \\
=& N\mathrm{Cov}\left(Z^{(1)}_{im,ij}, Z^{(1)}_{im',ij'}\right) \\
=& N\mathrm{Cov}\left\{\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im,ij;k} - \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij,im;k}\right], \left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im',ij';k} - \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij',im';k}\right]\right\} \\
=& N\mathrm{Cov}\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im,ij;k}, \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im',ij';k}\right] - N\mathrm{Cov}\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im,ij;k}, \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij',im';k}\right] - \\
& N\mathrm{Cov}\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij,im;k}, \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{im',ij';k}\right] + N\mathrm{Cov}\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij,im;k}, \frac{1}{n_i}\sum_{k=1}^{n_i}\xi^{(1)}_{ij',im';k}\right] \\
=& \frac{N}{n_i^2}\cdot n_i \cdot \mathrm{Cov}\left[\xi^{(1)}_{im,ij;1}, \xi^{(1)}_{im',ij';1}\right] - \frac{N}{n_i^2}\cdot n_i \cdot \mathrm{Cov}\left[\xi^{(1)}_{im,ij;1}, \xi^{(1)}_{ij',im';1}\right] - \\
& \frac{N}{n_i^2}\cdot n_i \cdot \mathrm{Cov}\left[\xi^{(1)}_{ij,im;1}, \xi^{(1)}_{im',ij';1}\right] + \frac{N}{n_i^2}\cdot n_i \cdot \mathrm{Cov}\left[\xi^{(1)}_{ij,im;1}, \xi^{(1)}_{ij',im';1}\right] \\
=& \frac{N}{n_i}\mathrm{E}\left[\xi^{(1)}_{im,ij}\cdot\xi^{(1)}_{im',ij'}\right] - \frac{N}{n_i}\mathrm{E}\left[\xi^{(1)}_{im,ij}\cdot\xi^{(1)}_{ij',im'}\right] - \frac{N}{n_i}\mathrm{E}\left[\xi^{(1)}_{ij,im}\cdot\xi^{(1)}_{im',ij'}\right] + \frac{N}{n_i}\mathrm{E}\left[\xi^{(1)}_{ij,im}\cdot\xi^{(1)}_{ij',im'}\right] \\
=& \gamma^{(1,1)}_{ij,ij'}(im, im') - \gamma^{(1,1)}_{ij,im'}(im, ij') - \gamma^{(1,1)}_{im,ij'}(ij, im') + \gamma^{(1,1)}_{im,im'}(ij, ij').
\end{aligned}
$$

All other nonzero cases can be calculated analogously. Moreover, using the same technique as above and independence, the zero cases can also be calculated.

Proof of Theorem 4.3.2

Proof. Now we show the multivariate normality of the statistic $\sqrt{N}(\widehat{\boldsymbol{p}} - \boldsymbol{p})$. There are two aspects to show multivariate normality. They are

1. the Asymptotic Equivalence Theorem in Theorem 4.3.1 and

2. Cramer-Wold device.

As shown in (4.7) that

$$\sqrt{N}Z^{(s)}_{i'j',ij} = \sqrt{N}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\left[F^{(s)}_{i'j'}\left(X^{(s)}_{ijk}\right) - w^{(s)}_{i'j',ij}\right] - \frac{1}{n_{i'}}\sum_{k'=1}^{n_{i'}}\left[F^{(s)}_{ij}\left(X^{(s)}_{i'j'k'}\right) - w^{(s)}_{ij,i'j'}\right]\right\},$$

where $i, i' = 1, \ldots, a; j, j' = 1, \ldots, b$. For illustration purpose, we are showing the case when there are two variables. That is, $c = 2$. (For cases when $c > 2$, the proof follows the same idea.) By (4.3) and (4.4), we have the following relationship,

$$\boldsymbol{p} = \boldsymbol{E}\boldsymbol{w} = \boldsymbol{I}_{ab} \otimes \left(\frac{1}{ab}\boldsymbol{1}_{ab}^{\top}\right) \otimes \boldsymbol{I}_2\boldsymbol{w}.$$

To prove multivariate normality, we need to apply Cramer-Wold device. Let

$$\boldsymbol{v} = \left(\boldsymbol{v}_{11}^{\top}, \ldots, \boldsymbol{v}_{1b}^{\top}, \ldots, \boldsymbol{v}_{a1}^{\top}, \ldots, \boldsymbol{v}_{ab}^{\top}\right)^{\top} = \left(v_{11}^{(1)}, v_{11}^{(2)}, v_{12}^{(1)}, v_{12}^{(2)}, \ldots, v_{ab}^{(1)}, v_{ab}^{(2)}\right)^{\top}$$

be arbitrary vectors of constants with $\|\boldsymbol{v}\| = 1$ and $\boldsymbol{v}_{ij} = \left(v_{ij}^{(1)}, v_{ij}^{(2)}\right)^{\top}$, where $i = 1, \ldots, a; j = 1, \ldots, b$. By asymptotic equivalence stated in Theorem 4.3.1, we have

$$\sqrt{N}\boldsymbol{v}^{\top}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right) \doteqdot \sqrt{N}\boldsymbol{v}^{\top}\boldsymbol{E}\boldsymbol{Z} \quad \text{and}$$

$$\sqrt{N}\boldsymbol{v}^{\top}\boldsymbol{E}\boldsymbol{Z} = \sqrt{N}\boldsymbol{v}^{\top} \underbrace{\begin{bmatrix} \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,11}^{(1)} \\ \vdots \\ \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,ab}^{(2)} \end{bmatrix}}_{2ab\times 1}. \tag{4.34}$$

We calculate the first element in the vector of (4.34). The others can be calculated in the same way. First note that

$$\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,11}^{(1)}$$

$$= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\left\{\frac{1}{n_1}\sum_{k=1}^{n_1}\left[F_{ij}^{(1)}\left(X_{11k}^{(1)}\right) - w_{ij,11}^{(1)}\right] - \frac{1}{n_i}\sum_{k'=1}^{n_i}\left[F_{11}^{(1)}\left(X_{ijk'}^{(1)} - w_{11,ij}^{(1)}\right)\right]\right\}$$

$$= \frac{1}{n_1}\sum_{k=1}^{n_1} G_1\left(X_{11k}^{(1)}\right) - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} w_{ij,11}^{(1)} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k'=1}^{n_i} F_{11}^{(1)}\left(X_{ijk'}^{(1)}\right) +$$

$$\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} w_{11,ij}^{(1)}$$

$$= \frac{1}{n_1}\sum_{k=1}^{n_1} G_1\left(X_{11k}^{(1)}\right) - p_{11}^{(1)} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k'=1}^{n_i} F_{11}^{(1)}\left(X_{ijk'}^{(1)}\right) + \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(1 - w_{ij,11}^{(1)}\right)$$

$$= \frac{1}{n_1}\sum_{k=1}^{n_1} G_1\left(X_{11k}^{(1)}\right) - p_{11}^{(1)} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k'=1}^{n_i} F_{11}^{(1)}\left(X_{ijk'}^{(1)}\right) + 1 - p_{11}^{(1)}$$

$$= \frac{1}{n_1}\sum_{k=1}^{n_1} G_1\left(X_{11k}^{(1)}\right) - \underbrace{\frac{1}{ab}\cdot\frac{1}{n_1}\sum_{k'=1}^{n_1} F_{11}^{(1)}\left(X_{11k'}^{(1)}\right)}_{(i,j)=(1,1)} -$$

$$\underbrace{\frac{1}{ab}\sum_{\substack{i=1 \\ (i,j)\neq(1,1)}}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k'=1}^{n_i} F_{11}^{(1)}\left(X_{ijk'}^{(1)}\right)}_{(i,j)\neq(1,1)} + 1 - 2p_{11}^{(1)}$$

$$= \frac{1}{n_1}\sum_{k=1}^{n_1}\left[G_1\left(X_{11k}^{(1)}\right) - \frac{1}{ab}F_{11}^{(1)}\left(X_{11k}^{(1)}\right)\right] - \frac{1}{ab}\sum_{\substack{i=1 \\ (i,j)\neq(1,1)}}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k'=1}^{n_i} F_{11}^{(1)}\left(X_{ijk'}^{(1)}\right) + 1 - 2p_{11}^{(1)}.$$

Then (4.34) becomes,

$$
\begin{aligned}
\sqrt{N}\boldsymbol{v}^{\top}\boldsymbol{E}\boldsymbol{Z} =& \sqrt{N}\boldsymbol{v}^{\top} \underbrace{\begin{bmatrix} \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,11}^{(1)} \\ \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,11}^{(2)} \\ \vdots \\ \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,ab}^{(1)} \\ \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,ab}^{(2)} \end{bmatrix}}_{2ab\times 1} \\
=& \sqrt{N}\sum_{g=1}^{a}\sum_{t=1}^{b}\sum_{s=1}^{2} v_{gt}^{(s)} \cdot \left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Z_{ij,gt}^{(s)}\right) \\
=& \sqrt{N}\sum_{g=1}^{a}\sum_{t=1}^{b}\sum_{s=1}^{2} v_{gt}^{(s)} \cdot \left\{\frac{1}{n_g}\sum_{k=1}^{n_g}\left[G_s\left(X_{gtk}^{(s)}\right) - \frac{1}{ab}F_{gt}^{(s)}\left(X_{gtk}^{(s)}\right)\right] - \right. \\
& \left. \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{(i,j)\neq(g,t)} \frac{1}{n_i}\sum_{k'=1}^{n_i} F_{gt}^{(s)}\left(X_{ijk'}^{(s)}\right) + 1 - 2p_{gt}^{(s)}\right\} \\
=& \sqrt{N}\sum_{g=1}^{a}\sum_{t=1}^{b}\sum_{s=1}^{2} \left\{\frac{1}{n_g}\sum_{k=1}^{n_g} v_{gt}^{(s)} \cdot \left[G_s\left(X_{gtk}^{(s)}\right) - \frac{1}{ab}F_{gt}^{(s)}\left(X_{gtk}^{(s)}\right)\right] - \right. \\
& \left. \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{(i,j)\neq(g,t)} \frac{1}{n_i}\sum_{k'=1}^{n_i} v_{gt}^{(s)} \cdot F_{gt}^{(s)}\left(X_{ijk'}^{(s)}\right) + v_{gt}^{(s)} \cdot \left(1 - 2p_{gt}^{(s)}\right)\right\} \\
=& \sqrt{N}\sum_{g=1}^{a}\sum_{t=1}^{b}\sum_{s=1}^{2} \left\{\frac{1}{n_g}\sum_{k=1}^{n_g} v_{gt}^{(s)} \cdot \left[G_s\left(X_{gtk}^{(s)}\right) - \frac{1}{ab}F_{gt}^{(s)}\left(X_{gtk}^{(s)}\right)\right] - \right. \\
& \left. \left[\frac{1}{n_g}\sum_{k=1}^{n_g}\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{(i,j)\neq(g,t)} v_{ij}^{(s)} \cdot F_{ij}^{(s)}\left(X_{gtk}^{(s)}\right)\right] + v_{gt}^{(s)} \cdot \left(1 - 2p_{gt}^{(s)}\right)\right\} \\
=& \sum_{g=1}^{a}\frac{\sqrt{N}}{n_g}\sum_{k=1}^{n_g}\left\{\sum_{t=1}^{b}\sum_{s=1}^{2} v_{gt}^{(s)} \cdot \left[G_s\left(X_{gtk}^{(s)}\right) - \frac{1}{ab}F_{gt}^{(s)}\left(X_{gtk}^{(s)}\right)\right] - \right. \\
& \left. \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{(i,j)\neq(g,t)} v_{ij}^{(s)} \cdot F_{ij}^{(s)}\left(X_{gtk}^{(s)}\right) + v_{gt}^{(s)} \cdot \left(1 - 2p_{gt}^{(s)}\right)\right\} \\
=& \sum_{g=1}^{a}\frac{\sqrt{N}}{n_g}\sum_{k=1}^{n_g} Z_{gk}^{\star} = \sum_{g=1}^{a}\frac{\sqrt{N}}{\sqrt{n_g}}\sum_{k=1}^{n_g}\frac{Z_{gk}^{\star}}{\sqrt{n_g}},
\end{aligned}
$$

where

$$Z_{gk}^{\star} = \sum_{t=1}^{b}\sum_{s=1}^{2} v_{gt}^{(s)} \cdot \left[G_s\left(X_{gtk}^{(s)}\right) - \frac{1}{ab}F_{gt}^{(s)}\left(X_{gtk}^{(s)}\right)\right] -$$

$$\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{(i,j)\neq(g,t)} v_{ij}^{(s)} \cdot F_{ij}^{(s)}\left(X_{gtk}^{(s)}\right) + v_{gt}^{(s)} \cdot \left(1 - 2p_{gt}^{(s)}\right),$$

for $g = 1, \dots, a; k = 1, \dots, n_g$. By assumption in (4.6),

$$\frac{\sqrt{N}}{\sqrt{n_g}} \to \kappa > 0 \text{ for } g = 1, \dots, a.$$

As we can see $Z_{gk}^{\star}$'s are independent and uniformly bounded random variables,

$$\mathrm{E}\left(\sqrt{N}\boldsymbol{v}^{\top}\boldsymbol{E}\boldsymbol{Z}\right) = \mathrm{E}\left(\sum_{g=1}^{a}\frac{\sqrt{N}}{\sqrt{n_g}}\sum_{k=1}^{n_g}\frac{Z_{gk}^{\star}}{\sqrt{n_g}}\right) = 0.$$

Let

$$S_{n,g} = \sum_{k=1}^{n_g} Z_{gk}^{\star},$$

and

$$s_{n,g}^2 = \mathrm{Var}\left(\sum_{k=1}^{n_g} Z_{gk}^{\star}\right) = \sum_{k=1}^{n_g}\underbrace{\mathrm{Var}\left(Z_{gk}^{\star}\right)}_{>\eta>0} \geqslant \sum_{k=1}^{n_g}\eta = n_g\eta \to \infty.$$

That is, $s_{n,g}^2 \to \infty$. By Theorem 27.3 in Billingsley (2008),

$$\frac{S_{n,g}}{s_{n,g}} = \frac{\sum_{k=1}^{n_g} Z_{gk}^{\star}}{\sqrt{\mathrm{Var}\left(\sum_{k=1}^{n_g} Z_{gk}^{\star}\right)}} \xrightarrow{d} N(0,1).$$

That is,

$$\frac{\sum_{k=1}^{n_g} Z_{gk}^{\star}/\sqrt{n_g}}{\sqrt{\mathrm{Var}\left(\sum_{k=1}^{n_g} Z_{gk}^{\star}/\sqrt{n_g}\right)}} \xrightarrow{d} N(0,1).$$

Note that,

$$\mathrm{Var}\left(\sum_{k=1}^{n_g} Z_{gk}^{\star}/\sqrt{n_g}\right) = \frac{1}{n_g}\sum_{k=1}^{n_g}\mathrm{Var}\left(Z_{gk}^{\star}\right) = \mathrm{Var}\left(Z_{g1}^{\star}\right) = u^2 \text{ (a constant).}$$

Thus,

$$\sum_{k=1}^{n_g} Z_{gk}^{\star}/\sqrt{n_g} \xrightarrow{d} N\left(0, u^2\right).$$

We know any linear combination of independent normal random variables is still normal. Hence, $\sqrt{N}\boldsymbol{v}^\top \boldsymbol{E}\boldsymbol{Z} = \sum_{g=1}^{a} \frac{\sqrt{N}}{\sqrt{n_g}} \sum_{k=1}^{n_g} \frac{Z^\star_{gk}}{\sqrt{n_g}}$ converges in distribution to some univariate normal distribution.

So far we showed that for any $\boldsymbol{v} \in \mathbb{R}^{2ab}$, $\sqrt{N}\boldsymbol{v}^\top \boldsymbol{E}\boldsymbol{Z}$ is univariate normal. By Cramer-Wold device, $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$ has asymptotically multivariate normal distribution. By Theorem 4.3.1 that $\sqrt{N}\boldsymbol{E}\boldsymbol{Z}$ and $\sqrt{N}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right)$ are asymptotically equivalent, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right) \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{V}\right).$$

□

Proof of Theorem 4.3.3

Proof. To show the L_2 consistency of the estimators $\widehat{\gamma}_{ij,ij'}^{(s,s')}\left(lm, l'm'\right)$ in equation (4.29),

$$\mathrm{E}\left[\widehat{\gamma}_{ij,ij'}^{(s,s')}\left(lm, l'm'\right) - \gamma_{ij,ij'}^{(s,s')}\left(lm, l'm'\right)\right]^2 = \frac{N^2}{n_i^2}\cdot$$

$$\mathrm{E}\left\{\underbrace{\frac{1}{n_i - 1}\sum_{k=1}^{n_i} D_{ijk}^{(s)}(lm) \cdot D_{ij'k}^{(s')}\left(l'm'\right)}_{S_1} - \right.$$

$$\left.\underbrace{\mathrm{E}\left[\left(F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right) - w_{lm,ij}^{(s)}\right)\cdot\left(F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right) - w_{l'm',ij'}^{(s')}\right)\right]}_{S_2}\right\}^2.$$

Observe that

$$
\begin{aligned}
S_1 =& \frac{1}{n_i-1}\sum_{k=1}^{n_i} D_{ijk}^{(s)}(lm)\cdot D_{ij'k}^{(s')}(l'm') \\
=& \frac{1}{n_i-1}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-\widehat{w}_{lm,ij}^{(s)}\right]\cdot\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\widehat{w}_{l'm',ij'}^{(s')}\right] \\
=& \frac{n_i}{n_i-1}\cdot\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{w}_{l'm',ij'}^{(s')}-\right. \\
&\left.\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)+\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}\right] \\
=& \frac{n_i}{n_i-1}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\widehat{w}_{l'm',ij'}^{(s')}-\right. \\
&\left.\widehat{w}_{lm,ij}^{(s)}\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]+\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}\right\} \\
=& \frac{n_i}{n_i-1}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}\right\}.
\end{aligned}
$$

Note that

$$
\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)=\int\widehat{F}_{lm}^{(s)}\,d\widehat{F}_{ij}^{(s)}=\widehat{w}_{lm,ij}^{(s)}\quad\text{and}
$$

$$
\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)=\int\widehat{F}_{l'm'}^{(s')}\,d\widehat{F}_{ij'}^{(s')}=\widehat{w}_{l'm',ij'}^{(s')}.
$$

Further,

$$
\begin{aligned}
S_2 =& \mathrm{E}\left[\left(F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)-w_{lm,ij}^{(s)}\right)\cdot\left(F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)-w_{l'm',ij'}^{(s')}\right)\right] \\
=& \mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]\cdot w_{l'm',ij'}^{(s')}- \\
& w_{lm,ij}^{(s)}\cdot\mathrm{E}\left[F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]+w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')} \\
=& \mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]-w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')}.
\end{aligned}
$$

Note that $\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right] = w_{lm,ij}^{(s)}$ and $\mathrm{E}\left[F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right] = w_{l'm',ij'}^{(s')}$. Now,

$$S_1 - S_2 = \frac{n_i}{n_i-1}\left\{\underbrace{\overbrace{\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)}^{a_1} - \overbrace{\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}}^{a_2}}_{A}\right\} -$$

$$\underbrace{\overbrace{\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]}^{b_1} - \overbrace{w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')}}^{b_2}}_{B}.$$

Then,

$$\begin{aligned}
S_1 - S_2 &= \frac{n_i}{n_i-1}A - B\\
&= \frac{n_i}{n_i-1}(A-B) + \frac{1}{n_i-1}B\\
&= \frac{n_i}{n_i-1}\left[(a_1-a_2)-(b_1-b_2)\right] + \frac{1}{n_i-1}B\\
&= \frac{n_i}{n_i-1}\left[(a_1-b_1)-(a_2-b_2)\right] + \frac{1}{n_i-1}B.
\end{aligned}$$

Then by c_r-inequality, (i.e., $\mathrm{E}\mid X+Y\mid^r \leqslant c_r \mid X\mid^r + c_r\mid Y\mid^r$, where $c_r = 1$ for $0 < r < 1$, and $c_r = 2^{r-1}$ for $r \geqslant 1$),

$$\begin{aligned}
&\mathrm{E}\left[\widehat{\gamma}_{ij,ij'}^{(s,s')}(lm,l'm') - \gamma_{ij,ij'}^{(s,s')}(lm,l'm')\right]^2\\
&= \frac{N^2}{n_i^2}\mathrm{E}\left\{\frac{n_i}{n_i-1}\left[(a_1-b_1)-(a_2-b_2)\right] + \frac{1}{n_i-1}B\right\}^2\\
&\leqslant \frac{N^2}{n_i^2}\left\{2\mathrm{E}\left[\left(\frac{n_i}{n_i-1}\left[(a_1-b_1)-(a_2-b_2)\right]\right)^2\right] + \frac{2}{(n_i-1)^2}\underbrace{\mathrm{E}\left(B^2\right)}_{\leqslant 1}\right\}\\
&\leqslant \frac{N^2}{n_i^2}\left\{\frac{4n_i^2}{(n_i-1)^2}\left[\mathrm{E}(a_1-b_1)^2 + \mathrm{E}(a_2-b_2)^2\right] + \frac{2}{(n_i-1)^2}\right\}.
\end{aligned}$$

Note that

$$B = \mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right] - w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')},$$

where $F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right), F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right) \in [0,1]$ and $w_{lm,ij}^{(s)}, w_{l'm',ij'}^{(s')} \in [0,1]$, which makes $B \in [-1,1]$. Also, the last inequality holds because $(x-y)^2 = x^2 - 2xy + y^2 \leqslant 2x^2 + 2y^2$.

We show $\mathrm{E}\left[\widehat{\gamma}_{ij,ij'}^{(s,s')}(lm,l'm') - \gamma_{ij,ij'}^{(s,s')}(lm,l'm')\right]^2 \to 0$ by showing the following,

$$\mathrm{E}(a_1-b_1)^2 \to 0 \quad \text{and} \tag{4.35}$$

$$\mathrm{E}(a_2-b_2)^2 \to 0. \tag{4.36}$$

By assumption, $\min\limits_{1\leqslant i\leqslant a} n_i \to \infty$,

$$\frac{2}{(n_i-1)^2} \to 0. \tag{4.37}$$

To show (4.35),

$$\begin{aligned}
&\mathrm{E}\left(a_1-b_1\right)^2\\
&=\mathrm{E}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]\right\}^2\\
&=\mathrm{E}\left\{\underbrace{\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\frac{1}{n_i}\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)}_{C_1}+\right.\\
&\left.\underbrace{\frac{1}{n_i}\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]}_{C_2}\right\}^2.
\end{aligned}$$

By c_r-inequality,

$$\mathrm{E}\left(a_1-b_1\right)^2 \leqslant 2\mathrm{E}\left(C_1^2\right)+2\mathrm{E}\left(C_2^2\right). \tag{4.38}$$

Next, we show $\mathrm{E}(C_2^2)\to 0$. Let $S_{n_i}=\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)$ and $\mu = \mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'1}^{(s')}\right)\right]$. By the assumption of $\min\limits_{1\leqslant i\leqslant a} n_i\to\infty$ and L_2 weak law, we have the following,

$$\begin{aligned}
\mathrm{E}(C_2^2) &= \mathrm{E}\left(\frac{S_{n_i}}{n_i}-\mu\right)^2\\
&=\mathrm{Var}\left(\frac{S_{n_i}}{n_i}\right)\\
&=\frac{1}{n_i^2}\sum_{k=1}^{n_i}\mathrm{Var}\left(F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right)\\
&=\mathcal{O}\left(\frac{1}{n_i}\right)\to 0 \quad \text{as} \quad n_i\to\infty.
\end{aligned} \tag{4.39}$$

Finally, we show $\mathrm{E}\left(C_1^2\right)\to 0$. In order to do so, we first calculate the following inequality by adding and subtracting two terms. Note that for any fixed k in the fixed group i, we have

$$
\begin{aligned}
&\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\mid\\
=&\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\frac{1}{2}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-\\
&\frac{1}{2}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)+\\
&\frac{1}{2}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)+\frac{1}{2}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\mid\\
=&\mid \frac{1}{2}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]+\\
&\frac{1}{2}\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]+\\
&\frac{1}{2}F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]+\\
&\frac{1}{2}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]\mid\\
=&\mid \frac{1}{2}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)+F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]+\\
&\frac{1}{2}\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)+F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\mid\\
\leqslant&\mid \frac{1}{2}\underbrace{\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)+F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]}_{\leqslant 2}\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]\mid+\\
&\mid \frac{1}{2}\underbrace{\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)+F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]}_{\leqslant 2}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\mid\\
\leqslant&\mid \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\mid+\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\mid.
\end{aligned}
$$

That is,

$$
\begin{aligned}
&\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\mid\leqslant\\
&\mid \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)-F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\mid+\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\mid.
\end{aligned}
\tag{4.40}
$$

Now by the inequality in (4.40), we get

$$
\begin{aligned}
&\mathrm{E}\left(C_1^2\right)\\
&= \mathrm{E}\left(\mid C_1 \mid^2\right)\\
&= \frac{1}{n_i^2}\mathrm{E}\left\{\mid \sum_{k=1}^{n_i} \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid^2\right\}\\
&\leqslant \frac{1}{n_i^2}\mathrm{E}\left\{\left[\sum_{k=1}^{n_i} \mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid\right]^2\right\}\\
&= \mathrm{E}\left\{\left[\frac{1}{n_i}\sum_{k=1}^{n_i} \mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \cdot F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid\right]^2\right\}\\
&\leqslant \mathrm{E}\left\{\left[\frac{1}{n_i}\sum_{k=1}^{n_i} \mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \mid + \mid \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid\right]^2\right\}
\end{aligned}
$$

By c_r-inequality,

$$
\begin{aligned}
&\leqslant 2\mathrm{E}\left\{\left[\frac{1}{n_i}\sum_{k=1}^{n_i} \mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \mid\right]^2\right\} +\\
&\quad 2\mathrm{E}\left\{\left[\frac{1}{n_i}\sum_{k=1}^{n_i} \mid \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid\right]^2\right\}
\end{aligned}
$$

By proposition about bounds on the absolute moments of a sum/averaege of random variables,

$$
\begin{aligned}
&\leqslant \frac{2}{n_i}\sum_{k=1}^{n_i}\mathrm{E}\left\{\mid \widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) \mid^2\right\} +\\
&\quad \frac{2}{n_i}\sum_{k=1}^{n_i}\mathrm{E}\left\{\mid \widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) \mid^2\right\}\\
&= \frac{2}{n_i}\sum_{k=1}^{n_i}\underbrace{\mathrm{E}\left\{\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\}}_{C_{1.1}} +\\
&\quad \frac{2}{n_i}\sum_{k=1}^{n_i}\underbrace{\mathrm{E}\left\{\left[\widehat{F}_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right) - F_{l'm'}^{(s')}\left(X_{ij'k}^{(s')}\right)\right]^2\right\}}_{C_{1.2}}\\
&= \frac{2}{n_i}\sum_{k=1}^{n_i} C_{1.1} + \frac{2}{n_i}\sum_{k=1}^{n_i} C_{1.2}.
\end{aligned}
\tag{4.41}
$$

Next, we show $C_{1.1} \to 0$ and $C_{1.2} \to 0$. Firstly, for any fixed k in the fixed group i, we show $C_{1.1} \to 0$. Note that $\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) = \frac{1}{n_l}\sum_{k'=1}^{n_l} c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right)$.

$$
\begin{aligned}
C_{1.1} =& \mathrm{E}\left\{\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\} \\
=& \mathrm{E}\left\{\left[\frac{1}{n_l}\sum_{k'=1}^{n_l} c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\} \\
=& \frac{1}{n_l^2}\mathrm{E}\left\{\left[\sum_{k'=1}^{n_l}\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\right]^2\right\}
\end{aligned}
$$

By Fubini's Theorem,

$$
\begin{aligned}
=& \frac{1}{n_l}\sum_{k'=1}^{n_l}\sum_{k''=1}^{n_l}\mathrm{E}\left\{\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\cdot\right. \\
& \left.\left[c\left(X_{ijk}^{(s)} - X_{lmk''}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\right\}
\end{aligned}
$$

Apply conditional expectation $\mathrm{E}(X) = \mathrm{E}\left[\mathrm{E}(X|Y)\right]$ for $k' \neq k''$ cases,

$$
\begin{aligned}
=& \frac{1}{n_l^2}\sum_{\substack{k'=1 \\ k'\neq k''}}^{n_l}\sum_{k''=1}^{n_l}\mathrm{E}\left\{\mathrm{E}\left(\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]\cdot\right.\right. \\
& \left.\left.\left[c\left(X_{ijk}^{(s)} - X_{lmk''}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right] |X_{ijk}^{(s)}\right)\right\} \\
& + \frac{1}{n_l^2}\sum_{k'=k''=1}^{n_l}\mathrm{E}\left\{\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\} \\
=& \frac{1}{n_l^2}\sum_{\substack{k'=1 \\ k'\neq k''}}^{n_l}\sum_{k''=1}^{n_l}\mathrm{E}\left\{\underbrace{\mathrm{E}\left(\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right] |X_{ijk}^{(s)}\right)}_{=0}\cdot\right. \\
& \left.\underbrace{\mathrm{E}\left(\left[c\left(X_{ijk}^{(s)} - X_{lmk''}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right] |X_{ijk}^{(s)}\right)}_{=0}\right\} + \\
& \frac{1}{n_l^2}\sum_{k'=k''=1}^{n_l}\mathrm{E}\left\{\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\} \\
=& \frac{1}{n_l^2}\sum_{k'=k''=1}^{n_l}\mathrm{E}\left\{\left[c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]^2\right\} \leqslant \frac{1}{n_l} \to 0 \quad \text{as} \quad n_l \to \infty.
\end{aligned}
\tag{4.42}
$$

Note that $|c\left(X_{ijk}^{(s)} - X_{lmk'}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)|X_{ijk}^{(s)}| \leqslant 1$.

Similarly,

$$C_{1.2} \to 0 \quad \text{as } n_{l'} \to \infty. \tag{4.43}$$

Therefore, by (4.41), (4.42) and (4.43), we have

$$\mathrm{E}\left(C_1^2\right) \to 0 \quad \text{as } \min(n_l, n_{l'}) \to \infty. \tag{4.44}$$

Now, by (4.38), (4.39) and (4.44), we showed (4.35),

$$\mathrm{E}\left(a_1 - b_1\right)^2 \to 0 \quad \text{as } \min(n_i, n_l, n_{l'}) \to \infty. \tag{4.45}$$

Next, we show (4.36). In order to do so, we calculate the following inequality first, by adding and subtracting two terms. For any fixed k in the fixed group i, we get

$$\begin{aligned}
&|\widehat{w}^{(s)}_{lm,ij}\widehat{w}^{(s')}_{l'm',ij'} - w^{(s)}_{lm,ij}w^{(s')}_{l'm',ij'}| \\
&= |\widehat{w}^{(s)}_{lm,ij}\widehat{w}^{(s')}_{l'm',ij'} - \frac{1}{2}\widehat{w}^{(s)}_{lm,ij}w^{(s')}_{l'm',ij'} - \frac{1}{2}w^{(s)}_{lm,ij}\widehat{w}^{(s')}_{l'm',ij'} - \\
&\quad w^{(s)}_{lm,ij}w^{(s')}_{l'm',ij'} + \frac{1}{2}\widehat{w}^{(s)}_{lm,ij}w^{(s')}_{l'm',ij'} + \frac{1}{2}w^{(s)}_{lm,ij}\widehat{w}^{(s')}_{l'm',ij'}| \\
&= |\frac{1}{2}\widehat{w}^{(s)}_{lm,ij}\left(\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}\right) + \frac{1}{2}\widehat{w}^{(s')}_{l'm',ij'}\left(\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}\right) + \\
&\quad \frac{1}{2}w^{(s)}_{lm,ij}\left(\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}\right) + \frac{1}{2}w^{(s')}_{l'm',ij'}\left(\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}\right)| \\
&= |\frac{1}{2}\left(\widehat{w}^{(s)}_{lm,ij} + w^{(s)}_{lm,ij}\right)\left(\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}\right) + \frac{1}{2}\left(\widehat{w}^{(s')}_{l'm',ij'} + w^{(s')}_{l'm',ij'}\right)\left(\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}\right)| \\
&\leqslant |\frac{1}{2}\underbrace{\left(\widehat{w}^{(s)}_{lm,ij} + w^{(s)}_{lm,ij}\right)}_{0\leqslant\ldots\leqslant 2}\left(\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}\right)| + |\frac{1}{2}\underbrace{\left(\widehat{w}^{(s')}_{l'm',ij'} + w^{(s')}_{l'm',ij'}\right)}_{0\leqslant\ldots\leqslant 2}\left(\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}\right)| \\
&\leqslant |\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}| + |\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}|.
\end{aligned}$$

That is,

$$|\widehat{w}^{(s)}_{lm,ij}\widehat{w}^{(s')}_{l'm',ij'} - w^{(s)}_{lm,ij}w^{(s')}_{l'm',ij'}| \leqslant |\widehat{w}^{(s')}_{l'm',ij'} - w^{(s')}_{l'm',ij'}| + |\widehat{w}^{(s)}_{lm,ij} - w^{(s)}_{lm,ij}|. \tag{4.46}$$

By (4.46), we have

$$\begin{aligned}
\mathrm{E}\left(a_2-b_2\right)^2 =& \mathrm{E}\left(\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}-w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')}\right)^2\\
=&\mathrm{E}\left\{|\widehat{w}_{lm,ij}^{(s)}\cdot\widehat{w}_{l'm',ij'}^{(s')}-w_{lm,ij}^{(s)}\cdot w_{l'm',ij'}^{(s')}|\right\}^2\\
\leqslant&\mathrm{E}\left\{|\widehat{w}_{lm,ij}^{(s)}-w_{lm,ij}^{(s)}|+|\widehat{w}_{l'm',ij'}^{(s')}-w_{l'm',ij'}^{(s')}|\right\}^2\\
&\text{By } c_r\text{-inequality,}\\
\leqslant&2\underbrace{\mathrm{E}\left\{|\widehat{w}_{lm,ij}^{(s)}-w_{lm,ij}^{(s)}|^2\right\}}_{D_1}+2\underbrace{\mathrm{E}\left\{|\widehat{w}_{l'm',ij'}^{(s')}-w_{l'm',ij'}^{(s')}|^2\right\}}_{D_2}\\
=&2D_1+2D_2.
\end{aligned}\tag{4.47}$$

To show (4.36), we only need to show $D_1\to 0$ and $D_2\to 0$. Below, we first show $D_1\to 0$. Note that $w_{lm,ij}^{(s)}=\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]$ and $\widehat{w}_{lm,ij}^{(s)}=\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)$, where $\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)=\frac{1}{n_l}\sum_{k'=1}^{n_l}c\left(X_{ijk}^{(s)}-X_{lmk'}^{(s)}\right)$.

$$\begin{aligned}
&D_1=\mathrm{E}\left\{|\widehat{w}_{lm,ij}^{(s)}-w_{lm,ij}^{(s)}|^2\right\}\\
&=\mathrm{E}\left\{|\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]+\frac{1}{n_i}\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]|^2\right\}\\
&\leqslant\mathrm{E}\left\{\left[|\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]|+|\frac{1}{n_i}\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]|\right]^2\right\}\\
&\leqslant 2\underbrace{\mathrm{E}\left\{|\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]|^2\right\}}_{D_{1.1}}+\\
&\quad 2\underbrace{\mathrm{E}\left\{|\frac{1}{n_i}\sum_{k=1}^{n_i}F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)-\mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]|^2\right\}}_{D_{1.2}}\\
&=2D_{1.1}+2D_{1.2}.
\end{aligned}\tag{4.48}$$

We check $D_{1.2}$ first. Similar to the proof in (4.39), we get

$$\begin{aligned} D_{1.2} &= \mathrm{E}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i} F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - \mathrm{E}\left[F_{lm}^{(s)}\left(X_{ij1}^{(s)}\right)\right]\right\}^2 \\ &= \mathrm{Var}\left\{\frac{1}{n_i}\sum_{k=1}^{n_i} F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right\} \\ &= \frac{1}{n_i^2}\sum_{k=1}^{n_i} \mathrm{Var}\left(F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right) = \mathcal{O}\left(\frac{1}{n_i}\right) \to 0 \quad \text{as} \quad n_i \to \infty. \end{aligned} \tag{4.49}$$

Next, we check $D_{1.1}$. By (4.42) and the proposition about bounds on the absolute moments of a sum/average of random variables, we have

$$\begin{aligned} D_{1.1} =& \mathrm{E}\left\{|\frac{1}{n_i}\sum_{k=1}^{n_i}\left[\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)\right]|^2\right\} \\ \leqslant& \frac{1}{n_i}\sum_{k=1}^{n_i} \mathrm{E}\left\{|\widehat{F}_{lm}^{(s)}\left(X_{ijk}^{(s)}\right) - F_{lm}^{(s)}\left(X_{ijk}^{(s)}\right)|^2\right\} \leqslant \frac{1}{n_l} \to 0 \quad \text{as} \quad n_l \to \infty. \end{aligned} \tag{4.50}$$

So far, by (4.48), (4.49), and (4.50), we showed that

$$D_1 \to 0 \quad \text{as} \;\; \min\left(n_i, n_l\right) \to \infty. \tag{4.51}$$

Similarly, following the same idea, we can also show

$$D_2 \to 0 \quad \text{as} \;\; \min\left(n_i, n_{l'}\right) \to \infty. \tag{4.52}$$

Thus, by (4.47), (4.51) and (4.52), we showed (4.36),

$$\mathrm{E}\left(a_2 - b_2\right)^2 \to 0 \quad \text{as} \;\; \min\left(n_i, n_l, n_{l'}\right) \to \infty. \tag{4.53}$$

Hence, by (4.37),(4.45) and,(4.53), we showed

$$\mathrm{E}\left[\widehat{\gamma}_{ij,ij'}^{(s,s')}\left(lm, l'm'\right) - \gamma_{ij,ij'}^{(s,s')}\left(lm, l'm'\right)\right]^2 \to 0 \quad \text{as} \;\; \min\left(n_i, n_l, n_{l'}\right) \to \infty.$$

Therefore, $\widehat{\gamma}_{ij,ij'}^{(s,s')}\left(lm, l'm'\right)$ are L_2 consistent estimators of $\gamma_{ij,ij'}^{(s,s')}\left(lm\text{'}l'm'\right)$.

□

Chapter 5 Summary

In this dissertation, novel nonparametric testing approaches for analyzing multivariate growth curve data are introduced. Without assuming multivariate normality and homogeneity conditions, these methods have favorable numerical performance. Based on the generalized MANOVA, a modified Wilks' Lambda test statistic is proposed to mitigate the effect of potential heteroscedasticity. Moreover, the invariance property of the null distribution of the test statistic makes the method applicable to data that are skewed and heavy-tailed. We also propose Wald-type resampling-based test statistics for general factorial designs, where permutation, bootstrap and hybrid permutation-bootstrap procedures are investigated. For the permutation test, the studentization technique is used to correct the covariance of the permuted data leading to an asymptotically exact permutation test despite the time dependencies. The parametric and wild bootstrap tests are also shown to be asymptotically exact under the nonparametric settings. Additionally, we propose fully nonparametric tests where nonparametric treatment effects are characterized in terms of functionals of distribution functions with the only assumption of nondegenerate marginal distributions. The asymptotic properties of the nonparametric treatment effects are derived and, therefore, confidence intervals can be calculated. These rank-based tests can be applied to both metric and nonmetric data.

For applications in practice, we recommend practitioners to use the robust modified Wilks' Lambda test rather than the classical methods for multivariate growth curve data, especially, in unbalanced heteroscedastic designs with nonnormal data. Although the modified Wilks' Lambda test has satisfactory performance, it involves approximations. Our Wald-type resampling-based tests, on the other hand, are asymptotically exact and can be applied to skewed and heavy-tailed data with small samples in real-world applications. In cases of nonmetric data, the aforementioned tests are no longer appropriate. The fully nonparametric rank-based methods accommodate discrete, ordered categorical and continuous data that could possibly be skewed and heavy-tailed in a unified manner.

For future research, we plan to explore more parametric and nonparametric resampling schemes that are appropriate for nonmetric data. Also, based on the nonparametric effect measures, we plan to investigate other multiple contrast statistics. Moreover, we plan to study resampling methods in the presence of covariates for multivariate growth curve data.

Bibliography

Akritas, M. G. and Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs i: Multivariate repeated measures designs. *Journal of the American Statistical Association*, 89(425):336–343.

Akritas, M. G. and Brunner, E. (1997). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61(2):249–277.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. John Wiley & Sons, Hoboken, NJ.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. McGraw-Hill, New York.

Boik, R. J. (1988). The mixed model for multivariate repeated measures: validity conditions and an approximate test. *Psychometrika*, 53(4):469–486.

Bray, J. H. and Maxwell, S. E. (1985). *Multivariate analysis of variance*. Quantitative applications in the social sciences. Beverly Hills : Sage Publications.

Brunner, E. and Denker, M. (1994). Rank statistics under dependent observations and applications to factorial designs. *Journal of Statistical planning and Inference*, 42(3):353–378.

Brunner, E., Domhof, S., and Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*.

Brunner, E., Konietschke, F., Pauly, M., and Puri, M. L. (2017). Rank-based procedures in factorial designs: Hypotheses about non-parametric treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1463–1485.

Brunner, E. and Munzel, U. (2000). The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 42(1):17–25.

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.

Brunner, E. and Neumann, N. (1982). Rank tests for correlated random variables. *Biometrical Journal*, 24(4):373–389.

Brunner, E. and Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical papers*, 42(1):1–52.

Brunner, E., Puri, M. L., and Sun, S. (1995). Nonparametric methods for stratified two-sample designs with application to multiclinic trials. *Journal of the American Statistical Association*, 90(431):1004–1014.

Chung, E. and Romano, J. (2013). Exact and asymptotically robust permutation tests. *Annals of Statistics*, 41(2):484–507.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Academic press.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169.

Fadardi, M. S., Bathke, A. C., Harrar, S. W., and Abel, L. A. (2017). Task-induced changes in idiopathic infantile nystagmus vary with gaze. *Optometry and Vision Science*, 94(5):606–615.

Fan, C. and Zhang, D. (2014). Wald-type rank tests: A gee approach. *Computational Statistics & Data Analysis*, 74:1–16.

Fan, C. and Zhang, D. (2017). On power and sample size of the anova-type rank test. *Communications in Statistics-Simulation and Computation*, 46(4):3224–3241.

Fligner, M. A. and Policello, G. E. (1981). Robust rank procedures for the behrens-fisher problem. *Journal of the American Statistical Association*, 76(373):162–168.

Friedrich, Sarah, Brunner, Edgar, Pauly, and Markus (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, 153:255–265.

Gamage, J. and Mathew, T. (2008). Inference on mean sub-vectors of two multivariate normal populations with unequal covariance matrices. *Statistics & probability letters*, 78(4):420–425.

Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer Series in Statistics. Springer, New York.

Grender, J. M. and Johnson, W. D. (1994). Fitting multivariate polynomial growth curves in two-period crossover designs. *Statistics in medicine*, 13(9):931–943.

Harrar, S. W. and Bathke, A. C. (2012a). Erratum to: A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations aism (2012) 64:135-165. *Annals of the Institute of Statistical Mathematics*, 64(5):1087–1087.

Harrar, S. W. and Bathke, A. C. (2012b). A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations. *Annals of the Institute of Statistical Mathematics*, 64(1):135–165.

Huang, Y., Xu, H., Calian, V., and Hsu, J. C. (2006). To permute or not to permute. *Bioinformatics*, 22(18):2244–2248.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d, hypotheses and the generalized behrens-fisher problem. *Statistics and Probability Letters*, 36(1):9–12.

Janssen, A. (2005). Resampling student's t-type statistics. *Annals of the Institute of Statistical Mathematics*, 57(3):507–529.

Janssen, A. and Pauls, T. (2003). How do bootstrap and permutation tests work? *Annals of Statistics*, 31(3):768–806.

Johnson, R. A., Wichern, D. W., et al. (2002). *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ.

Konietschke, F., Bathke, A. C., Harrar, S. W., and Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general manova. *Journal of Multivariate Analysis*, 140:291–301.

Konietschke, F., Harrar, S. W., Lange, K., and Brunner, E. (2012a). Ranking procedures for matched pairs with missing data - asymptotic theory and a small sample approximation. *Computational Statistics & Data Analysis*, 56(5):1090–1102.

Konietschke, F., Hothorn, L. A., Brunner, E., et al. (2012b). Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of*

Statistics, 6:738–759.

Krishnamoorthy, K. and Lu, F. (2010). A parametric bootstrap solution to the manova under heteroscedasticity. *Journal of Statistical Computation and Simulation*, 80(8):873–887.

Krishnamoorthy, K. and Yu, J. (2004). Modified nel and van der merwe test for the multivariate behrens–fisher problem. *Statistics & probability letters*, 66(2):161–169.

Kruskal, W. H. et al. (1952). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23(4):525–540.

Li, B., Zhou, Z., Zhang, L., and Yang, Y. (2020). Testing treatment-by-period interaction in four-period crossover trials. *Pharmaceutical Statistics*, 19(2):145–163.

Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Annals of Statistics*, 7(2):381–394.

Munzel, U. (1999). Linear rank score statistics when ties are present. *Statistics & Probability Letters*, 41(4):389–395.

Naik, D. N. and Rao, S. S. (2001). Analysis of multivariate repeated measures data with a kronecker product structured covariance matrix. *Journal of Applied Statistics*, 28(1):91–105.

Neubert, K. and Brunner, E. (2007). A studentized permutation test for the non-parametric behrens–fisher problem. *Computational Statistics and Data Analysis*, 51(10):5192–5204.

Neuhaus, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Annals of Statistics*, 21(4):1760–1779.

O'Brien, P. N., Parenté, F. J., and Schmitt, C. J. (1982). A monte carlo study on the robustness of four manova criterion tests. *Journal of Statistical Computation and Simulation*, 15(2-3):183–192.

Omelka, M. and Pauly, M. (2012). Testing equality of correlation coefficients in two populations via permutation methods. *Journal of Statistical Planning and Inference*, 142(6):1396–1406.

Pauly, M., Brunner, E., and Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):461–473.

Pauly, M. et al. (2011). Weighted resampling of martingale difference arrays with applications. *Electronic Journal of Statistics*, 5:41–52.

Pavur, R. J. (1987). Distribution of multivariate quadratic forms under certain covariance structures. *Canadian Journal of Statistics*, 15(2).

Pesarin, F. (2001). *Multivariate permutation tests : with applications in biostatistics.* J. Wiley, New York.

Pesarin, F. and Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software.* Wiley Series in Probability and Statistics. J. Wiley, New York.

Rakočević, V. (1997). On continuity of the moore-penrose and drazin inverses. *Matematichki Vesnik*, 49(3-4):163–172.

Rao, C. R. and Mitra, S. K. (1971). *Generalized inverse of matrices and its applications.* Wiley, New York.

Rencher, A. C. (2001). *Methods of multivariate analysis.* Wiley series in probability and mathematical statistics. J. Wiley, New York.

Ruymgaart, F. H. (1980). A unified approach to the asymptotic distribution theory of certain midrank statistics. In *Statistique non Parametrique Asymptotique*, pages 1–18. Springer.

Schott, J. R. (2016). *Matrix analysis for statistics.* John Wiley & Sons.

Suo, C., Toulopoulou, T., Bramon, E., Walshe, M., Picchioni, M., Murray, R., and Ott, J. (2013). Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC bioinformatics*, 14(1):1–7.

Thompson, G. (1990). Asymptotic distribution of rank statistics under dependencies with multivariate application. *Journal of Multivariate Analysis*, 33(2):183–211.

Thompson, G. (1991). A unified approach to rank tests for multivariate and repeated measures designs. *Journal of the American Statistical Association*, 86(414):410–

419.

Timm, N. H. (1980). 2 multivariate analysis of variance of repeated measurements. *Handbook of statistics*, 1:41–87.

Timm, N. H. (2002). *Applied multivariate analysis*. Springer texts in statistics. Springer, New York.

Vallejo, G., Fidalgo, A., and Fernandez, P. (2001). Effects of covariance heterogeneity on three procedures for analyzing multivariate repeated measures designs. *Multivariate Behavioral Research*, 36(1):01–27.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295.

Xu, J. and Cui, X. (2008). Robustified manova with applications in detecting differentially expressed genes from oligonucleotide arrays. *Bioinformatics*, 24(8):1056–1062.

Xu, L. W. (2015). Parametric bootstrap approaches for two-way manova with unequal cell sizes and unequal cell covariance matrices. *Journal of Multivariate Analysis*, 133:291–303.

Xu, L. W., Yang, F. Q., Abula, A., and Qin, S. (2013). A parametric bootstrap approach for two-way anova in presence of possible interactions with unequal variances. *Journal of Multivariate Analysis*, 115:172–180.

Zeng, T. and Harrar, S. W. (2021a). Nonparametric tests for multivariate growth curve data: Practical procedures for finite samples [unpublished manuscript]. *University of Kentucky*.

Zeng, T. and Harrar, S. W. (2021b). Robust tests for trials with multiple endpoints [unpublished manuscript]. *University of Kentucky*.

Zhang, J. T. (2012). An approximate hotelling t2-test for heteroscedastic one-way manova. *Open Journal of Statistics*, 2(1):1–11.

Zhang, J. T. and Liu, X. (2013). A modified bartlett test for heteroscedastic one-way manova. *Metrika*, 76(1):135–152.

Vita

Ting Zeng

Education

- M.S. in Statistics, University of Kentucky
- B.S. in Mathematics, Centre College
- B.E. in Software Engineering, Sichuan University

Experience

- Teaching Assistant, Department of Statistics, University of Kentucky, 2016 - 2021
- Research Assistant, Applied Statistics Lab, University of Kentucky, 2017 - 2020
- Primary Instructor, Department of Statistics, University of Kentucky, 2018 - 2019

Publications

- **Zeng, T.** and Harrar, S. W. (2021). Robust tests for trials with multiple endpoints. *Pharmaceutical Statistics.* (Submitted)
- Elliott, C. F., Lambert, J. W., Stromberg, A. J., Wang, P., **Zeng, T.**, & Thompson, K. L. (2020). Feasibility as a mechanism for model identification and validation. *Journal of Applied Statistics*, 1-20.