2012

Inference for High-Dimensional Doubly Multivariate Data under General Conditions

John Zachary Hossler
The University of Montana

INFERENCE FOR HIGH-DIMENSIONAL DOUBLY MULTIVARIATE DATA UNDER GENERAL CONDITIONS

By

John Zachary Hossler

B.S., George Fox University, Newberg, Oregon, 2006
M.A., The University of Montana, Missoula, Montana, 2008

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctorate of Philosophy
in Mathematics

The University of Montana
Missoula, MT

May 2012

Approved by:

Dr. Sandy Ross, Associate Dean
Graduate School

Dr. Solomon W. Harrar, Committee Chair
Mathematical Sciences

Dr. Jon Graham
Mathematical Sciences

Dr. David Patterson
Mathematical Sciences

Dr. Johnathan Bardsley
Mathematical Sciences

Dr. Jesse V. Johnson
Computer Science

Hossler, John Z., Doctorate of Philosophy, May 2012 Mathematics

Inference for High-Dimensional Doubly Multivariate Data under General Conditions

Committee Chair: Solomon W. Harrar, Ph.D.

With technological, research, and theoretical advancements, the amount of data being generated for analysis is growing rapidly. In many cases, the number of subjects may be small, but the number of measurements taken on each subject may be very large. Consider, for example, two groups of patients. The subjects in one group are diseased and the other subjects are not. Over 9,000 relative fluorescent unit (RFU) signals, measures of the presence and abundance of proteins, are collected in a microarray or protoarray from each subject. Typically these kind of data show marked skewness (departure from normality) which invalidates standard multivariate normal-based theory. What is more, due to the cost involved, only a limited number of subjects can be included in the study. Therefore, standard large-sample asymptotic theory cannot be applied. It is of interest to determine if there are any differences in RFU signals between the two groups, and more importantly, if there are any RFU signal and group interaction effects. If such an interaction is detected, further research is warranted to identify any of these biological signals, commonly known as biomarkers.

To address these types of phenomena, we present inferential procedures in two-factor repeated measures multivariate analysis of variance (RM-MANOVA) models where the covariance structure is unknown and the number of measurements per subject tends to infinity. Both in the univariate case, in which the number of dimensions or response variables is one, and the multivariate case, in which there are several response variables, different sums of squares and cross product matrices are proposed to compensate for the unknown structure of the covariance matrix and unbalanced group sizes. Based on the new matrices, we present some multivariate test statistics, deriving their asymptotic distributions under fairly general conditions. We then use simulation results to assess the performance of the tests, and we analyze a real data set to demonstrate their applicability.

Keywords: MANOVA · Repeated measures · Longitudinal data · High-dimensionality · Unstructured variance/covariance · Non-normality · Stationarity · α-mixing · Central limit theorem · Kronecker product · Bootstrapping · Simulation study

Acknowledgments

First and foremost, for all of the wonderful blessings in my life, I would like to thank God our Father and the Lord Jesus Christ, without whom I could do and would be nothing. I would also like to thank my advisor, Dr. Solomon W. Harrar, for his many hours of assistance, patience, guidance, instruction, and conversation; the members of my Dissertation Committee: Jon, Dave, John, and Jesse; and all of the faculty and staff at The University of Montana Department of Mathematical Sciences. For each of you played a role in allowing me the opportunity to reach this point in my life and career. Finally, I would like to thank my wife, Jenny, for being "all my reasons," loving me without question or reservation, putting up with late nights and lengthy emails, encouraging me in my valleys, celebrating with me on my mountain tops, and giving me our beautiful daughter, Trinity Faith. To my family and friends without whom I would not be in this place today, thank you and I love you.

I dedicate this work to my wife, Jenny;
our daughter, Trinity;
and all of our future children.

Notation

$\boldsymbol{A}$	a matrix, denoted by boldface font
$\boldsymbol{A}_{jj'}$	the $(j,j')^{th}$ element of matrix $\boldsymbol{A}$
$a_{jj'ii'}$	the $(i,i')^{th}$ entry of the $(j,j')^{th}$ block of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{\cdot j}$	the j^{th} column of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i\cdot}$	the i^{th} row of matrix $\boldsymbol{A}$
$\boldsymbol{A}'$	the transpose of matrix $\boldsymbol{A}$
$\mathrm{E}(\cdot)$	the expected value function
$\mathrm{Var}(\cdot)$	the variance function
$\mathrm{Cov}(\cdot)$	the covariance function
$\mathrm{tr}(\cdot)$	the trace function (of a matrix)
$\mathrm{vec}(\cdot)$	the vectorization function for matrices
$\oplus$	the direct sum matrix operator
$\bigoplus_{i=1}^{m}$	the direct sum matrix operator for several (indexed) matrices
$\otimes$	the Kronecker product matrix operator
H_0 or $\mathcal{H}_0$	a null hypothesis (may have additional clarifying addenda)
SS	shorthand for "sums of squares"
SSCP	shorthand for "sums of squares and cross products"
CLT	shorthand for "Central Limit Theorem"
$\sum_{index}^{limit}$	the sum operator (compare to below)?
$\boldsymbol{\Sigma}$	a matrix, usually the variance-covariance matrix (boldface)
$\boldsymbol{I}_m$	$m \times m$ identity matrix
$\boldsymbol{J}_m$	$m \times m$ matrix of ones
$\boldsymbol{P}_m$	defined to be $\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{J}_m$
$\mathbf{1}_m$	$m \times 1$ vector of ones
$N(\mu,\sigma)$	normal distribution with parameters μ,σ
$MVN(\boldsymbol{\mu},\boldsymbol{\Sigma})$	multivariate normal distribution with parameters $\boldsymbol{\mu},\boldsymbol{\Sigma}$
$MVN_{n,p}(\boldsymbol{\mu},\boldsymbol{\Sigma})$	np-multivariate normal distribution with parameters $\boldsymbol{\mu},\boldsymbol{\Sigma}$
χ^2_n	χ^2-distribution with n degrees of freedom
F_{n_1,n_2}	F-distribution with degrees of freedom n_1, n_2
MVN	shorthand for "multivariate normal" (in distribution)
ANOVA	shorthand for "analysis of variance"
MANOVA	shorthand for "multivariate analysis of variance"
RM-ANOVA	shorthand for "repeated measures analysis of variance"
RM-MANOVA	shorthand for "repeated measures multivariate analysis of variance"

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In the dance of statistical academia, theory is often a step or two behind necessity. Booming advancements in technology and research over recent decades have created a need for statistical theory to address the real-life phenomena from which data so often arise. Traditionally, whether with proper justification or out of the need and desire for mathematical tractability, many assumptions are imposed in statistical methods. For instance, independence among subjects is the theoretical bedrock of the classical Central Limit Theorem first proposed over a century ago. Since that time much has changed. The aim of this dissertation is to present inferential procedures for two-factor repeated measurements analysis of variance, both when the number of response variables is one or when it is several, and when the number of measurements taken on each subject is very large (tends to infinity). To date, there is some research relinquishing assumptions regarding the covariance structure among such data, and there are many results dealing with dependent measurements within each subject. However, to the best of our knowledge, no work has been done in both of these areas with the added relaxation of the assumption on the underlying distribution. Thus far it has been assumed that the data arise from a normal or multivariate normal distribution, a condition which is dismissed in this paper seeking robustness. To address the scenario described above, this dissertation presents robust, formal tests of significance along with their asymptotic distributions, and real data are analyzed to illustrate the applicability of these new methods. For a more detailed account of these distinctions, please refer to Sections 2.1 and 4.1 beginning on pages 13 and 73, respectively.

In order that the reader may have sufficient background regarding some of the topics and methods used in this dissertation, this chapter gives some insight as regards said topics. Section 1.2.1 will include a discussion of the direct sum and some of its properties, and Section 1.2.2 will do the same for the Kronecker product. The vec$(\cdot)$ notation will be presented briefly in Section 1.2.3. Section 1.3 will discuss various modes of convergence, as well as "big-O" and "little-o" notation, which are used throughout the entire dissertation. Proofs or justification of many of the properties will be supplied in Appendix A. The topics discussed in this chapter vary in mathematical intensity; however, most (if not all) of the topics are uncommonly utilized in the usual education of the general mathematician or statistician. For this reason, it is my hope to achieve a balance between necessity and brevity. In the least I trust this will be both refreshing and a good exercise. In case the reader is already quite familiar with the aforementioned topics, he is advised to proceed direction to Chapter 2.

The remainder of the dissertation will be organized as follows. Chapter 2 will discuss the new techniques for addressing the univariate case. This will include some necessary technical lemmas and properties, a discussion of the adjusted sums of squares, and an overview of treating these sums of squares as quadratic forms even in the presence of heteroskedasticity and unbalanced group sizes. Furthermore, Chapter 2 will present test statistics with their asymptotic distributions, addressing any necessary assumptions along the way. Chapter 3 will present a large-scale simulation study to evaluate the finite sample performance of the asymptotic results in Chapter 2. Chapter 3 will include simulation results for the achieved size and power of the test statistics. Regarding the size of tests, many different scenarios are considered, and the modified test statistics will be compared with the traditional methods. Also, a brief exploration using bootstrapping techniques to estimate the variances of the test statistics will be included. Chapter 4 will closely mimic Chapter 2 as it is the multivariate extension to the univariate case. The differences in setup, design, hypotheses, and especially the forms of the test statistics will be addressed throughout Chapter 4. This chapter will also include a simulation study for one of the multivariate criteria commonly used in the statistical literature. Chapter 5 will give a practical example applying the new test statistics, and a summary discussion will be included here as well.

1.2 Some Matrix Operators

As linear algebra and matrix notation make many future results much more tractable and simple, there is reason to present background on some of its lesser known topics, such as the direct sum, the Kronecker product, and vec(·) notation.

1.2.1 Direct Sum

The direct sum operator provides shorthand notation for otherwise cumbersome matrices. When working with design matrices, quadratic forms, and other statistical topics, direct sum notation (when applicable) significantly limits the amount of sometimes detailed work, which often becomes too complex to follow.

Definition 1.2.1. *The direct sum is a matrix operation mapping two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, with dimensions $m \times n$ and $p \times q$, respectively, to a block matrix $\boldsymbol{C}$ with dimension $(m+p) \times (n+q)$ such that $\boldsymbol{A}$ and $\boldsymbol{B}$ are the diagonal block entries of $\boldsymbol{C}$, and all other entries are defined to be zero. It is denoted by the symbol $\oplus$, such that*

$$\oplus : \boldsymbol{A}, \boldsymbol{B} \rightarrow \boldsymbol{C},$$

where

$$\boldsymbol{C} = \boldsymbol{A} \oplus \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix}.$$

This definition can be extended to several matrices.

Definition 1.2.2. *The* <u>*direct sum of multiple matrices*</u> $\boldsymbol{A}_i, i = 1, \ldots, m$, *is*

$$\bigoplus_{i=1}^{m} \boldsymbol{A}_i = \boldsymbol{A}_1 \oplus \cdots \oplus \boldsymbol{A}_m = \operatorname{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m) = \begin{bmatrix} \boldsymbol{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{A}_m \end{bmatrix}.$$

Some properties of the direct sum are given below.

Property 1.2.3. *For any size matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, *and* $\boldsymbol{C}$,

$$(\boldsymbol{A} \oplus \boldsymbol{B}) \oplus \boldsymbol{C} = \boldsymbol{A} \oplus (\boldsymbol{B} \oplus \boldsymbol{C}) = \boldsymbol{A} \oplus \boldsymbol{B} \oplus \boldsymbol{C}.$$

The notion of associativity is intuitive based on the definition of direct sum. This concept clearly extends to several matrices, though it will not be shown here, by a recursive extension of the proof given in Appendix A.

Next we examine how scalars distribute across the direct sum of two matrices. The proof is very straightforward, and though somewhat trivial, serves the purpose of comparison between the direct sum and Kronecker product (discussed in the next section).

Property 1.2.4. *For any size matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$, *and scalar* k,

$$k(\boldsymbol{A} \oplus \boldsymbol{B}) = k\boldsymbol{A} \oplus k\boldsymbol{B}.$$

It should be noted that, provided $k \neq 1$ and $\boldsymbol{A}, \boldsymbol{B} \neq \boldsymbol{0}$,

$$k\boldsymbol{A} \oplus \boldsymbol{B} \neq k(\boldsymbol{A} \oplus \boldsymbol{B}) \neq \boldsymbol{A} \oplus k\boldsymbol{B}.$$

Again, the second statement in Property 1.2.4 is relatively obvious, but serves as a comparison between the direct sum and Kronecker product.

The trace of the direct sum of matrices is the sum of the traces of the individual matrices (Property 1.2.5). In working with the expected value of quadratic forms, this property is used heavily.

Property 1.2.5. *For any size matrices $\boldsymbol{A}_i, i = 1, \ldots, m$,*

$$\operatorname{tr}\left(\bigoplus_{i=1}^{m} \boldsymbol{A}_i\right) = \sum_{i=1}^{m} \operatorname{tr}\left(\boldsymbol{A}_i\right).$$

Property 1.2.6 gives a formula for evaluating the determinant of the direct sum of two matrices. As with Properties 1.2.3 and 1.2.4, this result easily extends to more than two matrices via a recursive argument.

Property 1.2.6. *For any size matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\det(\boldsymbol{A} \oplus \boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}).$$

While the next (and last) property presented here almost needs no proof, it is included to stress the importance of ordering and syntax. When working with sums of squares, basic linear models, and especially more complicated topics, the careless writer or reader may easily become entangled by statements that appear very innocent and similar to one another.

Property 1.2.7. *In general, $\oplus$ is not commutative; i.e., for generic matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\boldsymbol{A} \oplus \boldsymbol{B} \neq \boldsymbol{B} \oplus \boldsymbol{A}.$$

1.2.2 Kronecker Product

To further streamline future notation, we now introduce the notion of the Kronecker product. A special case of the tensor product, the Kronecker product, similar to the direct sum, is an operator used to ease the notation for matrices which can grow large quickly or to consolidate matrices which already have an appropriate structure. Recall that matrices are said to be *conformable* in an additive sense if they have the same dimension; matrices are conformable in a matrix-multiplicative sense if the number of columns of the left matrix is equal to the number of columns of the right matrix.

Definition 1.2.8. *The* <u>*Kronecker product*</u> *is a matrix operation mapping two matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, with dimensions* $m \times n$ *and* $p \times q$*, respectively, to a block matrix* $\boldsymbol{C}$ *with dimension* $(mp) \times (nq)$ *such that each element of* $\boldsymbol{A}$ *is expanded by being multiplied by the entire matrix* $\boldsymbol{B}$ *to form* mn $p \times q$ *block entries of* $\boldsymbol{C}$*. It is denoted by the symbol* $\otimes$*, such that*

$$\otimes : \boldsymbol{A}, \boldsymbol{B} \rightarrow \boldsymbol{C},$$

where

$$\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix},$$

where

$$a_{ij}\boldsymbol{B} = \begin{bmatrix} a_{ij}b_{11} & \cdots & a_{ij}b_{p1} \\ \vdots & \ddots & \vdots \\ a_{ij}b_{1q} & \cdots & a_{ij}b_{pq} \end{bmatrix}.$$

As in the previous section, we will list some major properties of the Kronecker product followed by their proofs. We begin with the way in which scalars interact with the Kronecker product.

Property 1.2.9. *For a constant k and matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$(k\boldsymbol{A}) \otimes \boldsymbol{B} = \boldsymbol{A} \otimes (k\boldsymbol{B}) = k(\boldsymbol{A} \otimes \boldsymbol{B}).$$

Next, we examine the associative nature of the Kronecker product. As is to be expected, the Kronecker product is associative.

Property 1.2.10. *For matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$,*

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}). \tag{1.1}$$

Property 1.2.10 says that the parentheses (1.1) can just as easily be omitted. Properties 1.2.11 and 1.2.12 display how the Kronecker product distributes over addition and multiplication.

Property 1.2.11. *For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$,*

$$(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$$

and

$$\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}.$$

Property 1.2.12. *For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$,*

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}.$$

It is useful to note that Property 1.2.12 naturally extends to

$$\left(\bigotimes_{i=1}^{m} \boldsymbol{A}_i\right)\left(\bigotimes_{j=1}^{m} \boldsymbol{B}_j\right) = \bigotimes_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{B}_j,$$

where $\bigotimes_{i=1}^{m} \boldsymbol{A}_i$ is defined analogously to $\bigoplus_{i=1}^{m} \boldsymbol{A}_i$ as in Definition 1.2.2.

The inverse of the Kronecker product of two matrices is equal to the Kronecker product of their respective inverses. The same is true for the transpose; both notions are given explicitly in the properties below.

Property 1.2.13. *For invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}.$$

Property 1.2.14. *For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'.$$

The trace of the Kronecker product of matrices is the product of the traces of the individual matrices.

Property 1.2.15. *For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) \operatorname{tr}(\boldsymbol{B}).$$

Again, this property can be extended recursively to several matrices such that for matrices $\boldsymbol{A}_i, i = 1, \ldots, m$,

$$\operatorname{tr}\left(\bigotimes_{i=1}^{m} \boldsymbol{A}_i\right) = \prod_{i=1}^{m} \operatorname{tr}(\boldsymbol{A}_i).$$

While the following is not as intuitive as the lack of commutativity of the direct sum, it can be easily seen that matrices do not commute over the Kronecker product. Again, though this may seem rather trivial, it is important to note in order to ensure the proper usage and syntax when working with both the direct sum and the Kronecker product.

Property 1.2.16. *In general, $\otimes$ is not commutative; i.e., for generic matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A}.$$

1.2.3 vec Notation

Test statistics used later in this dissertation are comprised of quadratic forms. While data are conveniently represented in matrices, manipulation of the data to compute the test statistics and to study their distributions is often simplified by reorganizing such matrices as vectors. An operator that is useful in this regard is the "vec" operator.

In order to preserve the notion of the quadratic form, we will now start with a matrix of observational values, then transform this matrix to a vector using the vec function.

Definition 1.2.17. *Let $\boldsymbol{A}$ be an $m \times n$ matrix, and denote by $\boldsymbol{A}_{\cdot j}$ the j^{th} column of $\boldsymbol{A}$. Then the* <u>*vectorization*</u> *of $\boldsymbol{A}$, written $\mathrm{vec}\,(\boldsymbol{A})$, produces the mn-column vector*

$$\mathrm{vec}\,(\boldsymbol{A}) = \begin{bmatrix} \boldsymbol{A}_{\cdot 1} \\ \vdots \\ \boldsymbol{A}_{\cdot n} \end{bmatrix}.$$

Essentially, the vec function stacks the columns of a matrix, one below the other going from left to right. Note that to stack the rows, or line up the rows and then transpose, simply take the vectorization of the transpose of the desired matrix. Now we are ready to interject important properties relating the trace, vectorization, and the Kronecker product.

Property 1.2.18. *For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{X}$,*

$$\mathrm{vec}\,(\boldsymbol{AXB}) = \left(\boldsymbol{B}' \otimes \boldsymbol{A}\right) \mathrm{vec}\,(\boldsymbol{X}).$$

Property 1.2.19. *For conformable matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, *and* $\boldsymbol{C}$,

$$\operatorname{tr}(\boldsymbol{ABC}) = \left(\operatorname{vec}\left(\boldsymbol{A}'\right)\right)' (\boldsymbol{I} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{C}).$$

Property 1.2.20. *For conformable matrices* $\boldsymbol{A}$, *and* $\boldsymbol{B}$,

$$\operatorname{tr}(\boldsymbol{AB}) = \left(\operatorname{vec}\left(\boldsymbol{A}'\right)\right)' \operatorname{vec}(\boldsymbol{B}).$$

Property 1.2.21. *For vector* $\boldsymbol{a}$,

$$\operatorname{vec}\left(\boldsymbol{a}\boldsymbol{a}'\right) = \boldsymbol{a} \otimes \boldsymbol{a}.$$

Property 1.2.22. *For conformable matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{X}$, *and* $\boldsymbol{Y}$,

$$(\operatorname{vec}(\boldsymbol{Y}))' (\boldsymbol{A} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{X}) = \operatorname{tr}\left(\boldsymbol{A}'\boldsymbol{Y}'\boldsymbol{B}\boldsymbol{X}\right).$$

The proof of this property is due to Proposition 31 in Broxson [13]. We will mainly use this in the context of quadratic forms, i.e., when $\boldsymbol{Y} = \boldsymbol{X}$. In this case, the result becomes

$$(\operatorname{vec}(\boldsymbol{X}))' (\boldsymbol{A} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{X}) = \operatorname{tr}\left(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right).$$

In later sections, we will have a quadratic form that can be written much like the LHS of the equation above, but it is far more efficient to calculate the RHS. Especially in simulation studies, this result saves precious time.

1.3 Convergence: Modes and Notation

Two main modes of convergence will be used in the dissertation: convergence in probability and convergence in distribution (law). If there is no specification, the usual convergence in limit should be assumed, which is denoted by $\to$. That is, $f(x)$ converges to L as $x \to a$, or $\lim_{x\to a} f(x) = L$, is written $f(x) \to L$ as $x \to a$.

First we define the two modes of convergence. In these definitions, let $|\cdot|$ represent the Euclidean norm and let $F(\cdot)$ represent the distribution function.

Definition 1.3.1. *We say a sequence of random vectors* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ *converges in probability to* $\boldsymbol{X}$, *written* $\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X} \quad \text{as} \quad n \to \infty$, *if for every* $\epsilon > 0$, $P\left(|\boldsymbol{X}_n - \boldsymbol{X}| > \epsilon\right) \to 0$ *as* $n \to \infty$.

Definition 1.3.2. *We say a sequence of random vectors* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ *converges in distribution (or converges in law, or converges weakly) to* $\boldsymbol{X}$, *written* $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \quad \text{as} \quad n \to \infty$, *if* $F_{\boldsymbol{X}_n}(\boldsymbol{x}) \to F_{\boldsymbol{X}}(\boldsymbol{x})$ *as* $n \to \infty$, *for every point* $\boldsymbol{x}$ *at which* $F_{\boldsymbol{X}}(\boldsymbol{x})$ *is continuous.*

It is well known that

$$\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X} \quad \Longrightarrow \quad \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}.$$

We will also use the "big-O" and "little-o"' notation throughout this dissertation. These are useful when discussing two sequences of numbers.

Definition 1.3.3. *Consider two functions* f *and* g. *If there exist* $K, M > 0$ *such that* $|f(x)| \leq K|g(x)|$ *whenever* $|x| > M$, *we write* $f(x) = O(g(x))$ *as* $x \to \infty$, *and we say that* f/g *is bounded for* x *large enough. If there exist* $K, \delta > 0$ *such that* $|f(x)| \leq K|g(x)|$ *as* $x \to 0$ *whenever* $|x| < \delta$, *we write* $f(x) = O(g(x))$ *as* $x \to 0$, *as we say that* f/g *is bounded for* x *small enough.*

In essence, $f(x) = O(g(x))$ means that f/g is bounded as x get sufficiently large or small, whichever is contextually applicable.

Definition 1.3.4. *Consider two functions f and g. We write* $\underline{f(x) = o(g(x))}$ *as* $x \to \infty$ *if* $\frac{f(x)}{g(x)} \to 0$ *as* $x \to \infty$. *We write* $\underline{f(x) = o(g(x))}$ *as* $x \to 0$ *if* $\frac{f(x)}{g(x)} \to 0$ *as* $x \to 0$.

In essence $f(x) = o(g(x))$ means that f is converging to 0 more rapidly than g.

Similar notation that will be used in the dissertation include "big-O-p," $O_p(\cdot)$, and "little-o-p," $o_p(\cdot)$. The definitions of $O_p(\cdot)$ and $o_p(\cdot)$ are akin to $O(\cdot)$ and $o(\cdot)$, excepting mainly that the mode of convergence is convergence in probability and functions are of random variables/vectors. Formal definitions (due to Serfling [40]) are given below.

Definition 1.3.5. *Consider a sequence of random variables $\{X_n\}$ with corresponding distribution functions $\{F_n\}$. We say $\{X_n\}$ is bounded in probability, and write* $\underline{X_n = O_p(1)}$, *if for every $\epsilon > 0$ there exist M_ϵ and N_ϵ such that*

$$F_n(M_\epsilon) - F_n(-M_\epsilon) > 1 - \epsilon \quad \textit{for all} \quad n > N_\epsilon.$$

More generally, considering another sequence of random variables $\{Y_n\}$, the notation $X_n = O_p(Y_n)$ means that the sequence $\{X_n/Y_n\}$ is $O_p(1)$.

We can see that

$$X_n \xrightarrow{D} X \quad \Longrightarrow \quad X_n = O_p(1).$$

Definition 1.3.6. *Consider a sequence of random variables $\{X_n\}$ and $\{Y_n\}$. We write* $\underline{X_n = o_p(Y_n)}$ *to denote that*

$$\frac{X_n}{Y_n} \xrightarrow{p} 0 \quad as \quad n \to \infty.$$

For example, we may write $\boldsymbol{\Sigma} - \boldsymbol{G} = o_p(1)$ to mean that $\boldsymbol{\Sigma} - \boldsymbol{G} \xrightarrow{p} \mathbf{0}$.

Chapter 2

Inference for a Large Number of Repeated Measures: Univariate Case

2.1 Introduction

In many experimental or observational studies, an outcome variable is observed from each subject repeatedly. The subjects are often grouped according to the treatment they received or the experimental conditions to which they were subjected. Sometimes the grouping occurs due to a natural phenomenon, such as the sex of the subject, and other times the grouping is imposed by the researcher. These data may be generated by longitudinal studies or crossover designs, among others, and they are appearing with great frequency. We will refer to such analysis as repeated measures analysis of variance, or repeated measures ANOVA or RM-ANOVA for short. The central idea of this chapter is to fuse to previous papers and make possible a more comprehensive analysis, which is contained in Chapter 4. The results in

Chapter 2 are very similar to those found in Wang and Akritas [45], though the methods and arithmetic techniques are more akin to what is found in Harrar and Bathke [26]; these papers do not make the same assumptions, and Chapter 2 will take a subset of assumptions from each paper. The specific assumptions will be discussed in greater detail in subsequent paragraphs and sections. The key assumption that is lifted is that of an underlying distribution. To the best of our knowledge, no work has been done removing the normality assumption for the data with the arithmetic methods used in Chapter 2; furthermore, as Chapter 4 will indicate, we can find no work in the case of a multivariate response variable dropping the assumption of normality.

The most commonly encountered repeated measures data can be viewed as arising from a two-factor crossed design. Consider the following scheme. Let X_{ijk} be independent random variables with mean μ_{ij} and covariance given by $\text{Cov}\left(X_{ijk}, X_{ij'k}\right) = \sigma_{jj'}$ (for now, unstructured), for $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, n_i$. The usual setting gives the interpretation that X_{ijk} is the response from the k^{th} subject in the i^{th} group at the j^{th} time point, though "group" and "time" merely offer one interpretation. Define here, for ease of later use,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_b \end{bmatrix}, \text{ where } \quad \boldsymbol{X}_j = \begin{bmatrix} \boldsymbol{X}_{1j} \\ \vdots \\ \boldsymbol{X}_{aj} \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_{ij} = \begin{bmatrix} X_{ij1} \\ \vdots \\ X_{ijn_i} \end{bmatrix}. \tag{2.1}$$

We shall consider the model $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where the unknown constants α_i, β_j, and γ_{ij} correspond to the main effects of factor A, the main effects of factor B, and the interaction effects between factors A and B, respectively. As usual, these shall be subject to the sum-to-zero identifiability constraints $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$. It is noteworthy that X_{ijk} and $X_{i'j'k'}$ are assumed to be independent only if $i \neq i'$ or $k \neq k'$. If we let $\epsilon_{ijk} = X_{ijk} - \mu_{ij}$, then $\epsilon_{i1k}, \epsilon_{i2k}, \ldots$ is considered as a sequence of dependent random variables.

The primary hypotheses of interest are

(i) $H_0^{\beta} \; : \; \beta_j = 0$ for $j = 1, \ldots, b$;

(ii) $H_0^{\gamma} \; : \; \gamma_{ij} = 0$ for $i = 1, \ldots, a$ and $j = 1, \ldots, b$; and

(iii) $H_0^{\alpha} \; : \; \alpha_i = 0$ for $i = 1, \ldots, a$.

Initially, attention will be given to the first two hypotheses. These translate to having no main effects of the levels of factor B (often the temporal factor) and no interaction effects between the levels of factor A and levels of factor B, respectively. The motivation for this order of presentation is that the main effects of factor B and the interaction effects both rely heavily on the same asymptotic theory, and their results can be consolidated with a suitable choice of appropriate notation. The asymptotic nature of the main effects of factor A is dealt with in the initial stages of the analysis, and then the theory is carried out similar to the usual case without relying on any asymptotic structure.

The asymptotic framework to be considered is that the number of levels of one of the factors, namely factor B, is large (tends to infinity) but that of the other factor and the sample size per treatment remain fixed [26]. In many current applications dealing with longitudinal data, such as fast functional magnetic resonance imaging or disease readings from thousands of signal intensities [45], vast quantities of data are being collected, often in a spatial or temporal context. Other applications include technological advancements in biotechnologies [26]. It should be noted that in general the fixed quantities a and n_i need not be small nor large.

For example, consider two groups of patients, one with Parkinson's disease and the other without. Protein microarrays are collected on each subject, and the response variable is a measure of protein. From each of these, over 9,000 relative fluorescent unit (RFU) signals are output giving the presence and abundance of human proteins [24]. We can think of factor A as a grouping factor (diseased or not) and factor B as a genome factor. It is especially of interest whether there are effects of genome or an interaction between group and genome. In these cases, if effects are found, more research is warranted to identify the specific proteins that are greatly influenced by group differences; these are known as biomarkers. However, if no effects are found, there is little reason to attempt to identify biomarkers, making this stage of analysis very pertinent. Given no interaction effects are present, a test for group effects should be conducted. If there are group effects, meaning protein levels are different among the groups, more research is warranted on the proteins.

Brunner et al. point out that aside from a very restrictive set of covariance structures, the usual repeated measures theory has failed to produce valid, high-dimensional inference procedures [14]. To date, classical multivariate analysis cannot be applied in the high-dimensional case where the number of repeated measurements is greater than the sample size since the covariance matrix will not be invertible [14]. Moreover, even the current methods not relying on a structured covariance matrix have a major source of concern. These methods assume normality of the responses from each subject [26]. This is often an unrealistic assumption. Furthermore, many asymptotic methods of analysis do not embrace the dependent nature of the repeated measurements collected from each subject. This dissertation addresses all three of these limitations: no covariance structure is assumed, no underlying distribution is assumed, and dependency is embraced. Following methods such as those found in Billingsley's central limit theorem for dependent random variables [8], this dissertation develops inferential procedures for high-dimensional data in a very robust setting, including when the number of response variables is greater than one. These tests are assessed via simulation, and the results are compared to traditional repeated measures methods. New sums of squares similar to those from Wang and Akritas [45] are considered to account for potential heteroskedasticity and unbalanced group sizes. An increasingly popular method of reorganizing the sums of squares into quadratic forms is employed, yet the algebra used in the proofs of the main results, though greatly simplified using this technique, remains somewhat arduous. Various technical assumptions must still be made, such as a mixing condition [11], and these do provide some limitations which deserve further attention.

The remainder of this section will have the following organization. Section 2.2 will further the problem of interest, addressing the new sums of squares and arguing why they can and should be put into quadratic forms. Also included in Section 2.2 will be some necessary lemmas regarding the moments of the SS quadratic forms and some properties of the specific middle matrices of the SS quadratic forms. Section 2.3 will address the test statistics for the hypotheses of interest, while Section 2.4 will discuss the asymptotic distributions of these test statistics. Section 2.4 will also include details regarding some assumptions needed for the asymptotic theory.

2.2 Preliminaries

2.2.1 Sums of Squares

As this chapter progresses, many assumptions will be needed for the sake of later proofs. We shall state them near the topics from which they arise. The first such assumption is given here.

Assumption 2.2.1. *The data X_{ijk} are random variables with mean μ_{ij}, where X_{ijk} and $X_{i'jk'}$ are independent for $i \neq i'$ or $k \neq k'$.*

It serves now to define the modified sums of squares (SS). The SS due to factor B will be $AM\beta$; the SS due to interaction will be $AM\gamma$; the SS due to error will be AME. Define

$$\begin{aligned} AM\beta &:= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widetilde{X}_{\cdot j\cdot} - \widetilde{X}_{\cdots}\right)^2, \\ AM\gamma &:= \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X}_{ij\cdot} - \widetilde{X}_{i\cdot\cdot} - \widetilde{X}_{\cdot j\cdot} + \widetilde{X}_{\cdots}\right)^2, \quad \text{and} \\ AME &:= \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}\left(X_{ijk} - \overline{X}_{ij\cdot} - \overline{X}_{i\cdot k} - \widetilde{X}_{i\cdot\cdot}\right)^2; \end{aligned} \tag{2.2}$$

where

$$\begin{aligned}
\widetilde{X}_{\cdots} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}X_{ijk},\\
\widetilde{X}_{i\cdot\cdot} &= \overline{X}_{i\cdot\cdot} = \frac{1}{bn_i}\sum_{j=1}^{b}\sum_{k=1}^{n_i}X_{ijk},\\
\widetilde{X}_{\cdot j\cdot} &= \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{k=1}^{n_i}X_{ijk},\\
\overline{X}_{ij\cdot} &= \frac{1}{n_i}\sum_{k=1}^{n_i}X_{ijk},\\
\overline{X}_{i\cdot k} &= \frac{1}{b}\sum_{j=1}^{b}X_{ijk}, \quad \text{and}\\
n &= \sum_{i=1}^{a}n_i. \qquad (2.3)
\end{aligned}$$

Since the covariance terms ($\sigma_{jj'}$) are different for each level of factor B but the same for different levels of factor A, and since the covariance is assumed to have no structure, it does not make sense to use the usual weighting structure as in homoskedastic ANOVA [45]; the SS do not have the same expectation under heteroskedasticity. There the data are combined when forming the hypothesis sum of squares. For instance, consider H_0^γ. If we let $\tilde{\boldsymbol{\mu}}_{1\times ab} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1b} & \mu_{21} & \cdots & \mu_{ab} \end{bmatrix}$, then we can write the null hypothesis of no interaction effect as $H_0^\gamma : \tilde{\boldsymbol{\mu}}\left(\boldsymbol{P}_a \otimes \boldsymbol{P}_b\right) = \boldsymbol{0}$. Defining

$$\overline{\boldsymbol{X}} = \begin{bmatrix} \overline{X}_{11\cdot} & \cdots & \overline{X}_{1b\cdot} & \overline{X}_{21\cdot} & \cdots & \overline{X}_{ab\cdot} \end{bmatrix},$$

it would be reasonable to use $K_\gamma := \overline{\boldsymbol{X}}\left(\boldsymbol{P}_a \otimes \boldsymbol{P}_b\right)\overline{\boldsymbol{X}}'$ to estimate (quantify) the variation due to the departure from H_0^γ. Then, similar algebra to that seen in the proof of Proposition 2.2.2 can show that $K_\gamma = (a-1)(b-1)AM\gamma$ as in (2.2). Analogous arguments would justify the other sums of squares.

Theoretical development of the results in this paper will be much more tractable and smooth if the SS are expressed as quadratic forms. This requires the use of the direct sum and the Kronecker product. Recall these operations and their properties are given in Chapter 1 and Appendix A. In general, let $\mathbf{1}_m$ be the $m \times 1$ vector of 1s, $\boldsymbol{I}_m$ be the $m \times m$ identity matrix, $\boldsymbol{J}_m$ be the $m \times m$ matrix of 1s, and $\boldsymbol{P}_m := \boldsymbol{I}_m - \frac{1}{m}\boldsymbol{J}_m$.

Proposition 2.2.2. *The sums of squares from (2.2) can be put in the following quadratic forms.*

$$\begin{aligned}
AM\beta &= \frac{1}{b-1}\boldsymbol{X}'\left(\boldsymbol{P}_b \otimes \left[\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\mathbf{1}_{n_i}\right)\frac{1}{a}\boldsymbol{J}_a\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\mathbf{1}'_{n_i}\right)\right]\right)\boldsymbol{X}; \\
AM\gamma &= \frac{1}{(a-1)(b-1)}\boldsymbol{X}'\left(\boldsymbol{P}_b \otimes \left[\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\mathbf{1}_{n_i}\right)\boldsymbol{P}_a\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\mathbf{1}'_{n_i}\right)\right]\right)\boldsymbol{X}; \\
AME &= \frac{1}{a(b-1)}\boldsymbol{X}'\left(\boldsymbol{P}_b \otimes \left(\bigoplus_{i=1}^{a}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}\right)\right)\boldsymbol{X}.
\end{aligned} \tag{2.4}$$

Proof: To rewrite the sums of squares from summation notation to quadratic forms requires building matrices and vectors that take repeated means from (2.3) in various fashions and then simplifying. First, for this proof only, let

$$\boldsymbol{G} = \bigoplus_{i=1}^{a}\frac{1}{n_i}\mathbf{1}_{n_i}.$$

Further, let

$$\begin{aligned}
\boldsymbol{A} &= \frac{1}{b}\boldsymbol{J}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right), \\
\boldsymbol{B} &= \boldsymbol{I}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right), \\
\boldsymbol{C} &= \frac{1}{b}\boldsymbol{J}_b \otimes \boldsymbol{G}', \quad \text{and} \\
\boldsymbol{D} &= \boldsymbol{I}_b \otimes \boldsymbol{G}'.
\end{aligned}$$

From here, we build the following vectors using the means from (2.3). Observe

$$
\begin{aligned}
\left[\ \widetilde{X}_{\cdots} \cdots \widetilde{X}_{\cdots}\ \right]' &= \left[\mathbf{1}_{ab}\left(\frac{1}{b}\mathbf{1}_b' \otimes \left[\frac{1}{a}\mathbf{1}_a' \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right]\right)\right]\boldsymbol{X} \\
&= \left(\frac{1}{b}\boldsymbol{J}_b \otimes \left[\frac{1}{a}\boldsymbol{J}_a \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right]\right)\boldsymbol{X} \ = \ \boldsymbol{AX}, \\
\left[\ \widetilde{X}_{\cdot 1\cdot} \cdots \widetilde{X}_{\cdot 1\cdot}\ \middle|\ \cdots\ \middle|\ \widetilde{X}_{\cdot b\cdot} \cdots \widetilde{X}_{\cdot b\cdot}\ \right]' &= \left[\left(\boldsymbol{I}_b \otimes \left[\frac{1}{a}\mathbf{1}_a' \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right]\right)\boldsymbol{X}\right] \otimes \mathbf{1}_a \\
&= \left(\boldsymbol{I}_b \otimes \left[\frac{1}{a}\boldsymbol{J}_a \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right]\right)\boldsymbol{X} \ = \ \boldsymbol{BX}, \\
\left[\ \widetilde{X}_{1\cdot\cdot} \cdots \widetilde{X}_{a\cdot\cdot}\ \middle|\ \cdots\ \middle|\ \widetilde{X}_{1\cdot\cdot} \cdots \widetilde{X}_{a\cdot\cdot}\ \right]' &= \left[\mathbf{1}_b \otimes \left[\frac{1}{b}\mathbf{1}_b' \otimes \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right]\right]\boldsymbol{X} \\
&= \left(\frac{1}{b}\boldsymbol{J}_b \otimes \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right)\boldsymbol{X} \ = \ \boldsymbol{CX}, \\
\left[\ \overline{X}_{11\cdot} \cdots \overline{X}_{a1\cdot}\ \middle|\ \cdots\ \middle|\ \overline{X}_{1b\cdot} \cdots \overline{X}_{ab\cdot}\ \right]' &= \left(\boldsymbol{I}_b \otimes \bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right)\boldsymbol{X} \ = \ \boldsymbol{DX}.
\end{aligned}
$$

Then, noting that the proper stacking is now achieved, we can write

$$
\begin{aligned}
(b-1)AM\beta &= (\boldsymbol{BX} - \boldsymbol{AX})'(\boldsymbol{BX} - \boldsymbol{AX}) \quad \text{and} \\
(a-1)(b-1)AM\gamma &= (\boldsymbol{DX} - \boldsymbol{CX} - \boldsymbol{BX} + \boldsymbol{AX})'(\boldsymbol{DX} - \boldsymbol{CX} - \boldsymbol{BX} + \boldsymbol{AX}).
\end{aligned}
$$

To ease the calculations later, let us note that we can write

$$
\begin{aligned}
\boldsymbol{B} - \boldsymbol{A} &= \boldsymbol{I}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right) - \frac{1}{b}\boldsymbol{J}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right) \\
&= \boldsymbol{P}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{D} - \boldsymbol{C} - \boldsymbol{B} + \boldsymbol{A} &= \boldsymbol{I}_b \otimes \boldsymbol{G}' - \frac{1}{b}\boldsymbol{J}_b \otimes \boldsymbol{G}' - \boldsymbol{I}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right) + \frac{1}{b}\boldsymbol{J}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right) \\
&= \left[\boldsymbol{P}_b \otimes (\boldsymbol{I}_a\boldsymbol{G}')\right] - \left[\boldsymbol{P}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a\boldsymbol{G}'\right)\right]
\end{aligned}
$$

$$= \boldsymbol{P}_b \otimes \left(\boldsymbol{P}_a \boldsymbol{G}'\right).$$

Continuing with the sums of squares, we now see that

$$\begin{aligned}
(b-1)AM\beta &= \boldsymbol{X}'\left(\boldsymbol{B}-\boldsymbol{A}\right)'\left(\boldsymbol{B}-\boldsymbol{A}\right)\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\boldsymbol{P}_b' \otimes \left(\boldsymbol{G} \cdot \frac{1}{a}\boldsymbol{J}_a'\right)\right]\left[\boldsymbol{P}_b \otimes \left(\frac{1}{a}\boldsymbol{J}_a \boldsymbol{G}'\right)\right]\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\left(\boldsymbol{P}_b\boldsymbol{P}_b\right) \otimes \left(\boldsymbol{G} \cdot \frac{1}{a}\boldsymbol{J}_a \cdot \frac{1}{a}\boldsymbol{J}_a \boldsymbol{G}'\right)\right]\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\boldsymbol{P}_b \otimes \left(\boldsymbol{G} \cdot \frac{1}{a}\boldsymbol{J}_a \boldsymbol{G}'\right)\right]\boldsymbol{X},
\end{aligned}$$

which coincides with (2.4). We also see that

$$\begin{aligned}
(a-1)(b-1)AM\gamma &= \boldsymbol{X}'\left(\boldsymbol{D}-\boldsymbol{C}-\boldsymbol{B}+\boldsymbol{A}\right)'\left(\boldsymbol{D}-\boldsymbol{C}-\boldsymbol{B}+\boldsymbol{A}\right)\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\boldsymbol{P}_b' \otimes \left(\boldsymbol{G}\boldsymbol{P}_a'\right)\right]\left[\boldsymbol{P}_b \otimes \left(\boldsymbol{P}_a \boldsymbol{G}'\right)\right]\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\left(\boldsymbol{P}_b\boldsymbol{P}_b\right) \otimes \left(\boldsymbol{G}\boldsymbol{P}_a\boldsymbol{P}_a\boldsymbol{G}'\right)\right]\boldsymbol{X} \\
&= \boldsymbol{X}'\left[\boldsymbol{P}_b \otimes \left(\boldsymbol{G}\boldsymbol{P}_a\boldsymbol{G}'\right)\right]\boldsymbol{X},
\end{aligned}$$

which also coincides with (2.4).

Since AME involves a weighted sum whose indices run from 1 to n_i, where n_i varies, its manipulation to a quadratic form is not as straight-forward as with $AM\beta$ and $AM\gamma$. The principles involved, however, are quite similar. Essentially there is a nested, weighted, unbalanced term in the sum of squares; as a result, there needs to be a nested term accounting for this weighted unbalance. Instead of defining terms outside of the calculation (as with $AM\beta$ and $AM\gamma$), we will take a more "brute-force" approach to this part of the proof.

To save space, terms that have duplicate appearances may be defined internally with underbraces. Similar to the techniques used above, observe that we can write

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}\left(X_{ijk}-\overline{X}_{ij\cdot}-\overline{X}_{i\cdot k}-\widetilde{X}_{i\cdot\cdot}\right)^2$$

$$= \Bigg[\underbrace{\boldsymbol{X} - \left[\begin{array}{c} \overline{X}_{11\cdot} \\ \vdots \\ \overline{X}_{11\cdot} \\ \hline \vdots \\ \hline \overline{X}_{a1\cdot} \\ \vdots \\ \overline{X}_{a1\cdot} \\ \hline \vdots \\ \hline \overline{X}_{ab\cdot} \\ \vdots \\ \overline{X}_{ab\cdot} \end{array}\right] - \left[\begin{array}{c} \widetilde{X}_{1\cdot 1} \\ \vdots \\ \widetilde{X}_{1\cdot n_1} \\ \hline \vdots \\ \hline \widetilde{X}_{a\cdot 1} \\ \vdots \\ \widetilde{X}_{a\cdot n_a} \\ \hline \widetilde{X}_{1\cdot 1} \\ \vdots \\ \widetilde{X}_{a\cdot n_a} \end{array}\right] + \left[\begin{array}{c} \widetilde{X}_{1\cdot\cdot} \\ \vdots \\ \widetilde{X}_{1\cdot\cdot} \\ \hline \vdots \\ \hline \widetilde{X}_{a\cdot\cdot} \\ \vdots \\ \widetilde{X}_{a\cdot\cdot} \\ \hline \widetilde{X}_{1\cdot\cdot} \\ \vdots \\ \widetilde{X}_{a\cdot\cdot} \end{array}\right]}_{:=\boldsymbol{A}_1} \Bigg]' \left(\boldsymbol{I}_b \otimes \left[\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{I}_{n_i}\right]\right) [\boldsymbol{A}_1]$$

$$= \boldsymbol{X}' \Bigg[\underbrace{\boldsymbol{I}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{I}_{n_i}\right) - \boldsymbol{I}_b \otimes \left(\bigoplus_{i=1}^{a} \frac{1}{n_i} \boldsymbol{J}_{n_i}\right) - \frac{1}{b}\boldsymbol{J}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{I}_{n_i}\right) + \frac{1}{b}\boldsymbol{J}_b \otimes \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{J}_{n_i}\right)}_{:=\boldsymbol{A}_2} \Bigg]'$$

$$\times \left(\boldsymbol{I}_b \otimes \left[\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{I}_{n_i}\right]\right) [\boldsymbol{A}_2 \boldsymbol{X}]$$

$$= \boldsymbol{X}' \Bigg[\underbrace{\boldsymbol{I}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}\right) - \frac{1}{b}\boldsymbol{J}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}\right)}_{:=\boldsymbol{A}_3} \Bigg]' \left(\boldsymbol{I}_b \otimes \left[\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{I}_{n_i}\right]\right) [\boldsymbol{A}_3 \boldsymbol{X}]$$

$$= \boldsymbol{X}' \left[\boldsymbol{P}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}\right)\right]' \left(\boldsymbol{I}_b \otimes \left[\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{I}_{n_i}\right]\right) \left[\boldsymbol{P}_b \otimes \left(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}\right)\right] \boldsymbol{X}$$

$$= \boldsymbol{X}' \left[\boldsymbol{P}_b \otimes \left(\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{P}_{n_i}\right)\right] \boldsymbol{X},$$

and this coincides with (2.4).

We can get the desired results by simply multiplying by the appropriate terms for each of

$AM\beta$, $AM\gamma$, and AME. □

Remark: Note, we rely heavily on the facts that $\boldsymbol{I}_m$, $\frac{1}{m}\boldsymbol{J}_m$, and $\boldsymbol{P}_m$ are symmetric and idempotent for any m; also, direct sums and direct products comprised of these matrices will also be symmetric and idempotent. Recall that a square matrix $\boldsymbol{A}$ is idempotent if $\boldsymbol{A}^2 = \boldsymbol{A}$.

Since it is of interest to examine the differences of sums of squares, we introduce the following notation. Define $\boldsymbol{C}_\beta$, $\boldsymbol{C}_\gamma$, and $\boldsymbol{C}_E$ to be the defining matrices of the sums of squares such that

$$\begin{aligned} AM\beta &= \boldsymbol{X}'\boldsymbol{C}_\beta\boldsymbol{X}, \\ AM\gamma &= \boldsymbol{X}'\boldsymbol{C}_\gamma\boldsymbol{X}, \\ AME &= \boldsymbol{X}'\boldsymbol{C}_E\boldsymbol{X}. \end{aligned} \tag{2.5}$$

Let ϕ be one of the factors under consideration: β or γ. Then we define

$$\boldsymbol{C}^*_{\phi_E} := \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{1}_{n_i}\right) c(\phi) \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{1}'_{n_i}\right) - \frac{1}{a}\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}, \tag{2.6}$$

where

$$c(\phi) = \begin{cases} \frac{1}{a}\boldsymbol{J}_a, & \phi = \beta \\ \frac{1}{a-1}\boldsymbol{P}_a, & \phi = \gamma \end{cases},$$

and

$$\boldsymbol{S}_{\phi_E} := \frac{1}{b(b-1)}(\boldsymbol{J}_b - \boldsymbol{I}_b) \otimes \boldsymbol{C}^*_{\phi_E}. \tag{2.7}$$

We wish to write the SS differences as the sum of two components. The first component will take on a form which is more tractable, a desirable trait in the modeling and hypothesis testings stages. The second component will be shown to be asymptotically negligible. For

$\phi \in \{\beta, \gamma\}$, we can now write

$$
\begin{aligned}
AM\phi - AME &= \boldsymbol{X}'(\boldsymbol{C}_\phi - \boldsymbol{C}_E)\boldsymbol{X} \\
&= \boldsymbol{X}'\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^*_{\phi_E}\right)\boldsymbol{X} \\
&= \boldsymbol{X}'\left(\frac{1}{b}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E} - \frac{1}{b(b-1)}(\boldsymbol{J}_b - \boldsymbol{I}_b) \otimes \boldsymbol{C}^*_{\phi_E}\right)\boldsymbol{X} \\
&= \boldsymbol{X}'\left(\frac{1}{b}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}\right)\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}.
\end{aligned}
\tag{2.8}
$$

2.2.2 Moments of Quadratic Forms

Consider $\boldsymbol{X}$ from (2.1). Denote the $(j,j')^{th}$ element of $\boldsymbol{\Sigma} := \text{Cov}\left([X_{i1k} \cdots X_{ibk}]'\right)$ by $\sigma_{jj'}$, which are the same for all i and k by assumption. That is,

$$
\boldsymbol{\Sigma} := \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1b} \\ \vdots & \ddots & \vdots \\ \sigma_{b1} & \cdots & \sigma_{bb} \end{bmatrix}. \tag{2.9}
$$

Furthermore, we now see

$$
\text{Var}\left(\boldsymbol{X}\right) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n. \tag{2.10}
$$

The following lemmas give the first and second moments of quadratic forms in $\boldsymbol{X}$.

Lemma 2.2.3. *Let $\boldsymbol{A}$ be a $(bn) \times (bn)$ matrix of constants. $\boldsymbol{A}$ can then be thought of as having b^2 blocks of size $n \times n$. Denote by $a_{jj'ii'}$ the $(i,i')^{th}$ element of the $(j,j')^{th}$ $n \times n$ block of matrix $\boldsymbol{A}$. Furthermore, let the vector of random variables $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1' & \cdots & \boldsymbol{X}_b' \end{bmatrix}'$, where $\boldsymbol{X}_j' = \begin{bmatrix} X_{1j} & \cdots & X_{nj} \end{bmatrix}$. Assume that X_{ij} and $X_{i'j'}$ are independent for $i \neq i'$. Moreover,* *assume* $\mathrm{E}(X_{ij}) = 0$ *and* $\mathrm{Cov}\left(X_{ij}, X_{ij'}\right) = \sigma_{jj'}$. *Then*

$$\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right) = \sum_{j,j'=1}^{b} \sum_{i=1}^{n} a_{jj'ii}\sigma_{jj'}.$$

Remark: The double-index notation for the summation, $\sum_{j,j'=1}^{b}$, is equivalent to the double sum $\sum_{j=1}^{b}\sum_{j'=1}^{b}$. Analogous extensions of this notation are also used throughout.

Proof: Observe that we can write

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right) &= \sum_{j,j',i,i'} a_{jj'ii'}\mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{i'j'}\right) \\
&= \sum_{j,j',i} a_{jj'ii}\mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{ij'}\right) \\
&= \sum_{j,j'=1}^{b}\sum_{i=1}^{n} a_{jj'ii}\sigma_{jj'}. \quad \square
\end{aligned}$$

Lemma 2.2.4. *Consider* $\boldsymbol{X}$ *as defined in Lemma 2.2.3. Let* $\boldsymbol{A}$ *be a matrix of the same form as in Lemma 2.2.3, and let* $\boldsymbol{B}$ *and* $b_{jj'ii'}$ *be defined similarly. Then,*

$$\begin{aligned}
\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X})\right] = \sum_{j,j',l,l'=1}^{b} \Bigg[& \sum_{i=1}^{n} \mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{ij'}\boldsymbol{X}'_{il}\boldsymbol{X}_{il'}\right) a_{jj'ii}b_{ll'ii} \\
& + \sigma_{jj'}\sigma_{ll'}\sum_{i\neq i'}^{n} a_{jj'ii}b_{ll'i'i'} + \sigma_{jl}\sigma_{j'l'}\sum_{i\neq i'}^{n} a_{jj'ii'}b_{ll'ii'} \\
& + \sigma_{jl'}\sigma_{j'l}\sum_{i\neq i'}^{n} a_{jj'ii'}b_{ll'i'i}\Bigg]. \qquad (2.11)
\end{aligned}$$

Proof: Observe

$$\begin{aligned}
\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X})\right] &= \mathrm{E}\left[\sum_{j,j',i,i'} a_{jj'ii'}\boldsymbol{X}'_{ij}\boldsymbol{X}_{i'j'}\sum_{l,l',k,k'} b_{ll'kk'}\boldsymbol{X}'_{kl}\boldsymbol{X}_{k'l'}\right] \\
&= \mathrm{E}\left[\sum_{j,j',i,i'}\sum_{l,l',k,k'} a_{jj'ii'}b_{ll'kk'}\boldsymbol{X}'_{ij}\boldsymbol{X}_{i'j'}\boldsymbol{X}'_{kl}\boldsymbol{X}_{k'l'}\right]
\end{aligned}$$

$$= \sum_{j,j',i,i'} \sum_{l,l',k,k'} a_{jj'ii'} b_{ll'kk'} \mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{i'j'}\boldsymbol{X}'_{kl}\boldsymbol{X}_{k'l'}\right) \tag{2.12}$$

$$\begin{aligned} &= \sum_{j,j',l,l'=1}^{b} \left[\sum_{i=1}^{n} \mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{ij'}\boldsymbol{X}'_{il}\boldsymbol{X}_{il'}\right) a_{jj'ii} b_{ll'ii} \right. \\ &\sigma_{jj'}\sigma_{ll'} \sum_{i \neq i'}^{n} a_{jj'ii} b_{ll'i'i'} + \sigma_{jl}\sigma_{j'l'} \sum_{i \neq i'}^{n} a_{jj'ii'} b_{ll'ii'} + \\ &\left. + \sigma_{jl'}\sigma_{j'l} \sum_{i \neq i'}^{n} a_{jj'ii'} b_{ll'i'i} \right]. \end{aligned} \tag{2.13}$$

Notice that the four terms in expression (2.13) come from a reorganization of indices from the previous expression (2.12). In order, the four terms in (2.13) correspond to the following relationships among i, i', k, k' in (2.12): (1) $i = i' = k = k'$; (2) $i = i' \neq k = k'$; (3) $i = k \neq i' = k'$; and (4) $i = k' \neq i' = k$. □

Remark: For most of the applications presented later in this chapter, we will have $\boldsymbol{A} = \boldsymbol{B}$.

Notice the second moment above depends on the fourth mixed moments of the data. This component vanishes if the diagonal entries of the blocks of either $\boldsymbol{A}$ or $\boldsymbol{B}$ are all zero.

2.2.3 Asymptotically Equivalent Forms of the Sums of Squares

The goal of this section will be to show that the second component of (2.8), $\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}$, vanishes as b tends to infinity. To begin to see how the Lemmas 2.2.3 and 2.2.4 can be applied to the framework of this chapter, we notice that $\boldsymbol{S}_{\phi_E}$ from (2.7) has the same form as $\boldsymbol{A}$ from the lemmas. Since for arbitrary m; $\boldsymbol{I}_m$, $\boldsymbol{J}_m$, and $\boldsymbol{P}_m$ are symmetric; using the properties of the direct sum and the Kronecker product we can see that $\boldsymbol{C}^*_{\phi_E}$ from (2.6) is symmetric for $\phi \in \{\beta, \gamma\}$. Furthermore, for $\boldsymbol{C}^*_{\beta_E}$ and $\boldsymbol{C}^*_{\gamma_E}$ defined in (2.6), $\boldsymbol{C}^*_{\beta_E}$ is an $n \times n$ matrix (a^2 blocks of size $n_i \times n_{i'}$), such that the

$$\begin{aligned} (i,i)^{th} \text{ block} &= \frac{1}{an_i(n_i-1)}(\boldsymbol{J}_{n_i} - \boldsymbol{I}_{n_i}), \text{ and the} \\ (i,i')^{th} \text{ block} &= \frac{1}{an_in_{i'}}\mathbf{1}_{n_i}\mathbf{1}'_{n_{i'}}, \end{aligned}$$

and $\boldsymbol{C}^*_{\gamma_E}$ is an $n \times n$ matrix (a^2 blocks of size $n_i \times n_{i'}$), such that the

$$\begin{aligned} (i,i)^{th} \text{ block} &= \frac{1}{an_i(n_i-1)}(\boldsymbol{J}_{n_i} - \boldsymbol{I}_{n_i}), \text{ and the} \\ (i,i')^{th} \text{ block} &= \frac{-1}{a(a-1)n_in_{i'}}\mathbf{1}_{n_i}\mathbf{1}'_{n_{i'}}. \end{aligned}$$

To see this, first let the use here (and throughout) of boldface zeros ($\mathbf{0}$) in the top-right and bottom-left corners of a matrix indicate that all non-specified elements are zero. Then, let $\boldsymbol{\delta} = \frac{1}{a}\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}$. Since $1 - \frac{1}{n_i} = \frac{n_i-1}{n_i}$,

$$\boldsymbol{\delta} = \begin{bmatrix} \frac{1}{n_1^2 a}\boldsymbol{I}_{n_1} + \frac{1}{n_1^2(n_1^2-1)a}(\boldsymbol{I}_{n_1} - \boldsymbol{J}_{n_1}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{n_a^2 a}\boldsymbol{I}_{n_a} + \frac{1}{n_a^2(n_a^2-1)a}(\boldsymbol{I}_{n_a} - \boldsymbol{J}_{n_a}) \end{bmatrix}.$$

Second, let $\boldsymbol{\delta}_1 = \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right)\frac{1}{a}\boldsymbol{J}_a\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}'_{n_i}\right)$. Then

$$\boldsymbol{\delta}_1 = \begin{bmatrix} \frac{1}{n_1} & & \mathbf{0} \\ \vdots & & \\ \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_a} \\ & & \vdots \\ \mathbf{0} & & \frac{1}{n_a} \end{bmatrix} \begin{bmatrix} \frac{1}{a} & \cdots & \frac{1}{a} \\ \vdots & \ddots & \vdots \\ \frac{1}{a} & \cdots & \frac{1}{a} \end{bmatrix} \begin{bmatrix} \frac{1}{n_1} & & \mathbf{0} \\ \vdots & & \\ \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_a} \\ & & \vdots \\ \mathbf{0} & & \frac{1}{n_a} \end{bmatrix}'$$

$$= \begin{bmatrix} \frac{1}{n_1^2 a}\mathbf{1}_{n_1}\mathbf{1}'_{n_1} & \frac{1}{n_1 n_2 a}\mathbf{1}_{n_1}\mathbf{1}'_{n_2} & \cdots & \frac{1}{n_1 n_a a}\mathbf{1}_{n_1}\mathbf{1}'_{n_a} \\ \frac{1}{n_1 n_2 a}\mathbf{1}_{n_2}\mathbf{1}'_{n_1} & \frac{1}{n_2^2 a}\mathbf{1}_{n_2}\mathbf{1}'_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{n_2 n_a a}\mathbf{1}_{n_2}\mathbf{1}'_{n_a} \\ \frac{1}{n_1 n_a a}\mathbf{1}_{n_a}\mathbf{1}'_{n_1} & \cdots & \frac{1}{n_{a-1} n_a a}\mathbf{1}_{n_a}\mathbf{1}'_{n_{a-1}} & \frac{1}{n_a^2 a}\mathbf{1}_{n_a}\mathbf{1}'_{n_a} \end{bmatrix}.$$

Third, let $\boldsymbol{\delta}_2 = \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right) \frac{1}{a-1}\boldsymbol{P}_a \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}'_{n_i}\right)$. Then

$$\boldsymbol{\delta}_2 = \begin{bmatrix} \frac{1}{n_1} & & \mathbf{0} \\ \vdots & & \\ \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_a} \\ & & \vdots \\ \mathbf{0} & & \frac{1}{n_a} \end{bmatrix} \begin{bmatrix} \frac{1}{a} & \frac{-1}{a(a-1)} & \cdots & \frac{-1}{a(a-1)} \\ \frac{-1}{a(a-1)} & \frac{1}{a} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{-1}{a(a-1)} \\ \frac{-1}{a(a-1)} & \cdots & \frac{-1}{a(a-1)} & \frac{1}{a} \end{bmatrix} \begin{bmatrix} \frac{1}{n_1} & & \mathbf{0} \\ \vdots & & \\ \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_a} \\ & & \vdots \\ \mathbf{0} & & \frac{1}{n_a} \end{bmatrix}'$$

$$= \begin{bmatrix} \frac{1}{n_1^2 a}\mathbf{1}_{n_1}\mathbf{1}'_{n_1} & \frac{-1}{n_1 n_2 a(a-1)}\mathbf{1}_{n_1}\mathbf{1}'_{n_2} & \cdots & \frac{-1}{n_1 n_a a(a-1)}\mathbf{1}_{n_1}\mathbf{1}'_{n_a} \\ \frac{-1}{n_1 n_2 a(a-1)}\mathbf{1}_{n_2}\mathbf{1}'_{n_1} & \frac{1}{n_2^2 a}\mathbf{1}_{n_2}\mathbf{1}'_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{-1}{n_2 n_a a(a-1)}\mathbf{1}_{n_2}\mathbf{1}'_{n_a} \\ \frac{-1}{n_1 n_a a(a-1)}\mathbf{1}_{n_a}\mathbf{1}'_{n_1} & \cdots & \frac{-1}{n_{a-1} n_a a(a-1)}\mathbf{1}_{n_a}\mathbf{1}'_{n_{a-1}} & \frac{1}{n_a^2 a}\mathbf{1}_{n_a}\mathbf{1}'_{n_a} \end{bmatrix}.$$

The results follow noticing that $\boldsymbol{C}^*_{\beta_E} = \boldsymbol{\delta}_1 - \boldsymbol{\delta}$ and $\boldsymbol{C}^*_{\gamma_E} = \boldsymbol{\delta}_2 - \boldsymbol{\delta}$.

Combining the forms of $\boldsymbol{C}^*_{\beta_E}$ and $\boldsymbol{C}^*_{\gamma_E}$ with Lemmas 2.2.3 and 2.2.4, we have the following corollary about the two moments of the quadratic forms $\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}$.

Proposition 2.2.5. *For $\boldsymbol{X}$ as given in (2.1), let*

$$\mathrm{Cov}\left(\boldsymbol{X}\right) = \mathrm{Var}\left(\boldsymbol{X}\right) := \boldsymbol{\Sigma}_x = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n.$$

For $\phi \in \{\beta, \gamma\}$,

$$\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}\right) = 0,$$

and

$$\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right] = \sum_{\substack{j\neq j'\\ l\neq l'}}^{b}(\sigma_{jl}\sigma_{j'l'}+\sigma_{jl'}\sigma_{j'l})\cdot\frac{1}{b^2(b-1)^2}\cdot\frac{1}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}+\alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_in_{i'}}\right),$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi=\beta \\ \frac{1}{(a-1)^2} & \text{if } \phi=\gamma \end{cases}.$

Proof: First, notice that

$$\boldsymbol{S}_{\phi_E} = \frac{1}{b(b-1)}(\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^*_{\phi_E}$$

is a matrix with all block diagonal entries equal to zero. This is due to the fact that $\boldsymbol{J}_b-\boldsymbol{I}_b$ has zero diagonal elements; then by the properties of scalar and Kronecker products, so does $\boldsymbol{S}_{\phi_E}$. Further, since $\boldsymbol{C}^*_{\phi_E}$ has zero diagonals, $\boldsymbol{S}_{\phi_E}$ will have zero block diagonals. Applying Lemma 2.2.3 then gives

$$\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}\right)=0.$$

To complete the proof, we need to show the second claim in the statement of the theorem holds. For the sake of notation, consider

$$\boldsymbol{A}=\boldsymbol{S}_{\phi_E}=\frac{1}{b(b-1)}(\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^*_{\phi_E},$$

and consider the notation given in Lemma 2.2.4, where now $\boldsymbol{B}=\boldsymbol{A}$. Similar to the arguments above, for all j, j', $a_{jj'ii}=0$. Thus,

$$\sum_{j,j',l,l'=1}^{b}\left[\sum_{i=1}^{n}\mathrm{E}\left(\boldsymbol{X}'_{ij}\boldsymbol{X}_{ij'}\boldsymbol{X}'_{il}\boldsymbol{X}_{il'}\right)a_{jj'ii}a_{ll'ii}+\sigma_{jj'}\sigma_{ll'}\sum_{i\neq i'}^{n}a_{jj'ii}a_{ll'i'i'}\right]=0.$$

Observe that for $j = j'$ or $l = l'$, $a_{jj'ii'} = a_{ll'ii'} = 0$, which would imply

$$\sum_{j,j',l,l'=1}^{b} \sigma_{jl}\sigma_{j'l'} \sum_{i \neq i'}^{n} a_{jj'ii'} a_{ll'ii'} = 0.$$

Consequently, consider $j \neq j'$ and $l \neq l'$. Since $\frac{1}{b(b-1)}(\boldsymbol{J}_b - \boldsymbol{I}_b)$ is constant everywhere except the diagonal elements, for all j, j', l, l' such that $j \neq j'$ and $l \neq l'$, $a_{jj'ii'} = a_{ll'ii'}$. Then, in the case when $\phi = \gamma$, we sum the appropriate elements of $\boldsymbol{A}$ to produce

$$\begin{aligned}
\sigma_{jl}\sigma_{j'l'} \sum_{i \neq i'}^{n} a_{jj'ii'} a_{ll'ii'} &= \sigma_{jl}\sigma_{j'l'} \left(\frac{1}{b(b-1)}\right)^2 \left[\frac{n_1(n_1-1)}{n_1^2(n_1-1)^2a^2} + \cdots + \frac{n_a(n_a-1)}{n_a^2(n_a-1)^2a^2}\right. \\
&\quad \left. + \frac{n_1n_2}{n_1^2(n_2)^2a^2(a-1)^2} + \cdots + \frac{n_{a-1}n_a}{n_{a-1}^2(n_a)^2a^2(a-1)^2}\right] \\
&= \sigma_{jl}\sigma_{j'l'} \cdot \frac{1}{b^2(b-1)^2} \cdot \frac{1}{a^2} \left[\sum_{i=1}^{a} n_i(n_i-1) + \frac{1}{(a-1)^2} \sum_{i \neq i'}^{a} n_i n_{i'}\right];
\end{aligned}$$

in the case when $\phi = \beta$, similar algebra shows

$$\sigma_{jl}\sigma_{j'l'} \sum_{i \neq i'}^{n} a_{jj'ii'} a_{ll'ii'} = \sigma_{jl}\sigma_{j'l'} \cdot \frac{1}{b^2(b-1)^2} \cdot \frac{1}{a^2} \left[\sum_{i=1}^{a} n_i(n_i-1) + \sum_{i \neq i'}^{a} n_i n_{i'}\right].$$

Also observe that for $j = j'$ or $l = l'$, $a_{jj'ii'} = 0 = a_{ll'i'i}$, which would imply

$$\sum_{j,j',l,l'=1}^{b} \sigma_{jl'}\sigma_{j'l} \sum_{i \neq i'}^{n} a_{jj'ii'} a_{ll'i'i} = 0.$$

Again, consider $j \neq j'$ and $l \neq l'$. Since $\boldsymbol{C}^*_{\phi_E}$ is symmetric, we know for all $j \neq j'$ and $l \neq l'$, $a_{jj'ii'} = a_{ll'i'i}$, for $i = 1, 2, \ldots, n$. Therefore, for all j, j', l, l' such that $j \neq j'$ and $l \neq l'$, $a_{jj'ii'} = a_{ll'i'i}$. Thus, by arguments similar to those above, in the case when $\phi = \gamma$,

$$\sigma_{jl'}\sigma_{j'l} \sum_{i \neq i'}^{n} a_{jj'ii'} a_{ll'i'i} = \sigma_{jl'}\sigma_{j'l} \cdot \frac{1}{b^2(b-1)^2} \cdot \frac{1}{a^2} \left[\sum_{i=1}^{a} n_i(n_i-1) + \frac{1}{(a-1)^2} \sum_{i \neq i'}^{a} n_i n_{i'}\right]; \quad (2.14)$$

in the case when $\phi = \beta$,

$$\sigma_{jl'}\sigma_{j'l}\sum_{i\neq i'}^{n} a_{jj'ii'}a_{ll'i'i} = \sigma_{jl'}\sigma_{j'l}\cdot\frac{1}{b^2(b-1)^2}\cdot\frac{1}{a^2}\left[\sum_{i=1}^{a} n_i(n_i-1)+\sum_{i\neq i'}^{a} n_i n_{i'}\right]. \tag{2.15}$$

It follows from combining (2.14) and (2.15) that

$$\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right] = \sum_{\substack{j\neq j'\\ l\neq l'}}^{b}(\sigma_{jl}\sigma_{j'l'}+\sigma_{jl'}\sigma_{j'l})\cdot\frac{1}{b^2(b-1)^2}\cdot\frac{1}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}+\alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right),$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi=\beta \\ \frac{1}{(a-1)^2} & \text{if } \phi=\gamma \end{cases}$. □

Since the first moment is 0, we see straightaway that $\mathrm{Var}(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}) = \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right]$. We also have an immediate corollary. In essence, if we assume that the covariance terms decay fast enough, then the second moment converges to zero, which, by Markov's Inequality, implies the quadratic form $\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X}$ converges to zero in probability [29].

Assumption 2.2.6. *For some $\delta > 0$,*

$$\sum_{\substack{j\neq j'\\ l\neq l'}}^{b}(\sigma_{jl}\sigma_{j'l'}+\sigma_{jl'}\sigma_{j'l}) = O(b^{3-\delta}) \quad as \quad b\to\infty.$$

Corollary 2.2.7. *Suppose Assumption 2.2.6 holds. Then, for $\phi \in \{\beta,\gamma\}$,*

$$\sqrt{b}\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X} = o_p(1) \quad as \quad b\to\infty.$$

Proof: Since a and n_i; $i = 1, 2, \ldots, n$; are fixed constants we recognize that

$$\mathrm{E}\left[(\sqrt{b}\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right] = b\left[\sum_{\substack{j\neq j'\\ l\neq l'}}^{b}(\sigma_{jl}\sigma_{j'l'}+\sigma_{jl'}\sigma_{j'l})\right]O(b^{-4}).$$

Then by Assumption 2.2.6,

$$\mathrm{E}\left[(\sqrt{b}\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right] = O(b^{3-\delta})O(b^{-3}) = O(b^{-\delta}) \quad \text{as} \quad b \to \infty.$$

Therefore,

$$\mathrm{E}\left[(\sqrt{b}\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X})^2\right] \to 0 \quad \text{as} \quad b \to \infty,$$

and since convergence in quadratic mean implies converges in probability [29],

$$\sqrt{b}\boldsymbol{X}'\boldsymbol{S}_{\phi_E}\boldsymbol{X} = o_p(1). \quad \square$$

2.3 Test Statistics

We now turn our attention to the development of the test statistics. From Corollary 2.2.7, we see that we can now consider

$$\sqrt{b}\,(AM\phi - AME) = \sqrt{b}\boldsymbol{X}'\left(\frac{1}{b}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}\right)\boldsymbol{X} + o_p(1), \tag{2.16}$$

which essentially means that the left-hand-side and the first term in the right-hand-side have the same asymptotic distribution by Slutsky's Theorem [20]. From here, we first define asymptotically equivalent versions of the test statistics. The rationale behind this is that the asymptotically equivalent versions are much more tractable.

For $\phi \in \{\beta, \gamma\}$, define asymptotically equivalent versions of the test statistics as

$$T_\phi := \sqrt{b}\left(\boldsymbol{X}'\left[\frac{1}{b}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}\right]\boldsymbol{X}\right) = \frac{1}{\sqrt{b}}\sum_{j=1}^{b}\boldsymbol{X}_j'\boldsymbol{C}^*_{\phi_E}\boldsymbol{X}_j. \tag{2.17}$$

We then have the following results regarding the mean and variance of T_ϕ.

Proposition 2.3.1. *For* $\phi \in \{\beta, \gamma\}$,

$$\mathrm{E}\left(T_\phi\right) = 0.$$

Proof: Since $\boldsymbol{C}^*_{\phi_E}$ has zeros in all of its diagonal elements, $\frac{1}{\sqrt{b}}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}$ will have zeros in all of its block diagonal elements. Then, by Lemma 2.2.3, $\mathrm{E}\left(T_\phi\right) = 0$. □

Proposition 2.3.2. *For* $\phi \in \{\beta, \gamma\}$,

$$\mathrm{Var}\left(T_\phi\right) = \mathrm{E}\left(T_\phi^2\right) = \frac{2}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + \alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right)\cdot\frac{1}{b}\left(\sum_{j,l=1}^{b}\sigma_{jl}^2\right),$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi = \beta \\ \frac{1}{(a-1)^2} & \text{if } \phi = \gamma \end{cases}.$

Proof: Since $\boldsymbol{C}^*_{\phi_E}$ has zeros in all of its diagonal elements, $\frac{1}{\sqrt{b}}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}$ will have zeros in all of its block diagonal elements. Then, by Lemma 2.2.4, the first two terms in (2.11) will be zero. Further, by the symmetry of $\boldsymbol{I}_b$, $\boldsymbol{C}^*_{\phi_E}$, and thereby $\frac{1}{\sqrt{b}}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}$, the third and fourth terms in the bracketed sum from Lemma 2.2.4 are equal. Thus we can simplify, letting $\boldsymbol{A} = \frac{1}{\sqrt{b}}\boldsymbol{I}_b \otimes \boldsymbol{C}^*_{\phi_E}$, to see

$$\mathrm{E}\left(T_\phi^2\right) = \sum_{j,j',l,l'}^{b}\left(\sigma_{jl}\sigma_{j'l'} + \sigma_{jl'}\sigma_{j'l}\right)\sum_{i\neq i'}^{n} a_{jj'ii'}a_{ll'ii'}.$$

Because of the properties of the identity matrix and the Kronecker product, $a_{jj'ii'}$ and $a_{ll'ii'}$ will only be nonzero when $j = j'$ and $l = l'$, respectively. Since all of the $b \times b$ diagonal blocks are identical, we can simplify further to see

$$\mathrm{E}\left(T_\phi^2\right) = \sum_{i\neq i'}^{n} a_{11ii'}^2 \sum_{j,l=1}^{b} 2\sigma_{jl}^2.$$

For $\phi = \beta$,

$$\begin{aligned}\sum_{i\neq i'}^{n} a_{11ii'}^2 &= \sum_{i=1}^{a}\left(\frac{1}{an_i(n_i-1)}\right)^2 (n_i(n_i-1)) + \sum_{i\neq i'}^{a}\left(\frac{1}{an_in_{i'}}\right)^2 (n_in_{i'}) \\ &= \sum_{i=1}^{a}\frac{1}{a^2n_i(n_i-1)} + \sum_{i\neq i'}^{a}\frac{1}{a^2n_in_{i'}}.\end{aligned}$$

For $\phi = \gamma$, the only difference is the first term in the second sum, which is $\frac{-1}{a-1}$ times the corresponding terms above. Since this term is squared, the negative is irrelevant, and $(a-1)^2$ is included in the denominator of the second sum. This gives

$$\sum_{i\neq i'}^{n} a_{11ii'}^2 = \frac{1}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + \alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_in_{i'}}\right),$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi = \beta \\ \frac{1}{(a-1)^2} & \text{if } \phi = \gamma \end{cases}$. Notice that this term multiplied by 2 is free of b. Thus, by Assumption 2.4.2, we can write

$$\mathrm{E}\left(T_\phi^2\right) = \frac{1}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + \alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_in_{i'}}\right)\cdot\frac{1}{b}\sum_{j,l=1}^{b}\sigma_{jl}^2,$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi = \beta \\ \frac{1}{(a-1)^2} & \text{if } \phi = \gamma \end{cases}$. Finally, since $\mathrm{E}\left(T_\phi\right) = 0$ by Proposition 2.3.1, $\mathrm{Var}\left(T_\phi\right) = \mathrm{E}\left(T_\phi^2\right)$, and the result is proven. □

The SS defined above, and consequently the test statistics based on them, would only be useful if they can detect departures from the null hypotheses. Since we are examining group and interaction effects, it is reasonable to want the proper sums of squares $(AM\phi)$ to be equal to the error sums of squares (AME) only under the appropriate null hypotheses. This is the subject of Proposition 2.3.3.

Proposition 2.3.3. *For $\phi \in \{\beta, \gamma\}$, $\mathrm{E}\left(AM\phi\right) = \mathrm{E}\left(AME\right)$ if and only if H_0^ϕ holds.*

Proof: Recall from (2.8) that, for $\phi \in \{\beta, \gamma\}$, we can write

$$AM\phi - AME = \boldsymbol{X}' \left[\left(\frac{1}{b}\boldsymbol{I}_b - \frac{1}{b(b-1)}(\boldsymbol{J}_b - \boldsymbol{I}_b) \right) \otimes \boldsymbol{C}^*_{\phi_E} \right] \boldsymbol{X}.$$

Let $\boldsymbol{D} = \frac{1}{b}\boldsymbol{I}_b - \frac{1}{b(b-1)}(\boldsymbol{J}_b - \boldsymbol{I}_b)$. Observe that, since $\boldsymbol{C}^*_{\phi_E}$ is symmetric, for $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$,

$$\begin{aligned}
\mathrm{E}(AM\phi - AME) &= \mathrm{tr}\left[(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)\right] + \boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu} \\
&= \mathrm{tr}\left(\boldsymbol{D\Sigma} \otimes \boldsymbol{C}^*_{\phi_E}\right) + \boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu} \\
&= \mathrm{tr}\left(\boldsymbol{D\Sigma}\right) \mathrm{tr}\left(\boldsymbol{C}^*_{\phi_E}\right) + \boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu} \\
&= \boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu},
\end{aligned}$$

since the zero diagonal elements of $\boldsymbol{C}^*_{\phi_E}$ imply $\mathrm{tr}\left(\boldsymbol{C}^*_{\phi_E}\right) = 0$.

It now suffices to show that $\boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu} = 0$ if and only if H_0^ϕ holds. Recall the μ_{ij}, the identifiability constraints, and the hypotheses H_0^β and H_0^γ from the beginning of Section 2.2. Analogous to (2.3), denote

$$\begin{aligned}
\tilde{\mu}_{\cdots} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}\mu_{ijk} = \mu, \\
\tilde{\mu}_{i\cdot\cdot} &= \frac{1}{b}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}\mu_{ijk} = \mu + \alpha_i, \\
\tilde{\mu}_{\cdot j\cdot} &= \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{k=1}^{n_i}\mu_{ijk} = \mu + \beta_j, \\
\bar{\mu}_{ij\cdot} &= \frac{1}{n_i}\sum_{k=1}^{n_i}\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \\
\bar{\mu}_{i\cdot k} &= \frac{1}{b}\sum_{j=1}^{b}\mu_{ijk} = \mu + \alpha_i.
\end{aligned} \tag{2.18}$$

Observe that since we consider the model $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, all instances of $\frac{1}{n_i}\sum_{k=1}^{n_i}\mu_{ijk}$ could be written μ_{ij}.

Notice from the third equality in (2.8) that

$$\boldsymbol{\mu}'(\boldsymbol{D} \otimes \boldsymbol{C}^*_{\phi_E})\boldsymbol{\mu} = \boldsymbol{\mu}'(\boldsymbol{C}_\phi - \boldsymbol{C}_E)\boldsymbol{\mu} = \boldsymbol{\mu}'\boldsymbol{C}_\phi\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{C}_E\boldsymbol{\mu}.$$

The proof will be complete if we show that $\boldsymbol{\mu}'\boldsymbol{C}_E\boldsymbol{\mu} = 0$ and that $\boldsymbol{\mu}'\boldsymbol{C}_\phi\boldsymbol{\mu} = 0$ if and only if H_0^ϕ holds. To that end, observe that

$$\begin{aligned}
\boldsymbol{\mu}'(\boldsymbol{C}_E)\boldsymbol{\mu} &= \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(\mu_{ijk} - \bar{\mu}_{ij\cdot} - \bar{\mu}_{i\cdot k} + \tilde{\mu}_{i\cdot\cdot})^2 \\
&= \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}\left(\mu_{ij} - \mu_{ij} - \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} + \frac{1}{b}\sum_{j=1}^{b}\mu_{ij}\right)^2 \\
&= \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}0^2 \;=\; 0,
\end{aligned}$$

where the second and third equalities follow from (2.18).

Finally, by (2.18),

$$\boldsymbol{\mu}'(\boldsymbol{C}_\beta)\boldsymbol{\mu} = \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\tilde{\mu}_{\cdot j\cdot} - \tilde{\mu}_{\cdot\cdot\cdot})^2 = \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_j^2,$$

and

$$\boldsymbol{\mu}'(\boldsymbol{C}_\gamma)\boldsymbol{\mu} = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{\mu}_{ij\cdot} - \tilde{\mu}_{i\cdot\cdot} - \tilde{\mu}_{\cdot j\cdot} + \tilde{\mu}_{\cdot\cdot\cdot})^2 = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}^2.$$

We can now see that $\boldsymbol{\mu}'\boldsymbol{C}_\phi\boldsymbol{\mu} = 0$ if and only if H_0^ϕ holds, and the theorem is proved. □

2.4 Asymptotic Distributions

We now wish to derive the asymptotic distributions of the asymptotically equivalent versions of the test statistics, T_ϕ, from (2.17), for $\phi \in \{\beta, \alpha\}$. In addition to the assumptions already made, this will require some general assumptions which are fairly weak. We will also need an arsenal of lemmas. These test statistics are used to test the main effects of factor B, which is high dimensional (i.e., there are b levels of factor B and b tends to infinity), and the interaction effects between factors A and B, which is also high-dimensional. In practice, factor B is often a temporal factor (longitudinal or repeated measurements), and factor A is often a grouping factor.

From there, we will present a test statistic for factor A. While it relies on the high dimensional nature of factor B, the asymptotic technique will be distinct from that of the statistics regarding factor B and the interaction.

First we must prove some basic properties regarding functions of stationary, α-mixing sequences (both to be defined soon), as well as an inequality regarding moments of the random variables used. Then we will derive the asymptotic distribution of T_ϕ with the aid of Theorem 27.5 from Billingsley [8].

2.4.1 Testing for the Main Effects of Factor B and the Interaction Effects

The goal of this section is to derive the asymptotic distribution of T_ϕ, which, in turn, will also allow us to define the asymptotic distribution of $\sqrt{b}(AM\phi - AME)$. This holds because we showed earlier that $\sqrt{b}(AM\phi - AME)$ can be written as T_ϕ plus an asymptotically negligible component, and the latter can be ignored as b tends to infinity.

First we need to define stationarity, a basic definition in the study of time series [41].

Definition 2.4.1. *Consider a sequence of RVs $\{Y_t\}$. We say that $\{Y_t\}$ is stationary if for any collection of indices $\{t_1, \dots, t_n\}$, $n = 1, 2, \dots$, the distribution of $\{Y_{t_1}, \dots, Y_{t_n}\}$ is the same as the distribution of the time-shifted series $\{Y_{t_1+m}, \dots, Y_{t_n+m}\}$ for any time-shift $m \in \mathbb{Z}$; that is,*

$$\{Y_{t_1}, \dots, Y_{t_n}\} \stackrel{d}{=} \{Y_{t_1+m}, \dots, Y_{t_n+m}\}$$

for $m \in \mathbb{Z}$ and all $n = 1, 2, \dots$. If $\{Y_t\}$ is stationary, we say that stationarity holds for $\{Y_t\}$.

Assumption 2.4.2. *For all i and k, $\{X_{ijk}\}_{j=1}^{b}$ is stationary.*

The assumption of stationarity (among others) allows for a simplified expression for the covariance terms that appear in the variance of T_ϕ. Under the assumption of stationarity, Proposition 2.3.2 has the following corollary.

Corollary 2.4.3. *Suppose Assumption 2.4.2 holds. Then, for $\phi \in \{\beta, \gamma\}$,*

$$\mathrm{Var}\left(T_\phi\right) = \mathrm{E}\left(T_\phi^2\right) = \frac{2}{a^2}\left(\sum_{i=1}^{a} \frac{1}{n_i(n_i-1)} + \alpha(\phi)\sum_{i \neq i'}^{a} \frac{1}{n_i n_{i'}}\right)\left[\sigma_{11}^2 + 2\sum_{j=1}^{b-1}\left(1-\frac{j}{b}\right)\sigma_{1,1+j}^2\right],$$

where $\alpha(\phi) = \begin{cases} 1 & \text{if } \phi = \beta \\ \frac{1}{(a-1)^2} & \text{if } \phi = \gamma \end{cases}.$

Proof: Continuing from the last line of the proof of Proposition 2.3.2, we see that under Assumption 2.4.2,

$$\begin{aligned}
\sum_{j,l=1}^{b} \sigma_{jl}^2 &= \frac{1}{b}\left[\sum_{j=1}^{b} \sigma_{jj}^2 + 2\sum_{j=1}^{b-1}(b-j)\sigma_{1,1+j}^2\right] \\
&= \frac{1}{b}\left[b\sigma_{11}^2 + 2\sum_{j=1}^{b-1}(b-j)\sigma_{1,1+j}^2\right] \\
&= \sigma_{11}^2 + 2\sum_{j=1}^{b-1}\left(1-\frac{j}{b}\right)\sigma_{1,1+j}^2. \quad \square
\end{aligned}$$

Next we need to define α-mixing. The concept was first discussed in the mid-1950s by Rosenblatt [38], who presented a new CLT, though we will take the technical definition from Bradley [11] and Billingsley [8].

Definition 2.4.4. *Consider a probability space $(\Omega, \mathcal{F}, P)$, and any two σ-fields (or σ-algebras)*

$\mathcal{A} \subset \mathcal{F}$ and $\mathcal{B} \subset \mathcal{F}$. Define

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A) P(B)|.$$

Consider a sequence of random variables $\{X_t\}$ which is not necessarily stationary. For $-\infty \leq K \leq L \leq \infty$, define the σ-field

$$\mathcal{F}_K^L := \sigma(X_t, K \leq t \leq L),$$

which is the σ-field generated by $X_k, \ldots, X_L$. For each $m \geq 1$, define the dependence coefficient

$$\alpha_X(m) := \sup_{k \in \mathbb{Z}} \alpha\left(\mathcal{F}_{\infty}^k, \mathcal{F}_{k+m}^{\infty}\right).$$

Then $\{X_t\}$ is said to be α-mixing with $\alpha_X(m)$ if $\alpha_X(m) \to 0$ as $m \to \infty$.

There are many types of mixing, known as mixing conditions, which all describe some measure of dependence. α-mixing, also know as strong mixing, is the most common of these mixing conditions; for a complete list see Bradley (2005) [11]. m-dependence, a common notion in many time series contexts, is the highest ranking of the mixing conditions, implying all the others, while α-mixing is the lowest in the hierarchy.

For a sequence to be α-mixing means that the dependence in the elements of the sequence decays as the lag between them increases. The purpose of the following lemma is to show that measurable functions of independent, stationary, α-mixing sequences are also stationary and α-mixing. This result is primarily needed because T_ϕ is a scaled average of functions of multiple independent sequences.

Lemma 2.4.5. *Suppose $\{X_t^{(1)}\}$, $\{X_t^{(2)}\}$, $\ldots$, $\{X_t^{(r)}\}$ are independent, stationary, α-mixing sequences with $\alpha_{X^{(i)}}(m)$ for $i = 1, \ldots, r$, and suppose g is a measurable function. Define the sequence $\{Z_t\}$ by $Z_t = g(X_t^{(1)}, \ldots, X_t^{(r)})$. Then $\{Z_t\}$ is stationary and α-mixing.*

Proof: To prove that stationarity holds, consider, for any integer $k \geq 1$, a collection of indices $\{t_1, \ldots, t_n\}$ where $n \leq k$. Consider any $m \in \mathbb{Z}$. Under the assumption of stationarity, we know that

$$\begin{aligned}
\{X_{t_1}^{(1)}, \ldots, X_{t_n}^{(1)}\} &\stackrel{d}{=} \{X_{t_1+m}^{(1)}, \ldots, X_{t_n+m}^{(1)}\} \\
&\vdots \\
\{X_{t_1}^{(r)}, \ldots, X_{t_n}^{(r)}\} &\stackrel{d}{=} \{X_{t_1+m}^{(r)}, \ldots, X_{t_n+m}^{(r)}\}.
\end{aligned}$$

Consider any set A contained in the support set of $(Z_{i_1}, \ldots, Z_{i_n})$, and denote its inverse image by B. Then $B = \{(X_{i_1}^{(1)}, \ldots, X_{i_n}^{(1)}, \ldots, X_{i_1}^{(r)}, \ldots, X_{i_n}^{(r)}) : (Z_{i_1}, \ldots, Z_{i_n}) \in A\}$.

Then, since g is measurable and the $\{X_t^{(i)}\}$ are stationarity,

$$\begin{aligned}
P\left[(Z_{i_1}, \ldots, Z_{i_n}) \in A\right] &= P\left[\left(X_{i_1}^{(1)}, \ldots, X_{i_n}^{(1)}, \ldots, X_{i_1}^{(r)}, \ldots, X_{i_n}^{(r)}\right) \in B\right] \\
&= P\left[\left(X_{i_1+m}^{(1)}, \ldots, X_{i_n+m}^{(1)}, \ldots, X_{i_1+m}^{(1)}, \ldots, X_{i_n+m}^{(1)}\right) \in B\right] \\
&= P\left[(Z_{i_1+m}, \ldots, Z_{i_n+m}) \in A\right].
\end{aligned}$$

Therefore, $\{Z_{t_1}, \ldots, Z_{t_n}\} \stackrel{d}{=} \{Z_{t_1+m}, \ldots, Z_{t_n+m}\}$; thus, $\{Z_t\}$ is stationary.

To prove that α-mixing holds, we use Theorem 5.2 from Bradley [11]. Here we have a finite number of sequences, r, whereas Bradley's Theorem 5.2 has infinitely many. To reconcile this, we simply define all sequences past the initial r sequences to be sequences of zeros [11]. Then we have

$$\alpha_Z(m) \leq \sum_{i=1}^{r} \alpha_{X^{(i)}}(m) \to 0 \quad \text{as} \quad r \to \infty,$$

so $\{Z_t\}$ is α-mixing. □

Using Lemma 2.4.5, we wish to show that the asymptotically equivalent version of the test statistic, T_ϕ, defined in (2.17) is stationary and α-mixing. To do this, we use the form in the last expression of (2.17). Recall that

$$\boldsymbol{X}_j = \begin{bmatrix} X_{1j1} & \cdots & X_{1jn_1} & \cdots & X_{aj1} & \cdots & X_{ajn_a} \end{bmatrix}'$$

is an $n \times 1$ vector whose elements are mutually independent. Then define

$$Z_j = \boldsymbol{X}_j' \left(\frac{1}{\sqrt{b}} \boldsymbol{C}_{\phi_E}^* \right) \boldsymbol{X}_j. \tag{2.19}$$

Proposition 2.4.6. *For Z_j defined in (2.19), $\{Z_j\}$ is stationary and α-mixing.*

For an arbitrary matrix $\boldsymbol{A}$, the notation $[\boldsymbol{A}]_{ij}$ will represent the $(i,j)^{th}$ element of $\boldsymbol{A}$; for a vector $\boldsymbol{a}$, the notation $[\boldsymbol{a}]_i$ will represent the i^{th} element of $\boldsymbol{a}$.

Proof: First observe that we can write

$$Z_j = \sum_{l,l'=1}^{n} \left(\frac{1}{\sqrt{b}} \left[\boldsymbol{C}_{\phi_E}^* \right]_{ll'} \right) [\boldsymbol{X}_j]_l [\boldsymbol{X}_j]_{l'}.$$

Then we can rearrange and write n independent, stationary, α-mixing sequences as follows. Let

$$\begin{aligned}
Y^{(1)} &= X_{111}, X_{121}, X_{131}, \ldots \\
&\vdots \\
Y^{(n_1)} &= X_{11n_1}, X_{12n_1}, X_{13n_1}, \ldots \\
Y^{(n_1+1)} &= X_{211}, X_{221}, X_{231}, \ldots \\
&\vdots \\
Y^{(n_1+n_2)} &= X_{21n_2}, X_{22n_2}, X_{23n_2}, \ldots \\
&\vdots \\
Y^{(n)} &= X_{a1n_a}, X_{a2n_a}, X_{a3n_a}, \ldots
\end{aligned}$$

where we recall that $n = \sum_{i=1}^{a} n_i$.

Let $Y_j^{(i)}$ be the j^{th} element of $Y^{(i)}$, $i = 1, \ldots, n$. Then we can write

$$Z_j = g\left(Y_j^{(1)}, \ldots, Y_j^{(n)}\right) = \sum_{l,l'=1}^{n} \left(\frac{1}{\sqrt{b}} \left[\boldsymbol{C}_{\phi_E}^*\right]_{ll'}\right) \begin{bmatrix} Y_j^{(1)} \\ \vdots \\ Y_j^{(n)} \end{bmatrix}_l \begin{bmatrix} Y_j^{(1)} \\ \vdots \\ Y_j^{(n)} \end{bmatrix}_{l'}.$$

Then by Lemma 2.4.5, $\{Z_j\} = \left\{\boldsymbol{X}_j' \left(\frac{1}{\sqrt{b}} \boldsymbol{C}_{\phi_E}^*\right) \boldsymbol{X}_j\right\}$, as in (2.19), is stationary and α-mixing. $\square$

We now state the next assumption. The following assumption is necessary to appeal to the dependent CLT used in the proof of main theorem of this chapter.

Assumption 2.4.7. *For all i and k, $\{X_{ijk}\}_{j=1}^{b}$ is α-mixing with $\alpha_X(m) = O(m^{-5})$.*

The major tool for finding the asymptotic distribution is the dependent CLT proved in Billingsley [8]. Since $\mathrm{E}(Z_j) = 0$ as shown in the proof of Proposition 2.4.8, we need $\mathrm{E}\left(Z_j^{12}\right) < \infty$ in order to apply Billingsley's theorem.

Proposition 2.4.8. *For Z_j defined in (2.19),*

$$\mathrm{E}(Z_j) = 0.$$

Proof: Recall that $\boldsymbol{C}_{\phi_E}^*$ has zeros in all of its diagonal elements, and thus $\frac{1}{\sqrt{b}} \boldsymbol{C}_{\phi_E}^*$ will as well. Then by Lemma 2.2.3,

$$\mathrm{E}(Z_j) = \mathrm{E}\left[\boldsymbol{X}_j' \left(\frac{1}{\sqrt{b}} C_{\phi_E}^*\right) \boldsymbol{X}_j\right] = 0. \quad \square$$

In order to appeal to the CLT found in Billingsley [8], we need to make an appropriate assumption regarding the 24^{th} moments of the random variable X_{ijk}. The summands in

Billingsley's CLT must have finite 12^{th} moments in order for the theorem to hold [8]. Proposition 2.4.9 shows that it is sufficient for the 24^{th} moment of X_{ijk} to be finite for all i and k.

Proposition 2.4.9. *For Z_j defined in (2.19), if* $\mathrm{E}\left(X_{ijk}^{24}\right) < \infty$ *for all i and k, then* $\mathrm{E}\left(Z_j^{12}\right) < \infty$.

Proof: In this proof, for the sake of notation, let $\boldsymbol{A} = \frac{1}{\sqrt{b}}\boldsymbol{C}^*_{\phi_E}$, where a_{ij} will denote the $(i,j)^{th}$ element of $\boldsymbol{A}$. Also, since $\boldsymbol{X}_j$ is an $n \times 1$ vector, we will use the notation $[\boldsymbol{X}_j]_l$ to denote the l^{th} element of $\boldsymbol{X}_j$, where $l = 1, \ldots, n$. However, note that given j, for a fixed l, $[\boldsymbol{X}_j]_l$ is simply some X_{ijk} for some i and some k. Then the hypothesis of the lemma can be written as $\mathrm{E}\left([\boldsymbol{X}_j]_l\right) < \infty$ for all l.

Suppose then that $\mathrm{E}\left([\boldsymbol{X}_j]_n\right) < \infty$ for all n. Then we know for $r \leq 24$, $\mathrm{E}\left(|\,[\boldsymbol{X}_j]_n\,|^r\right) < \infty$ for all n. Observe,

$$\begin{aligned}
\mathrm{E}\left[\left(\boldsymbol{X}_j'\boldsymbol{A}\boldsymbol{X}_j\right)^{12}\right] &= \mathrm{E}\left[\left(\sum_{l,l'=1}^{n} a_{ll'}\,[\boldsymbol{X}_j]_l\,[\boldsymbol{X}_j]_{l'}\right)^{12}\right] \\
&= \mathrm{E}\left[\left(\sum_{l=1}^{n} a_{ll}\,[\boldsymbol{X}_j]_l^2 + \sum_{\substack{l,l'=1\\ l\neq l'}}^{n} a_{ll'}\,[\boldsymbol{X}_j]_l\,[\boldsymbol{X}_j]_{l'}\right)^{12}\right] \\
&\leq \left(\max_{l,l'}|a_{ll'}|\right)^{12} \mathrm{E}\left[\left|\sum_{l=1}^{n} [\boldsymbol{X}_j]_l^2 + \sum_{\substack{l,l'=1\\ l\neq l'}}^{n} [\boldsymbol{X}_j]_l\,[\boldsymbol{X}_j]_{l'}\right|^{12}\right].
\end{aligned}$$

Let $\left(\max_{l,l'}|a_{ll'}|\right)^{12} = \kappa$. Then by the c_r-inequality twice applied [29], we proceed to see that

$$\mathrm{E}\left[\left(\boldsymbol{X}_j'\boldsymbol{A}\boldsymbol{X}_j\right)^{12}\right] \leq \kappa \cdot 2^{11}\left[\mathrm{E}\left(\left|\sum_{l=1}^{n} [\boldsymbol{X}_j]_l^2\right|^{12}\right) + \mathrm{E}\left(\left|\sum_{\substack{l,l'=1\\ l\neq l'}}^{n} [\boldsymbol{X}_j]_l\,[\boldsymbol{X}_j]_{l'}\right|^{12}\right)\right]$$

$$\begin{aligned}
&\leq \kappa_1 \left[\sum_{l=1}^{n} 2^{11} \cdot \mathrm{E}\left[\left([\boldsymbol{X}_j]_l^2 \right)^{12} \right] + \sum_{\substack{l,l'=1 \\ l \neq l'}}^{n} 2^{11} \cdot \mathrm{E}\left[\left([\boldsymbol{X}_j]_l \, [\boldsymbol{X}_j]_{l'} \right)^{12} \right] \right] \\
&= \kappa_2 \left[\sum_{l=1}^{n} \mathrm{E}\left([\boldsymbol{X}_j]_l^{24} \right) + \sum_{\substack{l,l'=1 \\ l \neq l'}}^{n} \mathrm{E}\left([\boldsymbol{X}_j]_l^{12} \, [\boldsymbol{X}_j]_{l'}^{12} \right) \right] \\
&= \kappa_2 \left[\sum_{l=1}^{n} \mathrm{E}\left([\boldsymbol{X}_j]_l^{24} \right) + \sum_{\substack{l,l'=1 \\ l \neq l'}}^{n} \mathrm{E}\left([\boldsymbol{X}_j]_l^{12} \right) \mathrm{E}\left([\boldsymbol{X}_j]_{l'}^{12} \right) \right],
\end{aligned}$$

where the last step is justified since $[\boldsymbol{X}_j]_l$ and $[\boldsymbol{X}_j]_{l'}$ are independent for $l \neq l'$, $l, l' = 1, \ldots, n$; and κ_1 and κ_2 absorb the necessary constants in the intermediate steps.

Since all of the expectations in the final expression are finite, we have

$$\mathrm{E}\left(Z_j^{12} \right) = \mathrm{E}\left[\left(\boldsymbol{X}_j' \boldsymbol{A} \boldsymbol{X}_j \right)^{12} \right] < \infty. \quad \square$$

We now have the last assumption necessary before presenting the major theorem of this chapter.

Assumption 2.4.10. *For all i and k,* $\mathrm{E}\left(X_{ijk}^{24} \right) < \infty$.

We use the previous lemmas to derive a test statistic, which we shall call T_ϕ^*, for testing the hypotheses of a main effect of factor B (H_0^β, often temporal) and an interaction effect between factors A and B (H_0^γ, often temporal versus group/level). As a reminder, let us first give a list of assumptions that we have hitherto made.

Assumption 2.2.1. *The data X_{ijk} are RVs with mean μ_{ij}, where $X_{ijk} \perp X_{i'jk'}$ for $i \neq i'$ or $k \neq k'$.*

Assumption 2.2.6. *For some* $\delta > 0$*,*

$$\sum_{\substack{j \neq j' \\ l \neq l'}}^{b} (\sigma_{jl}\sigma_{j'l'} + \sigma_{jl'}\sigma_{j'l}) = O(b^{3-\delta}) \quad as \quad b \to \infty.$$

Assumption 2.4.2. *For all* i *and* k*,* $\{X_{ijk}\}_{j=1}^{b}$ *is stationary.*

Assumption 2.4.7. *For all* i *and* k*,* $\{X_{ijk}\}_{j=1}^{b}$ *is* α*-mixing with* $\alpha_X(m) = O(m^{-5})$*.*

Assumption 2.4.10. *For all* i *and* k*,* $\mathrm{E}\left(X_{ijk}^{24}\right) < \infty$*.*

Now we can state the theorem regarding the asymptotic distributions of the test statistics for the main effect of factor B and the interaction effect. Although the assumption of stationarity is needed to appeal to Billingsley's theorem, in general this assumption is not needed as shown in the multivariate case (Chapter 4). The assumption of stationarity is also useful in finding a tractable consistent estimator of the variance of the test statistic.

Theorem 2.4.11. *Suppose Assumptions 2.2.1, 2.2.6, 2.4.2, 2.4.7, and 2.4.10 hold. For* $\phi \in \{\beta, \gamma\}$*, define*

$$T_\phi^* = \sqrt{b}\,(AM\phi - AME)\,. \tag{2.20}$$

Then, for $\sigma_\phi^2 = \mathrm{Var}\,(T_\phi)$ *from Proposition 2.3.2, and under* H_0^ϕ*,*

$$\frac{T_\phi^*}{\sigma_\phi} \xrightarrow{D} \mathcal{Z} \quad as \quad b \to \infty,$$

where $\mathcal{Z} \sim N(0,1)$*.*

Proof: By Corollary 2.2.7, we know that T_ϕ is asymptotically equivalent to T_ϕ^*, since T_ϕ^* is equal to T_ϕ plus some asymptotically negligible component. Therefore, the asymptotic results that hold for T_ϕ will also hold for T_ϕ^*.

Using Assumption 2.2.1, we know that the Z_js defined in (2.19) are independent and identically distributed. By Proposition 2.4.8, $\mathrm{E}\,(Z_j) = 0$. Using Assumption 2.4.10, we appeal to

Proposition 2.4.9 to see that $\mathrm{E}\left(Z_j^{12}\right) < \infty$. Using Assumptions 2.4.2 and 2.4.7, we see by Proposition 2.4.6 that Z_j is stationary and α-mixing.

We see that these correspond to the conditions of Theorem 27.5 from Billingsley [8], so the desired result is proved where T_ϕ replaces T_ϕ^*. However, since these are asymptotically equivalent as $b \to \infty$, the result is proved for T_ϕ^*. □

In the case where there is only one group, i.e., when $a = 1$, there is no need to test for interaction, yet we can still test for main effects of Factor B. This situation arises, for example, when interest lies in testing for no growth over time for a sample of subjects, or $H_0 : \mu_1 = \cdots = \mu_b$, where μ_i is the mean growth at time (or level) i. This hypothesis can be tested by T_β^* using the results of Theorem 2.4.11 with $a = 1$. Letting $a = 1$ and supposing Assumptions 2.2.1, 2.2.6, 2.4.2, 2.4.7, and 2.4.10 hold, define

$$T_\beta^* = \sqrt{b}\left(AM\beta - AME\right). \tag{2.21}$$

Then, for $\sigma_\beta^2 = \mathrm{Var}\left(T_\beta\right)$ from Proposition 2.3.2, and under H_0^β,

$$\frac{T_\beta^*}{\sigma_\beta} \xrightarrow{D} \mathcal{Z} \quad \text{as} \quad b \to \infty,$$

where $\mathcal{Z} \sim N(0,1)$.

In practice, we need a consistent estimator $\hat{\sigma}_\phi$ of σ_ϕ. Then, by Slutsky's Theorem [20], $\frac{T_\phi^*}{\hat{\sigma}_\phi} \xrightarrow{D} \mathcal{Z}$ as $b \to \infty$. Of more specific interest in σ_ϕ^2 from Theorem 2.4.11 (or Corollary 2.4.3) is

$$\sigma_{11}^2 + 2\sum_{j=1}^{b-1}\left(1 - \frac{j}{b}\right)\sigma_{1,1+j}^2, \tag{2.22}$$

for which we need a consistent estimator. Recall that $\sigma_{1,1+j} = \gamma(j),\ j = 0, \ldots, b-1$, are the values of the autocovariance function at lag j. Given the form of the model, under the null hypotheses these quantities will be the same for all n subjects. Therefore, it is reasonable to estimate $\gamma(j),\ j = 0, \ldots, b-1$, for all n subjects and take the average as the overall estimate of $\gamma(j)$.

Let $\hat{\gamma}(h)$ be the estimate of $\gamma(h) = \sigma_{1,1+h},\ h = 0, \ldots, b-1$; and let $\hat{\gamma}^{(i,k)}(h)$ be the estimate of $\gamma(h)$ based on the data for the k^{th} subject in the i^{th} level of factor A. We will use the usual time series methods for finding the estimate of the autocovariance function of a stationary time series [41]. This brings us to the estimates

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{a} \sum_{k=1}^{n_i} \hat{\gamma}^{(i,k)}(h) \quad \text{for} \quad h = 0, \ldots, b-1; \tag{2.23}$$

where

$$\hat{\gamma}^{(i,k)}(h) = \frac{1}{b} \sum_{j=1}^{b-h} \left(X_{ijk} - \overline{X}_{i\cdot k}\right)\left(X_{i,j+h,k} - \overline{X}_{i\cdot k}\right). \tag{2.24}$$

If we make the further assumption that for each i and k, $\{X_{ijk}\}_{j=1}^{\infty}$ is a linear process, we can state that this is a consistent estimator. Shumway and Stoffer [41] show in Theorem A.6 that for a fixed h, $\hat{\gamma}(h) \xrightarrow{p} \tilde{\gamma}(h)$ as $b \to \infty$ for linear processes, where

$$\tilde{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{a} \sum_{k=1}^{n_i} \frac{1}{b} \sum_{j=1}^{b-h} (X_{ijk} - \mu_{i\cdot k})(X_{i,j+h,k} - \mu_{i\cdot k}).$$

Their assumption that the fourth moments of the white noise variates are finite is weaker than Assumption 2.4.10. The value of n is fixed; this is also required in Theorem A.6. For the sake of formality, let us state formally the assumption.

Assumption 2.4.12. *For all i and k, $\{X_{ijk}\}_{j=1}^{\infty}$ is a linear process. More specifically, $X_{ijk} = \sum_{t=-\infty}^{\infty} \psi_t \epsilon_{i,j-t,k}$, where the ϵ_{ijk} are independent and identically distributed with $\mathrm{E}(\epsilon_{ijk}) = 0$ and $\mathrm{Var}(\epsilon_{ijk}) = \sigma_\epsilon^2$.*

One goal of Chapters 3 and 4 will be to show that this assumption is not necessary. For instance, in Chapter 3, a new bootstrapping technique will be discussed as a method for estimating σ_ϕ from Proposition 2.3.2. As it stands, we have the following proposition.

Proposition 2.4.13. *Let Assumption 2.4.12 hold. For $\hat{\gamma}(k)$, $k = 0, \ldots, b-1$, from (2.23), the estimator, $\hat{\sigma}^2_\phi$, given by*

$$\hat{\sigma}^2_\phi = \frac{2}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + \alpha(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right)\left[\hat{\gamma}(0)^2 + 2\sum_{j=1}^{[b^*]}\left(1-\frac{j}{b}\right)\hat{\gamma}(j)^2\right], \tag{2.25}$$

where $\alpha(\phi) = \left\{\begin{array}{ll} 1 & \text{if } \phi = \beta \\ \frac{1}{(a-1)^2} & \text{if } \phi = \gamma \end{array}\right\}$, *is a good estimator for* $\sigma^2_\phi = \mathrm{Var}\,(T_\phi)$. *(Note that b^* is defined to be $O(b)$ such that $b^* < b$.)*

Proof: We know that for a fixed h, $\hat{\gamma}(h) \xrightarrow{p} \gamma(h)$ as $b \to \infty$. Then, by the Continuous Mapping Theorem [7], $\hat{\gamma}^2(h) \xrightarrow{p} \gamma^2(h)$ as $b \to \infty$.

Now, observe that

$$\begin{aligned}
&\frac{1}{\hat{\gamma}^2(0)}\frac{1}{b}\sum_{h=1}^{[b^*]}\left(1-\frac{h}{b}\right)\hat{\rho}^2(h) - \frac{1}{\hat{\gamma}^2(0)}\frac{1}{b}\sum_{h=1}^{b}\left(1-\frac{h}{b}\right)\rho^2(h) \\
&= \frac{1}{\hat{\gamma}^2(0)}\left[\frac{1}{b}\sum_{h=1}^{[b^*]}\left(1-\frac{h}{b}\right)\left(\hat{\rho}^2(h)-\rho^2(h)\right) - \frac{1}{b}\sum_{h=[b^*]+1}^{b}\left(1-\frac{h}{b}\right)\rho^2(h)\right].
\end{aligned}$$

Assuming that $\sum_{h=1}^{\infty}\rho^2(h) < \infty$,

$$0 \le \lim_{b\to\infty}\frac{1}{b}\sum_{h=[b^*]+1}^{b}\left(1-\frac{h}{b}\right)\rho^2(h) = \lim_{b\to\infty}\sum_{h=[b^*]+1}^{b}\frac{\rho^2(h)}{b} = 0.$$

In practice, we can take $b^* = b^m$ for $m \in (0,1)$. Since $\left(\hat{\rho}^2(h)-\rho^2(h)\right) \in [-1,1]$, we know that $\hat{\rho}^2(h)-\rho^2(h) = o_p(1)$ uniformly in h. Continuous functions on compact sets are uniformly continuous, and the convergence follows from the Continuous Mapping Theorem. Thus,

$$0 \le \frac{1}{b}\sum_{h=1}^{[b^*]}\left(1-\frac{h}{b}\right)\left(\hat{\rho}^2(h)-\rho^2(h)\right) = \frac{1}{b}\sum_{h=1}^{[b^*]}\left(1-\frac{h}{b}\right)o_p(1) \to 0 \quad \text{as} \quad b\to\infty.$$

Therefore, (2.23) is a consistent estimator of $\gamma(h)$, and (2.25) is a consistent estimator of $\sigma_\phi^2 = \text{Var}\,(T_\phi)$. □

2.4.2 Testing for the Main Effects of Factor A

The main goal of this section is to present a test statistic for the main effects of factor A (the non-high dimensional factor) and derive its asymptotic distribution. As b tends to infinity, we can use subject means over the length of b to compare group means and test whether there are significant differences. We use quadratic forms in means in order to apply the theory of quadratic forms.

First we must establish necessary notation and the SS as quadratic forms. Begin by defining

$$\widetilde{\boldsymbol{X}} := \begin{bmatrix} \overline{X}_{1\cdot 1} \\ \vdots \\ \overline{X}_{1\cdot n_1} \\ \hline \vdots \\ \hline \overline{X}_{a\cdot 1} \\ \vdots \\ \overline{X}_{a\cdot n_a} \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{X}}_1 \\ \vdots \\ \widetilde{\boldsymbol{X}}_a \end{bmatrix}, \tag{2.26}$$

where $\widetilde{\boldsymbol{X}}_i$ is defined to be the $n_i \times 1$ vector containing the means over b of each subject treated with the i^{th} level of factor A.

Recall, the hypothesis of interest is that there is no effect of factor A, or more precisely

(iii) $H_0^\alpha \;:\; \alpha_i = 0$ for $i = 1, \ldots, a$,

which, in practice, is often thought of as there being no group effects.

Recall that for fixed i and k, $\{X_{ijk}\}_{j=1}^{b}$ is assumed to be stationary and α-mixing. Also, $\{X_{ijk}\}_{j=1}^{b}$ and $\{X_{i'jk'}\}_{j=1}^{b}$ are independent for $i \neq i'$ or $k \neq k'$. Therefore, we know that $\overline{X}_{i\cdot k}$ and $\overline{X}_{i'\cdot k'}$ are iid for $i \neq i'$ or $k \neq k'$.

By Theorem 27.5 in Billingsley [8], we see that under Assumptions 2.2.1, 2.2.6, 2.4.2, 2.4.7, and 2.4.10, for each combination of i and k, and under H_0^{α},

$$\frac{\sqrt{b}\,\overline{X}_{i\cdot k}}{\tilde{\sigma}} \xrightarrow{D} Z \quad \text{as} \quad b \to \infty, \tag{2.27}$$

for $Z \sim N(0,1)$, where $\tilde{\sigma}^2 = \text{Var}\left(\frac{1}{\sqrt{b}}\sum_{j=1}^{b} X_{ijk}\right) = O(1)$. It will be seen later why we are not concerned with the specific details of the form of $\tilde{\sigma}^2$. Then, again under Assumptions 2.2.1, 2.2.6, 2.4.2, 2.4.7, and 2.4.10, we see that under H_0^{α}

$$\sqrt{b}\,\widetilde{\boldsymbol{X}} \xrightarrow{D} MVN\left(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}\right), \text{ as } b \to \infty, \tag{2.28}$$

where $\widetilde{\boldsymbol{\Sigma}} = \tilde{\sigma}^2 \boldsymbol{I}_n$ and MVN denotes the multivariate normal distribution.

If we think of the levels of factor A as being groups, then we can define the between groups SS, H, and the within groups SS, G, in the usual way. The former compares each group mean to the overall mean whereas the latter compares the group means and the subject means for all groups. These SS can be defined as

$$H := b\sum_{i=1}^{a} n_i \left(\overline{X}_{i\cdot\cdot} - \widetilde{X}_{\cdot\cdot\cdot}\right)^2, \tag{2.29}$$

and

$$G := b\sum_{i=1}^{a}\sum_{k=1}^{n_i} \left(\overline{X}_{i\cdot k} - \overline{X}_{i\cdot\cdot}\right)^2. \tag{2.30}$$

The following propositions help us utilize the theory of quadratic forms.

Proposition 2.4.14. *The sums of squares from 2.29 and 2.30 can be put in the following* *quadratic forms.*

$$H = \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)' \left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)\left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right);$$

$$G = \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)' \left(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}\right)\left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right). \tag{2.31}$$

Proof: First, observe

$$\begin{aligned}
H &= \sum_{i=1}^{a}\sum_{j=1}^{b} n_i\left(\overline{X}_{i\cdot\cdot} - \widetilde{X}_{\cdot\cdot\cdot}\right)^2 \\
&= b\sum_{i=1}^{a} n_i\left(\overline{X}_{i\cdot\cdot}^2 - 2\overline{X}_{i\cdot\cdot}\widetilde{X}_{\cdot\cdot\cdot} + \widetilde{X}_{\cdot\cdot\cdot}^2\right) \\
&= b\left[\left(\sum_{i=1}^{a} n_i\overline{X}_{i\cdot\cdot}^2\right) - 2\widetilde{X}_{\cdot\cdot\cdot}\left(\sum_{i=1}^{a} n_i\overline{X}_{i\cdot\cdot}\right) + \widetilde{X}_{\cdot\cdot\cdot}^2\left(\sum_{i=1}^{a} n_i\right)\right] \\
&= b\left[\left(\sum_{i=1}^{a} n_i\overline{X}_{i\cdot\cdot}^2\right) - 2\widetilde{X}_{\cdot\cdot\cdot}\left(n\widetilde{X}_{\cdot\cdot\cdot}\right) + n\widetilde{X}_{\cdot\cdot\cdot}^2\right] \\
&= b\left[\left(\sum_{i=1}^{a} n_i\overline{X}_{i\cdot\cdot}^2\right) - n\widetilde{X}_{\cdot\cdot\cdot}^2\right] \\
&= b\left(\left[\sum_{i=1}^{a} n_i\left(\widetilde{\boldsymbol{X}}_i'\frac{1}{n_i}\boldsymbol{1}_{n_i}\right)\left(\frac{1}{n_i}\boldsymbol{1}_{n_i}'\widetilde{\boldsymbol{X}}_i\right)\right] - n\left(\widetilde{\boldsymbol{X}}\frac{1}{n}\boldsymbol{1}_n\right)\left(\frac{1}{n}\boldsymbol{1}_n'\widetilde{\boldsymbol{X}}\right)\right) \\
&= b\left[\widetilde{\boldsymbol{X}}'\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{X}}'\left(\frac{1}{n}\boldsymbol{J}_n\right)\widetilde{\boldsymbol{X}}\right] \\
&= \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)'\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)\left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right).
\end{aligned}$$

Second, observe

$$\begin{aligned}
G &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\left(\overline{X}_{i\cdot k} - \overline{X}_{i\cdot\cdot}\right)^2 \\
&= b\sum_{i=1}^{a}\sum_{k=1}^{n_i}\left(\overline{X}_{i\cdot k}^2 - 2\overline{X}_{i\cdot k}\overline{X}_{i\cdot\cdot} + \overline{X}_{i\cdot\cdot}^2\right) \\
&= b\sum_{i=1}^{a}\left[\left(\sum_{k=1}^{n_i}\overline{X}_{i\cdot k}^2\right) - 2n_i\overline{X}_{i\cdot\cdot}^2 + n_i\overline{X}_{i\cdot\cdot}^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= b\sum_{i=1}^{a}\left(\sum_{k=1}^{n_i}\overline{X}_{i\cdot k}^2 - n_i\overline{X}_{i\cdot\cdot}^2\right) \\
&= b\sum_{i=1}^{a}\widetilde{\boldsymbol{X}}_i'\left(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\left(\widetilde{\boldsymbol{X}}_i\right) \\
&= b\,\widetilde{\boldsymbol{X}}'\begin{bmatrix}\left(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\widetilde{\boldsymbol{X}}_1 \\ \vdots \\ \left(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\widetilde{\boldsymbol{X}}_a\end{bmatrix} \\
&= \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)'\begin{bmatrix}\boldsymbol{I}_{n_1} - \frac{1}{n_1}\boldsymbol{J}_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a}\end{bmatrix}\left(\sqrt{b}\widetilde{\boldsymbol{X}}\right) \\
&= \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)'\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)\left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right). \quad \square
\end{aligned}
$$

In order to appeal to the theory of quadratic forms, we first need some results regarding the matrices of the quadratic forms in Proposition 2.4.14. The following properties addresses these necessities. Observe, for $n = \sum_{i=1}^{a} n_i$,

(1) $\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\frac{1}{n}\boldsymbol{J}_n = \frac{1}{n}\boldsymbol{J}_n = \frac{1}{n}\boldsymbol{J}_n\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)$,

(2) $\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n$ is idempotent,

(3) $\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}$ is idempotent, and

(4) $\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i} = \boldsymbol{0}$.

To see (1), rewrite the LHS breaking up $\boldsymbol{J}_n$ to get the following.

$$\left(\begin{bmatrix} \frac{1}{n_1}\boldsymbol{J}_{n_1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_a}\boldsymbol{J}_{n_a} \end{bmatrix} \frac{1}{n}\boldsymbol{1}_n\right)\boldsymbol{1}_n' = \left(\frac{1}{n}\boldsymbol{1}_n\right)\boldsymbol{1}_n' = \frac{1}{n}\boldsymbol{J}_n.$$

To see that the first equality holds, notice that the $1/n$ remains, and for each row, n_i 1s are being added but then multiplied by $1/n_i$, leaving 1. Since $\frac{1}{n}\boldsymbol{J}_n$ is symmetric, the second equality in (1) holds since

$$\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)' = \left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}'\right) = \left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right).$$

To see (2), first observe that since $\frac{1}{n_i}\boldsymbol{J}_{n_i}$ is idempotent, so is

$$\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} = \bigoplus_{i=1}^{a}\left(\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)^2 = \left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)^2.$$

Appealing to (1), we can observe that

$$\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)^2 = \left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)^2 - \frac{2}{n}\boldsymbol{J}_n + \left(\frac{1}{n}\boldsymbol{J}_n\right)^2 = \bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n.$$

To see (3), use the fact that $\boldsymbol{P}_{n_i}$ is idempotent to see that

$$\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)^2 = \bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}^2 = \bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}.$$

To see (4), appeal to (1) and observe that

$$\begin{aligned}\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}-\frac{1}{n}\boldsymbol{J}_n\right)\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i} &= \bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\boldsymbol{P}_{n_i}-\frac{1}{n}\boldsymbol{J}_n\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)\\ &= \bigoplus_{i=1}^{a}\boldsymbol{0}-\frac{1}{n}\boldsymbol{J}_n\left(\bigoplus_{i=1}^{a}\boldsymbol{I}_{n_i}\right)+\frac{1}{n}\boldsymbol{J}_n\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\\ &= -\frac{1}{n}\boldsymbol{J}_n+\frac{1}{n}\boldsymbol{J}_n\\ &= \boldsymbol{0}.\end{aligned}$$

Using H and G as defined in (2.31), we can define two independent asymptotic χ-square random variables and use them to develop an asymptotic F random variable. The usual F-test is implemented by taking the ratio of independent χ-square random variables divided by their respective degrees of freedom, which gives a random variable with an F-distribution [37]. Since we have shown that $\widetilde{\boldsymbol{X}}$ has an asymptotic multivariate normal distribution, we can appeal to the Continuous Mapping Theorem to derive an asymptotic F-test [7], [37]. This is the subject of the following theorem.

Theorem 2.4.15. *Suppose Assumptions 2.2.1, 2.2.6, 2.4.2, 2.4.7, and 2.4.10 hold. For H in (2.29) and G in (2.30), define*

$$F_\alpha^* = \frac{H/(a-1)}{G/(n-a)}. \tag{2.32}$$

Then

$$F_\alpha^* \xrightarrow{D} F_{a-1,n-a} \quad \text{as} \quad b\to\infty,$$

where $F_{a-1,n-a}$ represents an F-distribution with degrees of freedom $a-1$ and $n-a$.

Proof: First, we can use (2.28) to see that

$$\sqrt{b}\,\widetilde{\boldsymbol{X}} \xrightarrow{D} \boldsymbol{Z}, \quad \text{as} \quad b\to\infty, \tag{2.33}$$

where $\boldsymbol{Z} \sim MVN\left(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}\right)$. Define the matrices $\boldsymbol{A}_H$ and $\boldsymbol{A}_G$ such that

$$\frac{H}{\tilde{\sigma}^2} = \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)' \frac{\boldsymbol{A}_H}{\tilde{\sigma}^2} \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right), \tag{2.34}$$

$$\frac{G}{\tilde{\sigma}^2} = \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right)' \frac{\boldsymbol{A}_G}{\tilde{\sigma}^2} \left(\sqrt{b}\,\widetilde{\boldsymbol{X}}\right). \tag{2.35}$$

We will show that $\frac{H}{\tilde{\sigma}^2}$ and $\frac{G}{\tilde{\sigma}^2}$ have asymptotic χ^2-distributions with degrees of freedom $a-1$ and $n-a$, respectively. Then we will argue that H is asymptotically independent of G, so that (2.32) will have an asymptotic F-distribution with numerator degrees of freedom $(a-1)$ and denominator degrees of freedom $(n-a)$ [37].

From the theory of linear models, we must therefore show that $\frac{\boldsymbol{A}_H}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}}$ and $\frac{\boldsymbol{A}_G}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}}$ are idempotent, which will imply the χ^2-distributions [37]. We must also show that $\frac{\boldsymbol{A}_H}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}}\frac{\boldsymbol{A}_G}{\tilde{\sigma}^2} = \mathbf{0}$ to verify the asymptotic independence. These allow us to take the ratio of the χ^2-random variables divided by their respective degrees of freedom to get the appropriate F-distribution. It is clear that $\frac{1}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{I}$, and thus $\left(\frac{1}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}}\right)^2 = \frac{1}{\tilde{\sigma}^2}\widetilde{\boldsymbol{\Sigma}}$. Therefore, it suffices to show that $\boldsymbol{A}_H$ and $\boldsymbol{A}_G$ are idempotent and that $\boldsymbol{A}_H\boldsymbol{A}_G = \mathbf{0}$. Recognizing that $\boldsymbol{A}_H = \bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n$ and $\boldsymbol{A}_G = \bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i}$, we see that these three conditions are exactly (2), (3), and (4) from aforementioned properties. Furthermore, the rank of $\boldsymbol{A}_H$ is $a-1$ and the rank of $\boldsymbol{A}_G$ is $\sum_{i=1} a(n_i - 1) = n - a$, so the degrees of freedom of $\boldsymbol{H}$ and $\boldsymbol{G}$ are $a-1$ and $n-a$, respectively. Therefore, we have

$$\frac{H}{\tilde{\sigma}^2} \xrightarrow{D} \chi^2_{a-1} \quad \text{and} \quad \frac{G}{\tilde{\sigma}^2} \xrightarrow{D} \chi^2_{n-a} \quad \text{as} \quad b \to \infty.$$

Concordantly, by the Continuous Mapping Theorem [7],

$$F^*_\alpha = \frac{H/(a-1)}{G/(n-a)} = \frac{\frac{H/\tilde{\sigma}^2}{a-1}}{\frac{G/\tilde{\sigma}^2}{n-a}} \xrightarrow{D} F_{a-1,n-a} \quad \text{as} \quad b \to \infty. \quad \Box$$

Under the usual assumptions of normality and independent measurements, the existing F-test would be similar to that given above. (For a complete description of the unbalanced F-test, see Davis [16].) However, relaxing the assumptions that the data follow a normal distribution and that the measurements on each subject are independent forces another route to be taken. In this scenario, we must first appeal to a CLT for dependent data [8]. After the means for each subjects are taken, they are shown to be asymptotically normal. Since the responses among subjects are assumed to be independent, this now reduces to the usual F-test procedure given the necessary conditions for quadratic forms hold.

Chapter 3

Simulation Study for the Univariate Case

3.1 Introduction

This chapter serves to present a simulation study for the test statistics in the univariate case. Section 3.2 gives the design framework of the simulation study, including the various cases considered. Section 3.3 discusses the main results of the achieved sizes of the various test statistics, including a comparison to traditional methods and a further study using bootstrapping techniques to estimate the variances of the test statistics. Section 3.3.3 gives a comparison of the achieved power of the new and traditional tests; these results are for one group size structure and one value of b.

3.2 Simulation Design

For an assessment of the quality of the asymptotic distributions of the test statistics in Theorems 2.4.11 and 2.4.15, we conducted a simulation study by generating data from multiple distributions with various covariance structures and sample sizes/size structures.

The three distributions used to generate data for the simulation were the standard normal distribution (labeled P_1 in the simulation); the skew-normal distribution with location parameter 0, scale parameter 1, and skewing parameter 1 (P_2); and the log-normal distribution with log-scale parameter 0 and shape parameter 1 (P_3). These three distributions increase in the level of their skewness and kurtosis. Table 3.1 gives the mean, variance, skewness and kurtosis of these three distributions. The three covariance matrices ($b \times b$ in dimension) used will be labeled Σ_1, Σ_2, and Σ_3 in the simulation. The first follows that of an ARMA(2,2) process, which decays exponentially; the second follows polynomial decay such that the $(i, j)^{th}$ element was given by $\rho|i - j|^{-5/2}$, where $\rho = 0.55$; the third was given by $1.5\boldsymbol{I}$. For graphical representation of the values in Σ_1 and Σ_2, see Figure 3.2 below. Two group size structures were used, labeled N_1 and N_2. The first structure included four groups, or $a = 4$, where the number of subjects in each group was $n_1 = 4$, $n_2 = 5$, $n_3 = 6$, and $n_4 = 7$. The second structure included three groups, or $a = 3$, where the number of subjects in each group was $n_1 = 10$, $n_2 = 12$, and $n_3 = 14$.

For each combination of distribution, covariance structure, and group structure, 10,000 simulated data sets were generated for various values of b. The values of b were 5, 10, 20, 50, 100, 200, and 400. In each simulation the test statistics T^*_β, T^*_γ, and F^*_α were computed. Each time, the statistic was compared to the 0.05 critical value for its corresponding standard normal or F-distribution, and a decision to reject or fail to reject was made according as the statistic was beyond or within the corresponding critical value. The average number of rejections, which is the estimated size of the test, is reported for each case in Table 3.2. The estimated sizes for the test based on T^*_β, which is the main effect of B (time effect), is given in a column labeled β. The results of the test for an interaction effect between factors A and B and the test for the main effect of A (group effect) are similarly labeled γ and α.

Function of Moment	Distribution		
	P_1	P_2	P_3
Mean	0	0.5642	1.6487
Variance	1	1.6817	4.6708
Skewness	0	0.1369	6.1849
Kurtosis	0	0.0617	110.9364

Table 3.1: Functions of the first four moments of the distributions for P_1, P_2, and P_3.

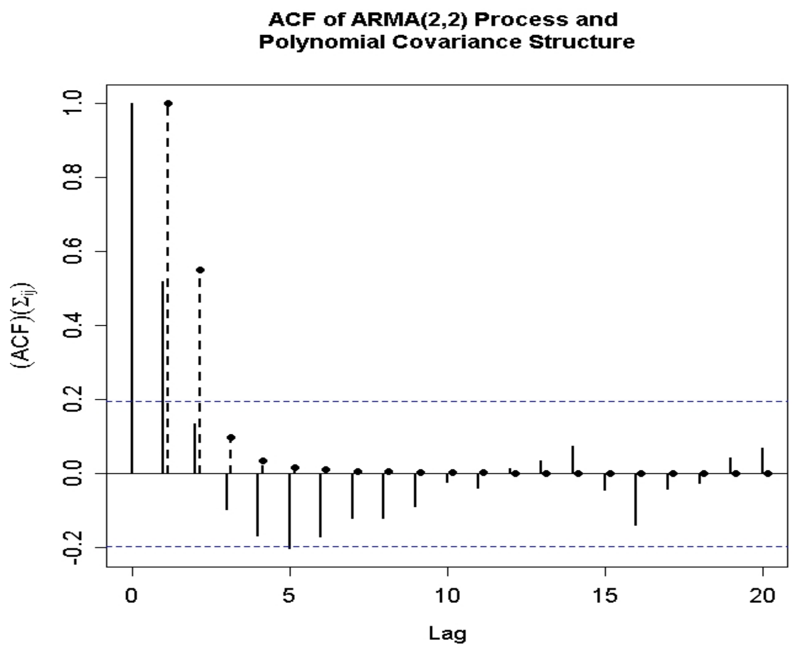

Figure 3.1: Autocorrelation function (ACF) for an ARMA(2,2) process and the $(i, j)^{th}$ element of the polynomial covariance matrix Σ_2. The ARMA(2,2) ACF values are plotted as solid lines and the polynomial ACF values are plotted as slightly offset dashed lines with points for emphasis.

3.3 Achieved Sizes of the Tests

3.3.1 The New Method

Looking at Table 3.2, the test of the main effect of factor A seems to be performing well regardless of the covariance structure, group size structure, choice of distribution, or value of b; however, it is slightly more variable and biased low for small b values such as $b = 5$ and $b = 10$. These sizes seems to be very close to 0.05 throughout, which is the desired size of the test.

The test for the main effect of factor B, as well as the test for the interaction effect, have somewhat erratic results for $b \leq 20$. The estimated sizes range from around 0.01 to 0.3 when ideally they would be around 0.05. This is not all that surprising or concerning as $b = 20$ is not sufficiently large for the asymptotic property of the test to take effect. These estimated sizes seem to stabilize some for $b = 50$; however, for P_3 and Σ_3 the results are still poor. For P_3, the estimated sizes range from 0.02870 to 0.04110 when $b = 400$; this could be a possible indication that for such skewness in a distribution (P_3 is the log-normal), $b = 400$ is not yet sufficiently large for the asymptotics to take effect.

In the cases of P_1 and P_2 and for $b = 100$, the results seem to begin to converge to the desired size of 0.05, taking on values from approximately 0.04 to slightly greater than 0.05. When b increases to 200, the estimated sizes are becoming more ideal, and even more still when $b = 400$. The best results for large values of b seem to be for P_1, which is the normal distribution. It seems that the the effect of the dependence seems to go away quickly in the absence of skewness.

Overall, the results are promising, though they do not, as a rule, seem to converge to 0.05 as fast as we would like. This may be due to the estimation technique used for $\text{Var}(T_\phi)$. This problem warrants further research and is also addressed numerically in Section 3.3.3.

3.3.2 Comparison with Traditional RM-ANOVA

To gauge the effectiveness of the tests from Theorems 2.4.11 and 2.4.15 we ran the same simulation for the usual test statistics used in traditional repeated measures ANOVA (RM-ANOVA). In this setting the SS are different from those mentioned earlier. The underlying assumptions

Pop	Cov	Samp	$b=5$			$b=10$			$b=20$			$b=50$		
			β	γ	α	β	γ	α	β	γ	α	β	γ	α
P_1	Σ_1	N_1	0.07290	0.08835	0.04920	0.06150	0.06570	0.04670	0.05255	0.05635	0.05045	0.04745	0.05110	0.05215
		N_2	0.07170	0.07270	0.05245	0.05750	0.05730	0.04865	0.05305	0.05470	0.04950	0.04505	0.04940	0.04875
	Σ_2	N_1	0.06715	0.07750	0.05340	0.05395	0.05575	0.04925	0.04585	0.04860	0.04965	0.04230	0.04830	0.05175
		N_2	0.06610	0.06920	0.04820	0.05350	0.05200	0.04765	0.04640	0.05010	0.05065	0.04300	0.04665	0.04690
	Σ_3	N_1	0.04955	0.05205	0.04805	0.03695	0.04140	0.04750	0.03535	0.03995	0.04855	0.03680	0.04055	0.04920
		N_2	0.05320	0.04745	0.05055	0.04105	0.03740	0.05020	0.03740	0.03570	0.04985	0.03925	0.03880	0.05080
P_2	Σ_1	N_1	0.19210	0.08195	0.04780	0.10805	0.06350	0.05175	0.07490	0.05400	0.04915	0.05590	0.05105	0.04850
		N_2	0.29425	0.07200	0.05190	0.15480	0.05690	0.05070	0.09995	0.05155	0.05080	0.06680	0.04705	0.05245
	Σ_2	N_1	0.21600	0.06805	0.04665	0.14125	0.05355	0.05030	0.10140	0.05045	0.04845	0.06795	0.04385	0.05420
		N_2	0.33225	0.06155	0.04965	0.22650	0.05035	0.05020	0.15445	0.04455	0.05000	0.09400	0.04255	0.05255
	Σ_3	N_1	0.05310	0.05120	0.05050	0.04000	0.04000	0.05105	0.03635	0.03970	0.05085	0.03830	0.03690	0.04875
		N_2	0.04745	0.04895	0.05025	0.03785	0.03760	0.04620	0.03850	0.03735	0.05175	0.03845	0.04070	0.04790
P_3	Σ_1	N_1	0.19360	0.01840	0.04130	0.06960	0.01750	0.04570	0.03770	0.01660	0.04635	0.02800	0.01770	0.04695
		N_2	0.28250	0.01655	0.04370	0.06960	0.01750	0.04570	0.05215	0.01420	0.04610	0.03390	0.01525	0.04635
	Σ_2	N_1	0.20560	0.01690	0.04180	0.10610	0.01440	0.04390	0.06500	0.01350	0.04670	0.04135	0.01775	0.04890
		N_2	0.29895	0.01590	0.04185	0.17805	0.01160	0.04475	0.10420	0.01115	0.04590	0.05415	0.01465	0.04970
	Σ_3	N_1	0.01360	0.01150	0.04235	0.01030	0.00840	0.04345	0.01080	0.01050	0.04815	0.01335	0.01440	0.04735
		N_2	0.01205	0.00885	0.04240	0.00785	0.00650	0.04550	0.00915	0.00710	0.04695	0.01155	0.01065	0.04940

Pop	Cov	Samp	$b=100$			$b=200$			$b=400$		
			β	γ	α	β	γ	α	β	γ	α
P_1	Σ_1	N_1	0.05045	0.05210	0.04830	0.04725	0.04850	0.05190	0.04750	0.05005	0.05075
		N_2	0.04815	0.04745	0.04815	0.04990	0.05065	0.05045	0.04700	0.04640	0.04870
	Σ_2	N_1	0.04650	0.04775	0.05085	0.04775	0.04665	0.04995	0.04875	0.04725	0.05095
		N_2	0.04590	0.04615	0.05135	0.04440	0.04560	0.05095	0.04840	0.04610	0.05060
	Σ_3	N_1	0.03955	0.04375	0.05390	0.04210	0.04225	0.05335	0.04215	0.04675	0.05030
		N_2	0.03965	0.04195	0.05100	0.04390	0.04130	0.05090	0.04420	0.04555	0.04865
P_2	Σ_1	N_1	0.05425	0.04925	0.05065	0.05000	0.04790	0.04655	0.05380	0.04890	0.04940
		N_2	0.05680	0.04965	0.04925	0.05375	0.05095	0.05155	0.05160	0.05075	0.05065
	Σ_2	N_1	0.05680	0.04755	0.05105	0.05565	0.04670	0.04975	0.05205	0.05055	0.04875
		N_2	0.07060	0.04540	0.05015	0.06200	0.04835	0.05205	0.05440	0.04915	0.05200
	Σ_3	N_1	0.04000	0.04210	0.04915	0.04185	0.04360	0.05090	0.04200	0.04535	0.04855
		N_2	0.04005	0.04245	0.05050	0.04295	0.04690	0.04960	0.04515	0.04305	0.04825
P_3	Σ_1	N_1	0.02720	0.02350	0.04770	0.03185	0.02855	0.04875	0.03515	0.03675	0.04870
		N_2	0.03210	0.02120	0.04795	0.03340	0.02715	0.04915	0.03615	0.03375	0.05205
	Σ_2	N_1	0.03615	0.02170	0.04770	0.03255	0.02760	0.04810	0.03925	0.03375	0.04780
		N_2	0.04335	0.02080	0.04870	0.04215	0.02690	0.0515	0.04110	0.03280	0.05240
	Σ_3	N_1	0.01940	0.01920	0.04560	0.02545	0.02555	0.05120	0.03155	0.03415	0.05005
		N_2	0.01785	0.01720	0.05100	0.02270	0.02405	0.04890	0.02870	0.03155	0.05110

Table 3.2: Simulated sizes for the tests with statistics T^*_β (β), T^*_γ (γ), and F^*_α (α) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, and under the two group structures N_1 and N_2. Results are for $b = 5$, $b = 10$, $b = 20$, $b = 50$, $b = 100$, $b = 200$, and $b = 400$.

also differ, mainly in that the data are assumed to arise from a normal distribution. While the traditional methods also do assume that the compound symmetric (exchangeable) covariance structure, they assume a restrictive covariance structure, namely that the covariance between any two measurements from the same subject is the same. Recall that under the assumptions of Theorem 2.4.11 the covariances need only to decay sufficiently fast. For a more complete description of the traditional RM-ANOVA design, see Davis [16].

Based on the models, it is not surprising to see the results in Table 3.3. We first consider the test for the main effect of factor B (time effect). Under the Σ_3 covariance structure, the sizes seem to be okay, i.e., near 0.05, for both sample size structures, for all three populations, and for any b. This comes as no surprise since the structure of Σ_3 fits the assumptions given in the traditional RM-ANOVA framework. However, under the other two covariance structures the results differ. In these cases, the simulated sizes never get sufficiently close to the desired level of 0.05. For P_3, the log-normal distribution, the simulated sizes range from just under 0.5 to just over 0.11 as b increases; for P_2, the skew-normal distribution, the simulated sizes range from just under 0.35 to just over 0.11 as b increases; for P_1, the normal distribution, the simulated sizes hover around 0.1 for all levels of b. The tests for the time effect do not have adequate size, and the results are worse as the distribution from which the data are generated becomes more skewed.

We next consider the tests for the interaction effect between factors A and B. Similar to the analysis above, under Σ_3 the simulated sizes are okay (under 0.08) for N_1 and good for N_2 (around 0.05) as b increases. The discrepancy may be due to the fact that the overall number of subjects is greater under N_2 than under N_1. However, under the other two covariance structures, the simulated sizes range from around 0.08 to 0.12 as b increases, getting worse as b increases. This is similar for any of the population distributions. Yet the sizes are not sufficiently low, and it is disconcerting that the simulated sizes increase as b increases, whereas under the tests from Theorem 2.4.11, the sizes are converging to 0.05 as b increases.

Overall, for both the tests of the main effect of factor B and the interaction effect, the test from the traditional RM-ANOVA framework do not seem to be equipped to handle the given conditions of the problem of interest, and the results from Theorem 2.4.11 are far superior.

However, the results are much different for the test of the main effect of factor A (group effect). In the case of the log-normal distribution, the achieved sizes are a little lower than the desired value of 0.05 when b is small, but this is not a problem as b increases. The results are very similar to the simulation results in Table 3.2. This is not surprising since the dependencies present in the time component, or longitudinal component, of the data become essentially washed out by taking the average response from each subject over time. After that, appealing to the Continuous Mapping Theorem [7], the basic CLT, and the usual F-test argument lead to the obvious, similar results [37].

Table 3.4 displays the simulated sizes from the tests from Theorems 2.4.11 and 2.4.15 where the data were generated according the to traditional RM-ANOVA framework; i.e., the compound symmetry (exchangeable) covariance structure was used. The purpose of this simulation was to asses whether the new method performs well under the traditional setup. The sizes are all appropriately close to 0.05, with the best results coming from the test of the main effect of factor A. This table indicates that there is no concern regarding the performance of the new method when the data are assumed to arise under the criteria of the traditional RM-ANOVA setting.

In conclusion, the simulation study clearly indicates that the traditional RM-ANOVA methods are not as sufficient as those in Chapter 2 except for testing of a group effect. If a time or interaction effect is to be tested, other methods, such as those in Chapter 2, should be implemented, if the conditions for RM-AVONA are not met.

3.3.3 Estimating the Variance via Bootstrapping

The theoretical form of the values $\dfrac{T_\phi^*}{\sigma_\phi}$, $\phi \in \{\beta, \gamma\}$, is a great starting point; however, in practice an estimate for σ_ϕ is needed. The estimation of the variance of T_ϕ^* can be troublesome, and it requires some stringent assumptions. Recall that to achieve a consistent estimator for σ_ϕ we need assumption 2.4.12, that the series from the responses of each subject is a linear process. This is a somewhat rigid assumption which is difficult to assess, and it should therefore be avoided if at all possible.

One possible solution to avoid such assumptions would be to use a bootstrapping technique. One major problem with bootstrapping time series data is capturing, or rather preserving, the dependence structure found amid the data [25]. The block bootstrap is one of the most prevalent methods for using bootstrapping techniques with time series data [19]. As with the standard bootstrap, bootstrap samples are generated from the data and statistics are

Pop	Cov	Samp	$b=5$			$b=10$			$b=20$			$b=50$		
			β	γ	α	β	γ	α	β	γ	α	β	γ	α
P_1	Σ_1	N_1	0.0747	0.0882	0.05	0.0938	0.0972	0.05	0.103	0.1091	0.0499	0.1014	0.1034	0.0513
		N_2	0.0776	0.0806	0.0526	0.0967	0.1007	0.0535	0.0982	0.1051	0.0504	0.1044	0.1138	0.0539
	Σ_2	N_1	0.0732	0.0852	0.0445	0.0799	0.0925	0.0499	0.0873	0.099	0.0511	0.0923	0.0969	0.0504
		N_2	0.0743	0.0772	0.0486	0.0867	0.0859	0.0528	0.0894	0.0976	0.0527	0.0923	0.0939	0.0491
	Σ_3	N_1	0.0509	0.0519	0.0516	0.0496	0.049	0.0515	0.0498	0.0537	0.0487	0.0474	0.0513	0.0492
		N_2	0.0508	0.054	0.0523	0.0489	0.0488	0.0488	0.0501	0.05	0.0482	0.048	0.0487	0.0483
P_2	Σ_1	N_1	0.2141	0.0883	0.048	0.1704	0.1017	0.0487	0.1491	0.1088	0.0497	0.133	0.1104	0.0547
		N_2	0.311	0.0863	0.0492	0.2241	0.0971	0.0473	0.1823	0.1022	0.0476	0.1585	0.109	0.0512
	Σ_2	N_1	0.2353	0.0802	0.052	0.2038	0.0903	0.0501	0.1821	0.095	0.0481	0.156	0.0922	0.0464
		N_2	0.3444	0.0741	0.0505	0.3056	0.0825	0.049	0.2555	0.0962	0.0487	0.1987	0.0906	0.054
	Σ_3	N_1	0.047	0.0496	0.0457	0.0484	0.0476	0.0507	0.0524	0.0475	0.0498	0.0507	0.0472	0.0483
		N_2	0.0504	0.0502	0.0497	0.0527	0.0514	0.0501	0.0521	0.0508	0.0489	0.0499	0.0518	0.0516
P_3	Σ_1	N_1	0.3256	0.0769	0.0404	0.2074	0.1012	0.0429	0.1637	0.1076	0.0425	0.1397	0.122	0.0466
		N_2	0.4536	0.0756	0.0416	0.2845	0.0949	0.0464	0.2156	0.0956	0.048	0.1658	0.1086	0.0444
	Σ_2	N_1	0.3481	0.0709	0.0442	0.2784	0.0911	0.0493	0.2132	0.0956	0.0448	0.1655	0.1114	0.049
		N_2	0.4877	0.0694	0.0432	0.4237	0.0833	0.0435	0.3295	0.0889	0.0476	0.2321	0.0948	0.0451
	Σ_3	N_1	0.032	0.0488	0.0427	0.0425	0.0552	0.0449	0.0489	0.0589	0.0468	0.0507	0.0752	0.0478
		N_2	0.0411	0.0397	0.0431	0.0436	0.0521	0.0459	0.0457	0.0474	0.0426	0.0514	0.0591	0.0423

Pop	Cov	Samp	$b=100$			$b=200$			$b=400$		
			β	γ	α	β	γ	α	β	γ	α
P_1	Σ_1	N_1	0.1044	0.1166	0.0488	0.1143	0.1136	0.0556	0.1124	0.1125	0.0515
		N_2	0.1062	0.1132	0.0498	0.1162	0.1172	0.0472	0.1126	0.1129	0.049
	Σ_2	N_1	0.0975	0.1029	0.0481	0.0973	0.0957	0.0475	0.0975	0.0985	0.0509
		N_2	0.0973	0.0962	0.0498	0.0933	0.0908	0.0472	0.0983	0.0926	0.053
	Σ_3	N_1	0.0528	0.0514	0.0494	0.0486	0.0484	0.0497	0.0492	0.0543	0.053
		N_2	0.0457	0.0471	0.0485	0.0528	0.0521	0.0517	0.0469	0.0467	0.0506
P_2	Σ_1	N_1	0.1278	0.1155	0.0495	0.1263	0.1154	0.0491	0.1168	0.1109	0.0458
		N_2	0.1443	0.1102	0.0496	0.1349	0.1144	0.0481	0.1327	0.1115	0.0491
	Σ_2	N_1	0.1455	0.0925	0.0515	0.126	0.0974	0.0513	0.1205	0.1002	0.0476
		N_2	0.163	0.0928	0.0462	0.1431	0.1029	0.049	0.1286	0.0998	0.0511
	Σ_3	N_1	0.0442	0.0516	0.0504	0.0546	0.0502	0.0511	0.0505	0.0511	0.0525
		N_2	0.0492	0.0492	0.0503	0.0549	0.0486	0.0501	0.0501	0.0537	0.0511
P_3	Σ_1	N_1	0.1331	0.1239	0.0483	0.1245	0.1257	0.0463	0.1184	0.1354	0.0478
		N_2	0.1424	0.1127	0.0464	0.1405	0.1151	0.0508	0.1343	0.1232	0.0518
	Σ_2	N_1	0.1481	0.1136	0.0482	0.1281	0.1212	0.0493	0.1174	0.1254	0.0493
		N_2	0.1876	0.0953	0.046	0.1606	0.0989	0.0471	0.1357	0.1043	0.0479
	Σ_3	N_1	0.0508	0.0808	0.0485	0.0534	0.0789	0.0511	0.0496	0.0833	0.0473
		N_2	0.0475	0.0561	0.0504	0.0501	0.0552	0.0444	0.0504	0.0602	0.0489

Table 3.3: Simulated sizes for the tests with F-statistics from traditional MANOVA (see Davis, [16]) for the effect of factor B (β), interaction effect between factors A and B (γ), and the effect of factor A (α) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, and under the two group structures N_1 and N_2. Results are for $b = 5$, $b = 10$, $b = 20$, $b = 50$, $b = 100$, $b = 200$, and $b = 400$.

calculated from each bootstrap sample. The standard deviation of these bootstrap sample statistics is used as an estimate of the standard error of whatever statistic is of interest. The standard bootstrap takes a sample with replacement from the original data to form each bootstrap sample, yet it is not so simple when working with dependent data.

To illustrate the block bootstrap, consider for the moment that there is only one subject whose b repeated measurements are a time series of unknown origin. The block bootstrap

Sample Size	Test	b: 5	10	20	50	100	200	400
N_1	β	0.0526	0.0410	0.0372	0.0392	0.0437	0.0397	0.0416
	γ	0.0531	0.0377	0.0390	0.0369	0.0425	0.0402	0.0414
	α	0.0461	0.0542	0.0486	0.0538	0.0482	0.0491	0.0489
N_2	β	0.0528	0.0380	0.0402	0.0404	0.0419	0.0415	0.0427
	γ	0.0472	0.0397	0.0366	0.0402	0.0434	0.0437	0.0463
	α	0.0516	0.0494	0.0496	0.0503	0.0549	0.0500	0.0517

Table 3.4: Simulated sizes for the tests with statistics T^*_β (β), T^*_γ (γ), and F^*_α (α) when sampling via the usual RM-ANOVA framework (from Davis [16]) and under the two group structures N_1 and N_2. Results are for $b = 5, 10, 20, 50, 100, 200, 400$.

breaks the data into q blocks of responses. From the q blocks, the bootstrap sample is created by sampling the blocks with replacement and aligning them end to end, preserving the order of each block. For blocks of length l, each bootstrap sample will be ql in length, where, ideally, $ql = b$. In the simulation discussion below, distinct (non-overlapping) blocks were chosen, although the blocks could be overlapping and even of randomly varying lengths [25]. It has been found that the difference in using non-overlapping or overlapping blocks for numerical results is often very small [3].

The justification for extending the bootstrap to multiple dimensions (i.e., multiple subjects) is the assumption that each subject is independent. It consequently stands to reason that individual block bootstraps taken on the time series of each subject individually could be combined to make one full bootstrap sample. The algorithm is as follows.

1. Begin with one subject.
2. Randomly sample q blocks, where $ql = b$.
3. Let the q blocks laid end to end be a block bootstrap of that subject.
4. Repeat steps 2 and 3, using the same q block locations, for each subject. This is now the first full bootstrap sample $\boldsymbol{X}^{*^1}$.
5. Repeat steps 1 though 4 B times to obtain the bootstrap samples $\boldsymbol{X}^{*^1}, \ldots, \boldsymbol{X}^{*^B}$.

6. Calculate the test statistic $(T_\phi^*)^{*^b}$, $b = 1, \ldots, B$, for each corresponding bootstrap sample.

7. Use the variance of the B bootstrap statistics as an estimate for $\sigma_\phi^2 = \text{Var}\left(T_\phi^*\right)$.

The original simulation was repeated using this method within each simulation to calculate an estimate of the standard deviation σ_ϕ instead of using the estimate from Proposition 2.4.13. Block lengths of 1, 2, 2, 2, 4, 4, 4 were chosen corresponding to b values of 5, 10, 20, 50, 100, 200, 400 in keeping with the suggestion from Härdle et al. [25] that the block length should be on the order of $b^{1/4}$. For each simulation, 200 bootstraps were run to estimate σ_ϕ. The estimated sizes of the tests are given in Tables 3.5 and 3.6.

For $b \leq 20$, the estimated sizes for all tests were not acceptably small, ranging from about 0.1 to 0.4, where the critial value was 0.05. The estimated sizes do decrease toward 0.05 as b increases, though the convergence is very slow. The size estimates seem to be smaller for Σ_3 across the board, which provides some promise, but only for that specific case. In fact, this makes sense since for Σ_3 the bootstrapping reduces to the usual case. The sizes are not overly large, but they are not an improvement over the original method for estimating the standard deviation. This could be due to the role block length has to play in the estimate, but as the simulations are expensive, only one setup was used. Further research in this area is needed.

In summary, the extended block bootstrap method did not seem to offer any improvement as far as the sizes of the tests are concerned. However, the results do not seem to depend on the population, and the sizes are relatively small and close to 0.05 for large values of b. Thus, further research to find a theory that explains this phenomenon may prove fruitful.

Pop	Cov	Samp	$b=5$		$b=10$		$b=20$	
			β	γ	β	γ	β	γ
P_1	Σ_1	N_1	0.3782	0.3278	0.3246	0.2765	0.2346	0.1882
		N_2	0.3923	0.3548	0.3312	0.3011	0.2361	0.2141
	Σ_2	N_1	0.3667	0.3112	0.2856	0.2419	0.1999	0.1637
		N_2	0.3839	0.3305	0.2978	0.2595	0.2150	0.1840
	Σ_3	N_1	0.2884	0.2149	0.2119	0.1552	0.1455	0.1118
		N_2	0.2999	0.2635	0.2147	0.1794	0.1526	0.1154
P_2	Σ_1	N_1	0.2468	0.3253	0.2431	0.2861	0.1912	0.1957
		N_2	0.1908	0.3514	0.2189	0.3045	0.1733	0.2134
	Σ_2	N_1	0.2345	0.3092	0.1978	0.2483	0.1415	0.1682
		N_2	0.2007	0.3401	0.1505	0.2622	0.1153	0.1813
	Σ_3	N_1	0.2823	0.2143	0.2071	0.1611	0.1349	0.1035
		N_2	0.2989	0.2417	0.2158	0.1832	0.1458	0.1213
P_3	Σ_1	N_1	0.2215	0.3189	0.2102	0.2731	0.1532	0.1829
		N_2	0.1780	0.3369	0.1983	0.2942	0.1456	0.1996
	Σ_2	N_1	0.2295	0.3071	0.1706	0.2254	0.1137	0.1550
		N_2	0.2055	0.3274	0.1367	0.2407	0.0934	0.1620
	Σ_3	N_1	0.2228	0.1723	0.1661	0.1335	0.1091	0.0965
		N_2	0.2546	0.2028	0.1854	0.1496	0.1169	0.1030

Table 3.5: Using bootstrapping to estimate the variance of the test statistics: Simulated sizes for the tests with statistics T^*_β (β), T^*_γ (γ), and F^*_α (α) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, and under the two group structures N_1 and N_2. Results are for $b = 5$, $b = 10$, and $b = 20$.

3.4 Power Simulation

3.4.1 Main Results and Comparison with Traditional RM-ANOVA

In order to further assess and compare the tests from Theorems 2.4.11 and 2.4.15, we ran simulations under N_2 and for $b = 100$ with representing perturbations signifying various alternative hypotheses and calculated the power of the test. The choices of N_2 and $b = 100$ are due to the fact that the simulated sizes of the tests from Theorem 2.4.11 were among the more desirable for all covariance structures and population distributions. Note that under P_3 the simulated sizes were not as close to 0.05 as desired for any sample size structure or value of b. Each alternative hypothesis was used for all three covariance structures and all three population distributions. For the test of the main effect of factor B (time effect), the alternatives $A_1, \ldots, A_9$ correspond to $\beta_1 = 0.01, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4$. The same values

Pop	Cov	Samp	$b = 50$		$b = 100$		$b = 200$		$b = 400$	
			β	γ	β	γ	β	γ	β	γ
P_1	Σ_1	N_1	0.1657	0.144	0.1194	0.1036	0.0985	0.0911	0.0876	0.085
		N_2	0.1734	0.1505	0.1198	0.1159	0.1045	0.0920	0.0918	0.0920
	Σ_2	N_1	0.1414	0.1165	0.1046	0.0922	0.0833	0.0822	0.0800	0.0679
		N_2	0.1416	0.1245	0.1099	0.0950	0.0916	0.0792	0.0775	0.0735
	Σ_3	N_1	0.0902	0.0765	0.0769	0.0729	0.0661	0.0617	0.0598	0.0572
		N_2	0.0924	0.0879	0.0823	0.0760	0.0645	0.0623	0.0591	0.0594
P_2	Σ_1	N_1	0.1407	0.1358	0.112	0.1051	0.0902	0.0939	0.0929	0.0835
		N_2	0.1386	0.1450	0.1026	0.1115	0.0915	0.0953	0.0872	0.0864
	Σ_2	N_1	0.1079	0.1212	0.0885	0.0902	0.0736	0.0839	0.0699	0.0672
		N_2	0.0934	0.1264	0.0775	0.0965	0.0721	0.0853	0.0702	0.0716
	Σ_3	N_1	0.0942	0.0788	0.0793	0.0644	0.0629	0.0606	0.0605	0.0552
		N_2	0.0935	0.0811	0.0786	0.0704	0.0616	0.0637	0.0589	0.0558
P_3	Σ_1	N_1	0.1221	0.1377	0.1052	0.1057	0.0930	0.0862	0.0858	0.0853
		N_2	0.1158	0.144	0.0937	0.1111	0.0862	0.0963	0.0798	0.0884
	Σ_2	N_1	0.0865	0.1192	0.0749	0.0881	0.0715	0.0719	0.0637	0.0733
		N_2	0.0774	0.1231	0.0666	0.0928	0.0630	0.0795	0.0630	0.0693
	Σ_3	N_1	0.0803	0.0709	0.0701	0.0696	0.0633	0.0636	0.0582	0.0572
		N_2	0.0861	0.0760	0.0764	0.0760	0.0637	0.0646	0.0583	0.0608

Table 3.6: Using bootstrapping to estimate the variance of the test statistics: Simulated sizes for the tests with statistics T^*_β (β), T^*_γ (γ), and F^*_α (α) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, and under the two group structures N_1 and N_2. Results are for $b = 50$, $b = 100$, $b = 200$, and $b = 400$.

for γ_{21} and α_2 correspond to the alternatives $A_1, \ldots, A_9$ when testing the interaction effect and the main effect of factor A (group effect), respectively. The simulated powers were calculated for the tests from Theorems 2.4.11 and 2.4.15, labeled *ARMU* (asymptotic repeated measures, univariate case), and the traditional RM-ANOVA design, labeled *RMA*. The results are given in Tables 3.7, 3.8, and 3.9.

We begin with the comparison of power for the tests of the main effect of factor B. For simplicity, we refer to these as the ARMU and RMA methods (or tests, or statistics) as mentioned above. Table 3.7 appears to indicate that the simulated power is as good or better in the RMA method for all populations and covariance structures. Under Σ_1 and Σ_2 the RMA method simulated powers are greater for all alternative hypotheses. However, this is not wholly indicative of a better test since the sizes of these tests are greater than the desired level of 0.05; in fact, they were simulated to be anywhere from 0.0975 to 0.1481. Tests with differing sizes are generally not comparable with respect to power. Under Σ_3 the RMA method has greater power in addition to having appropriate size. This is natural since the structure of Σ_3 aligns with the traditional RM-ANOVA assumptions found in Davis [16].

For the ARMU method, the power of the tests makes a significant jump under alternative hypotheses A_5 or A_6 for population distributions P_1 and P_2. These correspond to β_1 being 1 or 1.5. When β_2 is at least 2, which is alternative hypothesis A_7, the simulated power is nearly or exactly 1 under all conditions except for the combination P_1 and Σ_3, where it is 0.8918, which is still very large. These results indicate the sensitivity in the ARMU tests increases greatly when the deviation from the null hypothesis is at least 1 for one level of factor B. However, the sensitivity in the tests is low under P_3, which is the log-normal distribution. Here the power does not significantly increase until the deviation is very large, and the size of these tests is also small. It appears that the increased skewness of the underlying population is leading to poor test performance all around.

Overall, for the tests of the main effect of factor B, the RMA method seems preferable under Σ_3, but for all other cases the ARMU method appears better as the sizes are more preferable and the estimate power increases more quickly.

In comparing simulated powers for the tests of effect of interaction we first note that under P_3 none of the results are good. The size of the tests for the ARMU method are too low (around 0.02) while the size of the tests for the RMA method are too high (around 0.1). In all cases, the simulated power shows little to no increase as the deviation in the alternative hypotheses increases.

Both methods do not show a significant jump in power until the alternative A_8 or A_9, which corresponds to $\gamma_{21} = 3$ or $\gamma_{21} = 4$, respectively. Under covariance structures Σ_1 and Σ_2 and for populations P_1 and P_2, the simulated power in the ARMU method is much greater than in the RMA method when the deviation in the alternative hypothesis is great; when the deviation is at most 2 (A_7) the RMA method has slightly better simulated powers and is systematically better for detecting smaller deviation. However, the sizes of the RMA method tests are not near 0.05–they are closer to 0.1–so the tests are not directly comparable. Under Σ_3 the sizes are comparable. Here, though, the ARMU method has simulated powers at least as good as the RMA method for P_1 and P_2, and much greater for alternative hypotheses A_8 and A_9.

Overall, for the tests of the interaction effect of factors A and B, the ARMU method seems preferable under all cases, having good power and appropriate size; however, neither method performs well under the log-normal distribution, where the underlying population distribution is too skewed for the test of interaction to perform well.

In comparing simulated powers for the tests of the main effect of factor A, the results for the ARMU method and RMA method are not distinguishable. Both have appropriate simulated sizes and all tests are very sensitive to departures from the null hypotheses. For populations P_1 and P_2, the powers reach 1 or nearly 1 by A_3, which corresponds to $\alpha_2 = 0.5$, and are very high for A_2, which corresponds to $\alpha_2 = 0.25$. The simulated powers do not increase greatly for P_3 until A_3, and it is not until A_5, which corresponds to $\alpha_2 = 1$, when the simulated powers are all 1 or nearly 1. Again, this may be a byproduct of the log-normal distribution being extremely skewed.

To summarize the power comparison of the tests from the ARMU method and the RMA method, the ARMU method seems the better overall choice. The ARMU method does no worse than the RMA method across the board, and the RMA method has many instances where the simulated size was not ideal (away from 0.05) for $b = 100$. Both methods performed very well for the tests of the main effect of factor A. Both methods seemed to perform more poorly under P_3, the log-normal distribution, which is likely due to the fact that this is a highly skewed distribution compared to the normal and skew-normal distributions. The tests for the main effect of factor B were more sensitive than the tests for the effect of interaction.

Design	Hyp	P_1 Σ_1	P_1 Σ_2	P_1 Σ_3	P_2 Σ_1	P_2 Σ_2	P_2 Σ_3	P_3 Σ_1	P_3 Σ_2	P_3 Σ_3
ARMU	H_0	0.0505	0.0465	0.0396	0.0543	0.0568	0.0400	0.0272	0.0362	0.0194
	A_1	0.0510	0.0447	0.0387	0.0520	0.0530	0.0411	0.0286	0.0321	0.0206
	A_2	0.0532	0.0483	0.0444	0.0506	0.0484	0.0468	0.0247	0.0307	0.0183
	A_3	0.0879	0.0628	0.0583	0.0733	0.0495	0.0663	0.0247	0.0236	0.0214
	A_4	0.1753	0.1116	0.0871	0.1735	0.0743	0.1331	0.0312	0.0231	0.0244
	A_5	0.3954	0.2234	0.1662	0.4223	0.1356	0.2827	0.0322	0.0235	0.0261
	A_6	0.9263	0.6534	0.5117	0.9757	0.6011	0.7817	0.0564	0.0259	0.0480
	A_7	0.9999	0.9744	0.8918	1	0.9810	0.9910	0.1168	0.0379	0.0898
	A_8	1	1	0.9999	1	1	1	0.5142	0.1454	0.3146
	A_9	1	1	1	1	1	1	0.9271	0.4873	0.7079
RMA	H_0	0.1044	0.0975	0.0528	0.1278	0.1455	0.0442	0.1331	0.1481	0.0508
	A_1	0.1152	0.1007	0.0506	0.1219	0.1158	0.0512	0.1270	0.1360	0.0527
	A_2	0.1313	0.1101	0.0565	0.1191	0.1008	0.0575	0.1182	0.1260	0.0527
	A_3	0.2054	0.1522	0.0828	0.1899	0.1099	0.1023	0.1125	0.1055	0.0602
	A_4	0.3886	0.2538	0.144	0.3681	0.1645	0.2052	0.1235	0.0993	0.0660
	A_5	0.6520	0.4093	0.2512	0.6848	0.3015	0.4057	0.1362	0.1012	0.0808
	A_6	0.9848	0.8540	0.6488	0.9968	0.8123	0.8787	0.2117	0.1135	0.1247
	A_7	1	0.9955	0.9463	1	0.9977	0.9978	0.3716	0.1615	0.2111
	A_8	1	1	1	1	1	1	0.8432	0.4020	0.5518
	A_9	1	1	1	1	1	1	0.9957	0.8132	0.8970

Table 3.7: *Effect of factor B*: Simulated powers for the test (time effect) with statistic T^*_β (ARMU) and the corresponding traditional RM-ANOVA F-statistic (RMA, from Davis [16]) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, for the group structure N_1, and for $b = 100$. Results are for multiple alternative hypotheses A_i.

Design	Hyp	P_1 Σ_1	P_1 Σ_2	P_1 Σ_3	P_2 Σ_1	P_2 Σ_2	P_2 Σ_3	P_3 Σ_1	P_3 Σ_2	P_3 Σ_3
ARMU	H_0	0.0521	0.0478	0.0438	0.0493	0.0476	0.0421	0.0235	0.0217	0.0192
	A_1	0.0520	0.0492	0.0438	0.0476	0.0470	0.0449	0.0244	0.0225	0.0181
	A_2	0.0486	0.0465	0.0431	0.0497	0.0468	0.0415	0.0205	0.0227	0.0184
	A_3	0.0530	0.0457	0.0435	0.0529	0.0450	0.0436	0.0237	0.0220	0.0181
	A_4	0.0512	0.0469	0.0404	0.0562	0.0531	0.0448	0.0239	0.0237	0.0206
	A_5	0.0519	0.0517	0.0474	0.0626	0.0548	0.0443	0.0200	0.0231	0.0194
	A_6	0.0744	0.0581	0.0479	0.0988	0.0691	0.0562	0.0262	0.0228	0.0226
	A_7	0.1113	0.0754	0.0642	0.1837	0.1029	0.0836	0.0284	0.0276	0.0204
	A_8	0.3365	0.1739	0.1349	0.5839	0.3152	0.2303	0.0388	0.0304	0.0253
	A_9	0.7344	0.4173	0.3033	0.9512	0.7083	0.5400	0.0611	0.0431	0.0293
RMA	H_0	0.1166	0.1029	0.0514	0.1155	0.0925	0.0516	0.1239	0.1136	0.0808
	A_1	0.1131	0.1020	0.0496	0.1134	0.0973	0.0540	0.1300	0.1164	0.0790
	A_2	0.1107	0.1003	0.0537	0.1079	0.0918	0.0515	0.1287	0.1146	0.0751
	A_3	0.1203	0.0996	0.0506	0.1203	0.1040	0.0503	0.1288	0.1141	0.0775
	A_4	0.1214	0.1004	0.0499	0.1223	0.1040	0.0568	0.1302	0.1148	0.0775
	A_5	0.1242	0.1100	0.0591	0.1289	0.1080	0.0539	0.1337	0.1128	0.0761
	A_6	0.1456	0.1143	0.0634	0.1586	0.1221	0.0644	0.1367	0.1159	0.0803
	A_7	0.1708	0.1283	0.0727	0.2112	0.1507	0.0764	0.1347	0.1295	0.0827
	A_8	0.2777	0.1810	0.1010	0.3938	0.2478	0.1248	0.1563	0.1363	0.0884
	A_9	0.4691	0.2991	0.1523	0.6792	0.4323	0.2442	0.1827	0.1462	0.0915

Table 3.8: *Effect of interaction between factors A and B*: Simulated powers for the test with statistic T^*_γ (ARMU) and the corresponding traditional RM-ANOVA F-statistic (RMA, from Davis [16]) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, for the group structure N_1, and for $b = 100$. Results are for multiple alternative hypotheses A_i.

Design	Hyp	P_1 Σ_1	P_1 Σ_2	P_1 Σ_3	P_2 Σ_1	P_2 Σ_2	P_2 Σ_3	P_3 Σ_1	P_3 Σ_2	P_3 Σ_3
ARMU	H_0	0.0483	0.0509	0.0539	0.0507	0.0511	0.0492	0.0477	0.0477	0.0456
	A_1	0.2836	0.1476	0.2255	0.4060	0.2001	0.3103	0.0906	0.0750	0.0770
	A_2	0.9696	0.7062	0.9138	0.9975	0.8816	0.9830	0.3833	0.1949	0.2970
	A_3	1	0.9995	1	1	1	1	0.9403	0.6496	0.8564
	A_4	1	1	1	1	1	1	0.9991	0.9468	0.9949
	A_5	1	1	1	1	1	1	1	0.9973	0.9998
	A_6	1	1	1	1	1	1	1	1	1
	A_7	1	1	1	1	1	1	1	1	1
	A_8	1	1	1	1	1	1	1	1	1
	A_9	1	1	1	1	1	1	1	1	1
RMA	H_0	0.0488	0.0481	0.0494	0.0495	0.0515	0.0504	0.0483	0.0482	0.0485
	A_1	0.2862	0.1524	0.2228	0.4066	0.1990	0.3086	0.0948	0.0685	0.0810
	A_2	0.9695	0.7099	0.9055	0.9968	0.8764	0.9819	0.3846	0.1828	0.2962
	A_3	1	0.9999	1	1	1	1	0.9415	0.6411	0.8538
	A_4	1	1	1	1	1	1	0.9994	0.9519	0.9950
	A_5	1	1	1	1	1	1	1	0.9977	0.9999
	A_6	1	1	1	1	1	1	1	0.9999	1
	A_7	1	1	1	1	1	1	1	1	1
	A_8	1	1	1	1	1	1	1	1	1
	A_9	1	1	1	1	1	1	1	1	1

Table 3.9: *Effect of factor A*: Simulated powers for the test (group effect) with statistic F_α^* (ARMU) and the corresponding traditional RM-ANOVA F-statistic (RMA, from Davis [16]) when sampling from the normal (P_1), skew-normal (P_2), and log-normal (P_3) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, for the group structure N_1, and for $b = 100$. Results are for multiple alternative hypotheses A_i.

Chapter 4

Inference for Large Number of Repeated Measures: Multivariate Case

4.1 Introduction

In many experimental or observational studies, multiple outcome variables are observed from each subject repeatedly. The subjects are often grouped according to the treatment they received or the experimental conditions to which they were subjected. Sometimes the grouping occurs due to a natural phenomenon, such as the sex of the subject, and other times the grouping is imposed by the researcher. These data may be generated by longitudinal studies or crossover designs, among others, and are the natural multivariate extension to the univariate case presented in Chapter 2. Some authors refer to this as doubly multivariate data analysis [9], [43]. In keeping with previous chapters, we will refer to such analysis as repeated measures multivariate analysis of variance, or repeated measures MANOVA or RM-MANOVA for short.

One strategy for analyzing multivariate repeated measures data, in particular when the outcomes can be labeled as primary, secondary, and so forth (based on the advice of the researcher), is to analyze each outcome variable separately using readily available longitudinal data analysis methods [18], [44]. This approach has one major pitfall, however, in that it completely ignores the covariance structure across the outcome variables which can lead to invalid inference. Another approach is to conduct a univariate analysis only after establishing multivariate significance [36].

The most commonly encountered multivariate repeated measures data can be viewed as arising from a two-factor crossed design. This is merely an extension of the univariate case save that the responses are now random vectors rather than random variables. Taking the analog of the univariate case, let $\boldsymbol{X}_{ijk}$ be independent, $p \times 1$ random vectors with mean $\boldsymbol{\mu}_{ij} = \mathrm{E}\left(\boldsymbol{X}_{ijk}\right)$, for $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, n_i$. When listed individually, let $X_{ijk}^{(h)}$ be the h^{th} response variable from the k^{th} subject treated with the i^{th} level of factor A and the j^{th} level of factor B, for $h = 1 \ldots, p$. The usual setting gives the interpretation that $\boldsymbol{X}_{ijk}$ is the vector of responses from the k^{th} subject in the i^{th} group at the j^{th} time point. Assume the mean vectors $\boldsymbol{\mu}_{ij} = \mathrm{E}\left(\boldsymbol{X}_{ijk}\right)$ admit the decomposition $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}$, where the unknown vectors of constants $\boldsymbol{\mu}$, $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_j$, and $\boldsymbol{\gamma}_{ij}$ correspond to the overall mean, the main effects of the i^{th} level of factor A, the main effects of the j^{th} level of factor B, and the interaction effects of the i^{th} level of factor A and the j^{th} level of factor B, respectively. The interaction effect will be denoted by Γ. As usual, these shall be subject to the identifiability restrictions, or sum-to-zero constraints, $\sum_i \boldsymbol{\alpha}_i = \sum_j \boldsymbol{\beta}_j = \sum_i \boldsymbol{\gamma}_{ij} = \sum_j \boldsymbol{\gamma}_{ij} = \mathbf{0}$. An essential aspect of this model is that $\boldsymbol{X}_{ijk}$ and $\boldsymbol{X}_{i'j'k'}$ are assumed to be independent only if $i \neq i'$ or $k \neq k'$ since observations are correlated through time. More specifically, defining $\boldsymbol{\epsilon}_{ijk} = \boldsymbol{X}_{ijk} - \boldsymbol{\mu}_{ij}$, we consider $\boldsymbol{\epsilon}_{i1k}, \boldsymbol{\epsilon}_{i2k}, \ldots$ to be a sequence of dependent random vectors. Consequently, a matrix of correlated observations is made per subject.

The primary hypotheses of interest are

(i) $\mathcal{H}_0^{(B)} \ : \ \boldsymbol{\beta}_j = \mathbf{0}$ for $j = 1, \ldots, b$;

(ii) $\mathcal{H}_0^{(\Gamma)}$: $\boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $i = 1, \ldots, a$ and $j = 1, \ldots, b$; and

(iii) $\mathcal{H}_0^{(A)}$: $\boldsymbol{\alpha}_i = \mathbf{0}$ for $i = 1, \ldots, a$.

These null hypotheses correspond to no main effects of factor B, no interaction effects of level of factor A and levels of factor B, and no main effects of factor A, respectively. As in the univariate case, the asymptotic framework to be considered is that the number of levels of one of the factors, namely factor B, is large (tends to infinity) but that the number of levels of the other factor and the sample sizes within the levels of factors A and B remain fixed. See Harrar and Bathke [26] for a similar asymptotic framework.

There are two commonly used approaches for modeling multivariate repeated measures. These approaches are identified as the Doubly Multivariate Model (DMM) and the Multivariate Mixed Model (MMM) [26]. These two approaches differ in the way covariance over time is modeled. In a MMM, random effects are used to model the covariance in a way analogous to the univariate mixed model and the hypotheses of interest can be formulated as linear hypotheses. On the other hand, a DMM stacks the multivariate repeated observations from the same subject to form a vector of observations. Then the hypotheses of interest are formulated as general linear hypotheses. While a MMM inherently imposes a structure on the covariance over time, a DMM allows the most general structure for the covariance. For detailed accounts of these two approaches, the reader is advised to consult Timm [43] and Boik [9].

Some advances have been made for the DMM in the past decade, assuming normality and assuming the covariance matrix of the vector of observations per subject admits a Kronecker product structure; i.e., the covariance of the vector $\begin{bmatrix} \boldsymbol{X}'_{i1k} & \cdots & \boldsymbol{X}'_{ibk} \end{bmatrix}'$ is assumed to have the form $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are $b \times b$ and $p \times p$, respectively, and both are positive definite matrices. Assuming independence across subjects and multivariate normality for the vector of observations from each subject, Naik and Rao [34] proposed various tests for the three hypotheses of interest in a multivariate repeated measures design. Under similar assumptions, Srivastava et al. [42] recently have obtained likelihood-based results in the general multivariate linear model setup. There are a few authors who have addressed the issue of testing the validity of a Kronecker product structure for the covariance matrix (see, for example, Mitchell et al. [32], Lu and Zimmerman [30]).

To the best of our knowledge, no work has been done removing the assumption of normality. In this chapter, we develop methods for testing the time, treatment and interaction effects for the multivariate repeated measures design for the unstructured covariance matrix case under general assumptions on the underlying distribution generating the data. The reduction of the results to the special case of a Kronecker product-type covariance structure will be indicated later as appropriate. Another important contribution of this chapter is that it proposes modified versions of some multivariate test statistics, and their null distributions are derived for the situation when the dimension of the vector of observations per subject tends to infinity but replication size per group is limited. While this is similar to the univariate results from Chapter 2, here the scope of the problem is broadened by removing the assumption of stationarity. Similar asymptotic frameworks in a two-way cross classified layout have been considered by Gupta et. al. [22], Bathke and Harrar [4], and Harrar and Bathke [26]. However, the present work differs from these in important ways. In all three cases it is assumed that the observations under two different levels of factor B (the time factor in this case) are independent. The test statistics considered in the present work take the dependence into account. Furthermore, the derivation of the results in the face of dependence across the levels of factor B is much more involved.

Methods for high-dimensional analysis have become increasingly popular due to the type and amount of data arising in the present day. Recently, high throughput diagnostics and biotechnological advancements such as fMRI and microarrays have produced vast amounts of data to be analyzed [26], [24]. There are also agricultural and health sciences applications needing new methods for analyzing multivariate data [10], [1]. In many such scenarios, the number of levels of one factor is very large, but the sample size is often limited due to cost or availability. For instance, Holden et. al [27] explore climate data in which high and low temperatures are recorded daily for a long period of time across many sites in Montana. The number of independent data collection sites is limited due to terrain. Microarrays are another example in which sample sizes become limited as they are monetarily expensive to collect Furthermore, it may be unreasonable to assume a specific covariance structure or underlying distribution for the data. The new methods in this chapter address both of these possible limitations. Many current multivariate techniques assume a specific covariance structure, an underlying normal distribution, or availability of a large number of replications [36], [28].

The remainder of this chapter will have the following organization. Section 4.2 will include some background lemmas and concepts necessary for multivariate computations. Also included in Section 4.2 will be a brief account of a fairly weak dependence structure to be assumed and a central limit theorem needed to prove the main theorems in later sections. Section 4.3 will define the sum of squares matrices used to calculate the test statistics in both their intuitive and more tractable forms, the latter being akin to the univariate quadratic forms, and hence hereafter referred to as matrix quadratic forms. The test statistics will also be included here. Section 4.4 will contain a discussion of the asymptotic null distributions of the test statistics, which are the main results of this work. Section 4.5 will present a simulation study, and Section 4.6 will give reductions to certain results under some specific, known structures.

4.2 Preliminaries

In proving the main results of the dissertation, we need to compute the first and second moments of matrix quadratic forms. To aid these calculations, we establish some formulas in the following propositions.

Proposition 4.2.1. *Let* $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_b \end{bmatrix}$, *where* $\boldsymbol{X}_j = \begin{bmatrix} \boldsymbol{X}_{1j} & \cdots & \boldsymbol{X}_{nj} \end{bmatrix}$ *and* $\boldsymbol{X}_{ij} = \begin{bmatrix} X_{ij}^{(1)} & \cdots & X_{ij}^{(p)} \end{bmatrix}'$. *Assume that* $\boldsymbol{X}_{ij}$ *and* $\boldsymbol{X}_{i'j'}$ *are independent for* $i \neq i'$ *and* $\mathrm{E}\left(\left|X_{ij}^{(r_1)} X_{ij'}^{(r_2)} X_{il}^{(r_3)} X_{il'}^{(r_4)}\right|\right) < \infty$ *for* $j, j', l, l' = 1, \ldots, b$ *and* $r_1, r_2, r_3, r_4 = 1, \ldots, p$. *Further assume* $\mathrm{E}\left(\boldsymbol{X}\right) = \boldsymbol{M}$ *and* $\mathrm{Cov}\left(\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'}\right) = \boldsymbol{\Sigma}_{jj'}$, *which is* $p \times p$. *Let* $\boldsymbol{A}$ *be an* $(n \cdot b) \times (n \cdot b)$

symmetric matrix of constants such that $\boldsymbol{A}$ *is a block partitioned matrix with* $\boldsymbol{A}_{jj'}$ *being the* $n \times n$ *submatrices in the* $(j, j')^{th}$ *blocks of* $\boldsymbol{A}$*. Then*

$$\mathrm{E}\left(\boldsymbol{XAX'}\right) = \left(\sum_{j,j'=1}^{b}\sum_{i=1}^{n} a_{jj'ii}\boldsymbol{\Sigma}_{jj'}\right) + \boldsymbol{MAM'}, \tag{4.1}$$

where $a_{jj'ii'}$ *is the* $(i, i')^{th}$ *element of the* $(j, j')^{th}$ *block* $\boldsymbol{A}_{jj'}$ *of* $\boldsymbol{A}$*.*

Further, let $\boldsymbol{B}$*,* $\boldsymbol{B}_{jj'}$*, and* $b_{jj'ii'}$ *be defined similarly to* $\boldsymbol{A}$*,* $\boldsymbol{A}_{jj'}$*, and* $a_{jj'ii'}$*, respectively. Then*

$$\begin{aligned}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{XAX'}\right)\mathrm{vec}\left(\boldsymbol{XBX'}\right)'\right] = \sum_{j,j',l,l'}\Bigg\{\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{ij'}\right)\mathrm{vec}\left(\boldsymbol{X}_{il}\boldsymbol{X}'_{il'}\right)'\right]a_{jj'ii}b_{ll'ii} \\ +\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{jj'}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{ll'}\right)'\sum_{i\neq i'}^{n} a_{jj'ii}b_{ll'i'i'} \\ +\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{jl}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l'}\right)'\sum_{i\neq i'}^{n} a_{jj'ii'}b_{ll'ii'} +\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{jl'}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l}\right)'\sum_{i\neq i'}^{n} a_{jj'ii'}b_{ll'i'i}\Bigg\}.\end{aligned} \tag{4.2}$$

Proof: Let $\boldsymbol{\mu}_{ij} = \mathrm{E}\left(\boldsymbol{X}_{ij}\right)$. Observe,

$$\begin{aligned}\mathrm{E}\left(\boldsymbol{XAX'}\right) &= \mathrm{E}\left(\sum_{j,j'=1}^{b}\boldsymbol{X}_j\boldsymbol{A}_{jj'}\boldsymbol{X}_{j'}\right) \\ &= \sum_{i,i'=1}^{n}\sum_{j,j'=1}^{b} a_{jj'ii'}\mathrm{E}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{i'j'}\right) \\ &= \sum_{i,i'=1}^{n}\sum_{j,j'=1}^{b} a_{jj'ii'}\left[\mathrm{Cov}\left(\boldsymbol{X}_{ij},\boldsymbol{X}_{i'j'}\right)+\boldsymbol{\mu}_{ij}\boldsymbol{\mu}'_{i'j'}\right] \\ &= \left(\sum_{i=1}^{n}\sum_{j,j'=1}^{b} a_{jj'ii}\boldsymbol{\Sigma}_{jj'}\right) + \left(\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{b} a_{jj'ii'}\boldsymbol{\mu}_{ij}\boldsymbol{\mu}'_{i'j'}\right) \\ &= \left(\sum_{j,j'=1}^{b}\sum_{i=1}^{n} a_{jj'ii}\boldsymbol{\Sigma}_{jj'}\right) + \mathrm{E}\left(\boldsymbol{MAM'}\right) \\ &= \left(\sum_{j,j'=1}^{b}\sum_{i=1}^{n} a_{jj'ii}\boldsymbol{\Sigma}_{jj'}\right) + \boldsymbol{MAM'}.\end{aligned}$$

Notice that the fourth equality is justified by the assumption of independence of $\boldsymbol{X}_{ij}$ and $\boldsymbol{X}_{i'j'}$ for $i \neq i'$.

To prove (4.2), observe,

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\right)\mathrm{vec}\left(\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}'\right)'\right] \\
&= \mathrm{E}\left[\mathrm{vec}\left(\sum_{j,j'=1}^{b}\sum_{i,i'=1}^{n} a_{jj'ii'}\boldsymbol{X}_{ij}\boldsymbol{X}'_{i'j'}\right)\mathrm{vec}\left(\sum_{l,l'=1}^{b}\sum_{k,k'=1}^{n} b_{ll'kk'}\boldsymbol{X}_{kl}\boldsymbol{X}'_{k'l'}\right)'\right] \\
&= \sum_{j,j'=1}^{b}\sum_{i,i'=1}^{n}\sum_{l,l'=1}^{b}\sum_{k,k'=1}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{i'j'}\right)\mathrm{vec}\left(\boldsymbol{X}_{kl}\boldsymbol{X}'_{k'l'}\right)a_{jj'ii'}b_{ll'kk'}\right] \\
&= \sum_{j,j',l,l'=1}^{b}\left\{\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{ij'}\right)\mathrm{vec}\left(\boldsymbol{X}_{il}\boldsymbol{X}'_{il'}\right)\right]a_{jj'ii}b_{ll'ii}\right. \\
&\quad + \sum_{i\neq i'}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{ij'}\right)\mathrm{vec}\left(\boldsymbol{X}_{i'l}\boldsymbol{X}'_{i'l'}\right)\right]a_{jj'ii}b_{ll'i'i'} \\
&\quad + \sum_{i\neq i'}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{il}\right)\mathrm{vec}\left(\boldsymbol{X}_{i'j'}\boldsymbol{X}'_{i'l'}\right)\right]a_{jj'ii'}b_{ll'ii'} \\
&\quad \left. + \sum_{i\neq i'}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{il'}\right)\mathrm{vec}\left(\boldsymbol{X}_{i'j'}\boldsymbol{X}'_{i'l}\right)\right]a_{jj'ii'}b_{ll'i'i}\right\} \\
&= \sum_{j,j',l,l'=1}^{b}\left\{\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{ij'}\right)\mathrm{vec}\left(\boldsymbol{X}_{il}\boldsymbol{X}'_{il'}\right)'\right]a_{jj'ii}b_{ll'ii}\right. \\
&\quad + \mathrm{vec}\left(\boldsymbol{\Sigma}_{jj'}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{ll'}\right)'\sum_{i\neq i'}^{n}a_{jj'ii}b_{ll'i'i'} + \mathrm{vec}\left(\boldsymbol{\Sigma}_{jl}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l'}\right)'\sum_{i\neq i'}^{n}a_{jj'ii'}b_{ll'ii'} \\
&\quad \left. + \mathrm{vec}\left(\boldsymbol{\Sigma}_{jl'}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l}\right)'\sum_{i\neq i'}^{n}a_{jj'ii'}b_{ll'i'i}\right\}.
\end{aligned}$$

Note that the indices i, i', k, and k' are reduced to i and i' with multiple summations; this aids in recognizing the similarities between the terms in the final expression. □

It is noteworthy that the first term in the RHS of (4.2) involves fourth moments, which have the potential to cause difficulties. However, it is equally noteworthy that this term will vanish if the diagonals of the blocks of either $\boldsymbol{A}$ or $\boldsymbol{B}$ are filled with zeros. This is indeed the case in the asymptotic results presented later.

In Theorems 4.4.7 and 4.4.12, we establish the asymptotic distribution of the test statistics under the large number of repeated measures asymptotic framework. The central idea in the proofs is to decompose the test statistics into a sum of dependent matrix quadratic forms and an asymptotically negligible component. This is similar to the idea in the univariate case given in Corollary 2.2.7, (2.16), and Theorem 2.4.11. Then a CLT for dependent vectors will be applied on the former, non-negligible component. The CLT for dependent random vectors used in this paper requires asymptotic independence of the past and future observations when the separation is great enough.

Definition 4.2.2 gives a notion of weak dependence for a sequence of random vectors. It is the analogous extension to α-mixing for a sequence of random variables given in Definition 2.4.4.

Definition 4.2.2. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and let $\mathcal{A}$ and $\mathcal{B}$ be sub-σ-algebras of $\mathcal{F}$. Define the dependence coefficient $\widetilde{\alpha}$ by*

$$\widetilde{\alpha}(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|$$

as a measure of dependence between $\mathcal{A}$ and $\mathcal{B}$. For a sequence of random vectors $\{\boldsymbol{X}_t\}$, let $\mathcal{F}_a^b := \sigma(\boldsymbol{X}_a, \boldsymbol{X}_{a+1}, \ldots, \boldsymbol{X}_b)$ be the σ-algebra generated by $\boldsymbol{X}_{a+1}, \ldots, \boldsymbol{X}_{a+b}$ and define

$$\alpha(m) = \sup_{k \in \mathbb{Z}} \widetilde{\alpha}\left(\mathcal{F}_{-\infty}^{k}, \mathcal{F}_{k+m}^{\infty}\right).$$

Then the sequence $\{\boldsymbol{X}_t\}$ is said to be <u>α-mixing</u> with $\alpha_{\boldsymbol{X}}(m)$ if $\alpha_{\boldsymbol{X}}(m) \to 0$ as $m \to \infty$.

Notice that the quantity $\alpha(m)$ measures how much dependence exists between observations separated by at least m periods.

The following CLT for α-mixing sequences of random variables (with $p = 1$) is due to White and Domowitz [46]. The extension to $p > 1$ is straightforward by the Cramer-Wold device [15].

Theorem 4.2.3 (White and Domowitz, 1984). *Let $Z_1, Z_2, \ldots,$ be an α-mixing sequence of random variables with* $\mathrm{E}(Z_j) = 0$. *Assume*

(i) there exists $V^2 \in (0,\infty)$ such that $\mathrm{E}\left(T_k(n)^2\right) - V^2 \to 0$ as $n \to \infty$, uniformly in k, where $T_k(n) = n^{-1/2}\sum_{j=k+1}^{k+n} Z_j$;

(ii) $\mathrm{E}\left(|Z_j|^{2r}\right) \leq \Delta < \infty \ \forall\, j$ and some $r > 1$; and

(iii) $\alpha(m) = O\left(m^{-\lambda}\right)$ for some $\lambda > r/(r-1)$.

Then $(nV^2)^{-1/2}\sum_{j=1}^{n} Z_j \xrightarrow{D} N(0,1)$ as $n \to \infty$.

Theorem 4.2.3 was originally proved by Serfling [39] under some restrictions on the dependence which are not quite intuitive. White and Domowitz [46] have shown that the Theorem holds true under the α-mixing dependence structure, which is much more intuitive. In comparison to the CLT from Billingsley [8], this one requires fewer assumptions. More precisely, stationarity and existence of higher moments are not needed.

4.3 Sums of Squares and Test Statistics

4.3.1 Sums of Squares

Analogous to the univariate case in Chapter 2, where now we consider the $p \times 1$ random vector $\boldsymbol{X}_{ijk}$ rather than the random variable X_{ijk}, for $n = \sum_{i=1}^{a} n_i$, let

$$\begin{aligned}
\widetilde{\boldsymbol{X}}_{\cdots} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}, \\
\overline{\boldsymbol{X}}_{\cdots} &= \frac{1}{nb}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}, \\
\widetilde{\boldsymbol{X}}_{i\cdot\cdot} &= \overline{\boldsymbol{X}}_{i\cdot\cdot} = \frac{1}{bn_i}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk},
\end{aligned}$$

$$\begin{aligned}\widetilde{\boldsymbol{X}}_{\cdot j\cdot} &= \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk},\\ \overline{\boldsymbol{X}}_{ij\cdot} &= \frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{X}_{ijk}, \quad \text{and}\\ \overline{\boldsymbol{X}}_{i\cdot k} &= \frac{1}{b}\sum_{j=1}^{b}\boldsymbol{X}_{ijk}.\end{aligned} \tag{4.3}$$

The multivariate analog to (2.2) allows us to define the hypothesis sums of squares and cross products (SSCP) matrices for the main effects of factor B and the interaction effect Γ, respectively, by

$$\boldsymbol{H}^{(B)} := \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widetilde{\boldsymbol{X}}_{\cdot j\cdot}-\widetilde{\boldsymbol{X}}_{\cdots}\right)\left(\widetilde{\boldsymbol{X}}_{\cdot j\cdot}-\widetilde{\boldsymbol{X}}_{\cdots}\right)' \quad \text{and} \tag{4.4}$$

$$\boldsymbol{H}^{(\Gamma)} := \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{\boldsymbol{X}}_{ij\cdot}-\widetilde{\boldsymbol{X}}_{i\cdot\cdot}-\widetilde{\boldsymbol{X}}_{\cdot j\cdot}+\widetilde{\boldsymbol{X}}_{\cdots}\right)\left(\overline{\boldsymbol{X}}_{ij\cdot}-\widetilde{\boldsymbol{X}}_{i\cdot\cdot}-\widetilde{\boldsymbol{X}}_{\cdot j\cdot}+\widetilde{\boldsymbol{X}}_{\cdots}\right)', \tag{4.5}$$

and the within subjects error SSCP matrices by

$$\begin{aligned}\boldsymbol{G}^{(B)} := &\frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\\ &\times\sum_{k=1}^{n_i}\left(\boldsymbol{X}_{ijk}-\overline{\boldsymbol{X}}_{ij\cdot}-\overline{\boldsymbol{X}}_{i\cdot k}-\widetilde{\boldsymbol{X}}_{i\cdot\cdot}\right)\left(\boldsymbol{X}_{ijk}-\overline{\boldsymbol{X}}_{ij\cdot}-\overline{\boldsymbol{X}}_{i\cdot k}-\widetilde{\boldsymbol{X}}_{i\cdot\cdot}\right)',\end{aligned} \tag{4.6}$$

where $\boldsymbol{G}^{(\Gamma)} = \boldsymbol{G}^{(B)}$. Note that for concise notation we will write, for example, $\boldsymbol{H}^{(\phi)}$ or $\boldsymbol{G}^{(\phi)}$, where $\phi \in \{B, \Gamma, A\}$, depending on the scenario; B is used for factor B, Γ for the interaction, and A for factor A. Similar notation will also be utilized elsewhere.

These are different from the usual SSCP matrices to account for the unbalanced nature of the levels in factor A. Using these, we define hypothesis SSCP matrices and the within subjects error SSCP matrices for the main effect of factor B and the interaction, and then for the main effects of factor A. The justification for this is analogous to that given in Section

2.2.1.

Given no interaction effects, the hypothesis of no main effects $\mathcal{H}_0^{(A)}$ is equivalent to testing $H_0 : \bar{\boldsymbol{\mu}}_{1\cdot} = \ldots = \bar{\boldsymbol{\mu}}_{a\cdot}$. This later hypothesis suggests that it will be reasonable to conduct one-way MANOVA on the n means $\overline{\boldsymbol{X}}_{1\cdot 1}, \ldots, \overline{\boldsymbol{X}}_{1\cdot n_1}, \ldots, \overline{\boldsymbol{X}}_{a\cdot 1}, \ldots, \overline{\boldsymbol{X}}_{a\cdot n_a}$. In view of this, the between and within group SSCP matrices for testing the effects of factor A become

$$\boldsymbol{H}^{(A)} := \sum_{i=1}^{a}\sum_{j=1}^{b} n_i \left(\overline{\boldsymbol{X}}_{i\cdot\cdot} - \widetilde{\boldsymbol{X}}_{\cdot\cdot\cdot}\right)\left(\overline{\boldsymbol{X}}_{i\cdot\cdot} - \widetilde{\boldsymbol{X}}_{\cdot\cdot\cdot}\right)' \tag{4.7}$$

and

$$\boldsymbol{G}^{(A)} := \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i} \left(\overline{\boldsymbol{X}}_{i\cdot k} - \overline{\boldsymbol{X}}_{i\cdot\cdot}\right)\left(\overline{\boldsymbol{X}}_{i\cdot k} - \overline{\boldsymbol{X}}_{i\cdot\cdot}\right)', \tag{4.8}$$

respectively. All of the SSCP matrices given above are invariant to translation; i.e., they are invariant to adding a constant to all observations. Therefore, we can assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$.

Since the SSCP matrices $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$ are directly analogous to the univariate case, we have the following proposition above the matrix quadratic forms of $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$. First, let $\boldsymbol{X}$ have the following organization. Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_b \end{bmatrix};\ \boldsymbol{X}_j = \begin{bmatrix} \boldsymbol{X}_{1j} & \cdots & \boldsymbol{X}_{aj} \end{bmatrix};\ \boldsymbol{X}_{ij} = \begin{bmatrix} \boldsymbol{X}_{ij1} & \cdots & \boldsymbol{X}_{ijn_i} \end{bmatrix}. \tag{4.9}$$

Also, define $\widetilde{\boldsymbol{X}}$ by

$$\widetilde{\boldsymbol{X}} = \left[\begin{array}{ccc|c|ccc} \overline{\boldsymbol{X}}_{1\cdot 1} & \cdots & \overline{\boldsymbol{X}}_{1\cdot n_1} & \cdots & \overline{\boldsymbol{X}}_{a\cdot 1} & \cdots & \overline{\boldsymbol{X}}_{a\cdot n_a} \end{array}\right]. \tag{4.10}$$

Proposition 4.3.1. *We can rewrite $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$, for $\phi \in \{B, \Gamma, A\}$, as the following.*

$$\begin{aligned} \boldsymbol{H}^{(B)} &= \frac{1}{a(b-1)}\boldsymbol{X}\left(\boldsymbol{P}_b \otimes \left[\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right)\boldsymbol{J}_a\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right)\right]\right)\boldsymbol{X}'; \\ \boldsymbol{H}^{(\Gamma)} &= \frac{1}{(a-1)(b-1)}\boldsymbol{X}\left(\boldsymbol{P}_b \otimes \left[\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}\right)\boldsymbol{P}_a\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\mathbf{1}_{n_i}'\right)\right]\right)\boldsymbol{X}'; \end{aligned}$$

$$\begin{aligned}
\boldsymbol{G}^{(B)} &= \boldsymbol{G}^{(\Gamma)} = \frac{1}{a(b-1)}\boldsymbol{X}\left(\boldsymbol{P}_b \otimes \left(\bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}\right)\right)\boldsymbol{X}';\\
\boldsymbol{H}^{(A)} &= \left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)\left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)';\\
\boldsymbol{G}^{(A)} &= \left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)\left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)'. \qquad (4.11)
\end{aligned}$$

Proof: These results are obtained by straightforward extensions of (2.4) and (2.31), where $p = 1$, to the case where $p > 1$. □

Analogous to the univariate case, we will wish to display these matrix quadratic forms more succinctly. Thus, for $\phi \in \{B, \Gamma\}$, we define $\boldsymbol{C}_H^{(\phi)}$ and $\boldsymbol{C}_G^{(\phi)}$ so that we can write

$$\boldsymbol{H}^{(\phi)} = \boldsymbol{X}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}_H^{(\phi)}\right)\boldsymbol{X}' \quad \text{and} \quad \boldsymbol{G}^{(\phi)} = \boldsymbol{X}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}_G^{(\phi)}\right)\boldsymbol{X}'. \qquad (4.12)$$

We further define $\boldsymbol{C}^{(\phi)}$ such that $\boldsymbol{C}^{(\phi)} = \boldsymbol{C}_H^{\phi} - \boldsymbol{C}_G^{(\phi)}$.

4.3.2 Test Statistics

In order that the inference might be valid, we first wish to show that for each of $\boldsymbol{H}^{(B)}$, $\boldsymbol{H}^{(\Gamma)}$, and $\boldsymbol{H}^{(A)}$, the expected value under the corresponding null hypothesis is equal to that of $\boldsymbol{G}^{(B)}$, $\boldsymbol{G}^{(\Gamma)}$, and $\boldsymbol{G}^{(A)}$ up to a constant multiple for $\boldsymbol{H}^{(A)}$, respectively. This is necessary so that the tests of the hypotheses will detect departures from the null hypotheses. In fact, the following result holds.

Proposition 4.3.2. $\mathrm{E}\left(\boldsymbol{H}^{(\phi)}\right) = \mathrm{E}\left(\boldsymbol{G}^{(\phi)}\right)$ *if and only if* $\mathcal{H}_0^{(\phi)}$ *holds for each* $\phi \in \{B, \Gamma\}$, *and* $\mathrm{E}\left(\boldsymbol{H}^{(A)}/(a-1)\right) = \mathrm{E}\left(\boldsymbol{G}^{(A)}/(n-a)\right)$ *if and only if* $\mathcal{H}_0^{(A)}$ *holds.*

Proof: This proof will be organized into two cases. The first case will consider the main effect of factor B and the interaction effect Γ, or $\phi \in \{B, \Gamma\}$; the second case will consider the main effect of factor A, or $\phi \in \{A\}$.

Case 1: Since there are many similarities for $\phi = B$ and $\phi = \Gamma$, we treat them together. We begin by observing that we are under the general framework of Proposition 4.2.1. We let $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{M}$, and denote $\widetilde{\boldsymbol{\mu}}_{\cdot j \cdot}$ similar to $\widetilde{\boldsymbol{X}}_{\cdot j \cdot}$ as in (4.3), etc.

To show that $\mathrm{E}\left(\boldsymbol{H}^{(\phi)}\right) = \mathrm{E}\left(\boldsymbol{G}^{(\phi)}\right)$, we consider the difference in the two terms and show that it is $\mathbf{0}$ if and only if the corresponding null hypothesis holds. We will show that the first term of (4.1) vanishes since the submatrix block diagonals are zero, and then we will argue that the second term can equal $\mathbf{0}$ if and only if the corresponding null hypothesis holds.

Observe that

$$\mathrm{E}\left(\boldsymbol{H}^{(\phi)}\right) - \mathrm{E}\left(\boldsymbol{G}^{(\phi)}\right) = \mathrm{E}\left[\boldsymbol{X}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\right].$$

Now, recognize that $\boldsymbol{C}^{(\phi)}$ is simply $\boldsymbol{C}^*_{\phi_E}$ from (2.6), which has zero diagonals. Therefore, the submatrix blocks of $\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(\phi)}$ will have zero diagonals, and the first term in $\mathrm{E}\left[\boldsymbol{X}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\right]$ from Proposition 4.2.1 is $\mathbf{0}$.

It remains to show that $\boldsymbol{M}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(\phi)}\right)\boldsymbol{M}' = \mathbf{0}$ if and only if and only if $\mathcal{H}_0^{(\phi)}$ holds. Similar to (4.3), observe that we can write

$$\widetilde{\boldsymbol{\mu}}_{\cdots} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{\mu}_{ij} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}\right) = \boldsymbol{\mu}$$

and

$$\widetilde{\boldsymbol{\mu}}_{\cdot j \cdot} = \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{\mu}_{ij} = \frac{1}{ab}\sum_{i=1}^{a}\left(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}\right) = \boldsymbol{\mu} + \boldsymbol{\beta}_j$$

due to the identifiability constraints. Then we see we can write

$$\begin{aligned}\boldsymbol{M}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(B)}\right)\boldsymbol{M}' &= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widetilde{\boldsymbol{\mu}}_{\cdot j \cdot} - \widetilde{\boldsymbol{\mu}}_{\cdots}\right)\left(\widetilde{\boldsymbol{\mu}}_{\cdot j \cdot} - \widetilde{\boldsymbol{\mu}}_{\cdots}\right)' \\ &= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\boldsymbol{\mu} - \boldsymbol{\beta}_j - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu} - \boldsymbol{\beta}_j - \boldsymbol{\mu}\right)'\end{aligned}$$

$$= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\boldsymbol{\beta}_j\boldsymbol{\beta}_j',$$

which is clearly $\mathbf{0}$ if and only if $\boldsymbol{\beta}_j = \mathbf{0}\ \forall\ j$.

Similarly we see that

$$\begin{aligned}
&\boldsymbol{M}\left(\frac{1}{b-1}\boldsymbol{P}_b \otimes \boldsymbol{C}^{(\Gamma)}\right)\boldsymbol{M}' \\
&\quad= \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widetilde{\boldsymbol{\mu}}_{ij\cdot} - \widetilde{\boldsymbol{\mu}}_{i\cdot\cdot} - \widetilde{\boldsymbol{\mu}}_{\cdot j\cdot} + \widetilde{\boldsymbol{\mu}}_{\cdot\cdot\cdot}\right)\left(\widetilde{\boldsymbol{\mu}}_{ij\cdot} - \widetilde{\boldsymbol{\mu}}_{i\cdot\cdot} - \widetilde{\boldsymbol{\mu}}_{\cdot j\cdot} + \widetilde{\boldsymbol{\mu}}_{\cdot\cdot\cdot}\right)' \\
&\quad= \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} - (\boldsymbol{\mu} + \boldsymbol{\alpha}_i) - (\boldsymbol{\mu} + \boldsymbol{\beta}_j) + \boldsymbol{\mu}\right) \\
&\qquad\times\left(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} - (\boldsymbol{\mu} + \boldsymbol{\alpha}_i) - (\boldsymbol{\mu} + \boldsymbol{\beta}_j) + \boldsymbol{\mu}\right)' \\
&\quad= \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\boldsymbol{\gamma}_{ij}\boldsymbol{\gamma}_{ij}',
\end{aligned}$$

which is clearly $\mathbf{0}$ if and only if $\boldsymbol{\gamma}_{ij} = \mathbf{0}\ \forall\ i, j$.

Since the null hypotheses are $\mathcal{H}_0^{(B)} : \boldsymbol{\beta}_j = \mathbf{0}\ \forall\ j$ and $\mathcal{H}_0^{(\Gamma)} : \boldsymbol{\gamma}_{ij} = \mathbf{0}\ \forall\ i, j$, Case 1 is proved.

◇

Case 2: Consider matrices $\boldsymbol{A}_H$ and $\boldsymbol{A}_G$ such that $\boldsymbol{H}^{(A)}/(a-1) = \widetilde{\boldsymbol{X}}\boldsymbol{A}_H\widetilde{\boldsymbol{X}}'$ and $\boldsymbol{H}^{(G)}/(n-a) = \widetilde{\boldsymbol{X}}\boldsymbol{A}_G\widetilde{\boldsymbol{X}}'$, for $\widetilde{\boldsymbol{X}}$ defined in (4.10). For $\widetilde{\boldsymbol{X}}$ we have

$$\overline{\boldsymbol{\mu}}_{i\cdot} = \mathrm{E}\left(\overline{\boldsymbol{X}}_{i\cdot k}\right) = \frac{1}{b}\sum_{j=1}^{b}\left(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}\right) = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$$

due to the identifiability constraints.

Let the i^{th} column of $\widetilde{\boldsymbol{X}}$ be denoted $\widetilde{\boldsymbol{X}}_i$. Furthermore, let $\mathrm{Cov}\left(\widetilde{\boldsymbol{X}}_i, \widetilde{\boldsymbol{X}}_i\right)$ be denoted by $\boldsymbol{\Psi}$, which is the same for all i. Using Proposition 4.2.1 with $b = 1$ for an appropriate general matrix $\boldsymbol{A}$, we can write

$$\mathrm{E}\left(\widetilde{\boldsymbol{X}}\boldsymbol{A}\widetilde{\boldsymbol{X}}'\right) = \sum_{i,i'=1}^{n} a_{ii'}\left[\mathrm{Cov}\left(\widetilde{\boldsymbol{X}}_i, \widetilde{\boldsymbol{X}}_{i'}\right) + \overline{\boldsymbol{\mu}}_{i\cdot}\overline{\boldsymbol{\mu}}_{i'\cdot}'\right].$$

First, recognize that $\text{Cov}\left(\widetilde{\boldsymbol{X}}_i, \widetilde{\boldsymbol{X}}_{i'}\right) = \mathbf{0}$ when $i \neq i'$. With some slight manipulation of the subscripts, we see that this simplifies to

$$\text{E}\left(\widetilde{\boldsymbol{X}}\boldsymbol{A}\widetilde{\boldsymbol{X}}'\right) = \boldsymbol{\Psi}\sum_{i=1}^{n} a_{ii} + \sum_{i,i'=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_{i'}} a_{kk'ii'}\overline{\boldsymbol{\mu}}_{i\cdot}\overline{\boldsymbol{\mu}}'_{i'\cdot}, \tag{4.13}$$

where the two-subscript $a_{ii'}$ is the usual $(i,i')^{th}$ element of $\boldsymbol{A}$ and the four-subscript notation $a_{kk'ii'}$ is the $(i,i')^{th}$ element of the $(k,k')^{th}$ block of $\boldsymbol{A}$. (Note that the two-subscript notation is simpler and therefore more preferable if permitted.) For $\boldsymbol{A} = \boldsymbol{A}_H = b\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)/(a-1)$, the first term of (4.13) is $\boldsymbol{\Psi}b\left(\sum_{i=1}^{a}\frac{1}{n_i}n_i - \frac{1}{n}n\right)/(a-1) = b\boldsymbol{\Psi}$. Moreover, for $\boldsymbol{A} = \boldsymbol{A}_G = b\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)/(n-a)$, the first term of (4.13) is $\boldsymbol{\Psi}b\left(\sum_{i=1}^{a}\left(1-\frac{1}{n_i}\right)n_i\right)/(n-a) = b\boldsymbol{\Psi}\left(\sum_{i=1}^{a}n_i - 1\right)/(n-a) = b\boldsymbol{\Psi}$. Therefore, for $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G$, the first term of (4.13) is $\mathbf{0}$.

To prove $\text{E}\left(\boldsymbol{H}^{(A)}/(a-1)\right) - \text{E}\left(\boldsymbol{G}^{(A)}/(n-a)\right) = \mathbf{0}$ if and only if $\mathcal{H}_0^{(A)}$ holds, it remains to show that for $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G$ the second term of (4.13) is $\mathbf{0}$ if and only if $\mathcal{H}_0^{(A)}$ holds. Clearly, the elements $a_{kk'ii'}$ of $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G$ are not all zero. Observe that under $\mathcal{H}_0^{(A)}$, we can write $\boldsymbol{\mu} + \boldsymbol{\alpha}_i$ as $\boldsymbol{\mu} + \boldsymbol{\alpha}_1$ for all i. The second term of (4.13) can be written as

$$\sum_{i,i'=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_{i'}} a_{kk'ii'}\overline{\boldsymbol{\mu}}_{i\cdot}\overline{\boldsymbol{\mu}}'_{i'\cdot} = \sum_{i,i'=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_{i'}} a_{kk'ii'}\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_i\right)\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_i\right)'.$$

When $\mathcal{H}_0^{(A)}$ does not hold, this is not equal to $\mathbf{0}$ because the components after $a_{kk'ii'}$ cannot be pulled out in front of the summation and will add non-zero portions to the sum.

Now consider when $\mathcal{H}_0^{(A)}$ does hold. In that case, we can pull out the $\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_1\right)\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_1\right)'$ in front of the summation since all $\boldsymbol{\alpha}_i$ are equal under $\mathcal{H}_0^{(A)}$. Then the second term of (4.13) becomes

$$\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_1\right)\left(\boldsymbol{\mu}+\boldsymbol{\alpha}_1\right)'\sum_{i,i'=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_{i'}} a_{kk'ii'}.$$

Calculating this quantity reduces to summing all of the elements of $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G = \left[b\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)/(a-1)\right] - \left[b\left(\bigoplus_{i=1}^{a}\boldsymbol{P}_{n_i}\right)/(n-a)\right]$. The sum of the elements of $\frac{1}{n_i}\boldsymbol{J}_{n_i}$ is n_i, and the sum of the elements of $\frac{1}{n}\boldsymbol{J}_n$ is n. Also, the sum of the elements of $\boldsymbol{I}_{n_i}$ is n_i; therefore, the sum of the elements of $\boldsymbol{P}_{n_i}$ is 0. Considering each of the a^2 blocks, the overall sum of all of the elements of $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G$ would then become $\sum_{i=1}^{a} n_i - n + \sum_{i=1}^{a} 0 = n - n - 0 = 0$. Thus, for $\boldsymbol{A} = \boldsymbol{A}_H - \boldsymbol{A}_G$, the second term of (4.13) is

$$(\boldsymbol{\mu} + \boldsymbol{\alpha}_1)(\boldsymbol{\mu} + \boldsymbol{\alpha}_1)' \cdot 0 = 0.$$

This completes the proof for Case 2. $\diamond$ $\square$

Considering the conclusion of Proposition 4.3.2, we can compare the hypothesis SSCP matrices and the corresponding error SSCP matrices to obtain a meaningful test statistic. In traditional MANOVA, there are a multitude of test statistics from which to choose [28]. None of these statistics perform uniformly better than the others in the whole parameter space [36]. For this chapter, we consider the four most commonly used test statistics, namely the Dempster, Wilks' Lambda (likelihood ratio, or LR), Lawley-Hotelling (LH), and Bartlett-Nanda-Pillai (BNP) criteria. In what follows, we present the test statistics for the main effects and interaction in a unified manner. For $\phi \in \{B, \Gamma, A\}$, define

(a) Dempster's ANOVA Type criterion: $T_{\mathrm{D}}^{(\phi)} = \mathrm{tr}\left(\boldsymbol{H}^{(\phi)}\right) / \mathrm{tr}\left(\boldsymbol{G}^{(\phi)}\right)$,

(b) Wilks' Lambda (Likelihood Ratio) criterion: $T_{\mathrm{WL}}^{(\phi)} = -\log\left(\left|\boldsymbol{G}^{(\phi)}\right| / \left|\boldsymbol{H}^{(\phi)} + \boldsymbol{G}^{(\phi)}\right|\right)$,

(c) The Lawley-Hotelling criterion: $T_{\mathrm{LH}}^{(\phi)} = \mathrm{tr}\left(\boldsymbol{H}^{(\phi)}\left(\boldsymbol{G}^{(\phi)}\right)^{-1}\right)$ and

(d) The Bartlett-Nanda-Pillai criterion: $T_{\mathrm{BNP}}^{(\phi)} = \mathrm{tr}\left(\boldsymbol{H}^{(\phi)}\left(\boldsymbol{H}^{(\phi)} + \boldsymbol{G}^{(\phi)}\right)^{-1}\right)$.

Unlike the two-way cross classified design considered by Harrar and Bathke [26], the design under consideration in this chapter requires the use of different error SSCP matrices for testing the effects of factors A and B and the interaction between them.

4.4 Asymptotic Distributions

The asymptotic setup is that the number of levels of one of the factors, namely B, is large but the sample size and the number of levels of the other factor (A) remain fixed. Harrar and Bathke considered a similar asymptotic framework, but their results are applicable only when observations for different levels of factor B are independent [26]. In this asymptotic situation, because of the dependence across the levels of factor B, the within error SSCP matrix used for testing main effects of factor B and the interaction effects between factors A and B (labeled Γ) differ from that used for testing the main effects of factor A. Consequently, the results will be different and new. Furthermore, the derivations are more involved.

In the remainder of this section we obtain the asymptotic null distributions of the four test statistics for testing the main and interaction effects. To facilitate a succinct presentation of the results, we introduce the vectors of constants $\boldsymbol{\mu}_{ij}^{(\phi)}$, defined as $\boldsymbol{\mu}_{ij}^{(\phi)} = \boldsymbol{\beta}_j$, $\boldsymbol{\alpha}_i + \boldsymbol{\beta}_j$, or $\boldsymbol{\alpha}_i$ according as $\phi = B$, Γ, or A. Since the results for testing $\mathcal{H}_0^{(B)}$ and $\mathcal{H}_0^{(\Gamma)}$ are similar in form and their derivations proceed along the same lines, we group them under the same heading in the following subsection.

4.4.1 Testing for the Main Effect of Factor B and the Interaction Effect

The following theorem gives a probability limit for $\boldsymbol{G}^{(B)}$ as defined in (4.11).

Theorem 4.4.1. *Assume that a; n_i for $i = 1, \ldots, a$; and p are bounded as $b \to \infty$. Further suppose that* $\mathrm{E}\left(\left|X_{ijk}^{(r_1)} X_{ij'k}^{(r_2)} X_{ilk}^{(r_3)} X_{il'k}^{(r_4)}\right|\right) < \infty$ *for $j, j', l, l' = 1, \ldots, b$ and $r_1, r_2, r_3, r_4 = 1, \ldots, p$ and that $\boldsymbol{X}_{i1k}, \boldsymbol{X}_{i2k}, \cdots$ is an α-mixing sequence with $\alpha_{\boldsymbol{X}}(m) = O(m^{-5})$. Define*

$$\boldsymbol{\Sigma} = \frac{1}{ab} \sum_{i=1}^{a} \frac{1}{n_i} \sum_{j=1}^{b} \boldsymbol{\Sigma}_{jj},$$

and assume $\boldsymbol{\Sigma} = O(1)$ *as* $b \to \infty$*. Then*

$$\boldsymbol{G}^{(B)} - \boldsymbol{\Sigma} \xrightarrow{p} \boldsymbol{0}_{p\times p} \quad as \quad b \to \infty.$$

Remark: $\boldsymbol{\Sigma}$ defined above is different than $\boldsymbol{\Sigma}$ from Chapter 2.

Proof: Recall that $\boldsymbol{C}_G^{(B)} = \bigoplus_{i=1}^{a} \frac{1}{n_i(n_i-1)} \boldsymbol{P}_{n_i}$. Since $\frac{1}{b-1}\boldsymbol{P}_b = \frac{1}{b-1}\boldsymbol{I}_b - \frac{1}{b(b-1)}\boldsymbol{J}_b = \frac{1}{b}\boldsymbol{I}_b - \frac{1}{b(b-1)}\left(\boldsymbol{J}_b - \boldsymbol{I}_b\right)$, we can write

$$\boldsymbol{G}^{(B)} = \frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}' - \frac{1}{ab(b-1)}\boldsymbol{X}\left((\boldsymbol{J}_b - \boldsymbol{I}_b) \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}' \tag{4.14}$$

Since $\boldsymbol{J}_b - \boldsymbol{I}_b$ has zero diagonals, it follows from Proposition 4.2.1 that the second term of (4.14) is $o_p(b^{-1})$. Furthermore, since the sum of the diagonal elements of $\boldsymbol{P}_{n_i}$ is $n_i(1-1/n_i) = n_i - 1$, we again appeal to Proposition 4.2.1 to see that

$$\mathrm{E}\left[\frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}'\right] = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\boldsymbol{\Sigma}_{jj} = \boldsymbol{\Sigma}.$$

To complete the proof it remains to show that

$$\mathrm{Var}\left[\frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}'\right] \to \boldsymbol{0} \quad \text{as} \quad b \to \infty.$$

Observe that

$$\mathrm{Var}\left[\frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \left(\bigoplus_{i=1}^{a}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}\right)\right)\boldsymbol{X}'\right] = \frac{1}{(ab)^2}\sum_{i=1}^{a}\frac{1}{n_i^2}\sum_{j=1}^{b}\mathrm{Var}\left(\boldsymbol{X}_{ij}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij}\right)$$
$$+ \frac{1}{(ab)^2}\sum_{i=1}^{a}\frac{1}{n_i^2}\sum_{j\neq j'}^{b}\mathrm{Cov}\left(\boldsymbol{X}_{ij}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij}, \boldsymbol{X}_{ij'}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij'}\right).$$

It is obvious that the first term is of order $O(b^{-1})$. Now we can see that

$$\operatorname{Cov}\left(\boldsymbol{X}_{ij}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij},\boldsymbol{X}_{ij'}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij'}\right)=O\left(|j-j'|^{-5/2}\right), \tag{4.15}$$

which follows by noting that, for each fixed i, the random sequence of matrices $\{\boldsymbol{X}_{ij}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij}\}_{j=1}^{\infty}$ is an α-mixing sequence with dependence coefficient $\alpha_{\boldsymbol{X}}(m)=O(m^{-5})$ (see, for example, Theorem 5.2 of Bradley, p.126 [12]) and that the covariance decays at the rate $O(m^{-5/2})$ (see, for example, Lemma 3 of Billingsley, p.377 [8]). Summing (4.15) over the indices from above, we see that

$$\begin{aligned}\sum_{j\neq j'}^{b}\operatorname{Cov}\left(\boldsymbol{X}_{ij}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij},\boldsymbol{X}_{ij'}\frac{1}{n_i-1}\boldsymbol{P}_{n_i}\boldsymbol{X}'_{ij'}\right) &= \sum_{j\neq j'}^{b}O\left(|j-j'|^{-5/2}\right)\\ &= \sum_{t=1}^{b}2(b-t)O\left(t^{-5/2}\right)\\ &= O(b).\end{aligned}$$

Thus, the second term is also of order $O(b^{-1})$ and the result is proved. □

We know from Theorem 4.4.1 that, under two technical assumptions, $\boldsymbol{G}^{(\phi)}-\boldsymbol{\Sigma}=o_p(1)$ as $b\to\infty$, and it is established in Theorem 4.4.7 below that $\sqrt{b}\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Omega}=O_p(1)$ for $\phi\in\{B,\Gamma\}$ as $b\to\infty$ and for any matrix of constants $\boldsymbol{\Omega}$. In view of these, the expansions

$$\begin{aligned}T_{\mathrm{D}}^{(\phi)} &= 1+\frac{1}{\sqrt{b}}\left[\sqrt{b}\operatorname{tr}\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\cdot\frac{1}{\operatorname{tr}(\boldsymbol{\Sigma})}\right]+o_p(b^{-1/2}),\\ T_{\mathrm{WL}}^{(\phi)} &= \log\left|\boldsymbol{I}_p+\boldsymbol{H}^{(\phi)}\boldsymbol{G}^{(\phi)-1}\right|\\ &= p\log 2+\frac{1}{2\sqrt{b}}\left[\sqrt{b}\operatorname{tr}\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Sigma}^{-1}\right]+o_p(b^{-1/2}),\\ T_{\mathrm{LH}}^{(\phi)} &= p+\frac{1}{\sqrt{b}}\left[\sqrt{b}\operatorname{tr}\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Sigma}^{-1}\right]+o_p(b^{-1/2}),\quad\text{and}\\ T_{\mathrm{BNP}}^{(\phi)} &= \operatorname{tr}\left(\boldsymbol{H}^{(\phi)}\boldsymbol{G}^{(\phi)-1}\left(\boldsymbol{I}_p+\boldsymbol{H}^{(\phi)}\boldsymbol{G}^{(\phi)-1}\right)^{-1}\right)\\ &= 2p+\frac{1}{\sqrt{b}}\left[\sqrt{b}\operatorname{tr}\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Sigma}^{-1}\right]+o_p(b^{-1/2})\end{aligned} \tag{4.16}$$

can be easily verified [21]. According to these expansions, one can see that all four test statistics, scaled and centered suitably, can be expressed as

$$\sqrt{b}\ \left(\ell T_{\mathcal{G}}^{(\phi)} - q\right) = \sqrt{b}\ \mathrm{tr}\left[\left(\boldsymbol{H}^{(\phi)} - \boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Omega}\right] + o_p(1), \tag{4.17}$$

where $\ell = 1, 2, 1, 4$; $q = 1, 2p\log 2, p, 2p$; and $\mathcal{G} = \mathrm{D}, \mathrm{WL}, \mathrm{LH}, \mathrm{BNP}$, respectively; and $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\Sigma}^{-1}, [\mathrm{tr}\,(\boldsymbol{\Sigma})]^{-1}\boldsymbol{I}_p$, respectively [26]. In light of the expression (4.17), the null distributions of the four test statistics can be derived in a unified manner by obtaining the null distribution of $\sqrt{b}\ \mathrm{tr}\left[\left(\boldsymbol{H}^{(\phi)} - \boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Omega}\right]$ for any fixed matrix $\boldsymbol{\Omega}$. The null distribution of the latter quantity is given in Theorem 4.4.7.

Before presenting the main theorem of the section, we begin by listing a few assumptions. These are akin to the univariate assumptions needed for Theorem 2.4.11, save that we no longer need stationarity (Assumption 2.4.2) but we do have to make an assumption regarding the estimation of the second moment. Recall the notation $\mathrm{Cov}\left(\boldsymbol{X}_{ijk}, \boldsymbol{X}_{ij'k}\right) = \boldsymbol{\Sigma}_{jj'}$.

Assumption 4.4.2. *For each j, assume that the random vectors $\boldsymbol{X}_{ijk}$ are independently distributed with $\mathrm{E}\,(\boldsymbol{X}_{ijk}) = \boldsymbol{\mu}_{ij}^{(\phi)}$ and $\mathrm{Var}\,(\boldsymbol{X}_{ijk}) = \boldsymbol{\Sigma}_{jj}$ for $i = 1, \ldots, a$ and $k = 1, \ldots, n_i$.*

Assumption 4.4.3. $\left[\mathrm{tr}\left(\sum_{j,j'=1}^{b}\boldsymbol{\Sigma}_{jj'}\boldsymbol{\Omega}\right)\right]^2 = o(b^3)$ *as $b \to \infty$.*

Assumption 4.4.4. *For each (i,k), the sequence of random vectors $\boldsymbol{X}_{i1k}, \boldsymbol{X}_{i2k}, \cdots$ is an α-mixing sequence with $\alpha_X(m) = O(m^{-\lambda})$ with $\lambda > r/(r-1)$ for r as in Assumption 4.4.5.*

Assumption 4.4.5. $\mathrm{E}\left(\|\boldsymbol{\Sigma}_{jj}^{-1/2}(\boldsymbol{X}_{ijk} - \boldsymbol{\mu}_{ij}^{(\phi)})\|^{4r}\right) \le M < \infty$ *for some $r > 1$.*

Assumption 4.4.6. *There exists $\eta^2(\boldsymbol{\Omega}) \in (0,\infty)$ such that $b^{-1}\sum_{j,j'=m+1}^{m+b}\mathrm{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Sigma}_{jj'}\right)^2 - \eta^2(\boldsymbol{\Omega}) \to 0$ as $b \to \infty$, uniformly in m.*

Assumption 4.4.2 is analogous to Assumption 2.2.1 and simply supposes independence between subjects. Assumption 4.4.3 is analogous to Assumption 2.2.6 and is necessary for the convergence of the distribution; essentially, the variability must decrease sufficiently fast as the lag increases so the second term in the difference of the SSCP matrices will converge to zero.

Assumption 4.4.4 is analogous to Assumption 2.4.7 and supposes that the dependence among the data decays sufficiently quickly. Assumption 4.4.5 is roughly analogous to Assumption 2.4.10 and is simply a moment condition ensuring the existence of the expectations used in the proof of Theorem 4.4.7. Assumption 4.4.6 replaces the univariate assumption of stationarity. In the univariate case, stationarity was needed to find a consistent estimator of the variance of the test statistic. In the multivariate case, Assumption 4.4.6 drives at the same cause, and is a rather technical assumption needed to appeal to the Theorem from White and Domowitz [46]. In light of these assumptions, we can now present Theorem 4.4.7, the main result of this chapter.

Theorem 4.4.7. *Suppose that Assumptions 4.4.2, 4.4.5, 4.4.6, 4.4.4, and 4.4.3 hold. Then*

$$\sqrt{\frac{b}{\tau_\phi^2(\boldsymbol{\Omega})}}\,\mathrm{tr}\left[\left(\boldsymbol{H}^{(\phi)} - \boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Omega}\right] \xrightarrow{D} \mathcal{Z} \quad \textit{as} \quad b \to \infty,$$

where $\mathcal{Z} \sim N(0,1)$ *and*

$$\tau_\phi^2(\boldsymbol{\Omega}) = \frac{2\eta^2(\boldsymbol{\Omega})}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + c(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right), \quad c(B) = 1, \;\; \textit{and} \;\; c(\Gamma) = (a-1)^{-2}.$$

Proof: Recall the convenient re-expressions of $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$ from (4.12). Similar to the method in (4.14) in the proof of Theorem 4.4.1, we can decompose the difference between $\boldsymbol{H}^{(\phi)}$ and $\boldsymbol{G}^{(\phi)}$ as

$$\boldsymbol{H}^{(\phi)} - \boldsymbol{G}^{(\phi)} = \boldsymbol{X}\left[\frac{1}{b}\boldsymbol{I}_b \otimes \boldsymbol{C}^{(\phi)}\right]\boldsymbol{X}' - \boldsymbol{X}\left[\frac{1}{b(b-1)}\left(\boldsymbol{J}_b - \boldsymbol{I}_b\right)\otimes \boldsymbol{C}^{(\phi)}\right]\boldsymbol{X}', \tag{4.18}$$

where

$$\boldsymbol{C}^{(\phi)} = \begin{cases} \frac{1}{a}\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{1}_{n_i}\right)\boldsymbol{J}_a\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{1}'_{n_i}\right) - \frac{1}{a}\bigoplus_{i=1}^{a}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}, & \text{if } \phi = B \\ \frac{1}{a}\left(\bigoplus_{i=1}^{a-1}\frac{1}{n_i}\boldsymbol{1}_{n_i}\right)\boldsymbol{P}_a\left(\bigoplus_{i=1}^{a}\frac{1}{n_i}\boldsymbol{1}'_{n_i}\right) - \frac{1}{a}\bigoplus_{i=1}^{a}\frac{1}{n_i(n_i-1)}\boldsymbol{P}_{n_i}, & \text{if } \phi = \Gamma \end{cases}.$$

Then,

$$\begin{aligned}\sqrt{b}\,\mathrm{tr}\left[\left(\boldsymbol{H}^{(\phi)}-\boldsymbol{G}^{(\phi)}\right)\boldsymbol{\Omega}\right] &= \sqrt{b}\left\{\mathrm{tr}\left[\boldsymbol{X}\left(\boldsymbol{I}_b\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right]-\mathrm{tr}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right]\right\}\\ &= \frac{1}{\sqrt{b}}\sum_{j=1}^{b}Z_j^{(\phi)}-\frac{1}{\sqrt{b}(b-1)}\mathrm{tr}\left[(\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right],\end{aligned} \tag{4.19}$$

where $Z_j^{(\phi)}=\mathrm{tr}\left(\boldsymbol{X}_j\boldsymbol{C}^{(\phi)}\boldsymbol{X}_j'\boldsymbol{\Omega}\right)$.

Using Proposition 4.2, observe that

$$\begin{aligned}\mathrm{E}\left(\mathrm{tr}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right]\right) &= \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right]\right)\\ &= \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\right]\boldsymbol{\Omega}\right)\\ &= \mathrm{tr}\left(\boldsymbol{0}\cdot\boldsymbol{\Omega}\right)\\ &= 0\end{aligned}$$

and that

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{\sqrt{b}(b-1)}\mathrm{tr}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\boldsymbol{\Omega}\right]\right)\\ &\quad= \frac{1}{b(b-1)^2}\mathrm{Var}\left(\mathrm{vec}\left(\boldsymbol{\Omega}\right)'\mathrm{vec}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\right]\right)\\ &\quad= \frac{1}{b(b-1)^2}\mathrm{vec}\left(\boldsymbol{\Omega}\right)'\mathrm{Var}\left[\boldsymbol{X}\left((\boldsymbol{J}_b-\boldsymbol{I}_b)\otimes\boldsymbol{C}^{(\phi)}\right)\boldsymbol{X}'\right]\mathrm{vec}\left(\boldsymbol{\Omega}\right)\\ &\quad= \frac{1}{b(b-1)^2}\mathrm{vec}\left(\boldsymbol{\Omega}\right)'\\ &\qquad\times\left[\sum_{j,j',l,l'=1}^{b}\frac{2}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}+c(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_in_{i'}}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{jl}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l'}\right)'\right]\mathrm{vec}\left(\boldsymbol{\Omega}\right)\\ &\quad= \frac{1}{b(b-1)^2}\frac{2}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}+c(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_in_{i'}}\right)\\ &\qquad\times\left[\sum_{j,l=1}^{b}\mathrm{vec}\left(\boldsymbol{\Omega}\right)'\mathrm{vec}\left(\boldsymbol{\Sigma}_{jl}\right)\sum_{j',l'=1}^{b}\mathrm{vec}\left(\boldsymbol{\Sigma}_{j'l'}\right)'\mathrm{vec}\left(\boldsymbol{\Omega}\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{b(b-1)^2}\frac{2}{a^2}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + c(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right)\left[\sum_{j,l=1}^{b}\operatorname{tr}\left(\boldsymbol{\Sigma}_{jl}\boldsymbol{\Omega}\right)\right]^2 \\
&= o(b^{-4})o(b^3) \to 0 \quad \text{as} \quad b\to\infty,
\end{aligned}
$$

where the last equality is justified by Assumption 4.4.3.

Similarly, using Proposition 4.2 again, we see that

$$
\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{b}}\sum_{j=m+1}^{m+b} Z_j^{\phi}\right) &= \frac{2}{a^2 b}\left(\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)} + c(\phi)\sum_{i\neq i'}^{a}\frac{1}{n_i n_{i'}}\right)\left[\sum_{j,j'=1}^{b}\operatorname{tr}\left(\boldsymbol{\Sigma}_{jj'}\boldsymbol{\Omega}\right)\right]^2 \\
&\to \tau_{\phi}^2(\boldsymbol{\Omega}) \quad \text{as} \quad b\to\infty,
\end{aligned}
$$

where the convergence is due to Assumption 4.4.6.

Now, the proof will be complete by appealing to Theorem 4.2.3 for the first term of (4.19) if we show that assumptions (conditions) (i) and (ii) of the that theorem are satisfied. Condition (i) holds by Assumption 4.4.6. For condition (ii), note that $\mathrm{E}\left(\left|Z_j^{(\phi)}\right|^{2r}\right)$ is uniformly bounded by Assumption 4.4.5 because

$$
\mathrm{E}\left(\left|Z_j^{(\phi)}\right|^{2r}\right) \leq K_r\left[\max_{i,k}\mathrm{E}\left(\|\boldsymbol{\Sigma}_{jj}^{-1/2}(\boldsymbol{X}_{ijk}-\boldsymbol{\mu}_{ij}^{(\phi)})\|^{4r}\right)\right]\left(\operatorname{Var}\left(Z_j^{(\phi)}\right)\right)^r,
$$

which follows from Lemma 3 in Appendix A of Harrar and Bathke [26], where K_r is a constant that depends only on r. □

Theorem 4.4.7 is essentially a multivariate CLT for dependent random vectors provided at least the fourth mixed moments are bounded and the dependence is not so strong as to impede the asymptotic framework. In other words, the dependence over time must decay fast enough, and the dependence due to responses must be mild enough to allow for a consistent estimate of the overall covariance in the responses across parameters.

As in the univariate case, if there is only one level of factor A, the test statistic, more specifically τ_{Γ}^2, is not defined. In this case, these is no need to test for interaction, yet we can still test for a main effect of factor B. This is the subject of the following discussion, which we get quasi gratis from Theorem 4.4.7. In practice, we can think of this as testing for level profiles over the levels of factor B (which is often time). The random vectors from each subject would remain about the same in the case of no main effects of factor B. For example, if climate measurements were taken over time but no grouping was present among the sites, the values of the response vector for each site would remain roughly the same over time if there were no main effects present.

Letting $a = 1$ and supposing Assumptions 4.4.2, 4.4.3, 4.4.4, 4.4.5, and 4.4.6 hold, we can define

$$\tau_B^2 = \frac{2\eta^2(\boldsymbol{\Omega})}{n(n-1)}.$$

Then

$$\sqrt{\frac{b}{\tau_B^2(\boldsymbol{\Omega})}}\,\mathrm{tr}\left[\left(\boldsymbol{H}^{(B)} - \boldsymbol{G}^{(B)}\right)\boldsymbol{\Omega}\right] \xrightarrow{D} N(0,1) \quad \text{as} \quad b \to \infty.$$

4.4.2 Estimating the Asymptotic Variance

Since $\boldsymbol{G}^{(\phi)} - \boldsymbol{\Sigma} \xrightarrow{p} \boldsymbol{0}_{p\times p}$ as $b \to \infty$ by Theorem 4.4.1, the null distributions of the four multivariate test statistics can be obtained from Theorem 4.4.7 by choosing $\boldsymbol{\Omega}$ suitably, as mentioned after (4.17). For instance, the asymptotic null distributions of $T_{\mathrm{D}}^{(\phi)}$ for $\phi \in \{B, \Gamma\}$ are obtained by setting $\boldsymbol{\Omega}$ to $[\mathrm{tr}\,(\boldsymbol{\Sigma})]^{-1}\boldsymbol{I}_p$. For the other three test statistics, $\boldsymbol{\Omega}$ needs to be set to $\boldsymbol{\Sigma}^{-1}$ to get the asymptotic null distributions. However, to apply the theorem in practice, one needs a consistent estimator of $\eta^2(\boldsymbol{\Omega})$. Based on the idea of banding the empirical covariance matrix [6], it seems reasonable to estimate the covariance by its empirical version. For application, this would give

$$\hat{\eta}^2(\boldsymbol{\Omega}) = \frac{1}{b}\left(\sum_{j=1}^{b}\sum_{j'=L(j,h)}^{U(j,h)} \mathrm{tr}\left(\boldsymbol{\Omega}\widehat{\boldsymbol{\Sigma}}_{jj'}\right)^2\right)$$

as an estimate of

$$\eta^2(\boldsymbol{\Omega}) = \frac{1}{b}\left(\sum_{j,j'=1}^{b} \operatorname{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Sigma}_{jj'}\right)^2\right),$$

for some $0 < h < 1$ where $\widehat{\boldsymbol{\Sigma}}_{jj'} = (n-a)^{-1}\sum_{i=1}^{a}(\boldsymbol{X}_{ijk} - \overline{\boldsymbol{X}}_{ij\cdot})(\boldsymbol{X}_{ij'k} - \overline{\boldsymbol{X}}_{ij'\cdot})'$, and the quantities $L(j,h)$ and $U(j,h)$ are the integer parts of $\max\{1, j-b^h\}$ and $\min\{b, j+b^h\}$, respectively. Consistency of a similar estimator for the univariate case under slightly stronger moment conditions has already been proved. For details, see Wang and Akritras [45].

4.4.3 Testing for the Main Effects of Factor A

It is shown in Theorem 4.4.12 that $\boldsymbol{H}^{(A)}$ and $\boldsymbol{G}^{(A)}$ are asymptotically independent and each has a central Wishart distribution. The theorem has the important implication that the asymptotic distributions of our test statistics reduce to the well known distributions in multivariate statistical theory. Similar technical assumptions must first be made, given here.

Assumption 4.4.8. *For each j, assume that the random vectors $\boldsymbol{X}_{ijk}$ are independently distributed with* $\mathrm{E}\left(\boldsymbol{X}_{ijk}\right) = \boldsymbol{\mu}_{ij}^{(A)}$ *and* $\mathrm{Var}\left(\boldsymbol{X}_{ijk}\right) = \boldsymbol{\Sigma}_{jj}$ *for $i = 1, \ldots, a$ and $k = 1, \ldots, n_i$.*

Assumption 4.4.9. *For each (i,k), the sequence of random vectors $\boldsymbol{X}_{i1k}, \boldsymbol{X}_{i2k}, \cdots$ is an α-mixing sequence with $\alpha(m) = O(m^{-\lambda})$ with $\lambda > r/(r-1)$ for r in Assumption 4.4.10.*

Assumption 4.4.10. $\mathrm{E}\left(\|\boldsymbol{X}_{ijk} - \boldsymbol{\mu}_{ij}^{(A)}\|^{2r}\right) \leq M < \infty$ *for some $r > 1$.*

Assumption 4.4.11. *There exists a positive definite matrix, $\boldsymbol{\Gamma}$, such that* $b^{-1}\left(\sum_{j,j'=m+1}^{m+b} \boldsymbol{\Sigma}_{jj'} - \boldsymbol{\Gamma}\right) \to 0$ *as $b \to \infty$, uniformly in m, where* $\boldsymbol{\Sigma}_{jj'} = \mathrm{Cov}\left(\boldsymbol{X}_{1j1}, \boldsymbol{X}_{1j'1}\right)$.

We can now state the theorem.

Theorem 4.4.12. *Suppose that Assumptions 4.4.8, 4.4.9 , 4.4.10, and 4.4.11 hold. Then*

$$b^{-1/2}\boldsymbol{H}^{(A)} \xrightarrow{D} W_p(a-1, \boldsymbol{\Gamma}) \quad \textit{and} \quad b^{-1/2}\boldsymbol{G}^{(A)} \xrightarrow{D} W_p(n-a, \boldsymbol{\Gamma}) \quad \textit{as} \quad b \to \infty,$$

where $W_p(q, \boldsymbol{\Psi})$ stands for $p \times p$ Wishart distribution with degrees of freedom q and expected value $q\boldsymbol{\Psi}$. Moreover, $\boldsymbol{H}^{(A)}$ and $\boldsymbol{G}^{(A)}$ are asymptotically independent as $b \to \infty$.

Proof: Recall the definition of $\widetilde{\boldsymbol{X}}$ from (4.10) and that we can write $\boldsymbol{H}^{(A)} = \left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)\left(\bigoplus_{i=1}^{a} \frac{1}{n_i}\boldsymbol{J}_{n_i} - \frac{1}{n}\boldsymbol{J}_n\right)\left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)'$ and $\boldsymbol{G}^{(A)} = \left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)(\bigoplus_{i=1}^{a} \boldsymbol{P}_{n_i})\left(\sqrt{b}\widetilde{\boldsymbol{X}}\right)'$. It is clear that the columns of $\widetilde{\boldsymbol{X}}$ are mutually independent and, under Assumptions 4.4.10, 4.4.11, and 4.4.9, we have $b^{-1/2}\overline{\boldsymbol{X}}_{i.k} \sim MVN_{p,n}(0, \boldsymbol{I}_n, \boldsymbol{\Gamma})$ by the multivariate version of Theorem 4.2.3, where, for a $p \times n$ random matrix, the notation $\boldsymbol{X} \sim MVN_{p,n}(\boldsymbol{M}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$ means $\text{vec}\,(\boldsymbol{X})$ has an np-variate normal distribution with mean vector $\text{vec}\,(\boldsymbol{M})$ and covariance matrix $\boldsymbol{\Xi} \otimes \boldsymbol{\Sigma}$ (see, for example, Gupta and Nagar [23]). That $\boldsymbol{H}^{(A)}$ and $\boldsymbol{G}^{(A)}$ are asymptotically independent Wishart matrices follows by the matrix variate versions of Cochran's Theorem (see Gupta and Nagar [23], Fujikoshi et al. [21]). □

Given the results of Theorem 4.4.12, the asymptotic null distributions of the $T_{\text{D}}^{(A)}, T_{\text{WL}}^{(A)}, T_{\text{LH}}^{(A)}$ and $T_{BNP}^{(A)}$ statistics do not have closed forms except in some special cases [2]. Good approximations based on F-distributions were proposed by Dempster [17], Rao [35], McKoen [31], and Muller [33] for the Dempster's, Wilks'-Lamda, Lawley-Hotelling's and Bartlett-Nanda-Pillai's statistics, respectively (see also Bathke et al. [5]). It needs to be pointed out that the approximation for Dempster's criterion involves estimating the degrees of freedom from the sample, and Wilks' Lambda criteron is defined without a "$-\log$" in front of it.

4.5 Simulation Study

This section presents simulated size results for the asymptotic likelihood ratio test statistics. For an assessment of the quality of the asymptotic distributions of the test statistics in Theorems 4.4.7 and 4.4.12, we conducted a simulation study by generating data from multiple distributions with various covariance structures due to factor B and sample sizes. In each of these, the dimensions $p = 2$ and $p = 3$, i.e., the number of response variables, were explored. For each set of criteria, 1,000 simulations were run. The discussion below is restricted to the Wilks' Lambda criterion; this likelihood ratio test is very common in multivariate analysis, and the scale of the simulation and discussion would be too lengthy for the scope of this section.

4.5.1 Setup

The two distributions used to generate data for the simulation were the multivariate standard normal distribution (labeled P_1 in the simulation); and the multivariate skew-normal distribution with location vector $\mathbf{0}$, covariance matrix $\boldsymbol{I}$, and skewing matrix $\boldsymbol{I}$ (P_2). The level of skewness is greater for the multivariate skew-normal distribution. We assume that $\boldsymbol{\Sigma}_{jj'} = \boldsymbol{\Psi}_{jj'}\boldsymbol{\Sigma}$; i.e., the Kronecker product covariance structure $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ is assumed for each subject. Since the null distributions of the test statistics are invariant under the transformation $\boldsymbol{X} \to \boldsymbol{\Sigma X}$, we can assume WLOG that $\boldsymbol{\Sigma} = \boldsymbol{I}$. The three covariance matrices due to factor B ($b \times b$ in dimension) used will be labeled Σ_1, Σ_2, and Σ_3 in the simulation. The first follows an ARMA(2,2) process, which decays exponentially. The second follows polynomial decay, and the $(i, j)^{th}$ element was given by $\rho|i - j|^{-5/2}$, where $\rho = 0.55$. The third was given by $1.5\boldsymbol{I}$. Note that this is identical to those used in Chapter 3, the difference here being that the covariance for the matrix $\boldsymbol{X}$ in Section 4.1 also accounts for the covariance due to the p response variables. Since it can be shown that the likelihood ratio test statistic is invariant to the covariance matrix due to the response variables, we use the identity matrix; thus, we have $\text{Cov}\,(\boldsymbol{X}) = \boldsymbol{\Psi}_{b\times b} \otimes \boldsymbol{I}_n \otimes \boldsymbol{I}_p$.

Furthermore, two group structures were used, labeled N_1 and N_2. The first structure included four groups, or $a = 4$, where the numbers of subjects in each group were $n_1 = 4$, $n_2 = 5$, $n_3 = 6$, and $n_4 = 7$. The second structure included three groups, or $a = 3$, where the number of subjects in each group were $n_1 = 10$, $n_2 = 12$, and $n_3 = 14$. This is again identical to Chapter 3.

For each combination of distribution, covariance structure, and group structure, 1000 simulations were run for various values of b. The values of b were 10, 20, 50, and 100, the limitations due to lengthy computing times. In each simulation the test statistics $T_{\mathrm{WL}}^{(B)}$, $T_{\mathrm{WL}}^{(\Gamma)}$, and $T_{\mathrm{WL}}^{(A)}$ were computed. Each time, the statistic was compared to the 0.05 critical value for its corresponding asymptotic distribution, and a decision to reject or fail to reject was made according as the statistic was beyond or within the corresponding critical value. The average number of rejections, which is the estimated actual (achieved) size of the test, is reported for each case in Table 4.1. The estimated sizes for the test based on $T_{\mathrm{WL}}^{(B)}$, which is the main of effect of B (often a time effect), is given in a column labeled B. The results of the test for the interaction effect, Γ, and the test for the main effect of A (often a group effect) are similarly labeled Γ and A.

4.5.2 Results

The test for the main effect of factor A seems to be performing fairly well regardless of the value of b. However, for smaller values of b the simulated sizes are a little more variable. For instance, the simulated sizes range from 0.031 to 0.059 for $b = 5$ and from 0.032 to 0.070 for $b = 10$, whereas they range from 0.027 to 0.049 for $b = 50$ and from 0.025 to 0.06 for $b = 100$. These sizes seem to be very close to 0.05 throughout, which is the desired size of the test, and there appears to be no strong pattern based on the other conditions of the simulation.

When looking at the simulated sizes for the effect of factor B and the interaction effect, both tests show similar patterns. For all values of b and p, and under both populations and all covariance structures, the simulated sizes under N_1 are less than under N_2. In general, these simulated sizes seem a bit erratic, but acceptably small. For $b = 10$, the simulated sizes range from 0.059 to 0.332 when $p = 2$ and from 0.097 to 0.395 when $p = 3$. For $b = 20$, the simulated sizes range from 0.034 to 0.282 when $p = 2$ and from 0.052 to 0.301 when $p = 3$. For $b = 50$, the simulated sizes range from 0.006 to 0.178 when $p = 2$ and from 0.020 to 0.217 when $p = 3$. For $b = 100$, the simulated sizes range from 0.001 to 0.098 when $p = 2$ and from

0.002 to 0.136 when $p = 3$. We can see here that simulated sizes are consistently higher for $p = 3$ than they are for $p = 2$.

Dim	Pop	Cov	Samp	$b = 10$			$b = 20$			$b = 50$			$b = 100$		
				B	Γ	A	B	Γ	A	B	Γ	A	B	Γ	A
$p = 2$	P_1	Σ_1	N_1	0.095	0.126	0.050	0.064	0.095	0.054	0.027	0.025	0.046	0.006	0.008	0.045
			N_2	0.331	0.332	0.035	0.239	0.282	0.048	0.166	0.178	0.040	0.097	0.088	0.038
		Σ_2	N_1	0.059	0.083	0.059	0.028	0.047	0.050	0.006	0.011	0.042	0.003	0.005	0.054
			N_2	0.271	0.271	0.046	0.181	0.199	0.036	0.090	0.088	0.041	0.056	0.054	0.043
		Σ_3	N_1	0.096	0.089	0.047	0.054	0.040	0.051	0.006	0.010	0.047	0.006	0.002	0.052
			N_2	0.276	0.308	0.037	0.203	0.172	0.040	0.104	0.094	0.049	0.051	0.046	0.048
	P_2	Σ_1	N_1	0.109	0.130	0.046	0.063	0.080	0.070	0.023	0.024	0.049	0.008	0.013	0.045
			N_2	0.331	0.314	0.042	0.237	0.251	0.032	0.130	0.160	0.042	0.098	0.097	0.048
		Σ_2	N_1	0.063	0.086	0.050	0.034	0.034	0.054	0.013	0.018	0.046	0.005	0.001	0.042
			N_2	0.247	0.269	0.048	0.199	0.185	0.043	0.114	0.085	0.042	0.066	0.036	0.045
		Σ_3	N_1	0.097	0.074	0.043	0.050	0.055	0.042	0.012	0.004	0.042	0.003	0.002	0.060
			N_2	0.300	0.277	0.049	0.182	0.207	0.051	0.076	0.108	0.041	0.055	0.053	0.041
Dim	Pop	Cov	Samp	$b = 10$			$b = 20$			$b = 50$			$b = 100$		
				B	Γ	A	B	Γ	A	B	Γ	A	B	Γ	A
$p = 3$	P_1	Σ_1	N_1	0.153	0.163	0.046	0.085	0.103	0.040	0.044	0.034	0.044	0.014	0.019	0.033
			N_2	0.395	0.377	0.038	0.300	0.299	0.031	0.217	0.211	0.048	0.136	0.140	0.034
		Σ_2	N_1	0.097	0.105	0.039	0.052	0.063	0.033	0.024	0.027	0.039	0.009	0.005	0.037
			N_2	0.326	0.323	0.031	0.231	0.239	0.050	0.138	0.160	0.047	0.067	0.079	0.039
		Σ_3	N_1	0.133	0.130	0.052	0.072	0.061	0.043	0.023	0.032	0.027	0.003	0.005	0.042
			N_2	0.350	0.335	0.029	0.252	0.265	0.038	0.142	0.144	0.035	0.081	0.085	0.029
	P_2	Σ_1	N_1	0.150	0.168	0.045	0.090	0.103	0.036	0.039	0.030	0.039	0.020	0.022	0.037
			N_2	0.372	0.385	0.041	0.301	0.292	0.037	0.191	0.202	0.045	0.131	0.146	0.043
		Σ_2	N_1	0.097	0.117	0.047	0.062	0.080	0.041	0.021	0.028	0.036	0.004	0.011	0.025
			N_2	0.313	0.341	0.036	0.233	0.242	0.045	0.141	0.152	0.032	0.069	0.089	0.036
		Σ_3	N_1	0.125	0.122	0.041	0.067	0.075	0.041	0.020	0.016	0.042	0.002	0.002	0.056
			N_2	0.346	0.368	0.042	0.226	0.263	0.041	0.137	0.130	0.042	0.073	0.073	0.044

Table 4.1: Multivariate simulated sizes for the (Wilks' Lambda Likelihood Ratio) tests with statistics $T_{\mathrm{WL}}^{(B)}$ (denoted by B), $T_{\mathrm{WL}}^{(\Gamma)}$ (denoted by Γ), and $T_{\mathrm{WL}}^{(A)}$ (denoted by A) when sampling from the multivariate normal (P_1) and multivariate skew-normal (P_2) distributions under the three covariance structures Σ_1, Σ_2 and Σ_3, under the two group structures N_1 and N_2, and for the dimensions $p = 2$ and $p = 3$. Results are for $b = 10$, $b = 20$, $b = 50$, and $b = 100$.

4.5.3 Discussion and Conclusion

In general, there does not seem to be a clear pattern differentiating the test for the effects of factor B and the test for the interaction effects. It does seem that the tests are somewhat more stable for $p = 3$ than for $p = 2$, as the simulated sizes do not drop as quickly. Most of the extremely low simulated sizes (≤ 0.01) are under the Σ_3 covariance structure, and this matter warrants further investigation.

As multivariate analysis usually requires even greater dimension than univariate analysis to perform adequately, it is possible that the asymptotics are not adequately portrayed with b only reaching 100. Also, some of the rather unusual values in Table 4.1 could possibly be due to the relatively small number of simulations (1000). Both of these restrictions are results of lengthy computing time. For larger values of b, each simulation can last up to three minutes, so repetition across all criteria and for 1000 simulations severely limits the scope of the simulation analysis. We give a strong recommendation for further research in this area. To that end, the other common statistics from (4.16) could also be explored. In general, however, the simulation study is promising for the Wilks' Lambda criterion, as the simulated sizes are relatively small and stabilizing as b increases; also, they seem to be moving toward a low value as b increases, possibly near the critical value of 0.05. As expected, the results for the main effect of factor A are more pleasing since the test statistics do not directly depend on the value of b, rather vectors are first average over each value b, and then the usual multivariate theory is applied.

4.6 Reductions under Specific Covariance Structures

This final section will present some basic corollaries to some of the results when certain covariance structures are assumed. Certain instances arise in which making certain assumptions regarding the covariance structure of the data become beneficial or even intuitive. We consider two such cases. In both cases, the number of parameters is greatly reduced. For instance, given that each subject is independent, the covariance has $\frac{bp(bp+1)}{2}$ unknown parameters, which, when b tends to infinity, is a great number of parameters. Consider the following covariance reductions:

$$(1) \qquad \operatorname{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} \times \boldsymbol{I}_n \times \boldsymbol{\Psi}, \quad \text{and}$$

$$(2) \qquad \operatorname{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} \times \boldsymbol{I}_n \times \boldsymbol{I}_p.$$

In both cases, $\boldsymbol{\Sigma}$ is $p \times p$ as defined before, and $\boldsymbol{\Psi}$ is $b \times b$. We see that reduction (1) only has $\dfrac{b(b+1)}{2} + \dfrac{p(p+1)}{2}$ unknown parameters and reduction (2) only has $\dfrac{b(b+1)}{2}$ unknown parameters, both of which are substantially less than $\dfrac{bp(bp+1)}{2}$.

In general, we can think of $\boldsymbol{\Sigma}$ as the covariance due to time or the repeated measure. We can think of $\boldsymbol{\Psi}$ as the covariance due to the response variables. Consider reduction (1) and recall that the responses from the subjects are assumed to be independent. Then, for the k^{th} subject in the i^{th} level of factor A (the i^{th} group), we would have

$$\text{Cov}\left(\begin{bmatrix} \boldsymbol{X}'_{i1k} & \cdots & \boldsymbol{X}'_{ibk} \end{bmatrix}'\right) = \boldsymbol{\Sigma} \times \boldsymbol{\Psi}.$$

Still considering only one subject, we can reduce $\boldsymbol{X}$ to $\boldsymbol{X}^*$ and write

$$\boldsymbol{X}^* = \begin{bmatrix} \boldsymbol{X}_{(1)} & \cdots & \boldsymbol{X}_{(b)} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{X}^{(1)})' \\ \vdots \\ (\boldsymbol{X}^{(p)})' \end{bmatrix},$$

where $\boldsymbol{X}_{(j)}$ is the vector of responses from the j^{th} time point over the p variables and $\boldsymbol{X}^{(h)}$ is the vector of responses for the h^{th} variable over the b time points. Then we see that $\text{Cov}\left(\boldsymbol{X}_{(j)}, \boldsymbol{X}_{(j')}\right) = \sigma_{jj'}\boldsymbol{\Psi}$ and $\text{Cov}\left(\boldsymbol{X}^{(h)}, \boldsymbol{X}^{(h')}\right) = \psi_{hh'}\boldsymbol{\Sigma}$. More practically, the two sources of covariance are separated under reduction (1), which may be reasonable due to the specific situation. Note that reduction (2) is merely a further simplification of reduction (1), so its detail is not provided. The intuition behind reduction (2) is that there is no covariance between the response variables. This assumption may provide nice simplification, but it will often be unreasonable.

In light of these reductions, we can simplify some of the earlier results. (4.1) from Proposition 4.2.1 would reduce to

$$(1) \qquad \text{E}\left(\boldsymbol{XAX}'\right) = \boldsymbol{\Sigma} \sum_{j,j'=1}^{b} \sum_{i=1}^{n} a_{jj'ii}\psi_{jj'} + \boldsymbol{MAM}', \quad \text{and}$$

$$(2) \qquad \mathrm{E}\left(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\right) = \boldsymbol{I}_p \sum_{j,j'=1}^{b} \sum_{i=1}^{n} a_{jj'ii}\psi_{jj'} + \boldsymbol{M}\boldsymbol{A}\boldsymbol{M}'.$$

Moreover, (4.2) from Proposition 4.2.1 would reduce to

$$\begin{aligned}(1) \qquad & \mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\right)\mathrm{vec}\left(\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}'\right)'\right] \\ & = \boldsymbol{\Sigma} \sum_{j,j',l,l'}^{b} \sum_{i=1}^{n} \mathrm{E}\left[\mathrm{vec}\left(\boldsymbol{X}_{ij}\boldsymbol{X}'_{ij'}\right)\mathrm{vec}\left(\boldsymbol{X}_{il}\boldsymbol{X}'_{il'}\right)'\right] a_{jj'ii}b_{ll'ii} \\ & + \mathrm{vec}\left(\boldsymbol{\Sigma}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}\right)' \sum_{j,j',l,l'}^{b} \left(\psi_{jj'}\psi_{ll'} \sum_{i \neq i'}^{n} a_{jj'ii}b_{ll'i'i'} \right. \\ & \left. + \psi_{jl}\psi_{j'l'} \sum_{i \neq i'}^{n} a_{jj'ii}b_{ll'i'i'} + \psi_{jl'}\psi_{j'l} \sum_{i \neq i'}^{n} a_{jj'ii}b_{ll'i'i'} \right).\end{aligned}$$

In the case of reduction (2), $\mathrm{vec}\left(\boldsymbol{I}_p\right)\mathrm{vec}\left(\boldsymbol{I}_p\right)'$ simply replaces $\mathrm{vec}\left(\boldsymbol{\Sigma}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}\right)'$.

To find an estimate of the probability limit for $\boldsymbol{G}^{(B)}$ as in Theorem 4.4.1, the reductions would yield the following simplifications:

$$(1) \qquad \mathrm{E}\left[\frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}'\right] = \frac{1}{ab}\mathrm{tr}\left(\boldsymbol{\Psi}\right)\boldsymbol{\Sigma}\sum_{i=1}^{a}\frac{1}{n_i}, \quad \text{and}$$

$$(2) \qquad \mathrm{E}\left[\frac{1}{ab}\boldsymbol{X}\left(\boldsymbol{I}_b \otimes \boldsymbol{C}_G^{(B)}\right)\boldsymbol{X}'\right] = \frac{1}{ab}\mathrm{tr}\left(\boldsymbol{\Psi}\right)\boldsymbol{I}_p\sum_{i=1}^{a}\frac{1}{n_i}.$$

Note that the $\boldsymbol{\Sigma}$ notation used in the case of reduction (1) has replaced the earlier notation in the estimator.

Chapter 5

Application and Discussion

5.1 Introduction

This chapter serves to provide some practical examples, a summary overview, conclusions, and a discussion of further research needed. Section 5.2 will give a detailed example from the univariate framework as well as a brief example from the multivariate framework. Section 5.3 will include a broad overview of the work of the dissertation and provide ideas for further areas of research.

5.2 Practical Examples

Let us first consider an example from the univariate case from Chapter 2. Han et al. [24] collected data from three groups of subjects in a study exploring human proteins and Parkinson's disease. 29 subjects in the study had been diagnosed with Parkinson's disease (Diseased), and two control groups without the disease were considered. One control group was comprised of 20 younger subjects (Control Young) and the other was comprised of 20 older subjects

(Control Old); when only one control group is considered (Control), the young control group, there are 20 subjects total. From each subject, a protein microarray, or protoarray, was collected containing 9,480 relative fluorescent unit (RFU) signals indicating the presence and abundance of human proteins. When considering all three groups, this study yields $a = 3$, $n_1 = 29$, $n_2 = 20$, $n_3 = 20$, and $b = 9480$; examining only the control groups to see if there is any significant difference between control groups, this study yields $a = 2$, $n_1 = 20$, $n_2 = 20$, and $b = 9480$.

In the first case, a group effect would indicate that the levels of the proteins are different between the diseased and control group. A main effect of factor B would indicate the protein levels differ across the microarray. An interaction effect would indicate that the level of the proteins (or change therein) depends on the group, or presence of Parkinson's disease. Detection of a significant interaction effect is the first step in the goal of identifying biomarkers, or the proteins which may indicate the presence/absence of the disease. If there are no differences between the diseased and control groups for different proteins, there is little justification to put in extra expense to look for specific differences. When considering only the two control groups, we are primarily interested in the group effects. No effect would indicate that the age of the control group is irrelevant.

First, let us look at the following figures. Figure 5.1 gives average RFU signal response values for the group diagnosed with Parkinson's disease as well as the two control groups, whereas Figure 5.2 plots only the two control groups. Average windows were taken across the 9,480 signals such that 30 values are reported (for ease of visualization). Figure 5.2 seems to indicate that there is little difference in the RFU signals between the two control groups. Figure 5.1 seems to show that there may be a difference in the three groups as the responses for the diseased group seem greater than the other two.

The test statistics from Theorems 2.4.11 and 2.4.15 are calculated and given in Tables 5.1 and 5.2. Based on the corresponding asymptotic distributions, the p-values are also given. Table 5.1 considers the case when all three groups were analyzed, whereas Table 5.2 compares

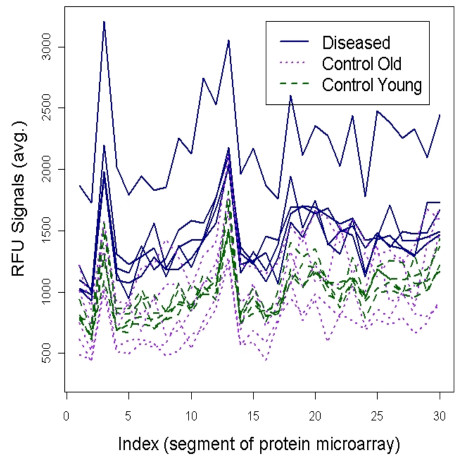

Figure 5.1: *Parkinson's: Disease vs. Controls*: RFU signal (protein) averages across the 9,480 responses, broken into 30 averages for visualization. Values for five random subjects for each of the two groups are plotted and the diseased group are plotted.

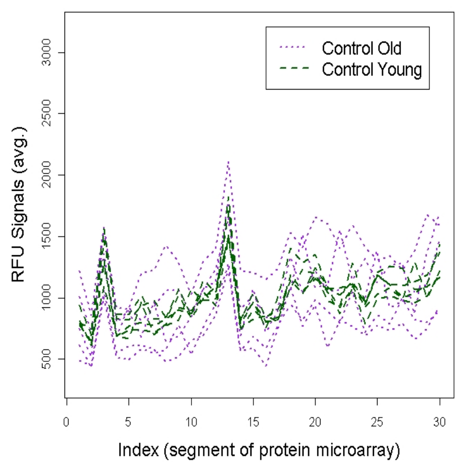

Figure 5.2: *Parkinson's: Control vs. Control*: RFU signal (protein) averages across the 9,480 responses, broken into 30 averages for visualization. Values for five random subjects for each of the two control groups are plotted.

only the two control groups. We can see from the p-value of 0.1588 that there is no significant difference in the control groups. Based on the p-value of 0.0001, we see that there is a significant group effect when the diseased group is included. Both scenarios indicate that there are significant main effects for RFU signal response and the interaction between RFU signal and group. This comes as no surprise based on Figures 5.1 and 5.2. If there were no main effects for RFU signal response, we would expect the lines to be roughly flat. The interaction effect is less pronounced in both cases, but still very present. This may be difficult to judge in Figures 5.1 and 5.2 as they are plotting average RFU signals. At any rate, it appears that further study to attempt to identify biomarkers would be warranted.

	T^*_β	T^*_γ	F^*_α
Statistic	2637.36	48.48	10.69
p-value	0.0000	0.0000	0.0001

Table 5.1: *Parkinson's: Disease vs. Controls*: Statistics and p-values for the tests of RFU signal (protein) main effect (T^*_β), interaction effect (T^*_γ), and group effect (F^*_α) when considering the Diseased, Control Young, and Control Old groups.

	T^*_β	T^*_γ	F^*_α
Statistic	1698.39	7.25	2.07
p-value	0.0000	0.0000	0.1588

Table 5.2: *Parkinson's: Control vs. Control*: Statistics and p-values for the tests of RFU signal (protein) main effect (T^*_β), interaction effect (T^*_γ), and group effect (F^*_α) when considering only the Control Young and Control Old groups.

Now let us first consider briefly an example from the multivariate case. Holden et al. [27] have collected climate data from the state of Montana over many years. In the interest of this analysis, let us consider two response variables, the minimum and maximum temperature for a day (so $p = 2$). The data were collected for multiple sites across Montana, and the sites have been separated into two groups based on elevation (so $a = 2$). The lower elevation group consists of twelve sites (so $n_1 = 12$) and the higher elevation group consists of ten sites (so $n_2 = 10$). Data were collected over a period of 145 days (so $b = 145$). It is known that the maximum and minimum temperatures often seem not to differ as much in higher elevations (peaks) as they do in lower elevations (valleys). It is of interest to test for an effect of elevation (i.e., a group main effect), an effect of time (a main effect), and an interaction effect.

The test statistics for time and interaction effects were calculated via Theorem 4.4.7. The test statistic for a time effect was 21.7729 with a p-value of < 0.00005, and the test statistic for a group-time interaction effect was -1.3179 with a p-value of 0.1875. This indicates that there is a significant time effect, meaning that the maximum and minimum temperatures do not remain the same over time. However, there is no evidence to support that these changes are due to elevation. The usual multivariate test corresponding to Theorem 4.4.12 showed no significant group effect, indicating that sites at higher elevations do not show significant differences in maximum and minimum temperatures compared to those at lower elevations.

5.3 Discussion

Overall, the results in this work make progress in developing tests for dependent, repeated measures when the number of measurements is large. The lack of covariance restrictions and distributional assumptions make the tests widely applicable. That the asymptotic distributions are free from moments of the data generating distribution establishes robustness for the existing theory derived under the assumption of normality. Also, since much of the theory to date has been dependent on the number of subjects being large, the asymptotic theory suggested in which only the number of measurements must be large helps to incorporate the demands of recent data in this technological age. The methods using the Kronecker product, direct sum, and vec notation while writing the sums of squares as quadratic forms (or matrix quadratic forms) also serve to make the mathematics much more tractable. Furthermore, the simulation study for the univariate case indicates that the tests perform at least as well as the traditional methods, and they perform very well when b becomes very large.

Further areas of research include topics from the the following discussion. A more extensive simulation study would help to ascertain under what circumstances the test statistics are performing best. More variations in the setup could be considered, and the power simulation study could be carried out to a greater extent. Also, a simulation study for the other multivariate criteria could be conducted. Faster programming functions and software may aid in this endeavor as the simulations grow very extensive as the dimension increases. Many of the assumptions are somewhat technical and could hopefully be relaxed, especially those regarding moments. Also, methods to assess whether the necessary assumptions are met in any given situation could be developed. The assessment of the α-mixing assumption would be especially useful, not only in this research but in other areas as well. The consistent estimation of the variance of the test statistics warrants further research. Moreover, the bootstrapping method proposed to estimate such variances deserves to be revisited in greater depth. Finally, as robustness was one of the main goals of the tests, it would be very interesting to explore the reductions and simplifications that occur when a more specific structure is imposed. For instance, if the data did arise from a multivariate distribution with a known covariance structure, it may prove very useful to see if better results could be achieved with the same or fewer assumptions. These reductions would likely be problem-specific, yet useful nonetheless.

Bibliography

[1] M. Akritas and S. Arnold, *Asymptotics for Analysis of Variance When the Number of Levels is Large*, Journal of the American Statistical Association **95** (2000), no. 449, 212–226.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Inc., New York, 2003.

[3] D. W. K. Andrews, *The Block-Block Bootstrap: Improved Asymptotic Refinements*, Cowles Foundation Discussion **1370** (2002).

[4] A. C. Bathke and S. W. Harrar, *Nonparametric Methods in Multivariate Factorial Designs for Large Number of Factor Levels*, Journal of Statistical Planning and Inference **138** (2008), no. 3, 588–610.

[5] A. C. Bathke, S. W. Harrar, and L. V. Madden, *How to Compare Small Multivariate Samples Using Nonparametric Tests*, Computational Statistics and Data Analysis **52** (2008), 4951–4965.

[6] P. J. Bickel and E. Levina, *Regularized Estimation of Large Covariance Matrices*, The Annals of Statistics **36** (2008), no. 1, 199–227.

[7] P. Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1968.

[8] ______, *Probability and Measure*, Second, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1994.

[9] R. J. Boik, *The Mixed Model for Multivariate Repeated Measures: Validity Condition and an Approximate Test*, Psychometrika **53** (1988), no. 4, 469–486.

[10] D. D. Boos and C. Brownie, *ANOVA and Rank Tests when Number of Treatments is Large*, Statistics and Probability Letters **23** (1995), 183–191.

[11] R. C. Bradley, *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions*, Probability Surveys **2** (2005), 107–144.

[12] R. C. Bradley Jr., *Central Limit Theorems under Weak Dependence*, Journal of Multivariate Analysis **11** (1981), no. 1, 1–16.

[13] B. J. Broxson, *The Kronecker Product*, UNF Theses and Dissertations **25** (2006), 1–65.

[14] E. Brunner, A. C. Bathke, and M. Placzek, *Estimation of Box's ϵ for Low- and High-Dimensional Repeated Measures Designs with Unequal Covariance Matrices*, Biometrical Journal **00** (2012), no. 0, 1–10.

[15] H. Cramer and H. Wold, *Some Theorems on Distribution Functions*, Journal London Mathematical Society **11** (1936), 290–295.

[16] C. S. Davis, *Statistical Methods for the Analysis of Repeated Measurements*, Springer Texts in Statistics, Springer-Verlag New York, Inc., New York, 2002.

[17] A. P. Dempster, *A High Dimensional Two Sample Significance Test*, Annals of Mathematical Statistics **29** (1958), 995–1010.

[18] P. J. Diggle, P. Heagerty, K.-Y. Liang, and S. L. Zeger, *Analysis of Longitudinal Data*, Second, Oxford Statistical Science Series, Oxford University Press, Oxford, UK, 2002.

[19] E. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, Inc., New York, 1993.

[20] T. S. Ferguson, *A Course in Large Sample Theory*, First, Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton, 1996.

[21] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu, *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Inc., New Jersey, 2002.

[22] A. K. Gupta, S. W. Harrar, and Y. Fujikoshi, *Asymptotics for Testing Hypothesis in Some Multivariate Components Model Under Non-Normality*, Journal of Multivariate Analysis **97** (2006), no. 1, 148–178.

[23] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Monographs and Surveys in Pure and Applied Mathematics, Chapman & Hall/CRC, Boca Raton, 2000.

[24] M. Han, E. Nagele, C. DeMarshall, N. Acharya, and R. Nagele, *Diagnosis of Parkinson's Disease based on Disease-Specific Autoantibody Profiles in Human Sera.*, PLoS One **7** (2012), no. 2, e32383.

[25] W. Härdle, J. Horowitz, and J.-P. Kreiss, *Bootstrap Methods for Time Series*, International Statistical Review **71** (2003), no. 2, 435–459.

[26] S. W. Harrar and A. C. Bathke, *A Modified Two-Factor Multivariate Analysis of Variance: Asymptotics and Small Sample Approximations*, Annals of the Institute of Statistical Mathematics **64** (2012), 135–165.

[27] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, *Empirical Downscaling of Daily Minimum Air Temperature at Very Fine Resolutions in Complex Terrain*, Agricultural and Forest Meteorology (2011).

[28] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Sixth, Pearson Education, Inc., Upper Saddle River, New Jersey, 2007.

[29] M. Loève, *Probability Theory I*, Fourth, Graduate Texts in Mathematics, vol. 45, Springer-Verlag Inc., New Jersey, 1977.

[30] N. Lu and D. L. Zimmerman, *The Likelihood Ratio Test for a Separable Covariance Matrix*, Statistics and Probability Letters **73** (2005), no. 4, 449–457.

[31] J. J. McKeon, *F Approximations to the Distribution of Hotelling's T_0^2 Statistic*, Biometrika **61** (1974), no. 2, 381–383.

[32] M. W. Mitchell, M. G. Genton, and M. L. Gumpertz, *A Likelihood Ratio Test for Separability of Covariances*, Journal of Multivariate Analysis **97** (2006), no. 5, 1025–1043.

[33] K. E. Muller, *A New F Approximation for the Pillai-Bartlett Trace under H_0*, Journal of Computational and Graphical Statistics **7** (1998), no. 1, 131–137.

[34] D. D. Naik and S. S. Rao, *Analysis of Multivariate Repeated Measures Data with a Kronecker Product Structured Covariance Matrix*, Journal of Applied Statistics **28** (2001), 91–105.

[35] C. R. Rao, *An Asymptotic Expansion of the Distributions of Wilks' Criterion*, Bulletin of the International Statistical Institute **33** (1951), no. 4, 177–180.

[36] A. C. Rencher, *Methods of Multivariate Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Inc., New York, 2002.

[37] A. C. Rencher and G. B. Schaalje, *Linear Models in Statistics*, Second, Statistics/Probability, John Wiley & Sons, Inc., Inc., New Jersey, 2008.

[38] M. Rosenblatt, *A Central Theorem and a Strong Mixing Condition*, Proceedings of the National Academy of Sciences of the United States of America **42** (1956), no. 1, 43–47.

[39] R. J. Serfling, *Contributions to Central Limit Theory for Dependent Variables*, Annals of the Institute of Statistical Mathematics **39** (1968), 1158–1175.

[40] ———, *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, 1980.

[41] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, Second, Springer Texts in Statistics, Springer Science+Business Media, LLC, New York, 2006.

[42] M. S. Srivastava, T. von Rosen, and D. von Rosen, *Estimation and Testing in General Multivariate Linear Models with Kronecker product Covariance Structure*, Sankhyā **71** (2009), no. 2, Ser. A, 137–163.

[43] N. H. Timm, *Applied Multivariate Analysis*, Springer Texts in Statistics, Springer, New York, 2002.

[44] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer Series in Statistics, Springer, New York, 2000.

[45] H. Wang and M. G. Akritas, *Inference from Heteroscedastic Functional Data*, Journal of Nonparametric Statistics **22** (2010), no. 2, 149–168.

[46] H. White and I. Domowitz, *Nonlinear Regression with Dependent Observations*, Econometrica **52** (1984), no. 1, 143–161.

Appendix A

Minor Proofs

A.1 Proofs from Section 1.2.1

The following contains the proofs of Properties 1.2.3 though 1.2.7.

Proof: Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be matrices and let k be a scalar. Observe,

$$\begin{aligned}
(\boldsymbol{A} \oplus \boldsymbol{B}) \oplus \boldsymbol{C} &= \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix} \oplus \boldsymbol{C} \\
&= \left[\begin{array}{cc|c} \boldsymbol{A} & \mathbf{0} & \multirow{2}{*}{$\mathbf{0}$} \\ \mathbf{0} & \boldsymbol{B} & \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \boldsymbol{C} \end{array}\right] \\
&= \begin{bmatrix} \boldsymbol{A} & & \mathbf{0} \\ & \boldsymbol{B} & \\ \mathbf{0} & & \boldsymbol{C} \end{bmatrix}
\end{aligned}$$

$$= \left[\begin{array}{c|cc} \boldsymbol{A} & \multicolumn{2}{c}{\boldsymbol{0}} \\ \hline \multirow{2}{*}{\boldsymbol{0}} & \boldsymbol{B} & \boldsymbol{0} \\ & \boldsymbol{0} & \boldsymbol{C} \end{array}\right]$$
$$= \boldsymbol{A} \oplus (\boldsymbol{B} \oplus \boldsymbol{C}).$$

Since the third equality is equivalent to $\boldsymbol{A} \oplus \boldsymbol{B} \oplus \boldsymbol{C}$, Property 1.2.3 is proven. $\diamond$

A simple argument shows Property 1.2.4 to be true. Observe,

$$k(\boldsymbol{A} \oplus \boldsymbol{B}) = k \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} k\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & k\boldsymbol{B} \end{bmatrix} = k\boldsymbol{A} \oplus k\boldsymbol{B}.$$

Clearly, unless $k = 1$, or $\boldsymbol{A} = \boldsymbol{0}$ or $\boldsymbol{B} = \boldsymbol{0}$, $k(\boldsymbol{A} \oplus \boldsymbol{B}) \neq k\boldsymbol{A} \oplus \boldsymbol{B}$ since $\boldsymbol{B} \neq k\boldsymbol{B}$, and $k(\boldsymbol{A} \oplus \boldsymbol{B}) \neq \boldsymbol{A} \oplus k\boldsymbol{B}$ since $\boldsymbol{A} \neq k\boldsymbol{A}$. Thus, Property 1.2.4 is proven. $\diamond$

Now, consider m matrices labeled $\boldsymbol{A}_i$, $i = 1, \ldots, m$. Observe,

$$\begin{aligned} \operatorname{tr}\left(\bigoplus_{i=1}^{m} \boldsymbol{A}_i\right) &= \operatorname{tr}\left(\begin{bmatrix} \boldsymbol{A}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{A}_m \end{bmatrix}\right) \\ &= \operatorname{tr}(\boldsymbol{A}_1) + \cdots + \operatorname{tr}(\boldsymbol{A}_m) \\ &= \sum_{i=1}^{m} \operatorname{tr}(\boldsymbol{A}_i). \end{aligned}$$

This proves Property 1.2.5. $\diamond$

Furthermore, consider matrices $\boldsymbol{C}$ and $\boldsymbol{D}$. Recall from Linear Algebra that if $\boldsymbol{A}$ and $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ are invertible, then

$$\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A}) \det(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}).$$

Therefore,

$$\begin{aligned}
\det(\boldsymbol{A} \oplus \boldsymbol{B}) &= \det \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix} \\
&= \det(\boldsymbol{A}) \det(\boldsymbol{B} - \mathbf{0}\boldsymbol{A}^{-1}\mathbf{0}) \\
&= \det(\boldsymbol{A}) \det(\boldsymbol{B}),
\end{aligned}$$

and Property 1.2.6 is proven. ◇

Finally, consider an easy counterexample to commutativity. Let $\boldsymbol{A} = \begin{bmatrix} 1 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 2 \end{bmatrix}$. Then

$$\boldsymbol{A} \oplus \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \neq \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{B} \oplus \boldsymbol{A},$$

proving Property 1.2.7. All five direct sum properties have now been shown to be true. ◇ □

A.2 Proofs from Section 1.2.2

The following contains the proofs of Properties 1.2.9 though 1.2.16.

Before proceeding to the proofs of these properties, a short argument must be made to justify some of the steps in the proofs. Since the individual parts of a partitioned matrix behave like elements of a usual matrix, the Kronecker product of a partitioned matrix with another matrix can be written as the partitioned matrix in which the parts are the Kronecker products of the original parts with the usual matrix. This is straightforward from Definition 1.2.8. More explicitly, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices and $\boldsymbol{A}$ is partitioned such that

$$\boldsymbol{A} = \left[\begin{array}{c|c} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \hline \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{array}\right],$$

then

$$\boldsymbol{A} \otimes \boldsymbol{B} = \left[\begin{array}{c|c} \boldsymbol{A}_1 \otimes \boldsymbol{B} & \boldsymbol{A}_2 \otimes \boldsymbol{B} \\ \hline \boldsymbol{A}_3 \otimes \boldsymbol{B} & \boldsymbol{A}_4 \otimes \boldsymbol{B} \end{array}\right].$$

Clearly, since the partitioning of a matrix is arbitrary, this will hold for matrices of more than four parts. However, this should suffice as the only argument necessary for the use of such a tool.

Proof: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices and let k be a scalar. Let $\boldsymbol{A}$ be an $m \times n$ matrix. Observe,

$$\begin{aligned}
(k\boldsymbol{A}) \otimes \boldsymbol{B} &= \begin{bmatrix} (ka_{11})\boldsymbol{B} & \cdots & (ka_{1n})\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ (ka_{m1})\boldsymbol{B} & \cdots & (ka_{mn})\boldsymbol{B} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}(k\boldsymbol{B}) & \cdots & a_{1n}(k\boldsymbol{B}) \\ \vdots & \ddots & \vdots \\ a_{m1}(k\boldsymbol{B}) & \cdots & a_{mn}(k\boldsymbol{B}) \end{bmatrix} \\
&= k \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \\
&= k(\boldsymbol{A} \otimes \boldsymbol{B}).
\end{aligned}$$

Since that the second equality is equivalent to $\boldsymbol{A} \otimes (k\boldsymbol{B})$, Property 1.2.9 is proven. ⋄

Furthermore, consider matrix $\boldsymbol{C}$. Using the aforementioned argument regarding partitioned matrices and Property 1.2.9, observe,

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \otimes \boldsymbol{C}$$

$$
\begin{aligned}
&= \begin{bmatrix} (a_{11}\boldsymbol{B}) \otimes \boldsymbol{C} & \cdots & (a_{1n}\boldsymbol{B}) \otimes \boldsymbol{C} \\ \vdots & \ddots & \vdots \\ (a_{m1}\boldsymbol{B}) \otimes \boldsymbol{C} & \cdots & (a_{mn}\boldsymbol{B}) \otimes \boldsymbol{C} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}(\boldsymbol{B} \otimes \boldsymbol{C}) & \cdots & a_{1n}(\boldsymbol{B} \otimes \boldsymbol{C}) \\ \vdots & \ddots & \vdots \\ a_{m1}(\boldsymbol{B} \otimes \boldsymbol{C}) & \cdots & a_{mn}(\boldsymbol{B} \otimes \boldsymbol{C}) \end{bmatrix} \\
&= \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}).
\end{aligned}
$$

This shows that the Kronecker product is associative and proves Property 1.2.10. ⋄

For now, let $\boldsymbol{B}$ also be of size $m \times n$. Observe,

$$
\begin{aligned}
(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} &= \begin{bmatrix} (a_{11} + b_{11})\boldsymbol{C} & \cdots & (a_{1n} + b_{1n})\boldsymbol{C} \\ \vdots & \ddots & \vdots \\ (a_{m1} + b_{m1})\boldsymbol{C} & \cdots & (a_{mn} + b_{mn})\boldsymbol{C} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\boldsymbol{C} + b_{11}\boldsymbol{C} & \cdots & a_{1n}\boldsymbol{C} + b_{1n}\boldsymbol{C} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{C} + b_{m1}\boldsymbol{C} & \cdots & a_{mn}\boldsymbol{C} + b_{mn}\boldsymbol{C} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\boldsymbol{C} & \cdots & a_{1n}\boldsymbol{C} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{C} & \cdots & a_{mn}\boldsymbol{C} \end{bmatrix} + \begin{bmatrix} b_{11}\boldsymbol{C} & \cdots & b_{1n}\boldsymbol{C} \\ \vdots & \ddots & \vdots \\ b_{m1}\boldsymbol{C} & \cdots & b_{mn}\boldsymbol{C} \end{bmatrix} \\
&= \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}.
\end{aligned}
$$

A very similar argument will prove the second statement (but is not necessary here); thus, Property 1.2.11 is proven. ⋄

Now, let $\boldsymbol{C}$ have size $n \times p$ (to be conformable with $\boldsymbol{A}$), and also consider a matrix $\boldsymbol{D}$ conformable with $\boldsymbol{B}$. Observe,

$$
(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D})
$$

$$= \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \begin{bmatrix} c_{11}\boldsymbol{D} & \cdots & c_{1p}\boldsymbol{D} \\ \vdots & \ddots & \vdots \\ c_{n1}\boldsymbol{D} & \cdots & a_{np}\boldsymbol{D} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}\boldsymbol{B}c_{11}\boldsymbol{D} + \cdots + a_{1n}\boldsymbol{B}c_{n1}\boldsymbol{D} & \cdots & a_{11}\boldsymbol{B}c_{1p}\boldsymbol{D} + \cdots + a_{1n}\boldsymbol{B}c_{np}\boldsymbol{D} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B}c_{11}\boldsymbol{D} + \cdots + a_{mn}\boldsymbol{B}c_{n1}\boldsymbol{D} & \cdots & a_{m1}\boldsymbol{B}c_{1p}\boldsymbol{D} + \cdots + a_{mn}\boldsymbol{B}c_{np}\boldsymbol{D} \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11}c_{11} + \cdots + a_{1n}c_{n1})\boldsymbol{BD} & \cdots & (a_{11}c_{1p} + \cdots + a_{1n}c_{np})\boldsymbol{BD} \\ \vdots & \ddots & \vdots \\ (a_{m1}c_{11} + \cdots + a_{mn}c_{n1})\boldsymbol{BD} & \cdots & (a_{m1}c_{1p} + \cdots + a_{mn}c_{np})\boldsymbol{BD} \end{bmatrix}$$

$$= \boldsymbol{AC} \otimes \boldsymbol{BD}.$$

This proves Property 1.2.12, and this argument can be repeated recursively to obtain to the extension

$$\left(\bigotimes_{i=1}^{m} \boldsymbol{A}_i\right)\left(\bigotimes_{j=1}^{m} \boldsymbol{B}_j\right) = \bigotimes_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{B}_j,$$

though that proof is not given here. ⋄

Moreover, assume now that $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible. Denote by $\boldsymbol{I}$ the identity matrix. For now, the size of $\boldsymbol{I}$ will be clear given the context of the situation. Also, realize quite trivially that $\boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}$, all three of which can be of differing sizes. Observe, by Property 1.2.12,

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}) = \boldsymbol{AA}^{-1} \otimes \boldsymbol{BB}^{-1} = \boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I},$$

and

$$(\boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1})(\boldsymbol{A} \otimes \boldsymbol{B}) = \boldsymbol{A}^{-1}\boldsymbol{A} \otimes \boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}.$$

Therefore, $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$, and Property 1.2.13 is proven. ⋄

Recall that $(k\boldsymbol{B})' = k\boldsymbol{B}'$. Observe,

$$
\begin{aligned}
(\boldsymbol{A} \otimes \boldsymbol{B})' &= \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}' \\
&= \begin{bmatrix} (a_{11}\boldsymbol{B})' & \cdots & (a_{m1}\boldsymbol{B})' \\ \vdots & \ddots & \vdots \\ (a_{1n}\boldsymbol{B})' & \cdots & (a_{mn}\boldsymbol{B})' \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\boldsymbol{B}' & \cdots & a_{m1}\boldsymbol{B}' \\ \vdots & \ddots & \vdots \\ a_{1n}\boldsymbol{B}' & \cdots & a_{mn}\boldsymbol{B}' \end{bmatrix} \\
&= \boldsymbol{A}' \otimes \boldsymbol{B}',
\end{aligned}
$$

which proves Property 1.2.14. $\diamond$

Continuing to the property regarding the trace, let us now assume that $\boldsymbol{A}$ has size $m \times m$. Observe,

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) &= \operatorname{tr}\left(\begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1m}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mm}\boldsymbol{B} \end{bmatrix} \right) \\
&= a_{11}\operatorname{tr}(\boldsymbol{B}) + \cdots + a_{mm}\operatorname{tr}(\boldsymbol{B}) \\
&= (a_{11} + \cdots + a_{mm})\operatorname{tr}(\boldsymbol{B}) \\
&= \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B}).
\end{aligned}
$$

This proves Property 1.2.15. $\diamond$

For the proof of noncommutativity, consider a simple counterexample. Let $\boldsymbol{A} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and let $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then, clearly,

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 0 & 2 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} = \boldsymbol{B} \otimes \boldsymbol{A},$$

proving Property 1.2.16. All eight Kronecker product properties have now been shown to be true. ◇ □

A.3 Proofs from Section 1.2.3

The following contains the proofs of Properties 1.2.18 though 1.2.22.

Proof: The proof of Property 1.2.22 can be found in Broxson (Propsition 31) [13], thus it is omitted here. ◇

Using Property 1.2.22, Property 1.2.19 is merely the simplification when we let $\boldsymbol{A}' = \boldsymbol{I}$, $\boldsymbol{Y}' = \boldsymbol{A}$, $\boldsymbol{B} = \boldsymbol{B}$, and $\boldsymbol{X} = \boldsymbol{C}$. This proves Property 1.2.19. ◇

Using Property 1.2.22, Property 1.2.20 is merely the simplification when we let $\boldsymbol{A}' = \boldsymbol{I}$, $\boldsymbol{Y}' = \boldsymbol{A}$, $\boldsymbol{B} = \boldsymbol{I}$, and $\boldsymbol{X} = \boldsymbol{B}$. This proves Property 1.2.20. ◇

To prove Property 1.2.18, first suppose that $\boldsymbol{B}$ is $m \times n$ and that $\boldsymbol{A}$ and $\boldsymbol{X}$ conform. Write $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_n \end{bmatrix}$ and $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m \end{bmatrix}$. Then, we can write the j^{th} column of $\boldsymbol{AXB}$ as

$$\begin{aligned} [\boldsymbol{AXB}]_{\cdot j} &= \boldsymbol{AXb}_k \\ &= \boldsymbol{A} \sum_{i=1}^{m} \boldsymbol{x}_i b_{ij} \\ &= \begin{bmatrix} b_{1j}\boldsymbol{A} & \cdots & b_{mj}\boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \end{bmatrix} \end{aligned}$$

$$= \left(\boldsymbol{b}_j' \otimes \boldsymbol{A}\right) \text{vec}\left(\boldsymbol{X}\right).$$

Then if stack the columns, we see

$$\begin{aligned}
\text{vec}\left(\boldsymbol{AXB}\right) &= \begin{bmatrix} [\boldsymbol{AXB}]_{\cdot 1} \\ \vdots \\ [\boldsymbol{AXB}]_{\cdot n} \end{bmatrix} \\
&= \begin{bmatrix} \left(\boldsymbol{b}_1' \otimes \boldsymbol{A}\right) \text{vec}\left(\boldsymbol{X}\right) \\ \vdots \\ \left(\boldsymbol{b}_n' \otimes \boldsymbol{A}\right) \text{vec}\left(\boldsymbol{X}\right) \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{b}_1' \otimes \boldsymbol{A} \\ \vdots \\ \boldsymbol{b}_n' \otimes \boldsymbol{A} \end{bmatrix} \text{vec}\left(\boldsymbol{X}\right) \\
&= \left(\boldsymbol{B}' \otimes \boldsymbol{A}\right) \text{vec}\left(\boldsymbol{X}\right),
\end{aligned}$$

and Property 1.2.18 is proven. $\diamond$

Using Property 1.2.18, Property 1.2.21 is merely the simplification when we let $\boldsymbol{A}' = \boldsymbol{a}$, $\boldsymbol{X} = \boldsymbol{I}_{1\times 1}$, and $\boldsymbol{B} = \boldsymbol{a}'$. Since $\text{vec}\left(\boldsymbol{I}_{1\times 1}\right) = 1$, Property 1.2.21 is proven. $\diamond$ $\square$

Appendix B

R Functions Implementing the Tests

B.1 Setup and Function Design

In order to properly and efficiently run the simulation code, the functions presented in Sections B.2 and B.3 were written in the computer program R. These functions give vectors and matrices such as $\mathbf{1}_m$, $\boldsymbol{I}_m$, $\boldsymbol{J}_m$, $\boldsymbol{P}_m$, and the direct sum of such functions. The last two functions were used specifically for running the simulation under the traditional RM-ANOVA assumptions found in Davis [16].

```
I = function(x){diag(1,x)}
one = function(x){rep(1,x)}
J = function(x){one(x)%*%t(one(x))}
Jn = function(x){1/x*one(x)%*%t(one(x))}
Pn = function(x){I(x)-Jn(x)}
onesD = function(x){ # Block diagonal of 1/ni*1_ni
A = matrix(1)
for(i in 1:(length(x))){
A = bdiag(A,(1/x[i]*as.matrix(one(x[i]))))
}
A = A[-1,-1]
```

```
return(A)
}
JD = function(x){ # Block diagonal of 1/ni*J_ni
A = matrix(1)
for(i in 1:(length(x))){
A = bdiag(A,(1/x[i]*one(x[i])%*%t(one(x[i]))))
}
A = A[-1,-1]
return(A)
}
PnD = function(x){ # Block diagonal of 1/(ni(ni-1))*P_ni
A = matrix(1)
for(i in 1:(length(x))){
A = bdiag(A,(1/(x[i]*(x[i]-1))*Pn(x[i])))
}
A = A[-1,-1]
return(A)
}
####
# Use the following for DAVIS
####
onesD2 = function(x){ # Block diagonal of 1/n*1_ni
A = matrix(1)
for(i in 1:(length(x))){
A = bdiag(A,(1/sum(x)*as.matrix(one(x[i]))))
}
A = A[-1,-1]
return(A)
}
PnD2 = function(x){ # Block diagonal of P_ni
A = matrix(1)
for(i in 1:(length(x))){
A = bdiag(A,Pn(x[i]))
}
A = A[-1,-1]
return(A)
}
```

B.2 Univariate Case Function

The function `HosslerMANOVAsim` is passed a set of arguments and returns a vector of rejection decisions, one decision corresponding to each test—the main effect of factor B (β), the interaction effect (γ), and the main effect of factor A (α). The first three arguments are wholly necessary and will be described. The remaining arguments could easily be defined internally, but for the sake of simulation speed, they are passed as arguments as to only be calculated once. As such, their definitions are commented out in the function below. All of these arguments are used in the middle matrix for the various sums of squares used to calculate the test statistics.

The argument `Xmat` is an $n \times b$ matrix of data organized in the following manner. Each row corresponds to one subject, and the b repeated measurements taken on each subject are given in the columns. The rows are organized such that n_1 subjects in the first level (group) of factor A are given, followed by the n_2 subjects in the second level, and so on for all a levels. This matrix is vectorized in the main function resulting in a vector as prescribed by (2.1). From the dimension of `Xmat`, the value of b is inferred.

Next, the argument `ni` is a vector of the number of subjects in each level of factor A, given in order from n_1 to n_a. The length of this vector implies the value of a and the sum of its elements imply the value of n. Finally, the argument `critVal` is the critical value used for each of the tests. This will be between 0 and 1, and, for example, was 0.05 for the simulation study.

After the three test statistics are calculated, they are compared to the p-value corresponding to `critVal` and the appropriate asymptotic distribution. The decision to reject or not is made, and a 3×1 vector, called `decisionVec`, is returned containing True/False elements. The simplicity of the function and its single output vector is due only to the nature of the simulation study. To calculate and return fewer items significantly speeds up the simulation process. For marketing or practical purposes, the test statistics, p-values, and decisions would be reported.

```
HosslerMANOVAsim = function(Xmat,ni,critVal,Cbeta,Cgamma,gBeta,gGamma,mid1,mid2,mid3,mid4){

a = length(ni)
b = dim(Xmat)[2]
n = sum(ni)

Xmat = as.matrix(Xmat)
X = as.matrix(as.vector(Xmat))
Xtilde = rowMeans(Xmat)

# C_beta_E^star matrix
# Cbeta = (bdiag(onesD(ni)) %*% (Jn(a)) %*% t(bdiag(onesD(ni)))) - 1/a*PnD(ni)
CbetaEstar = Cbeta

# C_gamma_E^star matrix
# Cgamma = (bdiag(onesD(ni)) %*% (1/(a-1)*Pn(a)) %*% t(bdiag(onesD(ni)))) - 1/a*PnD(ni)
CgammaEstar = Cgamma

####
# The middle matrices (the kronecker product) and  quadratic forms
# for tests of factor B effect and interaction effect
# (Appeals to trace theorem)
####

# Time and interaction effect statistics
# mid1 = (b/(b-1))*Pn(b)
# mid2 = I(b)
Mid1 = mid1
Mid2 = mid2
Tbeta = sum(diag(Mid1%*%t(Xmat)%*%CbetaEstar%*%Xmat)) / sqrt(b)
Tgamma = sum(diag(Mid1%*%t(Xmat)%*%CgammaEstar%*%Xmat)) / sqrt(b)

bnew = ceiling(b^(1/2))
sigEstPieces = rep(0,bnew)
for(i in 1:n){
sigEstPieces = sigEstPieces + (acf(Xmat[i,],type="covariance",lag.max=(bnew-1),plot=FALSE)$acf)^2
}
sigEst = 1/n * (sigEstPieces %*% (1/bnew*c(bnew,(2*((bnew-1):1)))))

# gBeta = 2/a^2*( sum(1/(ni*(ni-1))) + (sum(1/(ni%*%t(ni)))-sum(diag(1/(ni%*%t(ni))))) )
```

```
# gGamma = 2/a^2*( sum(1/(ni*(ni-1))) + 1/(a-1)^2*(sum(1/(ni%*%t(ni)))-sum(diag(1/(ni%*%t(ni))))) )
gFuncBeta = gBeta
gFuncGamma = gGamma

TbetaStat = Tbeta/sqrt(gFuncBeta*sigEst)
TgammaStat = Tgamma/sqrt(gFuncGamma*sigEst)

####
# Middle matrices and quadratic forms for test of factor A effect
####
# mid3 = I(n)-JD(ni)
# mid4 = JD(ni)-Jn(n)
Mid3 = mid3
Mid4 = mid4
G = b*t(Xtilde)%*%(Mid3)%*%Xtilde
H = b*t(Xtilde)%*%(Mid4)%*%Xtilde
FalphaStat = (H/(a-1))/(G/(n-a))

####
# p-value calculations with rejection decisions put into a vector and returned
####
betaPval = 2*pnorm(as.vector(-1*abs(TbetaStat)))
gammaPval = 2*pnorm(as.vector(-1*abs(TgammaStat)))
alphaPval = 1-pf(as.vector(FalphaStat),(a-1),(n-a))

betaR = betaPval<critVal
gammaR = gammaPval<critVal
alphaR = alphaPval<critVal

decisionVec = c(betaR,gammaR,alphaR)
return(decisionVec)
}
```

The function `DavisMANOVAsim` is very similar to `HosslerMANOVAsim`; however, the former is code running simulations under the traditional assumptions. `DavisMANOVAsim` uses the functions `onesD2` and `OnD2` from the first segments of code.

```
DavisMANOVAsim = function(Xmat,ni,critVal,Cbeta,Cgamma,gBeta,gGamma,mid1,mid2,mid3,mid4){
```

```
a = length(ni)
b = dim(Xmat)[2]
n = sum(ni)

Xmat = as.matrix(Xmat)
X = as.matrix(as.vector(Xmat))
Xtilde = rowMeans(Xmat)
# gGamma = Jn(b)
# mid1 = Pn(b)
# mid2 = (n*a^2) * (bdiag(onesD2(ni)) %*% (Jn(a)) %*% t(bdiag(onesD2(ni))))
# mid3 = (n*a^2) * (bdiag(onesD2(ni)) %*% (Pn(a)) %*% t(bdiag(onesD2(ni))))
# mid4 = as.matrix(PnD2(ni))
Davis1_1 = gGamma
Davis1 = mid1
Davis2 = mid2
Davis3 = mid3
Davis4 = mid4

####
# Define SS for DAVIS
####
SS_R_Davis = sum(diag(Davis1%*%t(Xmat)%*%Davis4%*%Xmat))
SS_SG_Davis = sum(diag(Davis1_1%*%t(Xmat)%*%Davis4%*%Xmat))

temp1 = rowMeans(Xmat)
temp2 = c(0,cumsum(ni))
temp3 = rep(0,a)
for (k in 1:a){
temp3[k] = mean(temp1[(temp2[k]+1):(temp2[k+1])])
}
# temp 3 is Xbar_{i..}, group mean
temp4 = rep(mean(Xmat),a)
# Xbar_{...}, overall mean repeated a times
SS_G_Davis = b*sum((temp3-temp4)^2*ni)

temp5 = colMeans(Xmat)
# temp 5 is Xbar_{.j.}, time mean
temp6 = rep(mean(Xmat),b)
# Xbar_{...}, overall mean repeated b times
```

```
SS_T_Davis = n*sum((temp5-temp6)^2)

# Need temp3 (Xbar_{i..}) and temp5 (Xbar_{.j.})
temp7 = c(0,cumsum(rep(ni,b)))
temp8 = rep(0,a*b)
for (k in 1:(a*b)){
temp8[k] = mean(X[(temp7[k]+1):(temp7[k+1])])
}
# temp 8 is Xbar_{ij.}, subject mean
temp9 = rep(temp8,rep(ni,b))
temp10 = rep(rep(temp3,ni),b)
temp11 = rep(temp5,rep(n,b))
temp12 = rep(mean(Xmat),n*b)
SS_GT_Davis = sum((temp9-temp10-temp11+temp12)^2)

####
# Time and interaction effect statistics
####
Fbeta_Davis = (SS_T_Davis/(b-1))/(SS_R_Davis/((n-a)*(b-1)))
Fgamma_Davis = (SS_GT_Davis/((a-1)*(b-1)))/(SS_R_Davis/((n-a)*(b-1)))
Falpha_Davis = (SS_G_Davis/(a-1))/(SS_SG_Davis/(n-a))

####
# p-value calculations with rejection decisions put into a vector and returned
####
betaPval = 1-pf(as.vector(Fbeta_Davis),(b-1),((n-a)*(b-1)))
gammaPval = 1-pf(as.vector(Fgamma_Davis),((a-1)*(b-1)),((n-a)*(b-1)))
alphaPval = 1-pf(as.vector(Falpha_Davis),(a-1),(n-a))

betaR = betaPval<critVal
gammaR = gammaPval<critVal
alphaR = alphaPval<critVal

decisionVec = c(betaR,gammaR,alphaR)
return(decisionVec)
}
```

B.2.1 Bootstrapping

The function tsBOOTmanova was implemented when using bootstrapping to estimate the variance of the test statistics. It returns a vector of estimated standard deviations for the two test statistics T_β and T_γ. The arguments mid1BOOT, CbetaBOOT, and CgammaBOOT are not wholly necessary. They are middle matrices for the quadratic forms and are only not included internally for the sake of efficiency.

The argument dataBOOT is the matrix of data. The number of bootstrap samples conducted is given by the value of the repsBOOT argument. The block length for the block bootstrap method is given by the blockLengthBOOT argument, and bBOOT is the value of b (the number of levels of factor B.

```
tsBOOTmanova = function(dataBOOT,repsBOOT,mid1BOOT,CbetaBOOT,CgammaBOOT,blockLengthBOOT,bBOOT){
beta_stat_storage = rep(0,repsBOOT)
gamma_stat_storage = rep(0,repsBOOT)

for (j in 1:repsBOOT){
tempData = tsBOOTsample(dataBOOT,blockLengthBOOT)
beta_stat_storage[j] = sum(diag(mid1BOOT%*%t(tempData)%*%CbetaBOOT%*%tempData)) / sqrt(bBOOT)
# TbetaBOOT
gamma_stat_storage[j] = sum(diag(mid1BOOT%*%t(tempData)%*%CgammaBOOT%*%tempData)) / sqrt(bBOOT)
# gammaBOOT
}

st_error_stats = c(sd(beta_stat_storage),sd(gamma_stat_storage))
return(st_error_stats)
}
```

In order to conduct the bootstrap samples, the function tsBOOTsample was implemented. It takes in a matrix of data and the block length, and it returns a bootstrap sample of the data.

```
tsBOOTsample = function(Data,blockLength){
```

```
p = 1/4
# Exponent for block length, l, where l~n^p
data_length = dim(Data)[2]
#data_length
##block_length = round(data_length)^(p))
block_length = blockLength
#block_length
tempBootMat = as.matrix(cbind(Data,Data[,1:(block_length-1)]))
#tempBootMat
num_blocks = data_length/block_length
#num_blocks
boot_index = sample(1:data_length,num_blocks,replace=TRUE)
#boot_index

bootMat = c()
for (i in 1:num_blocks){
bootMat = cbind(bootMat,tempBootMat[,((boot_index[i]):(boot_index[i]+block_length-1))])
}
bootMat = as.matrix(bootMat)
return(bootMat)
}
```

B.3 Multivariate Case Function

The code to conduct the multivariate simulation is very similar to the code for the univariate simulations. The arguments are similar, so their explanation is omitted. Since the simulation was expensive and only the likelihood ratio test was considered, the output of the function only includes the decisions for $T_{\mathrm{WL}}^{(\phi)}$. However, note that there are commented lines of code that could be used for the other three statistics.

```
HosslerMANOVAsimMV = function(Xmat,ni,critVal,Cbeta,Cgamma,gBeta,gGamma,mid1,mid2,mid3,mid4){

a = length(ni)
n = sum(ni)
b = dim(Xmat)[2]/n
```

```
p = dim(Xmat)[1]

Xmat = as.matrix(Xmat)
Xtilde = Xmat %*% (as.matrix(1/b*one(b)) %x% I(n))
Xmat2 = reorg(Xmat,p,n,b)

gFuncBeta = gBeta
gFuncGamma = gGamma

# mid1 = (1/(b-1))*Pn(b)
# mid2 = I(b)
# mid3 = I(n)-JD(ni)
# mid4 = JD(ni)-Jn(n)
Mid1 = mid1

# H^(B) middle matrix
# H_B_mid = (1/(b-1))*Pn(b) KRON (bdiag(onesD(ni)) %*% (Jn(a)) %*% t(bdiag(onesD(ni))))
H_B_mid = Mid1 %x% Cbeta

# H^(Gamma) middle matrix
#H_AB_mid = (1/(b-1))*Pn(b) KRON (bdiag(onesD(ni)) %*% (1/(a-1)*Pn(a)) %*% t(bdiag(onesD(ni))))
H_AB_mid = Mid1 %x% Cgamma

# G ^(B) = G^(Gamma) middle matrices
#G_B_mid = (1/(b-1))*Pn(b) KRON 1/a*PnD(ni)
G_B_mid = Mid1 %x% mid2
G_B_mid = Mid1 %x% mid2
G_AB_mid = Mid1 %x% mid2

# H^(A) middle matrix
# H_A_mid = JD(ni)-Jn(n)
H_A_mid = mid4

# G^(A) middle matrix
# G_A_mid = I(n)-JD(ni)
G_A_mid = mid3

####
# Matrices for test statistics
####
```

```
H_B = as.matrix( Xmat %*% H_B_mid %*% t(Xmat) )
H_AB = as.matrix( Xmat %*% H_AB_mid %*% t(Xmat) )
G_B = as.matrix( Xmat %*% G_B_mid %*% t(Xmat) )
G_B = as.matrix( Xmat %*% G_B_mid %*% t(Xmat) )
G_AB = as.matrix( Xmat %*% G_AB_mid %*% t(Xmat) )
H_A = as.matrix( Xtilde %*% (b*H_A_mid) %*% t(Xtilde) )
G_A = as.matrix( Xtilde %*% (b*G_A_mid) %*% t(Xtilde) )

####
# Wilk's Lambda (likelihood ratio) case test statistics with others for future work
###################
# Wilks' Lambda (WL)
TbetaStat_WL = -log(det(G_B)/det(H_B+G_B))
# -log( prod(eigen(G_B)$values)/prod(eigen(H_B+G_B)$values) )
TgammaStat_WL = -log(det(G_AB)/det(H_AB+G_AB))
TalphaStat_WL = -log(det(G_A)/det(H_A+G_A))
# Lawley-Hotelling (LH)
# TbetaStat_LH = sum(diag(H_B%*%solve(G_B)))
# TgammaStat_LH = sum(diag(H_AB%*%solve(G_AB)))
# TalphaStat_LH = sum(diag(H_A%*%solve(G_A)))
# Bartlett-Nanda-Pillai (BNP)
# TbetaStat_BNP = sum(diag(H_B%*%solve(H_B+G_B)))
# TgammaStat_BNP = sum(diag(H_AB%*%solve(H_AB+G_AB)))
# TalphaStat_BNP = sum(diag(H_A%*%solve(H_A+G_A)))
# Dempster's ANOVA Type (D)
# TbetaStat_D = sum(diag(H_B))/sum(diag(G_B))
# TgammaStat_D = sum(diag(H_AB))/sum(diag(G_AB))
# TalphaStat_D = sum(diag(H_A))/sum(diag(G_A))
###################

STATS_B = sqrt(b)*(c(2)*c(TbetaStat_WL)-c(2*p*log(2)))
STATS_AB = sqrt(b)*(c(2)*c(TgammaStat_WL)-c(2*p*log(2)))
STATS_A = c(TalphaStat_WL)

####
# Calculate variance component tau via eta
####
eta_beta = sqrt(sum(diag( (J(b)%x%solve(G_B)) %*% (Xmat2%*%PnD2(ni)%*%t(Xmat2)) )))
eta_gamma = sqrt(sum(diag( (J(b)%x%solve(G_AB)) %*% (Xmat2%*%PnD2(ni)%*%t(Xmat2)) )))
tau_beta = sqrt(eta_beta*gFuncBeta)
```

```
tau_gamma = sqrt(eta_gamma*gFuncGamma)

####
# Calculate test statistics and p-values, make rejection decisions and put into vector
####
TbetaStat = STATS_B/tau_beta
TgammaStat = STATS_AB/tau_gamma
TalphaStat = (n-a+(-(p-(a-1)+1)/2))*STATS_A

alphaCritVal = qchisq((1-critVal),p*(a-1)) + 1/(n-a)*(((p-(a-1)+1)/2-(-(p-(a-1)+1)/2))
   *(qchisq((1-critVal),p*(a-1))))

betaPval = 2*pnorm(as.vector(-1*abs(TbetaStat)))
gammaPval = 2*pnorm(as.vector(-1*abs(TgammaStat)))
#alphaPval = 1-pf(as.vector(FalphaStat),(a-1),(n-a))

betaR = betaPval<critVal
gammaR = gammaPval<critVal
alphaR = TalphaStat>alphaCritVal

decisionVec = c(betaR,gammaR,alphaR)
return(decisionVec)
}
```