

AN ABSTRACT OF THE DISSERTATION OF

Waseem S. Alnosaier for the degree of Doctor of Philosophy in Statistics presented on April 25, 2007.

Title: Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models.

Abstract approved:

David S. Birkes

Clifford B. Pereira

For data following a balanced mixed Anova model, the standard Anova method typically leads to exact F tests of the usual hypotheses about fixed effects. However, for most unbalanced designs with most hypotheses, the Anova method does not produce exact tests, and approximate methods are needed. One approach to approximate inference about fixed effects in mixed linear models is the method of Kenward and Roger (1997), which is available in SAS and widely used. In this thesis, we strengthen the theoretical foundations of the method by clarifying and weakening the assumptions, and by determining the orders of the approximations used in the derivation. We present two modifications of the K-R method which are comparable in performance but simpler in derivation and computation. It is known that the K-R method reproduces exact F tests in two special cases, namely for Hotelling T^2 and for Anova F ratios in fixed-effects models. We show that the K-R and proposed methods reproduce exact F tests in three more general models which include the two special cases, plus tests in all fixed effects linear models, many balanced mixed Anova models, and all multivariate regressions models. In addition, we present some theorems on the K-R, proposed, and Satterthwaite methods, investigating conditions under which they are identical. Also, we show the

difficulties in developing a K-R type method using the conventional, rather than adjusted, estimator of the variance-covariance matrix of the fixed-effects estimator. A simulation study is conducted to assess the performance of the K-R, proposed, Satterthwaite, and Containment methods for three kinds of block designs. The K-R and proposed methods perform quite similarly and they outperform other approaches in most cases.

©Copyright by Waseem S. Alnosaier

April 25, 2007

All Rights Reserved

Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models

by

Waseem S. Alnosaier

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirement for the
degree of

Doctor of Philosophy

Presented April 25, 2007

Commencement June 2007

Doctor of Philosophy dissertation of Waseem S. Alnosaier presented on April 25, 2007.

APPROVED:

Co-Major Professor, representing Statistics

Co-Major Professor, representing Statistics

Chair of the Department of Statistics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Waseem S. Alnosaier, Author

ACKNOWLEDGMENTS

I am deeply indebted to my advisors, Dr. David Birkes and Dr. Clifford Pereira, for giving me the opportunity to work with them. I especially thank Dr. Birkes who patiently and tirelessly helped me throughout my doctoral program. Without his encouragement and constant guidance, this dissertation would not have been finished. I really can not thank him enough, and it is an honor to work with him. I also greatly thank Dr. Pereira for his help and suggestions. I appreciate his attention, and valuable comments that improved the work of this thesis.

I would also like to thank the entire faculty, and staff in the Department of Statistics. I especially thank Dr. Alix Gitelman and Dr. Lisa Madsen for serving on my committee, and for their concern and comments. I thank Professor Robert Smythe and Professor Daniel Schafer for their help and concern. I also thank Professor Philippe Rossignol from the Department of Fisheries and Wildlife for serving on my committee.

I am thankful to the faculty in the Department of Statistics at University of Pittsburgh, especially Professor Leon Gleser, Professor Satish Iyengar, and Professor David Stoffer who helped me during my studies there.

Special thanks go to all current and former students, of whom which there are too many to mention, who influenced my statistical education. In particular, I thank my office mate, Yonghai Li, for the unforgettable memories and valuable discussions we had.

I would like to express my sincere appreciation to my parents for their compassion, encouragement, and support. They taught me at a young age how to respect education, and offered me love and support through the course of my life.

Thanks are due foremost to my wonderful wife, Sohailah, and my lovely children, Sarah, Kasim and Yaser, for supporting me throughout the entire process. Thank you Sohailah for your patience, sacrifices and believing in me. You helped me in so many ways, and I will never forget that. Your inspiration, understanding, and love made this work possible.

TABLE OF CONTENTS

	<u>Page</u>
1. INTRODUCTION	1
1.1 Previous Results	1
1.2 Contributions and Summary of Results	4
2. VARIANCE-COVARIANCE MATRIX OF THE FIXED EFFECTS	
ESTIMATOR	7
2.1 The Model	7
2.2 Notation	7
2.3 Assumptions	8
2.4 Estimating $\text{Var}(\hat{\beta})$	11
3. TESTING THE FIXED EFFECTS	21
3.1 Constructing a Wald-Type Pivot	21
3.2 Estimating the Denominator Degrees of Freedom and the Scale Factor	36
4. MODIFYING THE ESTIMATES OF THE DENOMINATOR DEGREES OF	
FREEDOM AND THE SCALE FACTOR	38
4.1 Balanced One-Way Anova Model	38
4.2 The Hotelling T^2 model	41
4.3 Kenward and Roger's Modification	48
4.4 Modifying the Approach with the Usual Variance-Covariance Matrix	54
5. TWO PROPOSED MODIFICATIONS FOR THE KENWARD-ROGER	
METHOD	55
5.1 First Proposed Modification	55
5.2 Second Proposed Modification	58

TABLE OF CONTENTS (Continued)

	<u>Page</u>
5.3 Comparisons among the K-R and the Proposed Modifications	61
6. KENWARD-ROGER APPROXIMATION FOR THREE GENERAL MODELS	67
6.1 Modified Rady's Model	67
6.2 A General Multivariate Linear Model	77
6.3 A Balanced Multivariate Model with a Random Group Effect	83
7. THE SATTERTHWAITE APPROXIMATION	97
7.1 The Satterthwaite Method	97
7.2 The K-R, the Satterthwaite and the Proposed Methods	98
8. SIMULATION STUDY FOR BLOCK DESIGNS	105
8.1 Preparing Formulas for Computations	105
8.2 Simulation Results	109
8.3 Comments and Conclusion	124
9. CONCLUSION	127
9.1 Summary	127
9.2 Future Research	128
Bibliography	130

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1.1 PBIB1 ($t = s = 15$, $k = 4$, and $n = 60$)	110
1.2 Simulated size of nominal 5% Wald F- tests for PBIB1	111
1.3 Mean of estimated denominator degrees of freedom for PBIB1	111
1.4 Mean of estimated scale for PBIB1	111
1.5 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for PBIB1	111
1.6 PBIB2 design ($t = 16$, $s = 48$, $k = 2$, and $n = 96$)	112
1.7 Simulated size of nominal 5% Wald F- tests for PBIB2	112
1.8 Mean of estimated denominator degrees of freedom for PBIB2	112
1.9 Mean of estimated scales for PBIB2	113
1.10 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for PBIB2	113
2.1 BIB1 ($t = 6$, $s = 15$, $k = 2$, and $n = 30$)	114
2.2 Simulated size of nominal 5% Wald F- tests for BIB1	114
2.3 Mean of estimated denominator degrees of freedom for BIB1	114

LIST OF TABLES (Continued)

<u>Table</u>	<u>Page</u>
2.4 Mean of estimated scales for BIB1	114
2.5 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for BIB1	115
2.6 BIB2 ($t = 6$, $s = 10$, $k = 3$, and $n = 30$)	115
2.7 Simulated size of nominal 5% Wald F- tests for BIB2	115
2.8 Mean of estimated denominator degrees of freedom for BIB2	116
2.9 Mean of estimated scales for BIB2	116
2.10 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency for BIB2	116
2.11 BIB3 ($t = 7$, $s = 7$, $k = 4$, and $n = 28$)	117
2.12 Simulated size of nominal 5% Wald F- tests for BIB3	117
2.13 Mean of estimated denominator degrees of freedom for BIB3	117
2.14 Mean of estimated scales for BIB3	117
2.15 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency for BIB3	118

LIST OF TABLES (Continued)

<u>Table</u>	<u>Page</u>
2.16 BIB4 ($t = 9$, $s = 36$, $k = 2$, and $n = 72$)	118
2.17 Simulated size of nominal 5% Wald F- tests for BIB4	118
2.18 Mean of estimated denominator degrees of freedom for BIB4	119
2.19 Mean of estimated scales for BIB4	119
2.20 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for BIB4	119
2.21 BIB5 design ($t = 9$, $s = 18$, $k = 4$, and $n = 72$)	120
2.22 Simulated size of nominal 5% Wald F- tests for BIB5	120
2.23 Mean of estimated denominator degrees of freedom for BIB5	120
2.24 Mean of estimated scales for BIB5	120
2.25 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for BIB5	121
2.26 BIB6 ($t = 9$, $s = 12$, $k = 6$, and $n = 72$)	121
2.27 Simulated size of nominal 5% Wald F- tests for BIB6	121
2.28 Mean of estimated denominator degrees of freedom for BIB6	121

LIST OF TABLES (Continued)

<u>Table</u>	<u>Page</u>
2.29 Mean of estimated scales for BIB6	122
2.30 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for BIB6	122
3.1 Simulated size of nominal 5% Wald F- tests for RCB1	122
3.2 Mean of estimated denominator degrees of freedom for RCB1	123
3.3 Mean of estimated scales for RCB1	123
3.4 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for RCB1	123
3.5 Simulated size of nominal 5% Wald F- tests for RCB2	123
3.6 Mean of estimated denominator degrees of freedom for RCB2	124
3.7 Mean of estimated scales for RCB2	124
3.8 Percentage relative bias in the variance estimates, convergence rate (CR), and efficiency (E) for RCB2	124

*To my wife, Sohailah,
for her significant support.*

KENWARD-ROGER APPROXIMATE F TEST FOR FIXED EFFECTS IN MIXED LINEAR MODELS

1. INTRODUCTION

When testing fixed effects in a mixed linear model, an exact F test may not exist. Several approaches are available to perform approximate F tests, and one of the most common approaches is the method derived by Kenward and Roger (1997). Since the Kenward-Roger method has been implemented in the MIXED procedure of the SAS system, it has become well known and widely used. In this thesis, we investigate the Kenward-Roger approach, and suggest two other modifications.

1.1 Previous Results

In testing a hypothesis about fixed effects, it is desirable to find an exact test. In balanced mixed linear models, the standard Anova method leads to optimal exact F tests for the fixed effects in most cases. In such models, Seifert (1979) showed that the standard Anova F tests are uniformly most powerful invariant unbiased tests (UMPIU). Rady (1986) studied the needed assumptions for the mixed linear model so that the Anova method produces optimal exact F tests for the fixed effects. Seely and Rady (1988) studied conditions where the random effects can be treated as fixed to construct an exact F test for the fixed effects. VanLeeuwen *et al.* (1998) introduced the concept of error orthogonal (EO) designs. In models with EO structure, we have an exact F test for certain standard hypotheses. Utlaut *et al.* (2001) introduced the concept of simple error orthogonal (SEO) designs, in which optimal exact F tests can be obtained. If a mixed effects Anova model has a type of “partial balance” called b&r balanced (VanLeeuwen *et al.*, 1999), then it is SEO. A typical normal balanced Anova model is b&r balanced, and therefore, has an optimal exact F test as indicated above. In particular, when the smallest random effect that contains a fixed effect is unique, there is an optimal exact F test for the fixed effect which coincides with the Anova F test (Birkes, 2004). In addition, exact F tests have been constructed for certain other models. For instance, an exact F test

can be obtained for testing certain hypotheses about fixed effects in a general multivariate model (Mardia *et al.*, 1992). Also, an exact F test can be obtained to test fixed effects in a balanced multivariate model with a random group effect (Birkes, 2006).

In many cases, constructing an exact test seems hard and approximations need to be employed. When no two mean squares in the Anova table have the same expectations under the null hypothesis, the Anova method may still be used to obtain an approximate F test. If a mean square can be created as a linear function of other mean squares such that it has the same expectation as the mean square of the fixed effects under the null hypothesis, then the created mean square is used in the denominator of an F test where its degrees of freedom can be approximated by Satterthwaite (1946).

Another common approach to produce an approximate F test for the fixed effects is based on the Wald-type statistic. When using a Wald-type test, the numerator degrees of freedom equal the number of contrasts being tested, but the denominator degrees of freedom needs to be estimated. In the Containment method, which is the default method in SAS's MIXED procedure, the denominator degrees of freedom are chosen be the smallest rank contribution of the random effects that contain the fixed effects to the design matrix. If no such effects are found, the estimate is the residual degrees of freedom.

Giesbrecht and Burns (1985) proposed a method based on Satterthwaite (1941) to determine the denominator degrees of freedom for an approximate Wald-type t test for a single contrast of the fixed effects. Fai and Cornelius (1996) extended Giesbrecht and Burns method where they approximate the Wald-type test for a multidimensional hypothesis by an F distribution. Both approaches proposed by Giesbrecht and Burns, and Fai and Cornelius use the conventional estimator of the variance-covariance matrix of the fixed effects estimator, which is the variance-covariance matrix of the generalized least squares (GLS) estimator of the fixed effects, replacing the variance components with the REML estimates. The Fai and Cornelius approach is referred to as the Satterthwaite approach in some literature and in this thesis as well. Bellavance *et al.* (1996) suggested a method to improve the Anova F test by using a scaled F distribution with modified

degrees of freedom.

It is known that the conventional estimator of the variance-covariance matrix of the fixed effects estimator underestimates (Kackar and Harville, 1984). Indeed, Kackar and Harville expressed the variance-covariance matrix of the fixed effects estimator as a sum of two components. The first component is the variance-covariance matrix of the GLS estimator of the fixed effects, and the second component represents what the first component underestimates. The second component was approximated by Kackar and Harville (1984) and Prasad and Rao (1990), and estimation of the first component was addressed by Harville and Jeske (1992) and later by Kenward and Roger (1997). In fact, Kenward and Roger (1997) combined estimators of the two components to produce an adjusted estimator for the variance-covariance matrix of the fixed effects estimator. They plug the adjusted estimator into the Wald-type statistic where a scaled form of the statistic follows an F distribution approximately. Unlike the Satterthwaite-based approaches where only the degrees of freedom need to be estimated, in Kenward and Roger's approximation, two quantities need to be estimated from the data: the denominator degrees of freedom and the scaling factor.

After deriving approximate expressions for the expectation and the variance of the Wald-type statistic, Kenward and Roger then match these with the first and second moments of the F distribution to determine the estimates of the denominator degrees of freedom and the scale. Moreover, Kenward and Roger modified the approximation in such a way that the estimates match the known values for two special cases where a scaled form of the Wald-type statistic has an exact F test. In fact, the idea of modifying the approach in a way to produce the exact values was adopted previously by Graybill and Wang (1980) when they modified approximate confidence intervals on certain functions of the variance components in a way to make them exact for some special cases.

Lately, the performance of Kenward and Roger's approximation has become a subject for some simulation studies. For example, Schaalje *et al.* (2002) compared the K-R and the Satterthwaite approaches for a split plot design with repeated measures for

several sample sizes and covariance structures. They found that the K-R method performs as well as or better than the Satterthwaite approach in all situations. They considered three factors in the comparisons: the complexity of the covariance matrix, imbalance, and the sample size, and they found that these factors affect the Satterthwaite method more than the K-R method. The Satterthwaite method was found to work well only when the sample size is moderately large and the covariance matrix was compound symmetric. The K-R method was found to have a tendency toward inflated levels when the sample size was small, except when the covariance structure was compound symmetric. Chen and Wei (2003) compared the Kenward-Roger approach and a modified Anova method suggested by Bellavance *et al.* (1996) for some crossover designs. They recommended using the K-R approach when the sample size is at least 24. For smaller sample sizes, they found the modified Anova method works better than the K-R approach. Savin *et al.* (2003) found the K-R approach reliable to construct a confidence interval for the common mean in interlaboratory trials. Spike *et al.* (2004) investigated the K-R and the Satterthwaite methods to estimate the denominator degrees of freedom for contrasts of the fixed subplot effects. Like the conclusion drawn by Schaalje *et al.* (2002) above, they suggested Kenward and Roger's approximation be preferred in small datasets, and the two methods are comparable for large datasets. Valderas *et al.* (2005) studied the performance of the K-R approach when AIC and BIC are used as criteria to select the covariance structure. They found the K-R method's level much higher than the target values. Even with the correct covariance structure, the level was found to be higher than the target for many cases in which the covariance structure is not compound symmetric.

1.2 Contributions and Summary of Results

Since some of the detailed derivation for Kenward and Roger's approach was absent from their original work, we provide the detailed theoretical derivation of the method which includes clarifying the assumptions to justify the theoretical derivation. Also, we weaken some of the assumptions that were imposed by Kenward and Roger, and determine the orders of the approximations used in the derivation. We present two

modifications of the K-R method which are comparable in performance but simpler in derivation and computation. Kenward and Roger modified their approach in such a way that their method reproduces exact F tests in two special cases, namely for Hotelling T^2 and for Anova F ratios. We show that the K-R and the two proposed methods reproduce exact F tests in three more general models, two of which are generalizations of the two special cases. We explore relationships among the K-R, proposed and Satterthwaite methods by specifying cases where the approaches produce the same estimate of the denominator degrees of freedom or are even identical. Also, we show the difficulties in developing a K-R type method using the conventional, rather than adjusted, estimator of the variance-covariance matrix of the fixed-effects estimator.

In chapter 2, using Taylor series expansions, matrix derivatives and invariance arguments, we derive the adjusted estimator for the variance -covariance matrix of the fixed effects estimator that was provided by Kenward and Roger (1997). Besides clarifying all assumptions that justify the theoretical derivation, we weaken some of the assumptions imposed by Kenward and Roger. In addition, we determine the orders of the approximations used in the derivation.

In chapter 3, we derive the approximate expectation and variance of the Wald-type statistics where they are constructed by using the conventional and the adjusted estimator of the variance-covariance matrix of the fixed effects estimator. In addition, we match the first and second moments of F distribution with these for the scaled form of the Wald-type statistic to obtain the Kenward-Roger approximation for the denominator degrees of freedom and scale before the modification.

Two special cases: the balanced one-way Anova model, and the Hotelling T^2 model where the Wald-type statistic have an exact F distribution are considered in chapter 4 to establish the modification of the expectation and the variance of the Wald-type statistic proposed by Kenward and Roger so the approach produces the right and known values for these two special cases. Kenward and Roger (1997) mentioned that it can be argued that the conventional estimator of the variance-covariance matrix of the fixed effects estimator can be used instead of the adjusted estimator. We discuss the

difficulties in modifying the approach by using the conventional estimator instead of the adjusted estimator.

The K-R approach was derived based on modifying the approximated expressions for the expectation and the variance of the Wald-type statistic. This modification is not unique, and hence we introduce two other modifications for the K-R method in chapter 5. We keep the modification of the expectation of the statistic as modified by Kenward and Roger; however, instead of modifying the variance, we modify other related quantities. The proposed modifications for the K-R method are comparable in performance and simpler in derivation and computation.

As mentioned above, the special cases used by Kenward and Roger are not the only cases where the K-R and proposed methods produce the exact values. Indeed, the Kenward-Roger and the proposed modifications produce the exact values for three general models where there is an optimal exact F test. The models studied in chapter 6 are: (1) Rady's model (with a slight modification) which includes a wide class of balanced mixed classification models and is more general than the balanced one-way Anova model, (2) a general linear multivariate model which is more general than the Hotelling T^2 , and (3) a balanced multivariate model with a random group effect. We show that the estimate of the denominator degrees of freedom and the scale factor match the known values for those models.

The Satterthwaite, the Kenward-Roger and the proposed methods perform similarly in some situations. Chapter 7 is devoted to study the cases where these methods produce the same estimate for the denominator degrees of freedom. Moreover, we study the cases where the approaches become identical to each other.

In chapter 8, we provide a simulation study for three types of block designs: partially balanced incomplete block designs, balanced incomplete block designs and complete block designs with some missing data. In the simulation study, the sample size, the ratio of the variance components, and the efficiency factor are considered to see how they affect the performance of the Kenward-Roger, the proposed, the Satterthwaite and the Containment methods.

2. VARIANCE-COVARIANCE MATRIX OF THE FIXED EFFECTS ESTIMATOR

For a multivariate mixed linear model, statisticians used to estimate the precision of the fixed effects estimates based on the asymptotic distribution. However, this estimate was known to be biased and underestimate the variance of the fixed effects estimate. Kenward and Roger (1997) proposed an adjustment for the estimator of the variance of the fixed effects estimator which is investigated in this chapter.

2.1 The Model

Consider n observations \mathbf{y} following a multivariate normal distribution,

$$\mathbf{y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

where $\mathbf{X}(n \times p)$ is a full column rank matrix of known covariates, $\boldsymbol{\beta}(p \times 1)$ is a vector of unknown parameters and $\boldsymbol{\Sigma}(n \times n)$ is an unknown variance-covariance matrix whose elements are assumed to be functions of r parameters, $\boldsymbol{\sigma}(r \times 1) = (\sigma_1, \dots, \sigma_r)'$. The generalized least squares estimator of $\boldsymbol{\beta}$ is $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$, and the matrix $\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ is the variance-covariance matrix of this estimator. The REML estimator of $\boldsymbol{\sigma}$ is denoted by $\hat{\boldsymbol{\sigma}}$, and the REML-based estimated generalized least squares estimator of $\boldsymbol{\beta}$ (EGLSE) is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{y}$. The center of our study is $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$ is a nuisance to be addressed in the analysis. We are interested in testing $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, for \mathbf{L}' an $(\ell \times p)$ fixed matrix.

2.2 Notation

Throughout the thesis, we use the following notation. For a matrix \mathbf{A} , we use \mathbf{A}' , $\Re(\mathbf{A})$, $\mathfrak{N}(\mathbf{A})$, $r(\mathbf{A})$, $\text{tr}(\mathbf{A})$, $|\mathbf{A}|$, to denote transpose, range, null space, rank, trace, and determinant of \mathbf{A} respectively. $\Re(\mathbf{A})^\perp$ is used to denote the orthogonal complement of $\Re(\mathbf{A})$. We use the abbreviation p.d. for positive definite and n.n.d. for nonnegative definite. The notation p.o. is used for projection operator and o.p.o. for orthogonal

projection operator. \mathbf{P}_A is o.p.o. on $\Re(\mathbf{A})$. In addition, we use the following

$$\mathbf{V} = \text{Var}(\hat{\boldsymbol{\beta}})$$

$$\mathbf{W} = \text{Var}(\hat{\boldsymbol{\sigma}}), \quad w_{ij} = \text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j)$$

$$\mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}$$

$$\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} \mathbf{L}'$$

$$\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$$

$$\mathbf{P}_i = -\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \boldsymbol{\Sigma}^{-1} \mathbf{X}$$

$$\mathbf{Q}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j} \boldsymbol{\Sigma}^{-1} \mathbf{X}$$

$$\mathbf{R}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \sigma_i \partial \sigma_j} \boldsymbol{\Sigma}^{-1} \mathbf{X}$$

$$\boldsymbol{\Lambda} = \text{Var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$$

$$\tilde{\boldsymbol{\Lambda}} = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \boldsymbol{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j) \boldsymbol{\Phi}$$

$$\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\boldsymbol{\Phi}} \hat{\mathbf{P}}_j - \frac{1}{4} \hat{\mathbf{R}}_{ij}) \right\} \hat{\boldsymbol{\Phi}}$$

$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_i \boldsymbol{\Phi}) \text{tr}(\boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_j \boldsymbol{\Phi})$$

$$A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_i \boldsymbol{\Phi} \boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_j \boldsymbol{\Phi}),$$

$$A_3 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\boldsymbol{\Theta} \boldsymbol{\Phi} (\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j - \frac{1}{4} \mathbf{R}_{ij}) \boldsymbol{\Phi}]$$

$$F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L}(\mathbf{L}' \hat{\boldsymbol{\Phi}}_A \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}}, \text{ unless otherwise mentioned.}$$

For a matrix $\mathbf{H}(\boldsymbol{\sigma})$, we use $\hat{\mathbf{H}} = \mathbf{H}(\hat{\boldsymbol{\sigma}})$.

2.3 Assumptions

For chapters 2 and 3, we impose the following assumptions about the model.

(A1) The expectation of $\hat{\boldsymbol{\beta}}$ exists.

(A2) $\boldsymbol{\Sigma}$ is a block diagonal and nonsingular matrix. Also, we assume that the elements of

$\Sigma_k, \Sigma_k^{-1}, \frac{\partial \Sigma_k}{\partial \sigma_i}, \frac{\partial^2 \Sigma_k}{\partial \sigma_i \partial \sigma_j}$, and \mathbf{X}_k are bounded, where $\Sigma = \text{diag}_{1 \leq k \leq m}(\Sigma_k)$, and

$\mathbf{X} = \text{col}_{1 \leq k \leq m}(\mathbf{X}_k)$, and $\sup n_i < \infty$, where n_i are the blocks sizes.

(A3) $E[\hat{\boldsymbol{\sigma}}] = \boldsymbol{\sigma} + O(n^{-3/2})$.

(A4) The possible dependence between $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\beta}}$ is ignored.

(A5) $\sum_{i=1}^r \sum_{j=1}^r \text{Cov} \left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right) \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right)', (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \right] = O(n^{-1/2})$.

(A6) $\boldsymbol{\Phi} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} = O(n^{-1})$, $(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} = O(n)$, $\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} = O(n^{-1/2})$, $\frac{\partial^2 \tilde{\boldsymbol{\beta}}}{\partial \sigma_i^2} = O(n^{-1/2})$

(A7) If $\mathbf{T} = O_p(n^\alpha)$, then $E[\mathbf{T}] = O(n^\alpha)$

Remarks

(i) Σ_k is said to be bounded when $|\max(\text{elements of } \Sigma_k)| \leq b(\boldsymbol{\sigma})$ for some constant b .

Since $\sup n_i < \infty$, then $n \rightarrow \infty \Leftrightarrow m \rightarrow \infty$.

(ii) Even though Kenward and Roger required the bias of the REML estimator to be ignored, we only require $E[\hat{\boldsymbol{\sigma}}] = \boldsymbol{\sigma} + O(n^{-3/2})$ as stated as assumption (A3). In fact, for the model mentioned in section 2.1, $E[\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}] = \mathbf{b}(\boldsymbol{\sigma}) + O(n^{-2})$, where $\mathbf{b}(\boldsymbol{\sigma})$ is of order $O(n^{-1})$ (Pase and Salvan, 1997, expression 9.62). Moreover, when the covariance structure is linear, $\mathbf{b}(\boldsymbol{\sigma}) = \mathbf{0}$, and hence $E[\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}] = O(n^{-2})$ which is stronger than what we need.

(iii) Assumption (A4) was also imposed by Kenward and Roger. In fact, we did investigate some models, like Hotelling T^2 model, the fixed effects model, and the one way Anova model with fixed group effects and unequal group variances. In these models, $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\beta}}$ are independent exactly. Also, for those models, the sum of covariance in assumption (A5) is zero which is stronger than the assumption. This assumption is needed to derive approximate expressions for the expectation and the variance of the statistic by conditioning on $\hat{\boldsymbol{\sigma}}$ as we will see in chapter 3. However, we should mention

that the same derivation of the expectation and the variance of the statistic can be done without conditioning on $\hat{\boldsymbol{\sigma}}$, in which case assumption (A4) is not needed anymore.

(iv) One circumstance for the covariance in assumption (A5) to be exactly zero is when $\hat{\boldsymbol{\sigma}}$ is obtained from a previous sample, and $\tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma})$ from current data (Kackar and Harville, 1984). Also, Kackar and Harville argued that when $\hat{\boldsymbol{\sigma}}$ is unbiased and

$E \left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right) \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right)' \mid \hat{\boldsymbol{\sigma}} \right]$ can be approximated by first order Taylor series, then the covariance

in (A5) is expected to be zero. It appears that Kenward and Roger assumed one of the arguments mentioned. For us, it is enough to assume (A5).

(v) $\mathbf{X}'\mathbf{H}(\boldsymbol{\sigma})\mathbf{X} = O(n)$, where $\mathbf{H}(\boldsymbol{\sigma}) = \text{diag}_{1 \leq k \leq m} (\mathbf{H}_k(\boldsymbol{\sigma}))$, and $\mathbf{H}_k(\boldsymbol{\sigma})$ is a product of any combination of $\boldsymbol{\Sigma}_k, \boldsymbol{\Sigma}_k^{-1}, \frac{\partial \boldsymbol{\Sigma}_k}{\partial \sigma_i}$, and $\frac{\partial^2 \boldsymbol{\Sigma}_k}{\partial \sigma_i \partial \sigma_j}$. This is true because $\mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k = O(1)$

for $1 \leq k \leq m \Rightarrow |\max(\text{elements of } \mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k)| \leq a$ for $1 \leq k \leq m$

$$\Rightarrow \sum_{k=1}^m |\max(\text{elements of } \mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k)| \leq ma$$

$$\Rightarrow \left| \sum_{k=1}^m \max(\text{elements of } \mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k) \right| \leq \sum_{k=1}^m |\max(\text{elements of } \mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k)| \leq ma,$$

and hence $\mathbf{X}'\mathbf{H}(\boldsymbol{\sigma})\mathbf{X} = \sum_{k=1}^m \mathbf{X}_k' \mathbf{H}_k(\boldsymbol{\sigma}) \mathbf{X}_k = O(m) = O(n)$.

(vi) A general situation where the first two conditions in assumption (A6) hold is when all $\mathbf{X}_k' \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k$ are contained in a compact

$$\text{set } \mathbb{C} \text{ of p.d. matrices} \Rightarrow \frac{1}{m} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \in \mathbb{C} \Rightarrow m \boldsymbol{\Phi} \in \mathbb{C}^{-1}$$

and hence $\boldsymbol{\Phi} = O(n^{-1})$. Also, we have $m \mathbf{L}' \boldsymbol{\Phi} \mathbf{L} \in \mathbf{L}' \mathbb{C}^{-1} \mathbf{L}$ (p.d.) $\Rightarrow (\mathbf{L}' \boldsymbol{\Phi} \mathbf{L})^{-1} = O(n)$.

Also, this assumption holds when we suppose $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_1 \forall k$, and we suppose that

$\frac{1}{m} \sum_{k=1}^m \mathbf{X}_k' \boldsymbol{\Sigma}_1^{-1} \mathbf{X}_k \rightarrow \mathbf{A}$ (p.d.). This supposition is reasonable in two situations:

1) $\mathbf{X}_k = \mathbf{X}_1 \forall k$, and in this case, $\mathbf{A} = \mathbf{X}_1' \boldsymbol{\Sigma}_1^{-1} \mathbf{X}_1$.

2) \mathbf{X}_k are regarded as iid random covariate matrices, and by the weak law of large numbers,

$\mathbf{X}'_k \boldsymbol{\Sigma}_1^{-1} \mathbf{X}_k$ converges to $E[\mathbf{X}'_1 \boldsymbol{\Sigma}_1^{-1} \mathbf{X}_1]$.

Then, since inversion is a continuous operation, $m\boldsymbol{\Phi} \rightarrow \mathbf{A}^{-1}$. That is $\boldsymbol{\Phi} = O(n^{-1})$

Also, $\frac{1}{m}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} \rightarrow (\mathbf{L}'\mathbf{A}^{-1}\mathbf{L})^{-1}$. That is $(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} = O(n)$.

$$\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} = -\boldsymbol{\Phi} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \mathbf{G} \mathbf{y} \quad (\text{as we will see in lemma 2.4.7}).$$

Consider $\mathbf{X}'\mathbf{B}(\boldsymbol{\sigma})\mathbf{y}$, where $\mathbf{B}(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \mathbf{G}$, and $\mathbf{X}'\mathbf{B}(\boldsymbol{\sigma})\mathbf{y} = \sum_{k=1}^m \mathbf{X}'_k \mathbf{B}_k(\boldsymbol{\sigma}) \mathbf{y}_k$,

where $\mathbf{y} = \text{col}_{1 \leq k \leq m}(\mathbf{y}_k)$, and $\mathbf{B}(\boldsymbol{\sigma}) = \text{diag}_{1 \leq k \leq m}(\mathbf{B}_k(\boldsymbol{\sigma}))$. If $\mathbf{X}'_k \mathbf{B}_k(\boldsymbol{\sigma}) \mathbf{y}_k$ are iid, then by the weak law of large numbers, we have $\sqrt{m} \left[\frac{1}{m} \sum_{k=1}^m \mathbf{X}'_k \mathbf{B}_k(\boldsymbol{\sigma}) \mathbf{y}_k - E[\mathbf{X}'_k \mathbf{B}_k(\boldsymbol{\sigma}) \mathbf{y}_k] \right] = O(1)$.

However, in lemma (2.4.7), we have $E[\mathbf{X}'_k \mathbf{B}_k(\boldsymbol{\sigma}) \mathbf{y}_k] = \mathbf{0}$, and hence $\mathbf{X}'\mathbf{B}(\boldsymbol{\sigma})\mathbf{y} = O(n^{1/2})$.

$$\Rightarrow \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} = O(n^{-1})O(n^{1/2}) = O(n^{-1/2}).$$

(vii) If $\mathbf{H}(\boldsymbol{\sigma}) = O_p(n^\alpha)$, then $\mathbf{H}(\hat{\boldsymbol{\sigma}}) = O_p(n^\alpha)$. This result is obtained by employing a Taylor series expansion for $\mathbf{H}(\hat{\boldsymbol{\sigma}})$ about $\boldsymbol{\sigma}$.

(viii) For some random matrices \mathbf{T} , their expectations have higher order than the random matrices themselves. For instance, $(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)(\hat{\sigma}_k - \sigma_k) = O_p(n^{-3/2})$, as it will be shown in theorem (2.4.9); however, $E[(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)(\hat{\sigma}_k - \sigma_k)] = O(n^{-2})$, from expression (9.74) in Pace and Salvan. In assumption (A7), we assume that the expectation will reserve the order which does not conflict with the cases mentioned above. \square

2.4 Estimating $\text{Var}(\hat{\boldsymbol{\beta}})$

There are two main sources of bias in $\hat{\boldsymbol{\Phi}}$ when it is used as an estimator for $\text{Var}(\hat{\boldsymbol{\beta}})$: $\hat{\boldsymbol{\Phi}}$ underestimates $\boldsymbol{\Phi}$, and $\boldsymbol{\Phi} = \text{Var}(\tilde{\boldsymbol{\beta}})$ does not take into account the variability of $\hat{\boldsymbol{\sigma}}$ in $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\sigma}})$. Kackar and Harville (1984) proposed that $\text{Var}(\hat{\boldsymbol{\beta}})$ can be partitioned as $\text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi} + \boldsymbol{\Lambda}$, and they addressed the second source of bias by approximating $\boldsymbol{\Lambda}$. The first source of bias was discussed by Harville and Jeske (1992), and Kenward and Roger (1997) proposed an approximation to adjust the first bias, and they combined both

adjustments to calculate their proposed estimator of $\text{Var}(\hat{\beta})$ which is denoted by $\hat{\Phi}_A$. In this section, several lemmas are derived to lead to the expression for $\hat{\Phi}_A$.

Lemma 2.4.1 With $L_R(\sigma)$ being the likelihood function of $\mathbf{z} = \mathbf{K}'\mathbf{y}$, where $\mathbf{K}'\mathbf{X} = \mathbf{0}$,

With $L_R(\sigma)$ being the likelihood function of $\mathbf{z} = \mathbf{K}'\mathbf{y}$, where $\mathbf{K}'\mathbf{X} = \mathbf{0}$,

$$2\log L_R(\sigma) = 2\ell_R(\sigma) = \text{constant} - \log|\Sigma| - \log|\mathbf{X}'\Sigma^{-1}\mathbf{X}| \\ - \mathbf{y}'[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}]\mathbf{y}$$

Proof $\mathbf{y} \sim N(\mathbf{X}\beta, \Sigma)$, $f(\mathbf{y}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)\right]$

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{K}'\Sigma\mathbf{K}), \quad (\text{Birkes, 2004, theorem 7.2.2}).$$

$$\text{and } f(\mathbf{z}) = (2\pi)^{-q/2} |\mathbf{K}'\Sigma\mathbf{K}|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{z}'(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{z}\right], \text{ where } q = n - p.$$

The REML estimator for σ is the maximum likelihood estimator from the marginal likelihood of $\mathbf{z} = \mathbf{K}'\mathbf{y}$ where \mathbf{K}' is any $q \times n$ matrix of full column rank

$$2\ell_R(\sigma) = -q \log(2\pi) - \log(\mathbf{K}'\Sigma\mathbf{K}) - \mathbf{z}'(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{z} \\ = -q \log(2\pi) - \log(\mathbf{K}'\Sigma\mathbf{K}) - \mathbf{y}'\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'\mathbf{y}$$

To prove the lemma, it suffices to show that

$$(a) \mathbf{y}'\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'\mathbf{y} = \mathbf{y}'[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}]\mathbf{y}$$

$$(b) -q \log(2\pi) - \log|\mathbf{K}'\Sigma\mathbf{K}| = \text{constant} - \log|\Sigma| - \log|\mathbf{X}'\Sigma^{-1}\mathbf{X}|$$

For part (a), it suffices to show that

$$\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}' = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$$

Indeed, $\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'\Sigma = \mathbf{I} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'$ (Seely, 2002, problem (2.B.4))

$$\Rightarrow \mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}' = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$$

Observe that Σ is nonsingular (assumption A2), and n.n.d (Birkes, 2004), then Σ is p.d.

(Seely, 2002, corollary 1.9.3),

and hence $r(\mathbf{K}'\Sigma\mathbf{K}) = r(\mathbf{K}')$ (Birkes, 2004, a lemma on Jan 23).

Also, we have $\Re(\mathbf{K}) = \Re(\mathbf{X}')$ (because $\Re(\mathbf{K}) = \Re(\mathbf{X})^\perp$)

So, all conditions of problem B.4 from Seely notes are satisfied.

For part (b), choose $\mathbf{T} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}$, so $\mathbf{T}'\mathbf{T} = (\mathbf{X}'\mathbf{X})^{-1/2} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2} = \mathbf{I}$.

Also, notice that we can choose \mathbf{K}' to be orthogonal so $\mathbf{K}'\mathbf{K} = \mathbf{I}$.

$\mathbf{K}'\mathbf{T} = \mathbf{0}$, and this is $\mathbf{K}'\mathbf{X} = \mathbf{0}$.

$$\text{Let } \begin{bmatrix} \mathbf{T} & \mathbf{K} \end{bmatrix} = \mathbf{R}, \text{ so } \mathbf{R}'\mathbf{R} = \begin{bmatrix} \mathbf{T}' \\ \mathbf{K}' \end{bmatrix} \begin{bmatrix} \mathbf{T} & \mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{T}'\mathbf{T} & \mathbf{T}'\mathbf{K} \\ \mathbf{K}'\mathbf{T} & \mathbf{K}'\mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

Hence $|\mathbf{R}'\mathbf{R}| = 1$,

$$\begin{aligned} \text{and } |\Sigma| &= |\mathbf{R}||\Sigma||\mathbf{R}'| = |\mathbf{R}\Sigma\mathbf{R}'| = \begin{vmatrix} \mathbf{T}'\Sigma\mathbf{T} & \mathbf{T}'\Sigma\mathbf{K} \\ \mathbf{K}'\Sigma\mathbf{T} & \mathbf{K}'\Sigma\mathbf{K} \end{vmatrix} = |\mathbf{K}'\Sigma\mathbf{K}| |\mathbf{T}'\Sigma\mathbf{T} - \mathbf{T}'\Sigma\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'\Sigma\mathbf{T}| \\ &= |\mathbf{K}'\Sigma\mathbf{K}| |(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}| |\mathbf{X}'\mathbf{X}|. \end{aligned}$$

$$\text{So, } \log|\Sigma| = \log|\mathbf{K}'\Sigma\mathbf{K}| - \log|\mathbf{X}'\Sigma^{-1}\mathbf{X}| + \log|\mathbf{X}'\mathbf{X}|$$

$$\Rightarrow -\log|\mathbf{K}'\Sigma\mathbf{K}| = -\log|\Sigma| - \log|\mathbf{X}'\Sigma^{-1}\mathbf{X}| + \text{constant. } \square$$

Lemma 2.4.2 $\hat{\boldsymbol{\beta}}$ is symmetric about $\boldsymbol{\beta}$.

Proof Since $\hat{\boldsymbol{\sigma}}$ is reflection and translation invariant (Birkes, 2004), then $\hat{\boldsymbol{\beta}}$ is reflection and translation equivariant (Birkes, 2004, lemma 7.1).

Since $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \Sigma)$, then $\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \stackrel{d}{=} -(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

$$\Rightarrow \hat{\boldsymbol{\beta}}(\mathbf{y}) - \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(\mathbf{y}) + (-\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}(\mathbf{y} + \mathbf{X}(-\boldsymbol{\beta})) \quad (\hat{\boldsymbol{\beta}} \text{ is translation equivariant})$$

$$= \hat{\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \stackrel{d}{=} -\hat{\boldsymbol{\beta}}(-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))$$

$$= -\hat{\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad \hat{\boldsymbol{\beta}} \text{ is a reflection equivariant}$$

$$= -(\hat{\boldsymbol{\beta}}(\mathbf{y}) - \boldsymbol{\beta}). \quad \square$$

Lemma 2.4.3 Given the expectation exists (assumption A1), $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.

Proof By applying lemma 2.3.2, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \stackrel{d}{=} -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = E[-(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$

$$\Rightarrow E(\hat{\beta}) - \beta = -E(\hat{\beta}) + \beta \Rightarrow 2E(\hat{\beta}) = 2\beta \Rightarrow E(\hat{\beta}) = \beta. \quad \square$$

Lemma 2.4.4 $\tilde{\beta}$ and $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$ are independent.

$$\begin{aligned} \text{Proof} \quad \text{Cov}[\tilde{\beta}, (\mathbf{I} - \mathbf{P}_X)\mathbf{y}] &= \text{Cov}[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}, (\mathbf{I} - \mathbf{P}_X)\mathbf{y}] \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\Sigma(\mathbf{I} - \mathbf{P}_X) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{P}_X) \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_X \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}' = \mathbf{0} \end{aligned}$$

Since $\tilde{\beta}$ and $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$ are both normally distributed, then they are independent (Birkes, 2004, proposition 7.2.3). \square

Lemma 2.4.5 $\hat{\beta} - \tilde{\beta}$ is a function of $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$.

Proof Consider $(\hat{\beta} - \tilde{\beta})(\mathbf{y}) = \delta(\mathbf{y})$, and hence $\delta(\mathbf{y})$ is a translation-invariant function.

To show that $\delta(\mathbf{y})$ is a function of $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$, it is equivalent to show that:

$$\begin{aligned} (\mathbf{I} - \mathbf{P}_X)\mathbf{y}^{(1)} &= (\mathbf{I} - \mathbf{P}_X)\mathbf{y}^{(2)} \Rightarrow \delta(\mathbf{y}^{(1)}) = \delta(\mathbf{y}^{(2)}) \\ (\mathbf{I} - \mathbf{P}_X)\mathbf{y}^{(1)} &= (\mathbf{I} - \mathbf{P}_X)\mathbf{y}^{(2)} \Rightarrow \mathbf{y}^{(1)} - \mathbf{P}_X\mathbf{y}^{(1)} = \mathbf{y}^{(2)} - \mathbf{P}_X\mathbf{y}^{(2)} \\ \Rightarrow \mathbf{y}^{(2)} &= \mathbf{y}^{(1)} - \mathbf{P}_X\mathbf{y}^{(1)} + \mathbf{P}_X\mathbf{y}^{(2)} = \mathbf{y}^{(1)} + \mathbf{P}_X(\mathbf{y}^{(2)} - \mathbf{y}^{(1)}) \\ \Rightarrow \delta(\mathbf{y}^{(2)}) &= \delta(\mathbf{y}^{(1)} + \mathbf{X}\mathbf{b}) \quad \text{for some } \mathbf{b} \\ &= \delta(\mathbf{y}^{(1)}) \quad (\delta \text{ is translation invariant}). \quad \square \end{aligned}$$

Lemma 2.4.6 $\text{Var}(\hat{\beta}) = \Phi + \Lambda$, where $\Lambda = \text{Var}(\hat{\beta} - \tilde{\beta})$

Proof From lemma 2.4.4, $\tilde{\beta}$ and $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$ are independent, and from lemma 2.4.5,

$\hat{\beta} - \tilde{\beta}$ is a function of $(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$. Then, $\tilde{\beta}$ and $\hat{\beta} - \tilde{\beta}$ are independent.

$$\begin{aligned} \text{Write} \quad \hat{\beta} &= \tilde{\beta} + \hat{\beta} - \tilde{\beta} \\ \Rightarrow \text{Var}(\hat{\beta}) &= \text{Var}(\tilde{\beta}) + \text{Var}(\hat{\beta} - \tilde{\beta}) = \Phi + \Lambda. \quad \square \end{aligned}$$

Lemma 2.4.7 $E\left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}\right] = \mathbf{0}$, and $E\left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}\right)(\hat{\sigma}_i - \sigma_i)\right] = \mathbf{0}$ for $i = 1, \dots, r$.

Proof

$$\begin{aligned} \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} &= \frac{\partial}{\partial \sigma_i} \left[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \right] \\ &= \frac{\partial}{\partial \sigma_i} \left[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \right] \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \frac{\partial}{\partial \sigma_i} \left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \right] \\ &= -\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \mathbf{G}\mathbf{y} \quad \text{where } \mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \end{aligned} \quad (2.1)$$

$$\Rightarrow E\left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}\right] = E\left[-\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \mathbf{G}\mathbf{y}\right] = -\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} E(\mathbf{G}\mathbf{y}),$$

where $E[\mathbf{G}\mathbf{y}] = E[(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{y}]$
 $= \boldsymbol{\Sigma}^{-1}E[\mathbf{y} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}] = \boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}.$

$$E\left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}(\hat{\sigma}_i - \sigma_i)\right] = -\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} E[g(\mathbf{y})], \text{ where } g(\mathbf{y}) = \mathbf{G}\mathbf{y}(\hat{\sigma}_i(\mathbf{y}) - \sigma_i).$$

$g(-\mathbf{y}) = \mathbf{G}(-\mathbf{y})[\hat{\sigma}_i(-\mathbf{y}) - \sigma_i] = -\mathbf{G}\mathbf{y}[\hat{\sigma}_i(-\mathbf{y}) - \sigma_i] = -g(\mathbf{y})$, because $\hat{\sigma}_i$ is reflection equivariant. So, g is reflection-equivariant.

Since $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$

$$\Rightarrow \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \stackrel{d}{=} -(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \Rightarrow \mathbf{y} \stackrel{d}{=} -\mathbf{y} + 2\mathbf{X}\boldsymbol{\beta}$$

and hence $g(\mathbf{y}) \stackrel{d}{=} g(-\mathbf{y} + 2\mathbf{X}\boldsymbol{\beta})$

$$= -g[\mathbf{y} + \mathbf{X}(-2\boldsymbol{\beta})] \quad \text{because } g \text{ is reflection equivariant}$$

$$= -g(\mathbf{y}), \quad \text{because } g \text{ is translation invariant.}$$

Notice that $g(\mathbf{y} + \mathbf{X}\boldsymbol{\beta}) = \mathbf{G}(\mathbf{y} + \mathbf{X}\boldsymbol{\beta})[\hat{\sigma}_i(\mathbf{y} + \mathbf{X}\boldsymbol{\beta}) - \sigma_i]$

$$= (\mathbf{G}\mathbf{y} + \mathbf{G}\mathbf{X}\boldsymbol{\beta})[\hat{\sigma}_i(\mathbf{y} + \mathbf{X}\boldsymbol{\beta}) - \sigma_i] = \mathbf{G}\mathbf{y}[\hat{\sigma}_i(\mathbf{y}) - \sigma_i] = g(\mathbf{y}),$$

and hence g is a translation invariant function.

Since $g(\mathbf{y}) = -g(\mathbf{y})$, then $E[g(\mathbf{y})] = E[-g(\mathbf{y})] = -E[g(\mathbf{y})]$

$$\Rightarrow 2E[g(\mathbf{y})] = \mathbf{0} \Rightarrow E[g(\mathbf{y})] = \mathbf{0}, \text{ and hence } E\left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}\right)(\hat{\sigma}_i - \sigma_i)\right] = \mathbf{0} \text{ for } i = 1, \dots, r \quad \square$$

Lemma 2.4.8 (a) $\frac{\partial \Phi}{\partial \sigma_i} = -\Phi \mathbf{P}_i \Phi$

(b) $\frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} = \Phi (\mathbf{P}_i \Phi \mathbf{P}_j + \mathbf{P}_j \Phi \mathbf{P}_i - \mathbf{Q}_{ij} - \mathbf{Q}_{ji} + \mathbf{R}_{ij}) \Phi,$

where $\mathbf{P}_i = -\mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \mathbf{X}$, $\mathbf{Q}_{ij} = \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{X}$,

and $\mathbf{R}_{ij} = \mathbf{X}' \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \sigma_i \partial \sigma_j} \Sigma^{-1} \mathbf{X}$

Proof (a) $\frac{\partial \Phi}{\partial \sigma_i} = -(\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \frac{\partial (\mathbf{X}' \Sigma^{-1} \mathbf{X})}{\partial \sigma_i} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}$
 $= -(\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} = -\Phi \mathbf{P}_i \Phi,$

(b) $\frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} = -\frac{\partial}{\partial \sigma_i} \left\{ (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \right\}$
 $= (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}$
 $- (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial^2 \Sigma^{-1}}{\partial \sigma_i \partial \sigma_j} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}$
 $+ (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}$
 $= \Phi \left\{ \mathbf{P}_i \Phi \mathbf{P}_j - \mathbf{X}' \frac{\partial^2 \Sigma^{-1}}{\partial \sigma_i \partial \sigma_j} \mathbf{X} + \mathbf{P}_j \Phi \mathbf{P}_i \right\} \Phi. \quad (2.2)$

Observe that $\mathbf{X}' \frac{\partial^2 \Sigma^{-1}}{\partial \sigma_i \partial \sigma_j} \mathbf{X} = \mathbf{X}' \frac{\partial}{\partial \sigma_i} \left(-\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \right) \mathbf{X}$
 $= \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{X} - \mathbf{X}' \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \sigma_i \partial \sigma_j} \Sigma^{-1} \mathbf{X} + \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \mathbf{X}$
 $= \mathbf{Q}_{ij} - \mathbf{R}_{ij} + \mathbf{Q}_{ji} \quad (2.3)$

Combining expressions 2.3 and 2.3, we obtain

$$\frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} = \Phi(\mathbf{P}_i \Phi \mathbf{P}_j + \mathbf{P}_j \Phi \mathbf{P}_i - \mathbf{Q}_{ij} - \mathbf{Q}_{ji} + \mathbf{R}_{ij})\Phi. \quad \square$$

Before proceeding, we present asymptotic orders for some terms that will be used often in chapters 2 and 3.

Lemma 2.4.9 (a) $\Phi = O(n^{-1})$, (b) $\mathbf{P}_i = O(n)$, (c) $\mathbf{Q}_{ij} = O(n)$, (d) $\mathbf{R}_{ij} = O(n)$,

$$(e) \Theta = O(n), \quad (f) w_{ij} = \text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j) = O(n^{-1}), \quad (g) \tilde{\Lambda} = O(n^{-2}), \quad (h) \frac{\partial \Phi}{\partial \sigma_i} = O(n^{-1}),$$

$$(i) \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} = O(n^{-1}), \quad (j) \frac{\partial w_{ij}}{\partial \sigma_i} = O(n^{-1}), \quad (k) \frac{\partial \tilde{\Lambda}}{\partial \sigma_i} = O(n^{-2}), \quad (l) \frac{\partial^2 w_{ij}}{\partial \sigma_i^2} = O(n^{-1}),$$

$$(m) \frac{\partial^2 \tilde{\Lambda}}{\partial \sigma_i} = O(n^{-2}), \quad (n) \hat{\sigma}_i - \sigma_i = O_p(n^{-1/2})$$

Proof Parts (b), (c), and (d) are direct from remark(v) above.

Parts (a) and (e) are direct from assumption (A6).

$w_{ij} = i^{ij} + a(\sigma)$, where i^{ij} is the (i, j) entry of the inverse of the expected information matrix (Pace and Salvan, 1997, expression 9.73). Pace and Salvan, showed that $i^{ij} = O(n^{-1})$, and $a(\sigma) = O(n^{-2})$.

From parts (a), (b), (c), and (f), we have $\tilde{\Lambda} = O(n^{-2})$.

$$\text{Since } \frac{\partial \Phi}{\partial \sigma_i} = -\Phi \mathbf{P}_i \Phi, \quad \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} = \Phi(\mathbf{P}_i \Phi \mathbf{P}_j + \mathbf{P}_j \Phi \mathbf{P}_i - \mathbf{Q}_{ij} - \mathbf{Q}_{ji} + \mathbf{R}_{ij})\Phi \text{ (lemma 2.4.8),}$$

then from previous parts, results (h) and (i) are obtained.

By using expression (9.17) in Pace and Salvan(1997), we have $\frac{\partial i^{ij}}{\partial \sigma_i} = O(n^{-1})$,

$$\frac{\partial^2 i^{ij}}{\partial \sigma_i^2} = O(n^{-1}), \quad \frac{\partial a(\sigma)}{\partial \sigma_i} = O(n^{-2}), \quad \text{and} \quad \frac{\partial^2 a(\sigma)}{\partial \sigma_i^2} = O(n^{-2}) \text{ hence results (j) and (l) are}$$

obtained. By computing the derivative of $\tilde{\Lambda}$, and using previous parts, results (k) and (m) are obtained. Finally, expression (n) is a direct result of the asymptotic normality of $\sqrt{n}(\hat{\sigma}_i - \sigma_i)$ (Pace and Salvan, expression 3.38). \square

Lemma 2.4.10 $\Lambda = \text{Var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \tilde{\Lambda} + O(n^{-5/2}),$

Proof Using a Taylor series expansion about $\boldsymbol{\sigma}$, we have

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\sigma}}) = \tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma}) + \sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}).$$

$$\Rightarrow \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} = \sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}),$$

$$\begin{aligned} \text{and } \Lambda = \text{Var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) &= \text{E} \left[\left(\sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}) \right) \left(\sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}) \right)' \right] \\ &\quad - \text{E} \left[\left(\sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}) \right) \right] \text{E} \left[\left(\sum_{i=1}^r \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + O_p(n^{-3/2}) \right)' \right] \end{aligned}$$

Applying lemma 2.4.7, we obtain

$$\begin{aligned} \Lambda &= \text{E} \left[\sum_{i=1}^r \sum_{j=1}^r \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right) \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right)' (\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j) \right] + O(n^{-5/2}) \\ &= \sum_{i=1}^r \sum_{j=1}^r \text{E} \left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right) \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right)' \right] \times \text{E} [(\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j)] + O(n^{-5/2}) \quad (\text{assumption A5}). \end{aligned} \quad (2.4)$$

By assumption (A3), we have

$$\text{E} [(\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j)] = \text{E}(\hat{\sigma}_i \hat{\sigma}_j) - \sigma_i \sigma_j + O(n^{-3/2}) = w_{ij} + O(n^{-3/2}) \quad (2.5)$$

$$\text{Since } \text{E} \left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right] = \mathbf{0} \quad (\text{lemma 2.4.7}), \text{ then } \text{E} \left[\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i} \right) \left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right)' \right] = \text{Cov} \left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}, \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right].$$

$$\begin{aligned} \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}' &= [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] \boldsymbol{\Sigma} [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] \\ &= \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] [\mathbf{I} - \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] \\ &= \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] = \mathbf{G}, \end{aligned}$$

Then, using expression (2.1), we obtain

$$\begin{aligned} \text{Cov} \left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}, \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j} \right] &= \boldsymbol{\Phi} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}] \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j} \boldsymbol{\Sigma}^{-1} \mathbf{X} \boldsymbol{\Phi} \\ &= \boldsymbol{\Phi} (\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j) \boldsymbol{\Phi} \end{aligned} \quad (2.6)$$

Combining expressions 2.4, 2.5 and 2.6 we have

$$\Lambda = \tilde{\Lambda} + O(n^{-5/2}), \quad \text{where } \tilde{\Lambda} = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi(\mathbf{Q}_{ij} - \mathbf{P}_i \Phi \mathbf{P}_j) \Phi. \quad \square$$

Lemma 2.4.11

- (a) $E[\hat{\Phi}] = \Phi - \tilde{\Lambda} + \mathbf{R}^* + O(n^{-5/2})$
- (b) $E[\hat{\Lambda}] = \tilde{\Lambda} + O(n^{-5/2})$
- (c) $E[\hat{\mathbf{R}}^*] = \mathbf{R}^* + O(n^{-5/2})$

$$\text{where } \mathbf{R}^* = \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi \mathbf{R}_{ij} \Phi, \quad \text{and } \hat{\Lambda} = \tilde{\Lambda}(\hat{\sigma}).$$

Proof (a) Using a Taylor series expansion about σ ,

$$\hat{\Phi} = \Phi + \sum_{i=1}^r \frac{\partial \Phi}{\partial \sigma_i} (\hat{\sigma}_i - \sigma_i) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) + O_p(n^{-5/2})$$

By assumption (A3), we have

$$\begin{aligned} E[\hat{\Phi}] &= \Phi + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} w_{ij} + O(n^{-5/2}) \\ &= \Phi + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi(\mathbf{P}_i \Phi \mathbf{P}_j + \mathbf{P}_j \Phi \mathbf{P}_i - \mathbf{Q}_{ij} + \mathbf{R}_{ij} - \mathbf{Q}_{ji}) \Phi + O(n^{-5/2}) \quad (\text{lemma 2.4.8}) \quad (2.7) \\ &= \Phi + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi(\mathbf{P}_i \Phi \mathbf{P}_j - \mathbf{Q}_{ij}) \Phi + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi(\mathbf{P}_j \Phi \mathbf{P}_i - \mathbf{Q}_{ji}) \Phi \\ &\quad + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi \mathbf{R}_{ij} \Phi + O(n^{-5/2}) \\ &= \Phi - \frac{1}{2} \tilde{\Lambda} - \frac{1}{2} \tilde{\Lambda} + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \Phi \mathbf{R}_{ij} \Phi + O(n^{-5/2}) = \Phi - \tilde{\Lambda} + \mathbf{R}^* + O(n^{-5/2}) \end{aligned}$$

(b) Using a Taylor series expansion about σ , we obtain

$$\hat{\Lambda} = \tilde{\Lambda} + O(n^{-5/2}), \quad \text{and then, } E[\hat{\Lambda}] = \tilde{\Lambda} + O(n^{-5/2}).$$

(c) Using a Taylor series expansion about σ , we obtain

$$\hat{\mathbf{R}}^* = \mathbf{R}^* + O(n^{-5/2}), \quad \text{and then, } E[\hat{\mathbf{R}}^*] = \mathbf{R}^* + O(n^{-5/2}).$$

Proposition 2.4.12 $E[\hat{\Phi}_A] = \text{Var}[\hat{\beta}] + O(n^{-5/2}),$

where $\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\Phi} \hat{\mathbf{P}}_j - \frac{1}{4} \hat{\mathbf{R}}_{ij}) \right\} \hat{\Phi}.$

Proof Notice that $\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda} - \hat{\mathbf{R}}^*,$

and hence $E[\hat{\Phi}_A] = \Phi - \tilde{\Lambda} + \mathbf{R}^* + 2\tilde{\Lambda} - \mathbf{R}^* + O(n^{-5/2}) = \Phi + \tilde{\Lambda} + O(n^{-5/2})$ (lemma 2.4.11)

$$= \Phi + \Lambda + O(n^{-5/2}) \quad (\text{lemma 2.4.9})$$

$$= \text{Var}(\hat{\beta}) + O(n^{-5/2}). \quad \square$$

Comments

From proposition 2.4.12, we obtain

$$E[\hat{\Phi}_A] = \text{Var}(\hat{\beta}) + O(n^{-5/2}),$$

and from lemma 2.4.10, we obtain

$$E[\hat{\Phi}] = \text{Var}(\hat{\beta}) + O(n^{-2}).$$

This shows that the adjusted estimator for $\text{Var}(\hat{\beta})$ has less bias than the conventional estimator.

3. TESTING THE FIXED EFFECTS

Consider the model as described in section 2.1. Suppose that we are interested in making inferences about ℓ linear combinations of the elements of $\boldsymbol{\beta}$. In other words, we are interested in testing $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ where \mathbf{L}' a fixed matrix of dimension $(\ell \times p)$. A common statistic often used is the Wald statistic:

$$F = \frac{1}{\ell} (\mathbf{L}'\hat{\boldsymbol{\beta}})' [\mathbf{L}'\hat{\mathbf{V}}\mathbf{L}]^{-1} (\mathbf{L}'\hat{\boldsymbol{\beta}}).$$

In fact, even though we call this statistic F , it does not necessarily have an F-distribution. Kenward and Roger (1997) approximate the distribution of F by choosing a scale λ and denominator degrees of freedom m such that $\lambda F \sim F(\ell, m)$ approximately.

3.1 Constructing a Wald-Type Pivot

The construction of a Wald-type pivot can be approached through either the adjusted estimator for the variance-covariance matrix of the fixed effects $\hat{\boldsymbol{\Phi}}_A$ (this is what was done by Kenward and Roger, 1997), or through the conventional estimator $\hat{\boldsymbol{\Phi}}$. Both approaches are to be considered in this section.

3.1.1 Constructing a Wald-Type Pivot Through $\hat{\boldsymbol{\Phi}}$

The Wald type pivot is $F = \frac{1}{\ell} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \mathbf{L}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

In this section, we will derive formulas for the $E[F]$ and $\text{Var}[F]$ approximately.

3.1.1.1 Deriving An Approximate Expression for $E[F]$

$$E[F] = E[E[F|\hat{\boldsymbol{\sigma}}]],$$

$$E[F|\hat{\boldsymbol{\sigma}}] = E\left[\frac{1}{\ell} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \mathbf{L}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \mid \hat{\boldsymbol{\sigma}}\right]$$

$$= \frac{1}{\ell} \left\{ E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L}] (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} E[\mathbf{L}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] + \text{tr} \left((\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \text{Var}[\mathbf{L}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \right) \right\},$$

using assumption (A4), and (Schott, 2005, theorem 10.18).

Since $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$ (lemma 2.4.3),

$$\begin{aligned} \text{then } \ell E[F|\hat{\boldsymbol{\sigma}}] &= \text{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})], \quad \text{where } \mathbf{V} = \text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi} + \boldsymbol{\Lambda} \\ &= \text{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})] + \text{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})]. \end{aligned}$$

Using a Taylor series expansion for $(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$ about $\boldsymbol{\sigma}$, we have

$$\begin{aligned} (\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1} &= (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial \sigma_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} + O_p(n^{-1/2}), \end{aligned}$$

and

$$\begin{aligned} (\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) &= \mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial \sigma_i} (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) \\ &\quad + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + O_p(n^{-3/2}). \end{aligned}$$

Since

$$\begin{aligned} \text{tr} \left[\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) \right] &= \\ 2 \text{tr} \left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} \mathbf{L}) (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \boldsymbol{\Phi}}{\partial \sigma_j} \mathbf{L}) \right] &- \text{tr} \left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} \mathbf{L}) \right], \quad (3.1) \end{aligned}$$

then by assumption (A3), we have

$$\begin{aligned} E[\text{tr}\{(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\}] &= \ell + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} \mathbf{L}) (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \boldsymbol{\Phi}}{\partial \sigma_j} \mathbf{L}) \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} \mathbf{L}) \right] + O(n^{-3/2}) \\ &= \ell + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_i \boldsymbol{\Phi} \boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{P}_j \boldsymbol{\Phi}] - \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\boldsymbol{\Theta} \frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + O(n^{-3/2}), \end{aligned}$$

where $\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}'$, and noticing that $\frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} = -\boldsymbol{\Phi} \mathbf{P}_i \boldsymbol{\Phi}$ (lemma 2.4.8).

Similar steps lead to $E[\text{tr}\{(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})\}] = \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + O(n^{-2})$.

Hence,

$$\ell E[F] = \ell + \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) - \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + O(n^{-3/2})$$

and,

$$\begin{aligned} E[F] &= 1 + \frac{1}{\ell} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) - \frac{1}{2\ell} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + \frac{1}{\ell} \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + O(n^{-3/2}) \\ &= 1 + \frac{1}{\ell} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) + \frac{2}{\ell} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\mathbf{\Theta}\mathbf{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i \mathbf{\Phi}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\mathbf{\Phi}] + O(n^{-3/2}). \end{aligned}$$

Proposition 3.1.1.1.1 $E[F] = 1 + \frac{A_2 + 2A_3}{\ell} + O(n^{-3/2}),$

where

$$A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}),$$

and $A_3 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\mathbf{\Theta}\mathbf{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i \mathbf{\Phi}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\mathbf{\Phi}]$

3.1.1.2 Deriving an Approximate Expression for $\text{Var}[F]$

$$\text{Var}[F] = E[\text{Var}[F | \hat{\mathbf{\sigma}}]] + \text{Var}[E[F | \hat{\mathbf{\sigma}}]]$$

To derive the right-hand side, we consider each term at a time. The first term will be derived in part A, and the second term will be derived in part B.

A. $\text{Var}[F | \hat{\mathbf{\sigma}}] = \frac{1}{\ell^2} \text{Var}[(\hat{\mathbf{\beta}} - \mathbf{\beta})' \mathbf{L}(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1} \mathbf{L}'(\hat{\mathbf{\beta}} - \mathbf{\beta}) | \hat{\mathbf{\sigma}}]$

The expression above is a quadratic form, and hence by assumption (A4),

$$\text{Var}[F | \hat{\mathbf{\sigma}}] = \frac{2}{\ell^2} \text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})] \quad (\text{Schott, 2005, theorem 10.22}).$$

Using a Taylor series expansion for $(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}$ about $\mathbf{\sigma}$, we have

$$\begin{aligned}
(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) &= (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad + \frac{1}{2} \sum_{j=1}^r \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) + O_p(n^{-3/2})
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow [(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 = [(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 \\
&\quad + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad + \frac{1}{2} (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad + \left(\sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \right) (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad + \left(\sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \right) \left(\sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \right) \\
&\quad + \frac{1}{2} \left(\sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) \right) (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) + O_p(n^{-3/2}),
\end{aligned}$$

and

$$\begin{aligned}
\text{tr}[(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 &= \text{tr} \left\{ [(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 \right. \\
&\quad + 2(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&\quad \left. + \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) \right\} + O_p(n^{-3/2}).
\end{aligned}$$

Taking the expectation and by assumption (A3), we obtain

$$\begin{aligned}
\mathbb{E}[\text{Var}[F \mid \hat{\boldsymbol{\sigma}}]] &= \frac{2}{\ell^2} \left\{ \text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 \right. \\
&\quad + \text{tr} \left((\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} \right) \\
&\quad \left. + \text{tr} \left((\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L}) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_j} \right) \right\} + O(n^{-3/2})
\end{aligned}$$

To proceed, we need to derive each term mentioned in the expression above.

$$\begin{aligned}
(i) \quad & [(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2 \\
&= (\mathbf{L}'\Phi\mathbf{L})^{-1}[(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})](\mathbf{L}'\Phi\mathbf{L})^{-1}[(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})] \\
&= [\mathbf{I} + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L})][\mathbf{I} + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L})] \\
&= \mathbf{I} + 2(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) \\
&= \mathbf{I} + 2(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) + O(n^{-2})
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
(ii) \quad & (\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) = [(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})](\mathbf{L}'\Phi\mathbf{L})^{-1}[(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})] \\
&= (\mathbf{L}'\Phi\mathbf{L}) + 2(\mathbf{L}'\Lambda\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L})
\end{aligned}$$

$$\begin{aligned}
\text{So, } \text{tr} \left[(\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\partial^2 (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} \right] \\
= \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2 (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\Phi\mathbf{L}) \right] + 2 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2 (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\Lambda\mathbf{L}) \right] \\
+ \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2 (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\Lambda\mathbf{L})(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) \right] \\
= \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2 (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}'\Phi\mathbf{L}) \right] + O(n^{-2}) \\
= 2 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi \Theta \mathbf{P}_j \Phi) - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\Theta \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} \right) + O(n^{-2}),
\end{aligned} \tag{3.3}$$

where the last expression above is obtained by utilizing (3.1).

Alternatively, we can utilize lemma 2.4.8, to rewrite expression (3.3) as

$$\begin{aligned}
& 2A_2 - 2 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\Theta \Phi (\mathbf{P}_i \Phi \mathbf{P}_j - \mathbf{Q}_{ij} + \frac{1}{2} \mathbf{R}_{ij}) \Phi] + O(n^{-2}). \\
(iii) \quad & \frac{\partial (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_i} (\mathbf{L}'\mathbf{V}\mathbf{L}) \frac{\partial (\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial \sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&= (\mathbf{L}'\Phi\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \Phi}{\partial \sigma_i} \mathbf{L}) (\mathbf{L}'\Phi\mathbf{L})^{-1} [(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})] \times \\
&\quad (\mathbf{L}'\Phi\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \Phi}{\partial \sigma_j} \mathbf{L}) (\mathbf{L}'\Phi\mathbf{L})^{-1} [(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\Lambda\mathbf{L})] \\
&= (\mathbf{L}'\Phi\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \Phi}{\partial \sigma_i} \mathbf{L}) (\mathbf{L}'\Phi\mathbf{L})^{-1} (\mathbf{L}' \frac{\partial \Phi}{\partial \sigma_j} \mathbf{L})
\end{aligned}$$

$$\begin{aligned}
& +(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_j}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L}) \\
& +(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_j}\mathbf{L}) + O(n^{-2})
\end{aligned}$$

And hence,

$$\begin{aligned}
& \text{tr}\left(\sum_{i=1}^r\sum_{j=1}^rw_{ij}\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})\right) \\
& = \text{tr}\left[\sum_{i=1}^r\sum_{j=1}^rw_{ij}\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j\mathbf{\Phi})\right] + O(n^{-2}) \\
& = A_2 + O(n^{-2})
\end{aligned} \tag{3.4}$$

From (i), (ii), and (iii), we obtain

$$\begin{aligned}
& \mathbb{E}[\text{Var}[F \mid \hat{\boldsymbol{\sigma}}]] = \\
& \frac{2}{\ell^2}\left\{\ell + 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + 3A_2 - 2\sum_{i=1}^r\sum_{j=1}^rw_{ij}\text{tr}[\mathbf{\Theta}\mathbf{\Phi}(\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j - \mathbf{Q}_{ij} + \frac{1}{2}\mathbf{R}_{ij})\mathbf{\Phi}]\right\} + O(n^{-\frac{3}{2}})
\end{aligned} \tag{3.5}$$

$$\begin{aligned}
\text{B. } \text{Var}[\mathbb{E}[F \mid \hat{\boldsymbol{\sigma}}]] &= \frac{1}{\ell^2} \text{Var}[\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]] \\
&= \frac{1}{\ell^2}\left\{\mathbb{E}\left(\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 - \left(\mathbb{E}\left\{\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right\}\right)^2\right\}
\end{aligned}$$

For the first term above, and by using a Taylor series expansion for $(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}$ about $\boldsymbol{\sigma}$,

$$\begin{aligned}
\left(\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 &= \text{tr}\left[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) + \sum_{i=1}^r(\hat{\sigma}_i - \sigma_i)\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\right. \\
&\quad \left.+ \frac{1}{2}\sum_{i=1}^r\sum_{j=1}^r(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L}) + O_p(n^{-\frac{3}{2}})\right] \times (\text{same terms}) \\
&= \left(\text{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 + 2\text{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\sum_{i=1}^r(\hat{\sigma}_i - \sigma_i)\text{tr}\left[\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\right] \\
&\quad + \text{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\sum_{i=1}^r\sum_{j=1}^r(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\text{tr}\left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]
\end{aligned}$$

$$+ \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \text{tr} \left[\frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] \text{tr} \left[\frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] + O_p(n^{-3/2})$$

Taking the expectation, and by assumption (A4), we have

$$\begin{aligned} E\left(\text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 &= \left(\text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 \\ &+ \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] \text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})] \\ &+ \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] \text{tr} \left[\frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] + O(n^{-3/2}) \end{aligned}$$

To proceed, we need to derive each term mentioned in the expression above.

$$\begin{aligned} (i) \quad & (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) = \mathbf{I} + (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) \\ \Rightarrow & \text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})] = \ell + \text{tr}(\Theta\Lambda) \\ \Rightarrow & \left(\text{tr}[(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 = \ell^2 + 2\ell\text{tr}(\Theta\Lambda) + O(n^{-2}) \end{aligned} \quad (3.6)$$

$$\begin{aligned} (ii) \quad & \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{V}\mathbf{L}) = \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\Phi\mathbf{L}) + \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\Lambda\mathbf{L}) \\ & = -(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}' \frac{\partial\Phi}{\partial\sigma_i} \mathbf{L} - (\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}' \frac{\partial\Phi}{\partial\sigma_i} \mathbf{L} (\mathbf{L}'\Phi\mathbf{L})^{-1} (\mathbf{L}'\Lambda\mathbf{L}) \end{aligned}$$

$$\text{So, } \text{tr} \left[\frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] = -\text{tr}(\Theta\Phi\mathbf{P}_i\Phi) - \text{tr}(\Theta\Phi\mathbf{P}_i\Phi\Theta\Lambda), \quad (3.7)$$

$$(iii) \quad \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) = \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\Phi\mathbf{L}) + \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\Lambda\mathbf{L}),$$

$$\begin{aligned} \text{and } & \text{tr} \left[\frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\mathbf{V}\mathbf{L}) \right] \\ & = \text{tr} \left[\frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\Phi\mathbf{L}) \right] + \text{tr} \left[\frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i \partial\sigma_j} (\mathbf{L}'\Lambda\mathbf{L}) \right] \\ & = 2\text{tr}(\Theta\Phi\mathbf{P}_i\Phi\Theta\Phi\mathbf{P}_j\Phi) - \text{tr} \left(\Theta \frac{\partial^2\Phi}{\partial\sigma_i \partial\sigma_j} \right) + 2\text{tr}(\Theta\Phi\mathbf{P}_i\Phi\Theta\Phi\mathbf{P}_j\Phi\Theta\Lambda) \\ & \quad - \text{tr} \left(\Theta \frac{\partial^2\Phi}{\partial\sigma_i \partial\sigma_j} \Theta\Lambda \right) \quad (\text{from expression 3.1}). \end{aligned} \quad (3.8)$$

From (i), (ii), and (iii),

$$\begin{aligned}
& \mathbb{E} \left(\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})] \right)^2 = \ell^2 + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) \\
& + \sum_{i=1}^r \sum_{j=1}^r w_{ij} [\ell + \text{tr}(\mathbf{\Theta}\mathbf{\Lambda})] \left\{ -\text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + 2\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) \right. \\
& + 2\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) - \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \mathbf{\Theta}\mathbf{\Lambda} \right) \left. \right\} \\
& + \sum_{i=1}^r \sum_{j=1}^r w_{ij} [\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}) + \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda})] [\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) + \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda})] + O(n^{-\frac{3}{2}}) \\
& = \ell^2 + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \left\{ 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) - \ell \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right. \\
& + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) - \ell \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \mathbf{\Theta}\mathbf{\Lambda} \right) + 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) \\
& - \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) - \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \mathbf{\Theta}\mathbf{\Lambda} \right) \\
& + \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) + 2\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) + \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Lambda}) \left. \right\} \\
& + O(n^{-\frac{3}{2}}) \\
& = \ell^2 + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \left\{ 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) - \ell \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right\} \\
& + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \left\{ \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}) \right\} + O(n^{-\frac{3}{2}}) \\
& = \ell^2 + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + 2\ell A_2 + A_1 - \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) + O(n^{-\frac{3}{2}}), \tag{3.9}
\end{aligned}$$

where
$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i \mathbf{\Phi}) \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_j \mathbf{\Phi}).$$

Also, from expression 3.2,

$$\left(\mathbb{E} \left[\text{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})] \right] \right)^2 = \ell^2 + 2\ell(A_2 + 2A_3) + O(n^{-2}) \tag{3.10}$$

From expressions (3.9) and (3.10), we obtain

$$\begin{aligned}
& \text{Var}[E[F|\hat{\mathbf{\Theta}}]] \\
&= \frac{1}{\ell^2} \left\{ \ell^2 + 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + 2\ell A_2 + A_1 - \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) - \ell^2 - 2\ell(A_2 + 2A_3) \right\} + O(n^{-\frac{3}{2}}) \\
&= \frac{1}{\ell^2} \left\{ 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + A_1 - \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) - 4\ell A_3 \right\} + O(n^{-\frac{3}{2}})
\end{aligned}$$

From parts (A) and (B), we are able to find an expression for $\text{Var}[F]$.

$$\begin{aligned}
\text{Var}[F] &= \frac{2}{\ell^2} \left\{ \ell + 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + 3A_2 - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right\} \\
&\quad + \frac{1}{\ell^2} \left\{ 2\ell \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) + A_1 - \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) - 4\ell A_3 \right\} + O(n^{-\frac{3}{2}}) \\
&= \frac{2}{\ell} \left\{ 1 + \frac{3A_2}{\ell} + \frac{A_1}{2\ell} - 2A_3 + \left(\frac{2}{\ell} + 1 \right) \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - \left(\frac{1}{\ell} + \frac{1}{2} \right) \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right\} + O(n^{-\frac{3}{2}}) \\
&= \frac{2}{\ell} \left\{ 1 + \frac{1}{2\ell} \left[A_1 + 6A_2 - 4\ell A_3 + 2(\ell + 2) \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - (\ell + 2) \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right] \right\} + O(n^{-\frac{3}{2}}) \\
&= \frac{2}{\ell} \left\{ 1 + \frac{1}{2\ell} \left[A_1 + 6A_2 - 4\ell A_3 + (\ell + 2) \left\{ 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) \right\} \right] \right\} + O(n^{-\frac{3}{2}}) \\
&= \frac{2}{\ell} \left\{ 1 + \frac{1}{2\ell} \left[A_1 + 6A_2 - 4\ell A_3 + (\ell + 2) \left\{ 4\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{R}_{ij}\mathbf{\Phi}) \right\} \right] \right\} + O(n^{-\frac{3}{2}}),
\end{aligned}$$

by noticing that

$$\begin{aligned}
& - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left(\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right) = \\
& 2\text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{R}_{ij}\mathbf{\Phi}) + O(n^{-\frac{5}{2}}) \quad (\text{lemmas 2.4.8 and 2.4.10}),
\end{aligned}$$

and
$$A_3 = \text{tr}(\mathbf{\Theta}\mathbf{\Lambda}) - \frac{1}{4} \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{R}_{ij}\mathbf{\Phi}) + O(n^{-\frac{5}{2}}).$$

Proposition 3.1.1.2.1
$$\text{Var}[F] = \frac{2}{\ell} (1 + B_1) + O(n^{-\frac{3}{2}}),$$

where
$$B_1 = \frac{1}{2\ell} [A_1 + 6A_2 + 8A_3]$$

3.1.2 Constructing a Wald-Type Pivot Through $\hat{\Phi}_A$

The Wald type pivot is $F = \frac{1}{\ell} (\hat{\beta} - \beta)' L (L' \hat{\Phi}_A L)^{-1} L' (\hat{\beta} - \beta)$. Similar to section

3.1.1, we derive expressions for $E[F]$ and $\text{Var}[F]$ approximately

3.1.2.1 Deriving an Approximate Expression for $E[F]$

First, we derive a useful lemma that will be needed in this section and chapter 4.

Lemma 3.1.2.1.1 For a matrix $\mathbf{B} = O_p(n^{-1})$, $(\mathbf{I} + \mathbf{B})^{-1} = \mathbf{I} - \mathbf{B} + O_p(n^{-2})$

Proof $(\mathbf{I} + \mathbf{B})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{B})^{-1} \mathbf{B}$ (Schott, 2005, theorem 1.7)

$$= \mathbf{I} - [\mathbf{I} - (\mathbf{I} + \mathbf{B})^{-1} \mathbf{B}] \mathbf{B}, \quad \text{applying the theorem again for } (\mathbf{I} + \mathbf{B})^{-1}$$

Notice that since $\mathbf{B} \xrightarrow{p} \mathbf{0}$, then $(\mathbf{I} + \mathbf{B})^{-1} \xrightarrow{p} (\mathbf{I} + \mathbf{0})^{-1} = \mathbf{I}$, and hence $(\mathbf{I} + \mathbf{B})^{-1} = O_p(1)$.

Therefore, $(\mathbf{I} + \mathbf{B})^{-1} = \mathbf{I} - \mathbf{B} + (\mathbf{I} + \mathbf{B})^{-1} \mathbf{B} \mathbf{B} = \mathbf{I} - \mathbf{B} + O(n^{-2})$. \square

$$E[F|\hat{\sigma}] = \frac{1}{\ell} \text{tr} \left((L' \hat{\Phi}_A L)^{-1} \text{Var}[L'(\hat{\beta} - \beta)] \right), \text{ by assumption (A4).}$$

$$\Rightarrow \ell E[F] = E \left(\text{tr}[(L' \hat{\Phi}_A L)^{-1} (L' \mathbf{V} L)] \right), \quad \text{where } \mathbf{V} = \text{Var}(\hat{\beta}) = \Phi + \Lambda$$

$$= \text{tr} \left(E[(L' \hat{\Phi}_A L)^{-1} (L' \mathbf{V} L)] \right)$$

$$\text{Since } \hat{\Phi}_A = \hat{\Phi} + \hat{\mathbf{A}}^* \quad \text{where } \hat{\mathbf{A}}^* = 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\Phi} \hat{\mathbf{P}}_j - \frac{1}{4} \hat{\mathbf{R}}_{ij}) \right\} \hat{\Phi},$$

$$\begin{aligned} \text{then } (L' \hat{\Phi}_A L)^{-1} (L' \mathbf{V} L) &= \left[(L' \hat{\Phi} L) + (L' \hat{\mathbf{A}}^* L) \right]^{-1} (L' \mathbf{V} L) \\ &= \left\{ (L' \hat{\Phi} L) \left[\mathbf{I} + (L' \hat{\Phi} L)^{-1} (L' \hat{\mathbf{A}}^* L) \right] \right\}^{-1} (L' \mathbf{V} L) \\ &= \left[\mathbf{I} + (L' \hat{\Phi} L)^{-1} (L' \hat{\mathbf{A}}^* L) \right]^{-1} (L' \hat{\Phi} L)^{-1} (L' \mathbf{V} L). \end{aligned}$$

Since $(L' \hat{\Phi} L)^{-1} (L' \hat{\mathbf{A}}^* L) = O_p(n^{-1})$, then by employing the lemma 3.1.2.1.1,

$$\begin{aligned}
(\mathbf{L}'\hat{\Phi}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) &= \left[\mathbf{I} - (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) + O_p(n^{-2}) \right] (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&= (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) - (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) + O_p(n^{-2}) \quad (3.11)
\end{aligned}$$

Notice that $E\left(\text{tr}[(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right) = \ell + A_2 + 2A_3 + O(n^{-3/2})$, (3.12)
(propagation 3.1.1.1.1)

$$\begin{aligned}
&\text{and } (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) \\
&= (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Phi\mathbf{L}) + (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Lambda\mathbf{L}) \\
&= (\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Phi\mathbf{L}) + O_p(n^{-2}).
\end{aligned}$$

Using a Taylor series expansion for $(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}$ about $\boldsymbol{\sigma}$, we have

$$\begin{aligned}
&(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Phi\mathbf{L}) + O_p(n^{-2}) \\
&= \left[(\mathbf{L}'\Phi\mathbf{L})^{-1} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} \right] (\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) \\
&\times \left[(\mathbf{L}'\Phi\mathbf{L})^{-1} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i} + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\Phi\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} \right] (\mathbf{L}'\Phi\mathbf{L}) + O_p(n^{-2}) \\
&= (\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) + O_p(n^{-3/2}) \quad (3.13)
\end{aligned}$$

From expressions (3.12 and 3.13), we obtain

$$\ell E[F] = \ell + A_2 + 2A_3 - E\left[\text{tr}(\Theta\hat{\mathbf{A}}^*)\right] + O(n^{-3/2}) \quad (3.14)$$

Lemma $E[\hat{\mathbf{A}}^*] = \mathbf{A}^* + O(n^{-3/2})$

Proof Direct from Lemmas 2.4.10 and 2.4.11 \square

Applying the lemma above on expression (3.14), we have

Proposition 3.1.2.1.2 $E[F] = 1 + \frac{A_2}{\ell} + O(n^{-3/2})$

3.1.2.2 Deriving an Approximate Expression for $\text{Var}[F]$

$$\text{Var}[F] = E[\text{Var}[F | \hat{\boldsymbol{\sigma}}]] + \text{Var}[E[F | \hat{\boldsymbol{\sigma}}]]$$

$$\text{A. } \text{Var}[F | \hat{\boldsymbol{\sigma}}] = \frac{1}{\ell^2} \text{Var}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}\mathbf{L}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) | \hat{\boldsymbol{\sigma}}]$$

The expression above is a quadratic form and hence by assumption (A4),

$$\text{Var}[F | \hat{\boldsymbol{\sigma}}] = \frac{2}{\ell^2} \text{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2] \quad (\text{Schott, 2005, theorem 10.22}).$$

Recall that from expression (3.11), we had

$$(\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) = (\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) - \mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) + O_p(n^{-2}).$$

$$\begin{aligned} \text{and hence, } \text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]^2 &= \text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]^2 \\ &\quad - 2\text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right] + O_p(n^{-2}) \\ &= \text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + (\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}))(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + (\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}))\right] \\ &\quad - 2\text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + (\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}))(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + (\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}))\right] + O_p(n^{-2}) \\ &= \text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] + 2\text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right] \\ &\quad - 2\text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] + O_p(n^{-2}). \end{aligned}$$

Using a Taylor series expansion for $(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$ about $\boldsymbol{\sigma}$, we have

$$\begin{aligned} \text{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]^2 &= \\ \text{tr}\left\{\left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] \right. \\ &\quad \times \left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] \Big\} \\ &\quad + 2\text{tr}\left\{\left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] \right. \\ &\quad \times \left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}) + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right] \Big\} \\ &\quad - 2\text{tr}\left\{\left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] \right. \\ &\quad \times \left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i^2}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] \Big\} \end{aligned}$$

$$\begin{aligned}
& \times \left[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) \right. \\
& \quad \left. + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} (\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L}) \right] \Big\} + O_p(n^{-2}) \\
& = \text{tr} \left[\mathbf{I} + 2 \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} (\mathbf{L}'\mathbf{\Phi}\mathbf{L}) \right] \\
& \quad + \text{tr} \left[\sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j) \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i} (\mathbf{L}'\mathbf{\Phi}\mathbf{L}) \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j} (\mathbf{L}'\mathbf{\Phi}\mathbf{L}) \right] \\
& \quad + 2\text{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})] - 2\text{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})] + O_p(n^{-\frac{3}{2}}).
\end{aligned}$$

By taking the expectation and assumption (A3),

$$\begin{aligned}
\mathbb{E}[\text{Var}[F \mid \hat{\boldsymbol{\sigma}}]] &= \frac{2}{\ell^2} \left\{ \ell + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \right] \right. \\
& \quad \left. + \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \right] + 2\text{tr}[\mathbf{\Theta}\mathbf{\Lambda}] - 2\text{tr}[\mathbf{\Theta}\mathbf{A}^*] \right\} + O(n^{-\frac{3}{2}}).
\end{aligned}$$

$$\text{Notice that } \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \right] = A_2,$$

$$\text{and from (3.1), } \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j} \mathbf{L}'\mathbf{\Phi}\mathbf{L} \right] = 2A_2 - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta} \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})}{\partial\sigma_i\partial\sigma_j} \right].$$

Hence,

$$\mathbb{E}[\text{Var}[F \mid \hat{\boldsymbol{\sigma}}]] = \frac{2}{\ell^2} \left\{ \ell + 3A_2 + 2\text{tr}[\mathbf{\Theta}\mathbf{\Lambda}] - 2\text{tr}[\mathbf{\Theta}\mathbf{A}^*] - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta} \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})}{\partial\sigma_i\partial\sigma_j} \right] \right\} + O(n^{-\frac{3}{2}}).$$

In addition, by lemmas 2.4.8 and 2.4.10, we have

$$\begin{aligned}
& 2\text{tr}[\mathbf{\Theta}\mathbf{\Lambda}] - 2\text{tr}[\mathbf{\Theta}\mathbf{A}^*] - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta} \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})}{\partial\sigma_i\partial\sigma_j} \right] \\
& = 2\text{tr}[\mathbf{\Theta}\tilde{\mathbf{\Lambda}}] - 2\text{tr}[\mathbf{\Theta}\mathbf{A}^*] - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta} \frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})}{\partial\sigma_i\partial\sigma_j} \right] + O(n^{-\frac{3}{2}}) \\
& = 2 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta}\mathbf{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\mathbf{\Phi} \right] - 4 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr} \left[\mathbf{\Theta}\mathbf{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\mathbf{\Phi} \right]
\end{aligned}$$

$$-2 \sum_{i=1}^r \sum_{j=1}^r w_{ij} \operatorname{tr} \left[\mathbf{\Theta} \mathbf{\Phi} (\mathbf{P}_i \mathbf{\Phi} \mathbf{P}_j - \mathbf{Q}_{ij} + \frac{1}{2} \mathbf{R}_{ij}) \mathbf{\Phi} \right] + O(n^{-3/2}) = 0 + O(n^{-3/2}) = O(n^{-3/2}) \quad (3.15)$$

$$\Rightarrow \mathbb{E}[\operatorname{Var}[F \mid \hat{\boldsymbol{\sigma}}]] = \frac{2}{\ell^2} \{ \ell + 3A_2 \} + O(n^{-3/2}) \quad (3.16)$$

$$\begin{aligned} \text{B. } \operatorname{Var}[\mathbb{E}[F \mid \hat{\boldsymbol{\sigma}}]] &= \frac{1}{\ell^2} \operatorname{Var}[\operatorname{tr}[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L})]], \text{ by assumption (A4)} \\ &= \frac{1}{\ell^2} \left\{ \mathbb{E} \left(\operatorname{tr}[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L})] \right)^2 - \left(\mathbb{E} \left\{ \operatorname{tr}[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L})] \right\} \right)^2 \right\}, \end{aligned}$$

$$\begin{aligned} (i) \quad & \left(\operatorname{tr}[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L})] \right)^2 \\ &= \left\{ \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) - (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}}) (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) + O_p(n^{-2}) \right] \right\}^2 \quad (\text{from 3.11}) \\ &= \left\{ \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) \right] \right\}^2 - 2 \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) \right] \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}}) (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) \right] \\ &\quad + O_p(n^{-2}) \\ &= \left\{ \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L}) \right] \right\}^2 - 2 \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}}) (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] \\ &\quad + O_p(n^{-2}) \quad (\text{recall that } \mathbf{V} = \mathbf{\Phi} + \mathbf{\Lambda}). \end{aligned}$$

$$\begin{aligned} \text{From expression (3.9), } & \mathbb{E} \left(\operatorname{tr}[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{V} \mathbf{L})] \right)^2 \\ &= \ell^2 + 2\ell \operatorname{tr}[\mathbf{\Theta} \mathbf{\Lambda}] + 2\ell A_2 + A_1 - \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \operatorname{tr} \left[\mathbf{\Theta} \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} \right] + O(n^{-3/2}). \end{aligned}$$

(ii) Using a Taylor series expansion for $(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1}$ about $\boldsymbol{\sigma}$, we have

$$\begin{aligned} & \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] \operatorname{tr} \left[(\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{L}})^{-1} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] = \\ & \operatorname{tr} \left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j) \frac{\partial^2 (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] \\ & \times \operatorname{tr} \left\{ \left[\mathbf{I} + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j) \frac{\partial^2 (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}' \mathbf{\Phi} \mathbf{L}) \right] \right. \\ & \times \left[(\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1} (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) + \sum_{i=1}^r (\hat{\sigma}_i - \sigma_i) \frac{\partial (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i} (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) \right. \\ & \left. \left. + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r (\hat{\sigma}_i - \sigma_i) (\hat{\sigma}_j - \sigma_j) \frac{\partial^2 (\mathbf{L}' \mathbf{\Phi} \mathbf{L})^{-1}}{\partial \sigma_i \partial \sigma_j} (\mathbf{L}' \hat{\mathbf{A}}^* \mathbf{L}) \right] \right\} + O_p(n^{-3/2}). \end{aligned}$$

By assumption (A3), we obtain

$$E\left(\text{tr}\left[(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Phi\mathbf{L})\right]\text{tr}\left[(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\Phi\mathbf{L})\right]\right) = \ell\text{tr}[\Theta\mathbf{A}^*] + O(n^{-3/2}).$$

From (i) and (ii),

$$E\left(\text{tr}[(\mathbf{L}'\hat{\Phi}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 = \ell^2 + 2\ell\text{tr}[\Theta\mathbf{A}] + 2\ell A_2 + A_1$$

$$- \sum_{i=1}^r \sum_{j=1}^r \ell w_{ij} \text{tr}\left[\Theta \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j}\right] - 2\ell\text{tr}[\Theta\mathbf{A}^*] + O(n^{-3/2}),$$

$$\text{and since } 2\text{tr}[\Theta\mathbf{A}] - \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}\left[\Theta \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j}\right] - 2\text{tr}[\Theta\mathbf{A}^*] = O(n^{-3/2}) \text{ (from 3.17),}$$

then

$$E\left(\text{tr}[(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2 = \ell^2 + 2\ell A_2 + A_1 + O(n^{-3/2}). \quad (3.17)$$

Recall that from expression (3.14), we have

$$\begin{aligned} E\left(\text{tr}[(\mathbf{L}'\hat{\Phi}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right) &= \ell E[F] = \ell + A_2 + O(n^{-3/2}) \\ \Rightarrow \left\{E\left(\text{tr}[(\mathbf{L}'\hat{\Phi}_A\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)\right\}^2 &= \ell^2 + 2\ell A_2 + O(n^{-3/2}). \end{aligned} \quad (3.18)$$

From expressions (3.17 and 3.18), we obtain

$$\begin{aligned} \text{Var}[E[F \mid \hat{\sigma}]] &= \frac{1}{\ell^2} \left\{ \ell^2 + 2\ell A_2 + A_1 - \ell^2 - 2\ell A_2 \right\} + O(n^{-3/2}) \\ &= \frac{A_1}{\ell^2} + O(n^{-3/2}), \end{aligned} \quad (3.19)$$

Therefore from expressions (3.16 and 3.19),

$$\textbf{Proposition 3.1.2.2.1} \quad \text{Var}[F] = \frac{2}{\ell}(1 + B) + O(n^{-3/2}),$$

$$\text{where} \quad B = \frac{1}{2\ell}(A_1 + 6A_2).$$

Summary

(i) When the adjusted estimate of the variance-covariance matrix is used in constructing the Wald-type statistic,

$$E[F] \approx \tilde{E} = 1 + \frac{A_2}{\ell},$$

and
$$\text{Var}[F] \approx \tilde{V} = \frac{2}{\ell}(1 + B),$$

where
$$B = \frac{1}{2\ell}(A_1 + 6A_2),$$

and
$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi), \quad A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi)$$

(ii) When the conventional estimate of the variance-covariance matrix is used,

$$E[F] \approx \tilde{E}_1 = 1 + \frac{A_2 + 2A_3}{\ell}$$

and
$$\text{Var}[F] \approx \tilde{V}_1 = \frac{2}{\ell}(1 + B_1),$$

where
$$B_1 = \frac{6A_2 + A_1 + 8A_3}{2\ell},$$

and
$$A_3 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}[\Theta \Phi (\mathbf{Q}_{ij} - \mathbf{P}_i \Phi \mathbf{P}_j - \frac{1}{4} \mathbf{R}_{ij}) \Phi]$$

Or alternatively,

$$E[F] \approx \tilde{E}_1 = 1 + \frac{S_2}{\ell}$$

and
$$\text{Var}[F] \approx \tilde{V}_1 = \frac{2}{\ell}(1 + B_1),$$

where
$$B_1 = \frac{6S_2 + S_1}{2\ell}, \quad \text{and} \quad S_2 = A_2 + 2A_3, \quad S_1 = A_1 - 4A_3$$

3.2 Estimating the Denominator Degrees of Freedom and the Scale Factor

In order to determine λ and m such that $F^* = \lambda F \sim F(\ell, m)$ approximately, we match the approximated first two moments of F^* with those of $F(\ell, m)$ distribution.

$$E[F(\ell, m)] = \frac{m}{m-2},$$

$$\text{Var}[F(\ell, m)] = 2 \left(\frac{m}{m-2} \right)^2 \frac{\ell + m - 2}{\ell(m-4)}.$$

In the previous section, we found expressions for \tilde{E} and \tilde{V} such that

$$E[F] \approx \tilde{E} = 1 + \frac{A_2}{\ell},$$

and
$$\text{Var}[F] \approx \tilde{V} = \frac{2}{\ell}(1 + B),$$

By matching the first moments,

$$E[F^*] \approx \lambda \tilde{E} = \frac{m}{m-2}$$

$$\Rightarrow \tilde{\lambda} = \frac{\tilde{m}}{\tilde{E}(\tilde{m}-2)}.$$

By matching the second moments,

$$\text{Var}[F^*] \approx \lambda^2 \tilde{V} = 2 \left(\frac{m}{m-2} \right)^2 \frac{\ell + m - 2}{\ell(m-4)}$$

$$\Leftrightarrow \tilde{V} = 2 \left[\frac{m}{(m-2)\lambda} \right]^2 \frac{\ell + m - 2}{\ell(m-4)} \quad \Leftrightarrow \frac{\tilde{V}}{2\tilde{E}^2} = \frac{\ell + m - 2}{\ell(m-4)}$$

and hence $\tilde{m} = 4 + \frac{2 + \ell}{\ell \tilde{\rho} - 1}$, where $\tilde{\rho} = \frac{\tilde{V}}{2\tilde{E}^2}$.

The scale and denominator degrees of freedom are to be estimated by substituting $\hat{\sigma}$ for σ in A_1 , A_2 , and A_3 . The quantity w_{ij} can be estimated by using the inverse of the expected information matrix.

4. MODIFYING THE ESTIMATES OF THE DENOMINATOR DEGREES OF FREEDOM AND THE SCALE FACTOR

As we saw in chapter three, the Wald-type statistic F does not have an F distribution usually; however, the scaled form of F will have an F distribution approximately. Indeed, there are number of special cases where the scaled form of F follows an F distribution exactly. Kenward and Roger (1997) considered two of these special cases; the Hotelling T^2 and the conventional Anova models, and then they modified the approximate expressions for the expectation and the variance of F so the modified expressions produce estimates of the denominator degree of freedom and the scale that match the exact and known values for the special cases. In this chapter, we derive Kenward and Roger (K-R) modification based on the Hotelling T^2 and the balanced one-way Anova models.

4.1 Balanced One-Way Anova Model

4.1.1 The Model and Assumptions

Consider the model: $y_{ij} = \mu + \tau_i + e_{ij}$ for $i = 1, \dots, t$, and $j = 1, \dots, m$,

where μ is the general mean, τ_i are the treatment effects, and $e_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$

Suppose that we are interested in testing $H_0: \tau_1 = \dots = \tau_t \Leftrightarrow \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$,

$$\text{where } \mathbf{L}' = \begin{bmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}, \quad \boldsymbol{\beta}' = [\mu \quad \tau_1 \quad \cdots \quad \tau_t]$$

$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, where $n = mt$. The design matrix; \mathbf{X} is not of full column rank. We can proceed with this parameterization, using the g-inverse, however, to simplify our computations, we will reparameterize the model so \mathbf{X} will be a full column rank matrix

as follows $\boldsymbol{\beta}^{*'} = [\mu^* \quad \tau_1^* \quad \cdots \quad \tau_{t-1}^*]$ with $\sum_{i=1}^t \tau_i^* = 0$, and the hypothesis becomes

$$H_0: \mathbf{L}^{*'} \boldsymbol{\beta}^* = \mathbf{0}, \quad \text{where} \quad \mathbf{L}^{*'} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

To simplify our notation, from now on, consider $\mathbf{X} = \mathbf{X}^*$, $\mathbf{L} = \mathbf{L}^*$, and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

4.1.2 The REML Estimates of σ

Since the design is balanced, then the REML estimates are the same as the Anova estimates (Searle *et al.*, 1992, section 3.8), and hence

$$\hat{\sigma}_{\text{REML}}^2 = \text{MSE} = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}}{n - t}.$$

4.1.3 Computing \mathbf{P}_1 , \mathbf{Q}_{11} and \mathbf{R}_{11}

$$\begin{aligned} \mathbf{P}_1 &= -\mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2} \boldsymbol{\Sigma}^{-1} \mathbf{X} = -\frac{1}{(\sigma^2)^2} \mathbf{X}'\mathbf{X}, \\ \mathbf{Q}_{11} &= \mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2} \boldsymbol{\Sigma}^{-1} \mathbf{X} = \frac{1}{(\sigma^2)^3} \mathbf{X}'\mathbf{X}, \\ \text{and} \quad \mathbf{R}_{11} &= \mathbf{X}'\boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial (\sigma^2)^2} \boldsymbol{\Sigma}^{-1} \mathbf{X} = \mathbf{0}, \text{ because } \frac{\partial^2 \boldsymbol{\Sigma}}{\partial (\sigma^2)^2} = \mathbf{0} \end{aligned}$$

In addition, since $\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$, then $\mathbf{Q}_{11} - \mathbf{P}_1\boldsymbol{\Phi}\mathbf{P}_1 = \mathbf{0}$,

and hence $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$. Recall that $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\boldsymbol{\Phi}} \hat{\mathbf{P}}_j - \frac{1}{4} \hat{\mathbf{R}}_{ij}) \right\} \hat{\boldsymbol{\Phi}}$.

4.1.4 Computing A_1 and A_2

$$A_1 = w_{11} [\text{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_1\boldsymbol{\Phi})]^2, \quad \text{and} \quad A_2 = w_{11} \text{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_1\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_1\boldsymbol{\Phi}).$$

Since $\hat{\sigma}_{\text{REML}}^2 = \text{MSE}$, and $\frac{(n-t)\text{MSE}}{\sigma^2} \sim \chi_{n-t}^2$, then $w_{11} = \text{Var}(\hat{\sigma}_{\text{REML}}^2) = \frac{2(\sigma^2)^2}{n-t}$.

Also, $\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_1\boldsymbol{\Phi} = -\frac{1}{\sigma^2} \mathbf{L} [\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

$$= -\frac{1}{\sigma^2} \mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}$$

Notice that $\mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}$ is a p.o. on $\mathfrak{R}(\mathbf{L})$

$$\begin{aligned} \Rightarrow \text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_1\mathbf{\Phi}) &= -\frac{1}{\sigma^2} \text{tr} \left(\mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \right) \\ &= -\frac{1}{\sigma^2} r \left(\mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \right) \quad (\text{Harville, 1997, corollary 10.2.2}) \\ &= -\frac{1}{\sigma^2} r(\mathbf{L}) = -\frac{\ell}{\sigma^2} \quad (\ell = t-1). \end{aligned}$$

Therefore,

$$\begin{aligned} A_1 &= \frac{2(\sigma^2)^2}{n-t} \frac{\ell^2}{(\sigma^2)^2} = \frac{2\ell^2}{n-t}, \\ A_2 &= \frac{2(\sigma^2)^2}{n-t} \frac{\ell}{(\sigma^2)^2} = \frac{2\ell}{n-t}. \end{aligned}$$

Accordingly, for this model, we have $A_1 = \ell A_2$,

and
$$B = \frac{1}{2\ell} (A_1 + 6A_2) = \frac{\ell + 6}{n-t}.$$

4.1.5 Estimating the Denominator Degrees of Freedom and the Scale Factor

Under the null hypothesis,

$$F = \frac{1}{\ell \text{MSE}} \mathbf{y}' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

Notice that $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ is an o.p.o (clear),

and

$$\begin{aligned} r(\mathbf{P}) &= \text{tr}(\mathbf{P}) = \text{tr} \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \\ &= \text{tr} \left(\mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \right) \end{aligned}$$

In addition, $\mathbf{L} \left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{L} \right]^{-1} \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}$ is a p.o. on $\mathfrak{R}(\mathbf{L})$, and hence $r(\mathbf{P}) = \ell$,

and
$$F = \frac{\mathbf{y}' \mathbf{P} \mathbf{y}}{r(\mathbf{P}) \text{MSE}}.$$

Since $\mathfrak{R}(\mathbf{P}) \subset \mathfrak{R}(\mathbf{X})$, then under the null hypothesis, $F \sim F(\ell, n-t)$ exactly,

where $n-t = r(\mathbf{I} - \mathbf{P}_X)$ (Birkes, 2003, lemm 4.4).

In other words, the denominator degrees of freedom should equal to $n-t$, and the scale should equal to one.

$$\text{Since } A_1 = \frac{2\ell^2}{n-t}, \quad A_2 = \frac{2\ell}{n-t}, \quad \text{and} \quad B = \frac{\ell+6}{n-t},$$

$$\text{then } E[F] \approx \tilde{E} = 1 + \frac{A_2}{\ell} = 1 + \frac{2}{n-t}, \quad \text{Var}[F] \approx \tilde{V} = \frac{2}{\ell}(1+B) = \frac{2}{\ell} \left(1 + \frac{\ell+6}{n-t}\right).$$

$$\Rightarrow \tilde{\rho} = \frac{\tilde{V}}{2\tilde{E}^2} = \frac{(n-t)}{\ell} \left[\frac{n-t+\ell+6}{(n-t+2)^2} \right],$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell\tilde{\rho}-1} = 4 + \frac{(n-t+2)^2(\ell+2)}{(n-t)(\ell-2)-4}, \quad \text{and} \quad \tilde{\lambda} = \frac{\tilde{m}}{\tilde{E}(\tilde{m}-2)} = \left(\frac{n-t}{n-t+2} \right) \left(\frac{\tilde{m}}{\tilde{m}-2} \right)$$

$$\begin{aligned} \text{It can be seen that } \tilde{m}-4 &= \frac{(n-t+2)^2(\ell+2)}{(n-t)(\ell-2)-4} > \frac{(n-t+2)^2(\ell+2)}{(n-t)(\ell+2)} \\ &= \frac{(n-t+2)^2}{n-t} = \frac{(n-t)^2 + 4(n-t) + 4}{n-t} = n-t+4 + \frac{4}{n-t} \end{aligned}$$

So, $\tilde{m} > n-t+8$, and this means that the approach overestimates the denominator degrees of freedom. The approach in this setting does not perform well especially when $n-t$ is a small value. Moreover, since $\tilde{m} > n-t+8$, then $2\tilde{m}-2(n-t) > 4$

$$\Rightarrow 2\tilde{m}-2(n-t)-4 + \tilde{m}(n-t) > \tilde{m}(n-t)$$

$$\Rightarrow \tilde{\lambda} = \frac{n-t}{n-t+2} \frac{\tilde{m}}{\tilde{m}-2} = \frac{\tilde{m}(n-t)}{2\tilde{m}-2(n-t)-4 + \tilde{m}(n-t)} < 1,$$

and again the approach doesn't perform in the way we wish.

4.2 The Hotelling's T^2 model

4.2.1 The Model and Assumptions

Suppose that $\mathbf{y}_1, \dots, \mathbf{y}_n$ are n independent observations from a p -variate population which has a normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$.

$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\mu}$ where \mathbf{X} is $np \times p$ design matrix, and $\boldsymbol{\mu}$ is $p \times 1$ vector.

$$\mathbf{X} = \begin{bmatrix} \mathbf{I}_p \\ \vdots \\ \mathbf{I}_p \end{bmatrix} = \mathbf{1}_n \otimes \mathbf{I}_p, \quad \Sigma = \begin{bmatrix} \Sigma_p & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma_p \end{bmatrix} = \mathbf{I}_n \otimes \Sigma_p$$

Also, suppose that we are interested in testing: $H_0: \boldsymbol{\mu} = \mathbf{0}$

Special Notation Since $\boldsymbol{\sigma} = \{\sigma_{if}\}_{\substack{p \times (p+1) \\ 2}} (i \leq f)$, then we will use the notation $\mathbf{P}_{if}, \mathbf{P}_{jg}$,

$\mathbf{Q}_{if,jg}$, $\mathbf{R}_{if,jg}$, and $w_{if,jg}$.

Lemma 4.2.2 In the Hotelling T^2 model, $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$.

Proof
$$\begin{aligned} \mathbf{X}'\Sigma^{-1}\mathbf{X} &= (\mathbf{1}_n \otimes \mathbf{I}_p)'(\mathbf{I}_p \otimes \Sigma_p^{-1})(\mathbf{1}_n \otimes \mathbf{I}_p) \\ &= \mathbf{1}_n' \mathbf{1}_p \mathbf{1}_n \otimes \mathbf{I}_p \Sigma_p^{-1} \mathbf{I}_p = n \Sigma_p^{-1} \quad (\text{Harville, 1997, chapter 16}). \end{aligned}$$

$$\Rightarrow \boldsymbol{\Phi} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} = (n \Sigma_p^{-1})^{-1} = \frac{\Sigma_p}{n}.$$

$$\mathbf{P}_{if} = -\mathbf{X}'\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{if}} \Sigma^{-1} \mathbf{X} = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_{if}} \mathbf{X} = n \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}},$$

$$\mathbf{P}_{if} \boldsymbol{\Phi} \mathbf{P}_{jg} = n \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \Sigma_p \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}},$$

$$\text{and } \mathbf{Q}_{if,jg} = \mathbf{X}'\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{if}} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{jg}} \Sigma^{-1} \mathbf{X} = \mathbf{X}' \left(\frac{\partial \Sigma^{-1}}{\partial \sigma_{if}} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{jg}} \right) \mathbf{X} = n \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \Sigma_p \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}}.$$

$$\Rightarrow \mathbf{Q}_{if,jg} - \mathbf{P}_{if} \boldsymbol{\Phi} \mathbf{P}_{jg} = \mathbf{0},$$

Since $\frac{\partial^2 \Sigma}{\partial \sigma_{if} \partial \sigma_{jg}} = \mathbf{0}$, then $\mathbf{R}_{if,jg} = \mathbf{0}$. We conclude that $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$.

Recall that $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\boldsymbol{\Phi}} \hat{\mathbf{P}}_j - \frac{1}{4} \hat{\mathbf{R}}_{ij}) \right\} \hat{\boldsymbol{\Phi}}$ \square

Theorem 4.2.3 The REML estimate of Σ_p is $\mathbf{S} = \frac{\mathbf{A}}{n-1}$,

$$\text{where } \mathbf{A} = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})', \quad \text{and } \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$$

Proof: When we are interested in deriving the REML estimate, we consider our data as

$\mathbf{z} = \mathbf{K}'\mathbf{y}$ such that $\mathbf{K}'\mathbf{X} = \mathbf{0}$.

Choose \mathbf{M} such that $\mathbf{M}'\mathbf{M} = \mathbf{I}_{n-1}$, $\mathbf{M}\mathbf{M}' = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$ (Birkes, 2004, chapter 8)

We can choose $\mathbf{K} = \mathbf{M} \otimes \mathbf{I}_p$, and this is true because

$$\mathbf{K}'\mathbf{X} = (\mathbf{M}' \otimes \mathbf{I}_p)(\mathbf{1}_n \otimes \mathbf{I}_p) = \mathbf{M}'\mathbf{1}_n \otimes \mathbf{I}_p = \mathbf{0} \text{ (notice that } \Re(\mathbf{M}') = \Re(\mathbf{1}_n)^\perp)$$

$$\text{Also, } \mathbf{K}'\mathbf{K} = (\mathbf{M}' \otimes \mathbf{I}_p)(\mathbf{M} \otimes \mathbf{I}_p) = \mathbf{I}_{n-1} \otimes \mathbf{I}_p = \mathbf{I}_{np-p}$$

$$\text{and } \mathbf{K}\mathbf{K}' = (\mathbf{M} \otimes \mathbf{I}_p)(\mathbf{M}' \otimes \mathbf{I}_p) = \mathbf{M}\mathbf{M}' \otimes \mathbf{I}_p = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}) \otimes \mathbf{I}_p = \mathbf{I}_{np} - \mathbf{P}_{\mathbf{1}_n} \otimes \mathbf{I}_p$$

$$\text{Observe that since } \mathbf{P}_X = (\mathbf{1}_n \otimes \mathbf{I}_p) \left[(\mathbf{1}_n' \otimes \mathbf{I}_p)(\mathbf{1}_n \otimes \mathbf{I}_p) \right]^{-1} (\mathbf{1}_n' \otimes \mathbf{I}_p) = \mathbf{P}_{\mathbf{1}_n} \otimes \mathbf{I}_p,$$

$$\text{then } \mathbf{K}\mathbf{K}' = \mathbf{I}_{np} - \mathbf{P}_X.$$

Since $\mathbf{z} = \mathbf{K}'\mathbf{y} = (\mathbf{M}' \otimes \mathbf{I}_p)\mathbf{y}$, then

$$\begin{aligned} \text{Var}[\mathbf{z}] &= \text{Var}[(\mathbf{M}' \otimes \mathbf{I}_p)\mathbf{y}] = (\mathbf{M}' \otimes \mathbf{I}_p)\Sigma(\mathbf{M} \otimes \mathbf{I}_p) \\ &= (\mathbf{M}' \otimes \mathbf{I}_p)(\mathbf{I}_n \otimes \Sigma_p)(\mathbf{M} \otimes \mathbf{I}_p) = \mathbf{M}'\mathbf{M} \otimes \Sigma_p. \end{aligned}$$

$$\text{So, } \text{Var}[\mathbf{z}] = \mathbf{M}'\mathbf{M} \otimes \Sigma_p = \mathbf{I}_{n-1} \otimes \Sigma_p.$$

In fact, $\text{Var}[\mathbf{z}]$ has the same construction as $\text{Var}[\mathbf{y}]$, but with $n-1$ blocks of Σ_p instead of n blocks. $\mathbf{z}_1, \dots, \mathbf{z}_{n-1}$ are iid $N_p(\mathbf{0}, \Sigma_p)$.

Lemma When $\mathbf{z}_1, \dots, \mathbf{z}_{n-1}$ are iid $N_p(\mathbf{0}, \Sigma_p)$, then $\frac{\sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i'}{n-1}$ is the ML estimator of Σ_p .

$$\text{Proof } \ell = \text{constant} - \frac{1}{2}(n-1)\log|\Sigma_p| - \frac{1}{2} \sum_{i=1}^{n-1} \mathbf{z}_i' \Sigma_p^{-1} \mathbf{z}_i$$

$$\text{Notice that } \sum_{i=1}^{n-1} \mathbf{z}_i' \Sigma_p^{-1} \mathbf{z}_i = \text{tr} \left(\sum_{i=1}^{n-1} \mathbf{z}_i' \Sigma_p^{-1} \mathbf{z}_i \right) = \text{tr} \left(\sum_{i=1}^{n-1} \Sigma_p^{-1} \mathbf{z}_i \mathbf{z}_i' \right) = \text{tr} \left(\Sigma_p^{-1} \sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i' \right)$$

$$\Rightarrow \ell = \text{constant} - \frac{1}{2}(n-1)\log|\Sigma_p| - \frac{1}{2} \text{tr}(\Sigma_p^{-1} \mathbf{A}), \text{ and hence the MLE of } \Sigma_p \text{ is } \frac{\sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i'}{n-1}$$

(Anderson, 2003, lemma 3.2.2), and (Rao, 2002, chapter 8). \square

Remark

Since $\mathbf{M}\mathbf{M}' = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$,

$$\sum_{i=1}^{n-1} m_{ji}^2 = \frac{n-1}{n} \quad \text{for } j=1, \dots, n$$

$$\sum_{i=1}^{n-1} m_{ji} m_{ki} = \frac{-1}{n} \quad \text{for } j \neq k, \quad j, k=1, \dots, n \quad \square$$

Now, since $\mathbf{z} = \mathbf{K}'\mathbf{y} = (\mathbf{M} \otimes \mathbf{I}_p)\mathbf{y}$, then $\mathbf{z}_i = \sum_{j=1}^n m_{ji} \mathbf{y}_j$,

$$\text{and} \quad \mathbf{z}_i \mathbf{z}_i' = \left(\sum_{j=1}^n m_{ji} \mathbf{y}_j \right) \left(\sum_{j=1}^n m_{ji} \mathbf{y}_j' \right) = \sum_{k=1}^n \sum_{j=1}^n m_{ji} m_{ki} \mathbf{y}_j \mathbf{y}_k'$$

$$\begin{aligned} \text{So, } \sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i' &= \sum_{i=1}^{n-1} \sum_{k=1}^n \sum_{j=1}^n m_{ji} m_{ki} \mathbf{y}_j \mathbf{y}_k' = \sum_{k=1}^n \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_k' \left(\sum_{i=1}^{n-1} m_{ji} m_{ki} \right) \\ &= \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' \left(\frac{n-1}{n} \right) + \sum_{\substack{j=1, \\ k \neq j}}^n \sum_{k=1}^n \mathbf{y}_j \mathbf{y}_k' \left(\frac{-1}{n} \right) \quad (\text{from the remark above}) \\ &= \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j \sum_{k=1}^n \mathbf{y}_k' \\ &= \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n} \left[n \bar{\mathbf{y}} n \bar{\mathbf{y}}' - \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' \right] = \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j' - n \bar{\mathbf{y}} \bar{\mathbf{y}}' \end{aligned}$$

$$\text{So, } \sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i' = \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' - n \bar{\mathbf{y}} \bar{\mathbf{y}}'$$

$$\text{and since } \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' + n \bar{\mathbf{y}} \bar{\mathbf{y}}' \quad (\text{Anderson, 2003, lemma 3.2.1}),$$

$$\text{then } \sum_{i=1}^{n-1} \mathbf{z}_i \mathbf{z}_i' = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'.$$

$$\text{We conclude that the REML estimate of } \Sigma_p = \mathbf{S} = \frac{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'}{n-1} = \frac{\mathbf{A}}{n-1}. \quad \square$$

4.2.4 Computing A_1 and A_2

Since we are testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$, then $\mathbf{L} = \mathbf{I}_p$, $\ell = p$, $\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}' = \boldsymbol{\Phi}^{-1}$,

$$A_1 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{\substack{j=1, \\ j \leq g}}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\mathbf{\Theta} \mathbf{\Phi} \mathbf{P}_{if} \mathbf{\Phi}) \text{tr}(\mathbf{\Theta} \mathbf{\Phi} \mathbf{P}_{jg} \mathbf{\Phi}) = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{\substack{j=1, \\ j \leq g}}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\mathbf{P}_{if} \mathbf{\Phi}) \text{tr}(\mathbf{P}_{jg} \mathbf{\Phi}), \quad (4.1)$$

$$A_2 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{\substack{j=1, \\ j \leq g}}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\mathbf{\Theta} \mathbf{\Phi} \mathbf{P}_{if} \mathbf{\Phi} \mathbf{\Theta} \mathbf{\Phi} \mathbf{P}_{jg} \mathbf{\Phi}) = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{\substack{j=1, \\ j \leq g}}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\mathbf{P}_{if} \mathbf{\Phi} \mathbf{P}_{jg} \mathbf{\Phi}) \quad (4.2)$$

$$\begin{aligned} \text{tr}(\mathbf{P}_{if} \mathbf{\Phi}) &= -2^{1-\delta_{if}} \sigma^{if}, \quad \text{and} \quad \text{tr}(\mathbf{P}_{jg} \mathbf{\Phi}) = -2^{1-\delta_{jg}} \sigma^{jg} \\ \Rightarrow \text{tr}(\mathbf{P}_{if} \mathbf{\Phi}) \text{tr}(\mathbf{P}_{jg} \mathbf{\Phi}) &= 2^{2-\delta_{if}-\delta_{jg}} \sigma^{if} \sigma^{jg}, \end{aligned} \quad (4.3)$$

$$\text{where} \quad \delta_{if} = \begin{cases} 1 & \text{for } i = f \\ 0 & \text{for } i \neq f \end{cases}, \quad \text{and} \quad \Sigma_p^{-1} = \{\sigma^{if}\}_{p \times p}$$

$\mathbf{S} \sim \text{Wishart} \left[\frac{\Sigma_p}{n-1}, n-1 \right]$, and $\mathbf{A} \sim \text{Wishart} [\Sigma_p, n-1]$ (Anderson, 2003, corollary 7.2.3).

Also, Anderson provided in section 7.3 that the characteristic function of $A_{11}, A_{22}, \dots, A_{pp}, 2A_{12}, 2A_{13}, \dots, 2A_{p-1,p}$ as

$$\mathbb{E} \left[e^{i \text{tr}(\mathbf{A} \mathbf{\Theta})} \right] = \left| \mathbf{I} - 2i \mathbf{\Theta} \Sigma_p \right|^{-N/2} = \Psi,$$

where $\mathbf{\Theta} = \{\theta_{ij}\}_{p \times p}$ with $\theta_{ij} = \theta_{ji}$, and $\mathbf{A} = \{A_{ij}\}_{p \times p}$, $N = n-1$

Finding the moments

For the first moments,

$$\begin{aligned} \mathbb{E}[A_{ij}] &= \frac{1}{i} \times \frac{\partial \Psi}{\partial \theta_{ij}} \Big|_{\mathbf{\Theta}=\mathbf{0}} \\ &= \frac{1}{i} \left(-\frac{1}{2} N \right) \left| \mathbf{I} - 2i \mathbf{\Theta} \Sigma_p \right|^{-\frac{1}{2} N-1} \left| \mathbf{I} - 2i \mathbf{\Theta} \Sigma_p \right| \text{tr} \left[\left(\mathbf{I} - 2i \mathbf{\Theta} \Sigma_p \right)^{-1} \frac{\partial}{\partial \theta_{ij}} \left(\mathbf{I} - 2i \mathbf{\Theta} \Sigma_p \right) \right] \Big|_{\mathbf{\Theta}=\mathbf{0}} \\ &= N \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{ij}} \Sigma_p \right) = \begin{cases} (n-1) \sigma_{ii} & \text{for } i = j \\ 2(n-1) \sigma_{ij} & \text{for } i \neq j \end{cases}, \quad \text{and hence } \mathbb{E}[\hat{\sigma}_{ij}] = \sigma_{ij} \end{aligned}$$

For the second moments,

$$2^{2-\delta_{if}-\delta_{jg}} \mathbb{E}[A_{if} A_{jg}] = \frac{1}{i^2} \times \frac{\partial^2 \Psi}{\partial \theta_{if} \partial \theta_{jg}} \Big|_{\mathbf{\Theta}=\mathbf{0}}$$

$$\begin{aligned}
&= \frac{1}{i^2} \frac{\partial}{\partial \theta_{jg}} \left\{ \frac{-N}{2} \left| \mathbf{I} - 2i\mathbf{\Theta}\mathbf{\Sigma}_p \right|^{-N/2} \text{tr} \left[(\mathbf{I} - 2i\mathbf{\Theta}\mathbf{\Sigma}_p)^{-1} \left(-2i \frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) \right] \right\} \Big|_{\mathbf{\Theta}=\mathbf{0}} \\
&= N^2 \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) + 2N \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) \\
\Rightarrow 2^{2-\delta_{if}-\delta_{jg}} \mathbb{E}[\mathbf{A}_{if} \mathbf{A}_{jg}] &= (n-1) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) + 2(n-1) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) \\
\Rightarrow \text{Cov}[\mathbf{A}_{if}, \mathbf{A}_{jg}] &= \mathbb{E}[\mathbf{A}_{if} \mathbf{A}_{jg}] - \mathbb{E}[\mathbf{A}_{if}] \mathbb{E}[\mathbf{A}_{jg}] \\
&= \left(\frac{1}{2^{2-\delta_{if}-\delta_{jg}}} \right) \left\{ (n-1)^2 \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) + 2(n-1) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) \right. \\
&\quad \left. - (n-1)^2 \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \right) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) \right\} \\
&= \left(\frac{2(n-1)}{2^{2-\delta_{if}-\delta_{jg}}} \right) \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) \\
\Rightarrow \text{Cov}[\hat{\sigma}_{if}, \hat{\sigma}_{jg}] &= \frac{2 \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right)}{(n-1) 2^{2-\delta_{if}-\delta_{jg}}} \tag{4.4}
\end{aligned}$$

$$\text{By noticing that } \text{tr} \left(\frac{\partial \mathbf{\Theta}}{\partial \theta_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Theta}}{\partial \theta_{jg}} \mathbf{\Sigma}_p \right) = 2^{1-\delta_{if}-\delta_{jg}} (\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij}), \tag{4.5}$$

$$\text{Cov}[\hat{\sigma}_{if}, \hat{\sigma}_{jg}] = \frac{\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij}}{(n-1)} \tag{4.6}$$

Combining expressions (4.1), (4.3), and (4.6), we obtain

$$\begin{aligned}
A_1 &= \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p \frac{\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij}}{(n-1)} 2^{2-\delta_{if}-\delta_{jg}} \sigma^{if} \sigma^{jg} \\
&= \sum_{i=1}^p \sum_{f=1}^p \sum_{j=1}^p \sum_{g=1}^p \frac{\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij}}{(n-1) 2^{2-\delta_{if}-\delta_{jg}}} 2^{2-\delta_{if}-\delta_{jg}} \sigma^{if} \sigma^{jg}
\end{aligned}$$

Note In the last expression above for A_1 , we divided by $2^{2-\delta_{if}-\delta_{jg}}$, and this is true because the summations repeat $w_{if,jg} 2^{2-\delta_{if}-\delta_{jg}}$ times.

$$\begin{aligned}
\Rightarrow A_1 &= \frac{1}{n-1} \sum_{g=1}^p \sum_{f=1}^p \sum_{j=1}^p \sum_{i=1}^p \sigma_{ig} \sigma^{if} \sigma_{jf} \sigma^{jg} + \sigma_{fg} \sigma^{jg} \sigma_{ij} \sigma^{if} \\
&= \frac{1}{n-1} \left\{ \sum_{f=1}^p \sum_{g=1}^p \left(\sum_{i=1}^p \sigma_{ig} \sigma^{if} \right) \left(\sum_{j=1}^p \sigma_{jf} \sigma^{jg} \right) + \sum_{j=1}^p \sum_{f=1}^p \left(\sum_{g=1}^p \sigma_{fg} \sigma^{jg} \right) \left(\sum_{i=1}^p \sigma_{ij} \sigma^{if} \right) \right\} \\
&= \frac{1}{n-1} \left\{ \sum_{f=1}^p \sum_{g=1}^p (\delta_{gf} \delta_{gf}) + \sum_{j=1}^p \sum_{f=1}^p (\delta_{ff} \delta_{ff}) \right\}
\end{aligned}$$

Therefore,

$$A_1 = \frac{2p}{n-1}$$

Also,

$$\begin{aligned}
\text{tr}(\mathbf{P}_{if} \mathbf{\Phi} \mathbf{P}_{jg} \mathbf{\Phi}) &= \text{tr} \left(n \mathbf{\Sigma}_p^{-1} \frac{\partial \mathbf{\Sigma}_p^{-1}}{\partial \sigma_{if}} \mathbf{\Sigma}_p \frac{\partial \mathbf{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \mathbf{\Sigma}_p \right) = \text{tr} \left(\mathbf{\Sigma}_p^{-1} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \mathbf{\Sigma}_p^{-1} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}} \right) \\
&= (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij}) 2^{1-\delta_{if}-\delta_{ij}}, \quad \text{by utilizing expression (4.5) above} \quad (4.7)
\end{aligned}$$

Combining expressions (4.2), (4.6), and (4.7), we obtain

$$\begin{aligned}
A_2 &= \sum_{i=1}^p \sum_{f=1, \substack{j=1, \\ i \leq f}}^p \sum_{j=1, \substack{g=1, \\ j \leq g}}^p \frac{(\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij})}{(n-1)} (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij}) 2^{1-\delta_{if}-\delta_{ij}} \\
&= \sum_{i=1}^p \sum_{f=1}^p \sum_{j=1}^p \sum_{g=1}^p \frac{(\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij})}{(n-1) 2^{2-\delta_{if}-\delta_{ij}}} (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij}) 2^{1-\delta_{if}-\delta_{ij}} \\
&= \frac{1}{2(n-1)} \left\{ \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{ig} \sigma^{ig} \sigma_{jf} \sigma^{jf} + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{ig} \sigma^{fg} \sigma_{jf} \sigma^{ij} \right. \\
&\quad \left. + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{fg} \sigma^{ig} \sigma_{ij} \sigma^{jf} + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{fg} \sigma^{fg} \sigma_{ij} \sigma^{ij} \right\} \\
&= \frac{1}{2(n-1)} \left\{ \sum_{g=1}^p \sum_{f=1}^p (1)(1) + \sum_{i=1}^p \sum_{f=1}^p \delta_{if} \delta_{if} + \sum_{i=1}^p \sum_{f=1}^p \delta_{if} \delta_{if} + \sum_{g=1}^p \sum_{j=1}^p (1)(1) \right\} \\
&= \frac{1}{2(n-1)} (p^2 + p + p + p^2)
\end{aligned}$$

Therefore,
$$A_2 = \frac{p(p+1)}{n-1}$$

4.2.5 Estimating the Denominator Degrees of Freedom and the Scale Factor

When testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$, a scaled Hotelling T^2 has an exact F-test (Krzanowski, 2000, chapter 8). Indeed, the denominator degrees of freedom m should equal $n - p$, and the scale λ should equal $\frac{n - p}{n - p + p - 1} = \frac{n - p}{n - 1}$.

Since $A_1 = \frac{2\ell}{n-1}$, $A_2 = \frac{\ell(\ell+1)}{n-1}$, and $B = \frac{3\ell+4}{n-1}$ (note that $\ell = p$), then

$$\tilde{E} = 1 + \frac{A_2}{\ell} = 1 + \frac{\ell+1}{n-1}, \quad \text{and} \quad \tilde{V} = \frac{2}{\ell}(1+B) = \frac{2}{\ell}\left(1 + \frac{3\ell+4}{n-1}\right)$$

$$\Rightarrow \tilde{\rho} = \frac{\tilde{V}}{2\tilde{E}^2} = \left(\frac{n+3\ell+3}{n+\ell}\right)\left(\frac{n-1}{\ell}\right),$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell\tilde{\rho}-1} = 4 + \frac{(n+\ell)(\ell+2)}{n^2 + n - 4\ell - 3 + 3n\ell},$$

$$\text{and} \quad \tilde{\lambda} = \frac{\tilde{m}}{\tilde{E}(\tilde{m}-2)} = \left(\frac{n-1}{n+\ell}\right)\left(\frac{\tilde{m}}{\tilde{m}-2}\right).$$

The estimates do not match the exact values for this model.

4.3 Kenward and Roger's Modification

From section 4.1, we found that for the balanced one-way Anova model,

$$A_1 = \frac{2\ell^2}{n-t}, \quad A_2 = \frac{2\ell}{n-t}, \quad B = \frac{\ell+6}{n-t}, \quad \text{and} \quad \frac{A_1}{A_2} = \ell,$$

$$\tilde{m} = 4 + \frac{\ell+2}{\frac{\tilde{V}}{\ell \frac{\tilde{V}}{2\tilde{E}^2} - 1}} \neq n-t,$$

$$\text{and} \quad \tilde{\lambda} = \frac{n-t}{\tilde{E}(n-t-2)} \neq 1$$

And from section 4.2, we found that for the Hotelling T^2 model,

$$A_1 = \frac{2\ell}{n-1}, \quad A_2 = \frac{\ell(\ell+1)}{n-1}, \quad B = \frac{3\ell+4}{n-1}, \quad \text{and} \quad \frac{A_1}{A_2} = \frac{2}{\ell+1},$$

$$\tilde{m} = 4 + \frac{\ell+2}{\frac{\tilde{V}}{\ell \frac{\tilde{V}}{2\tilde{E}^2} - 1}} \neq n-\ell,$$

and
$$\tilde{\lambda} = \frac{n-\ell}{\tilde{E}(n-\ell-2)} \neq \frac{n-\ell}{n-1}.$$

Instead of estimating $E[F]$ and $\text{Var}[F]$ by \tilde{E} and \tilde{V} , we will construct modified estimators E^* and V^* leading to

$$m^* = 4 + \frac{\ell+2}{\frac{\ell}{2E^{*2}} - 1}, \quad (4.8)$$

and
$$\lambda^* = \frac{m^*}{E^*(m^*-2)} \quad (4.9)$$

such that m^* and λ^* match the exact values for the denominator degrees of freedom and the scale respectively for the balanced one-way Anova and the Hotelling T^2 models.

Applying expression (4.9) on the balanced one-way Anova model,

$$\frac{n-t}{E^*(n-t-2)} = 1,$$

The expression above, can be rewritten in A_2 term as

$$\frac{\ell}{E^*(\ell - A_2)} = 1,$$

and hence
$$E^* = \frac{1}{1 - \frac{A_2}{\ell}} \quad (4.10)$$

Similarly, applying expression (4.9) on the Hotelling T^2 ,

$$\frac{n-\ell}{E^*(n-\ell-2)} = \frac{n-\ell}{n-1},$$

The expression above, can be rewritten in term of A_2 as

$$\begin{aligned} \frac{1}{E^* \left[\frac{\ell(\ell+1)}{A_2} - (\ell+1) \right]} &= \frac{A_2}{\ell(\ell+1)} \\ \Rightarrow \frac{1}{E^* [\ell(\ell+1) - (\ell+1)A_2]} &= \frac{1}{\ell(\ell+1)}, \end{aligned}$$

and hence
$$E^* = \frac{1}{1 - \frac{A_2}{\ell}} \quad (4.11)$$

From expressions (4.10 and 4.11), we can see that the approximate expression for $E[F]$ should be as

$$E^* = \frac{1}{1 - \frac{A_2}{\ell}} \quad (4.12)$$

Applying expression (4.8) on the balanced one-way Anova model,

$$4 + \frac{\ell + 2}{\ell \frac{V^*}{2E^{*2}} - 1} = n - t$$

The expression above can be rewritten in B term as

$$\begin{aligned} 4 + \frac{\ell + 2}{\ell \frac{V^*}{2E^{*2}} - 1} &= \frac{\ell + 6}{B} \\ \Rightarrow \frac{2(\ell + 2)E^{*2}}{\ell V^* - 2E^{*2}} &= \frac{\ell + 6}{B} - 4, \\ V^* &= \frac{2}{\ell} \left\{ E^{*2} \left(1 + \frac{\ell + 2}{\frac{\ell + 6}{B} - 4} \right) \right\}. \end{aligned} \quad (4.13)$$

Moreover, since for the balanced one-way Anova model,

$$B = \frac{\ell + 6}{n - t}, \quad \text{and } A_2 = \frac{2\ell}{n - t},$$

$$\text{then } E^* = \frac{1}{1 - \frac{A_2}{\ell}} = 1 - \frac{2}{\ell + 6} B,$$

$$\begin{aligned} \text{and hence } V^* &= \frac{2}{\ell} \left\{ \frac{\ell + 6 + (\ell - 2)B}{\left(1 - \frac{2}{\ell + 6} B\right)^2 (\ell + 6 - 4B)} \right\} \\ &= \frac{2}{\ell} \left\{ \frac{1 + \left(\frac{\ell - 2}{\ell + 6}\right)B}{\left(1 - \frac{2}{\ell + 6} B\right)^2 \left(1 - \frac{4B}{\ell + 6}\right)} \right\}, \end{aligned} \quad (4.14)$$

by dividing the numerator and the denominator by $\ell + 6$.

Similarly, applying expression (4.8) on the Hotelling T² model,

$$4 + \frac{\ell + 2}{\ell \frac{V^*}{2E^{*2}} - 1} = n - \ell$$

The expression above can be rewritten in B term as

$$\begin{aligned} 4 + \frac{\ell + 2}{\ell \frac{V^*}{2E^{*2}} - 1} &= \frac{3\ell + 4}{B} + 1 - \ell \\ \Rightarrow \frac{2(\ell + 2)E^{*2}}{\ell V^* - 2E^{*2}} &= \frac{3\ell + 4}{B} - 3 - \ell, \\ V^* &= \frac{2}{\ell} \left\{ E^{*2} \left(1 + \frac{\ell + 2}{\frac{3\ell + 4}{B} - 3 - \ell} \right) \right\} \end{aligned} \quad (4.15)$$

Moreover, Since for the Hotelling T² model,

$$B = \frac{3\ell + 4}{n - 1}, \quad \text{and } A_2 = \frac{\ell(\ell + 1)}{n - 1},$$

$$\text{then } E^* = \frac{1}{1 - \frac{A_2}{\ell}} = 1 - \frac{\ell + 1}{3\ell + 4} B,$$

$$\begin{aligned} \text{and hence } V^* &= \frac{2}{\ell} \left\{ \frac{3\ell + 4 - B}{\left(1 - \frac{\ell + 1}{3\ell + 4} B \right)^2 [3\ell + 4 - (\ell + 3)B]} \right\} \\ &= \frac{2}{\ell} \left\{ \frac{1 + \left(\frac{-1}{3\ell + 4} \right) B}{\left(1 - \frac{\ell + 1}{3\ell + 4} B \right)^2 \left(1 - \frac{\ell + 3}{3\ell + 4} B \right)} \right\}, \end{aligned} \quad (4.16)$$

by dividing the numerator and the denominator by $3\ell + 4$.

From expressions (4.14 and 4.16), we can see that the approximate expression for the variance should be as

$$V^* = \frac{2}{\ell} \left\{ \frac{1 + d_1 B}{(1 - d_2 B)^2 (1 - d_3 B)} \right\},$$

where

$$\left. \begin{aligned} d_1 &= \frac{\ell-2}{\ell+6} \\ d_2 &= \frac{2}{\ell+6} \\ d_3 &= \frac{4}{\ell+6} \end{aligned} \right\} \text{for the balanced one-way Anova model,} \quad \left. \begin{aligned} d_1 &= \frac{-1}{3\ell+4} \\ d_2 &= \frac{\ell+1}{3\ell+4} \\ d_3 &= \frac{\ell+3}{3\ell+4} \end{aligned} \right\} \text{for Hotelling T}^2 \text{ model}$$

So far we obtained a general expression for the variance but not for the coefficients d_i 's.

First, we determine some simple linear relationship among the d_i 's that are common between the two cases

- (i) The numerator of $d_2 = \ell$ - numerator of d_1
- (ii) The numerator of $d_3 =$ numerator of $d_2 + 2 = \ell$ - numerator of $d_1 + 2$
- (iii) The denominator of d_i 's is a linear function of the numerator of d_1 , where

$$\begin{aligned} c(\ell-2) + d &= \ell + 6 && \text{for the balanced one-way Anova model,} \\ -c + d &= 3\ell + 4 && \text{for the Hotelling T}^2 \text{ model.} \end{aligned}$$

Solving the two equations above for c and d , we have $c = -2$ and $d = 3\ell + 2$, and hence

$$\text{the denominator of } d_i \text{'s} = 3\ell + 2(1 - \text{the numerator of } d_1).$$

The ℓ 's involved in relationships (i), (ii), and (iii) for both special cases are fixed, whereas any other ℓ , which is involved in the numerator of d_1 , is accommodated as a function of the ratio of A_1 and A_2 .

$$\text{Let the numerator of } d_1 = g(A_1, A_2) = a \frac{A_1}{A_2} + b = g.$$

$$\text{Notice that } g = \ell - 2 \quad \text{for the balanced one-way Anova model,}$$

$$g = -1 \quad \text{for the Hotelling T}^2 \text{ model.}$$

$$\text{Since } \frac{A_1}{A_2} = \ell \text{ and } \frac{2}{\ell+1} \text{ for the balanced one-way Anova and Hotelling T}^2 \text{ respectively,}$$

$$\text{then } a\ell + b = \ell - 2,$$

$$\text{and } a \frac{2}{\ell+1} + b = -1 \tag{4.17}$$

Solving the equations above, we have

$$a = \frac{(\ell+1)(\ell-1)}{\ell(\ell+1)-2} = \frac{\ell+1}{\ell+2}, \quad \text{and} \quad b = -\frac{\ell+4}{\ell+2}.$$

From the expressions we obtained for both a and b , we have

$$g = \frac{\ell+1}{\ell+2} \frac{A_1}{A_2} - \frac{\ell+4}{\ell+2}$$

Therefore,
$$g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2} \quad \text{for both special cases.}$$

Summary

The Kenward and Roger approach that was derived in section 3.2 will be modified as

$$m^* = 4 + \frac{\ell+2}{\ell\rho^* - 1}, \quad \text{where} \quad \rho^* = \frac{V^*}{2E^{*2}},$$

and
$$\lambda^* = \frac{m^*}{E^*(m^* - 2)} \quad \text{such that}$$

$$E^* = \left(1 - \frac{A_2}{\ell}\right)^{-1},$$

and
$$V^* = \frac{2}{\ell} \left\{ \frac{1 + d_1 B}{(1 - d_2 B)^2 (1 - d_3 B)} \right\},$$

$$d_1 = \frac{g}{3\ell + 2(1 - g)}$$

$$d_2 = \frac{\ell - g}{3\ell + 2(1 - g)}$$

$$d_3 = \frac{\ell - g + 2}{3\ell + 2(1 - g)}, \quad \text{where} \quad g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2}.$$

Recall that before the modification, $E[F] \approx \tilde{E} = 1 + \frac{A_2}{\ell}$, and $\text{Var}[F] \approx \tilde{V} = \frac{2}{\ell}(1 + B)$,

where
$$B = \frac{1}{2\ell}(A_1 + 6A_2),$$

and
$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi), \quad A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi)$$

4.4 Modifying the Approach with the Usual Variance-Covariance Matrix

Recall that when we used the usual variance-covariance of the fixed effects estimates $\hat{\Phi}$ in section 3.1.1, we found that

$$E[F] \approx 1 + \frac{S_2}{\ell}, \quad \text{and} \quad \text{Var}[F] \approx \frac{2}{\ell}(1 + B_1),$$

$$\text{where} \quad B_1 = \frac{6S_2 + S_1}{2\ell}, \quad \text{and} \quad S_2 = A_2 + 2A_3, \quad S_1 = A_1 - 4A_3$$

Since for both special cases, $A_3 = 0$, then S_2 and S_1 are reduced to A_2 and A_1 respectively, and hence the modification based on the same two special cases considered for K-R modification should be analogous to K-R modification derived in section 4.3 (substitute S_2 and S_1 for A_2 and A_1 respectively). The problematic part about this modification is that $A_3 = 0$, where there is no clear guide for us to accomplish the modification for A_3 . Simulation studies (as we will see in chapter 8) show that this modification does not perform well to estimate the denominator degrees of freedom and the scale.

5. TWO PROPOSED MODIFICATIONS FOR THE KENWARD-ROGER METHOD

As we saw in chapter four, the modification applied by Kenward and Roger was based on modifying the approximate expressions for the expectation and the variance of the Wald type statistic F so the approximation produces the known values for the denominator degrees of freedom and the scale for the special cases where the distribution of F is exact. The modification for the approximate expression for the expectation was relatively simple and direct; whereas the modification applied by Kenward and Roger for the approximate expression for the variance was more complicated. In fact, the modification that applied by Kenward and Roger is not unique and is ad-hoc. In this chapter, we propose two other modifications for Kenward and Roger's method based on the same two special cases. The approximate expression for the expectation of F is modified as done by Kenward and Roger (1997); however, instead of modifying the approximate expression for the variance, we modify other quantities as will be seen shortly. These two proposed modifications are simpler than the K-R modification both in computation and in derivation. Moreover, like the K-R method, our proposed modifications produce the exact values for the denominator degrees of freedom and scale in the two special cases.

5.1 First Proposed Modification

In order to approximate the F test for the fixed effects, $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, we use the same form of Wald-type F statistic used in K-R

method,
$$F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}} .$$

Define $\tilde{h} = \frac{\ell + 2}{\ell \frac{\tilde{V}}{2\tilde{E}^2} - 1}$. Recall that $\tilde{E} = 1 + \frac{A_2}{\ell}$, and $\tilde{V} = \frac{2}{\ell}(1 + B)$

Also, recall that chapter 4, for the balanced one-way Anova model, we had

$$A_1 = \frac{2\ell^2}{n-t}, \quad A_2 = \frac{2\ell}{n-t}, \quad B = \frac{\ell+6}{n-t}, \quad \text{and} \quad \frac{A_1}{A_2} = \ell,$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell \frac{\tilde{V}}{2\tilde{E}^2} - 1} = 4 + \tilde{h} \neq n-t, \quad \text{and} \quad \tilde{\lambda} = \frac{n-t}{\tilde{E}(n-t-2)} \neq 1$$

And for the Hotelling T^2 model, we had

$$A_1 = \frac{2\ell}{n-1}, \quad A_2 = \frac{\ell(\ell+1)}{n-1}, \quad B = \frac{3\ell+4}{n-1}, \quad \text{and} \quad \frac{A_1}{A_2} = \frac{2}{\ell+1},$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell \frac{\tilde{V}}{2\tilde{E}^2} - 1} = 4 + \tilde{h} \neq n-\ell, \quad \text{and} \quad \tilde{\lambda} = \frac{n-\ell}{\tilde{E}(n-\ell-2)} \neq \frac{n-\ell}{n-1}$$

In this proposed modification, we keep the modified estimator for $E[F]$ to be E^* as was derived in section 4.3; however, instead of constructing a modified estimator for $\text{Var}[F]$, we will construct a modified estimator for \tilde{h} , say h_1 leading to

$$m_1 = 4 + h_1, \tag{5.1}$$

such that m_1 matches the exact values for the denominator degrees of freedom for the balanced one-way Anova and the Hotelling T^2 models.

Applying expression (5.1) on the balanced one way Anova model,

$$4 + h_1 = \frac{\ell+6}{B}, \quad (\text{from 4.13})$$

$$h_1 = \frac{\ell+6}{B} - 4$$

And, applying expression (5.1) on the Hotelling T^2 model,

$$4 + h_1 = \frac{3\ell+4}{B} + (1-\ell), \quad (\text{from 4.15})$$

$$h_1 = \frac{3\ell+4}{B} - 3 - \ell.$$

For both models,

$$h_1 = d_0 + \frac{d_1}{B}, \tag{5.2}$$

where $\left. \begin{array}{l} d_0 = -4 \\ d_1 = \ell + 6 \end{array} \right\}$ for the balanced one-way Anova model,

and $\left. \begin{array}{l} d_0 = -(\ell + 3) \\ d_1 = 3\ell + 4 \end{array} \right\}$ for the Hotelling T^2 model.

Also, notice that d_1 can be expressed as a linear function of d_0 for both models, that is

$$d_1 = a_1 d_0 + a_2,$$

where $\ell + 6 = -4a_1 + a_2$ for the balanced one way Anova model,

$3\ell + 4 = -(\ell + 3)a_1 + a_2$ for the Hotelling T^2 model.

Solving the equations above, we have $a_1 = -2$, and $a_2 = \ell - 2$. That is

$$d_1 = -2d_0 + \ell - 2. \quad (5.3)$$

Like the discussion we made for K-R modification in section 4.3, we need to distinguish between the fixed ℓ and the one to be considered as a function of the ratio of A_1 and A_2 .

The ℓ in expression 5.3 is a fixed one and the ℓ in the d_0 expression is considered to be a function of the ratio of A_1 and A_2 .

$$\text{Let } g(A_1, A_2) = d_0 = a + b \frac{A_1}{A_2}$$

$$\Rightarrow -4 = a + b\ell, \quad \text{for the balanced one way Anova model,}$$

$$-(3 + \ell) = a + b \frac{2}{\ell + 1} \quad \text{for the Hotelling } T^2 \text{ model.}$$

Solving the two equations above,

$$a = -\frac{\ell(\ell + 1)}{\ell + 2} - 4, \quad \text{and } b = \frac{\ell + 1}{\ell + 2}$$

$$\begin{aligned} \text{And hence, } d_0 &= -\frac{\ell(\ell + 1)}{\ell + 2} - 4 + \frac{\ell + 1}{\ell + 2} \left(\frac{A_1}{A_2} \right) \\ &= \frac{(\ell + 1)A_1 - (\ell^2 + 5\ell + 8)A_2}{(\ell + 2)A_2} \end{aligned} \quad (5.4)$$

Summary

Kenward and Rogers's approach that was derived in section 3.1 is modified as

$$m_1 = 4 + h_1, \quad \text{and} \quad \lambda_1 = \frac{m_1}{E^*(m_1 - 2)}$$

such that

$$E^* = (1 - A_2/\ell)^{-1} \quad \text{and} \quad h_1 = d_0 + \frac{\ell - 2 - 2d_0}{B},$$

where

$$d_0 = \frac{(\ell + 1)A_1 - (\ell^2 + 5\ell + 8)A_2}{(\ell + 2)A_2} \quad (\text{Recall that } B = \frac{6A_2 + A_1}{2\ell}).$$

5.2 Second Proposed Modification

In order to approximate the F test for the fixed effects, $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, we use the same form of Wald-type F statistic used in K-R

$$\text{method,} \quad F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}}_{\mathbf{A}} \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}}.$$

Recall that chapter 4, for the balanced one-way Anova model, we had

$$A_1 = \frac{2\ell^2}{n-t}, \quad A_2 = \frac{2\ell}{n-t}, \quad B = \frac{\ell+6}{n-t}, \quad \text{and} \quad \frac{A_1}{A_2} = \ell,$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell\tilde{\rho}-1} \neq n-t, \quad \text{where} \quad \tilde{\rho} = \frac{\tilde{V}}{2\tilde{E}^2}$$

$$\text{and} \quad \tilde{\lambda} = \frac{n-t}{\tilde{E}(n-t-2)} \neq 1$$

And for the Hotelling T^2 model, we had

$$A_1 = \frac{2\ell}{n-1}, \quad A_2 = \frac{\ell(\ell+1)}{n-1}, \quad B = \frac{3\ell+4}{n-1}, \quad \text{and} \quad \frac{A_1}{A_2} = \frac{2}{\ell+1},$$

$$\tilde{m} = 4 + \frac{\ell+2}{\ell\tilde{\rho}-1} \neq n-\ell,$$

$$\text{and} \quad \tilde{\lambda} = \frac{n-\ell}{\tilde{E}(n-\ell-2)} \neq \frac{n-\ell}{n-1}$$

In this proposed modification, we also keep the modified estimator for $E[F]$ to be E^* as for the Kenward and Roger modification and first proposed modification.

Since when $\lambda F \sim F(\ell, m)$ exactly,

then
$$\tilde{\rho} = \frac{V[F]}{2E[F]^2} = \frac{\ell + m - 2}{\ell(m - 4)},$$

we construct a modified estimator for $\zeta = \ell\tilde{\rho}$, say ζ_2 such that

$$\zeta_2 = \frac{\ell + m_2 - 2}{m_2 - 4}. \quad (5.5)$$

where m_2 are the exact values for the denominator degrees of freedom for the balanced one-way Anova and the Hotelling T^2 models.

For the balanced one-way Anova model, and since $F \sim F(\ell, n - t)$ exactly, then by (5.5),

$$\begin{aligned} \zeta_2 &= \frac{\ell + n - t - 2}{n - t - 4} \\ &= \frac{1 + \frac{\ell - 2}{n - t}}{1 - \frac{4}{n - t}} \quad (\text{dividing by } n - t) \end{aligned}$$

Similarly, since for the Hotelling T^2 model, $\frac{n - p}{n - 1} F \sim F(\ell, n - p)$ exactly, then by (5.5),

$$\begin{aligned} \zeta_2 &= \frac{n - 2}{n - \ell - 4} \\ &= \frac{1 - \frac{1}{n - 1}}{1 - \frac{\ell + 3}{n - 1}} \quad (\text{dividing by } n - 1) \end{aligned}$$

For both models,

$$\zeta_2 = \frac{a_0 + a_1 A_1 + a_2 A_2}{b_0 + b_1 A_1 + b_2 A_2},$$

Notice that unlike the first proposed modification where we express h_i in B term, in this proposed modification, we express ζ_2 in term of A_1 and A_2 .

For the balanced one-way Anova model,
$$\zeta_2 = \frac{a_0 + a_1 \frac{2\ell^2}{n - t} + a_2 \frac{2\ell}{n - t}}{b_0 + b_1 \frac{2\ell^2}{n - t} + b_2 \frac{2\ell}{n - t}} = \frac{1 + \frac{\ell - 2}{n - t}}{1 - \frac{4}{n - t}}$$

$$\Rightarrow a_0 = c, \quad 2\ell^2 a_1 + 2\ell a_2 = c(\ell - 2) \quad (5.6)$$

$$b_0 = c, \quad 2\ell^2 b_1 + 2\ell b_2 = c(-4) \quad (5.7)$$

for any $c \neq 0$

and for the Hotelling T^2 model,
$$\zeta_2 = \frac{a_0 + a_1 \frac{2\ell}{n-1} + a_2 \frac{\ell(\ell+1)}{n-1}}{b_0 + b_1 \frac{2\ell}{n-1} + b_2 \frac{\ell(\ell+1)}{n-1}} = \frac{1 - \frac{1}{n-1}}{1 - \frac{\ell+3}{n-1}}$$

$$\Rightarrow a_0 = c, \quad 2\ell a_1 + \ell(\ell+1)a_2 = -c \quad (5.8)$$

$$b_0 = c, \quad 2\ell b_1 + \ell(\ell+1)b_2 = -c(\ell+3) \quad (5.9)$$

By solving equations (5.6) and (5.8), we obtain

$$a_1 = \frac{c}{2\ell(\ell+2)}, \quad \text{and} \quad a_2 = -\frac{2c}{\ell(\ell+2)}.$$

By solving equations (5.7) and (5.9), we obtain

$$b_1 = -\frac{c}{\ell(\ell+2)}, \quad \text{and} \quad b_2 = -\frac{c(\ell+4)}{\ell(\ell+2)}.$$

Take $c = \ell(\ell+2)$

$$\Rightarrow a_0 = b_0 = \ell(\ell+2), \quad a_1 = \frac{\ell}{2}, \quad a_2 = -2, \quad b_1 = -1, \quad b_2 = -(\ell+4)$$

So, the modification that we apply for $\tilde{\zeta}$ will be

$$\zeta_2 = \frac{\ell(\ell+2) + \frac{\ell}{2}A_1 - 2A_2}{\ell(\ell+2) - A_1 - (\ell+4)A_2},$$

$$4 + \frac{\ell+2}{\zeta_2 - 1} = \frac{2\ell(\ell+2) + 2(A_1 - \ell A_2)}{A_1 + 2A_2}$$

Summary

The Kenward and Roger's approach that was derived in section 3.1 is modified as

$$m_2 = \frac{2\ell(\ell+2) + 2(A_1 - \ell A_2)}{A_1 + 2A_2},$$

and $\lambda_2 = \frac{m_2}{E^*(m_2 - 2)}$ such that $E^* = (1 - A_2/\ell)^{-1}$

5.3 Comparisons among the K-R and the Proposed Modifications

Based on the way we derived K-R and the proposed modifications, we should expect that these modifications give identical estimates of the denominator degrees of freedom when the ratio of A_1 and A_2 is the same as one of the ratios for the two special cases. In this section, we derive some useful results about the three modifications. From now on, when we say Kenward and Roger's approach, we mean after the modification that makes the approach produce the exact values for the two special cases.

Theorem 5.3.1 The two proposed and K-R methods give the same estimate of the denominator degrees of freedom when $A_1 = \ell A_2$ or $A_1 = \frac{2}{\ell+1} A_2$, and the estimate values are $\frac{2\ell}{A_2}$, and $\frac{\ell(\ell+1)}{A_2} - (\ell-1)$ respectively.

Proof (i) For the case where $A_1 = \ell A_2$

K-R approach

$$g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2} = \frac{(\ell+1)\ell A_2 - (\ell+4)A_2}{(\ell+2)A_2} = \ell - 2$$

$$\text{So, } d_1 = \frac{\ell-2}{3\ell+2(3-\ell)} = \frac{\ell-2}{\ell+6}, \quad d_2 = \frac{2}{\ell+6}, \quad d_3 = \frac{4}{\ell+6}.$$

$$\text{Since } E^* = (1 - A_2/\ell)^{-1}, \text{ and } B = \frac{1}{2\ell}(A_1 + 6A_2) = \frac{\ell+6}{2\ell} A_2,$$

$$\text{then } V^* = \frac{2}{\ell} \left\{ \frac{1 + d_1 B}{(1 - d_2 B)^2 (1 - d_3 B)} \right\} = \frac{2}{\ell} \left\{ \frac{1 + \frac{\ell-2}{2\ell} A_2}{(1 - A_2/\ell)^2 (1 - 2A_2/\ell)} \right\}.$$

$$\Rightarrow \rho^* = \frac{V^*}{2E^{*2}} = \frac{1}{\ell} \left\{ \frac{1 + \frac{\ell-2}{2\ell} A_2}{1 - \frac{2}{\ell} A_2} \right\}, \quad \ell \rho^* - 1 = \frac{1 + \frac{\ell-2}{2\ell} A_2}{1 - \frac{2}{\ell} A_2} - 1 = \frac{\frac{\ell+2}{2\ell} A_2}{1 - \frac{2}{\ell} A_2},$$

Therefore,

$$m^* = 4 + \frac{\ell+2}{\ell\rho^*-1} = 4 + \frac{\ell+2}{\frac{\ell+2}{2\ell}A_2} \left(1 - \frac{2}{\ell}A_2\right) = \frac{2\ell}{A_2}.$$

First proposed approach

$$\begin{aligned} d_0 &= \frac{(\ell+1)A_1 - (\ell^2 + 5\ell + 8)A_2}{(\ell+2)A_2} . \\ m_1 &= 4 + d_0 + \frac{\ell-2-2d_0}{B}, \\ &= 4 + d_0 \left(1 - \frac{2}{B}\right) + \frac{\ell-2}{B} \\ &= 4 + \frac{\ell(\ell+1)A_2 - (\ell^2 + 5\ell + 8)A_2}{(\ell+2)A_2} \left(1 - \frac{4\ell}{(6+\ell)A_2}\right) + \frac{2\ell(\ell-2)}{(6+\ell)A_2} \\ &= 4 + \frac{[\ell(\ell+1)A_2 - (\ell^2 + 5\ell + 8)A_2][(6+\ell)A_2 - 4\ell] + 2\ell(\ell^2 - 4)A_2}{(6+\ell)(\ell+2)A_2^2} \\ &= 4 + \frac{2\ell^3 + 16\ell^2 + 24\ell - (4\ell+8)(6+\ell)A_2}{(6+\ell)(\ell+2)A_2} \\ &= 4 + \frac{2\ell(6+\ell)(2+\ell)}{(6+\ell)(\ell+2)A_2} - \frac{(4\ell+8)(6+\ell)A_2}{(6+\ell)(\ell+2)A_2} = \frac{2\ell}{A_2} \end{aligned}$$

Second proposed approach

$$m_2 = \frac{2\ell(\ell+2) + 2(A_1 - \ell A_2)}{A_1 + 2A_2} = \frac{2\ell}{A_2}$$

(ii) For the case where $A_1 = \frac{2}{\ell+1}A_2$

K-R approach

$$g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2} = \frac{(\ell+1)}{(\ell+2)} \frac{A_1}{A_2} - \frac{\ell+4}{\ell+2} = \frac{(\ell+1)}{(\ell+2)} \frac{2}{\ell+1} - \frac{\ell+4}{\ell+2} = -1$$

$$\text{So, } d_1 = \frac{-1}{3\ell+4}, \quad d_2 = \frac{\ell+1}{3\ell+4}, \quad d_3 = \frac{\ell+3}{3\ell+4}.$$

$$\text{and since } E^* = (1 - A_2/\ell)^{-1}, \text{ and } B = \frac{1}{2\ell}(A_1 + 6A_2) = \frac{4+3\ell}{\ell(\ell+1)}A_2,$$

$$\begin{aligned}
\text{then } V^* &= \frac{2}{\ell} \left\{ \frac{1 + d_1 B}{(1 - d_2 B)^2 (1 - d_3 B)} \right\} \\
&= \frac{2}{\ell} \left\{ \frac{1 - \frac{1}{\ell(\ell+1)} A_2}{(1 - A_2/\ell)^2 (1 - \frac{\ell+3}{\ell(\ell+1)} A_2)} \right\}, \\
\rho^* &= \frac{V^*}{E^*} = \frac{1}{\ell} \left\{ \frac{1 - \frac{1}{\ell(\ell+1)} A_2}{1 - \frac{\ell+3}{\ell(\ell+1)} A_2} \right\} = \frac{1}{\ell} \left\{ \frac{\ell(\ell+1) - A_2}{\ell(\ell+1) - (\ell+3)A_2} \right\}, \\
\ell\rho^* - 1 &= \frac{\ell(\ell+1) - A_2}{\ell(\ell+1) - (\ell+3)A_2} - 1 = \frac{(\ell+2)A_2}{\ell(\ell+1) - (\ell+3)A_2},
\end{aligned}$$

and therefore

$$m^* = 4 + \frac{\ell+2}{\ell\rho-1} = 4 + \frac{\ell(\ell+1) - (\ell+3)A_2}{A_2} = 1 - \ell + \frac{\ell(\ell+1)}{A_2},$$

First proposed approach

$$\begin{aligned}
d_0 &= \frac{(\ell+1)A_1 - (\ell^2 + 5\ell + 8)A_2}{(\ell+2)A_2}, \\
m_1 &= 4 + d_0 + \frac{\ell-2-2d_0}{B} \\
&= 4 + d_0 \left(1 - \frac{2}{B} \right) + \frac{\ell-2}{B} \\
&= 4 + \frac{2A_2 - (\ell^2 + 5\ell + 8)A_2}{(\ell+2)A_2} \left(1 - \frac{2\ell(\ell+1)}{(4+3\ell)A_2} \right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2} \\
&= 4 + \frac{-(\ell^2 + 5\ell + 6)A_2}{(\ell+2)A_2} \left(\frac{(4+3\ell)A_2 - 2\ell(\ell+1)}{(4+3\ell)A_2} \right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2} \\
&= 4 - (\ell+3) \left(\frac{(4+3\ell)A_2 - 2\ell(\ell+1)}{(4+3\ell)A_2} \right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2} \\
&= 4 + \frac{\ell(\ell+1)(\ell-2) - (\ell+3)[(4+3\ell)A_2 - 2\ell(\ell+1)]}{(4+3\ell)A_2} \\
&= 4 + \frac{\ell(\ell+1)[2(\ell+3) + \ell - 2]}{(4+3\ell)A_2} - (\ell+3) = 1 - \ell + \frac{\ell(\ell+1)}{A_2}
\end{aligned}$$

Second proposed approach

$$m_2 = \frac{2\ell(\ell+2) + 2(A_1 - \ell A_2)}{A_1 + 2A_2} = 1 - \ell + \frac{\ell(\ell+1)}{A_2} \quad \square$$

Theorem 5.3.2 When $A_1 = \ell A_2$ or $A_1 = \frac{2}{\ell+1} A_2$, then the scale estimates are 1 and

$1 - \frac{\ell-1}{\ell(\ell+1)} A_2$ respectively for K-R and the proposed modifications as well.

Proof When $A_1 = \ell A_2$, or $A_1 = \frac{2}{\ell+1} A_2$, the three modifications give the same estimates

for the denominator degrees of freedom which are $\frac{2\ell}{A_2}$ and $1 - \ell + \frac{\ell(\ell+1)}{A_2}$ respectively

(theorem 5.3.1). In addition, since the $E[F]$ is modified in the same way for the three modifications, then the scale estimate is the same for the three modifications. In the following proof, we use the Kenward and Roger approach to find the scale estimate

(i) For the case where $A_1 = \ell A_2$, $m^* = \frac{2\ell}{A_2}$,

$$\text{and } \lambda^* = \frac{m^*}{E^*(m^* - 2)} = \frac{\frac{2\ell}{A_2}}{\left(\frac{1}{1 - A_2/\ell}\right)\left(\frac{2\ell}{A_2} - 2\right)} = 1.$$

(ii) For the case $A_1 = \frac{2}{\ell+1} A_2$, $m = 1 - \ell + \frac{\ell(\ell+1)}{A_2}$,

$$\begin{aligned} \text{and } \lambda^* &= \frac{m}{E^*(m - 2)} = \frac{1 - \ell + \frac{\ell(\ell+1)}{A_2}}{\left(\frac{1}{1 - A_2/\ell}\right)\left(\frac{\ell(\ell+1)}{A_2} - (\ell+1)\right)} \\ &= \frac{\frac{(1-\ell)}{\ell(\ell+1)} A_2 + 1}{\left(\frac{1}{1 - A_2/\ell}\right)\left(1 - \frac{A_2}{\ell}\right)}, \quad \text{multiplying by } \frac{A_2}{\ell(\ell+1)} \end{aligned}$$

$$= 1 - \frac{(\ell-1)}{\ell(\ell+1)} A_2 \quad \square$$

Corollary 5.3.3 If $A_1 = \ell A_2$ or $A_1 = \frac{2}{\ell+1} A_2$, then the K-R, and the proposed approaches are identical.

Proof From theorems (5.3.1 and 5.3.2), K-R, and the proposed approaches give the same estimate of the denominator degrees of freedom and the same estimate for the scale, and since the statistic used is the same, then all approaches are identical. \square

Theorem 5.3.4 When $\ell = 1$, then $A_1 = A_2$.

$$\textbf{Proof} \quad A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi), \quad A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi)$$

Since $\mathbf{L}'\Phi\mathbf{L}$ is a scalar, then $\Theta = \frac{\mathbf{L}\mathbf{L}'}{\mathbf{L}'\Phi\mathbf{L}}$,

$$\text{tr}(\Theta \Phi \mathbf{P}_i \Phi) = \text{tr}\left(\frac{\mathbf{L}\mathbf{L}'\Phi \mathbf{P}_i \Phi}{\mathbf{L}'\Phi\mathbf{L}}\right) = \frac{1}{\mathbf{L}'\Phi\mathbf{L}} \text{tr}(\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L}) = \frac{\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L}}{\mathbf{L}'\Phi\mathbf{L}}.$$

$$\text{and hence,} \quad A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L} \mathbf{L}'\Phi \mathbf{P}_j \Phi \mathbf{L}}{(\mathbf{L}'\Phi\mathbf{L})^2} \quad (5.10)$$

Also,

$$\begin{aligned} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi) &= \frac{1}{(\mathbf{L}'\Phi\mathbf{L})^2} \text{tr}(\mathbf{L}\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L} \mathbf{L}'\Phi \mathbf{P}_j \Phi \mathbf{L}) \\ &= \frac{\text{tr}(\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L} \mathbf{L}'\Phi \mathbf{P}_j \Phi \mathbf{L})}{(\mathbf{L}'\Phi\mathbf{L})^2} = \frac{\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L} \mathbf{L}'\Phi \mathbf{P}_j \Phi \mathbf{L}}{(\mathbf{L}'\Phi\mathbf{L})^2} \end{aligned}$$

$$\text{and hence,} \quad A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\mathbf{L}'\Phi \mathbf{P}_i \Phi \mathbf{L} \mathbf{L}'\Phi \mathbf{P}_j \Phi \mathbf{L}}{(\mathbf{L}'\Phi\mathbf{L})^2} \quad (5.11)$$

From expressions (5.10 and 5.11), we can see that $A_1 = A_2$. \square

Corollary 5.3.5 When $\ell = 1$, the K-R and the proposed approaches are identical.

The estimates of the denominator degrees of freedom and the scale are $\frac{2}{A_2}$ and 1

respectively.

Proof Since $A_1 = A_2$ (theorem 5.3.4), then by applying corollary (5.3.3), we have the K-R and the proposed approaches are identical. By using theorems (5.3.1 and 5.3.2), the proof is completed. \square

6. KENWARD AND ROGER'S APPROXIMATION FOR THREE GENERAL MODELS

In chapter 4, we saw how the K-R approach was modified so the approximation for the denominator degrees of freedom and the scale match the known values for the two special cases where F has an exact F distribution. In this chapter, we show that the K-R approximation produces the exact values for the denominator degrees of freedom and scale factor for more general models. Three models are considered: the modified Rady model which is more general than the balanced one-way Anova model, a general multivariate linear model which is more general than the Hotelling T^2 , and a balanced multivariate model with a random group effect.

6.1 Modified Rady's Model

A model that satisfies assumptions (A1)-(A5) will be called a Rady model. If the hypothesis satisfies conditions (A6) and (A7), and the model is a Rady model, the testing problem will be called a Rady testing problem. Rady (1986) proved that a Rady testing problem has an exact F test to test a linear hypothesis of the fixed effects. The model considered in this section is slightly different than Rady's model where we add two more assumptions (A8) and (A9) and the model will be called modified Rady's model. When the model is a modified Rady model, and the hypothesis satisfies conditions (A6) and (A7), then the testing problem is called a modified Rady testing problem.

6.1.1 Model and Assumptions

Consider the mixed linear model,

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \text{Cov}[\mathbf{y}] = \sigma_1^2 \boldsymbol{\Sigma}_1 + \cdots + \sigma_{r-1}^2 \boldsymbol{\Sigma}_{r-1} + \sigma_r^2 \mathbf{I}$, and $\boldsymbol{\Sigma}_i$ are n.n.d. matrices. \mathbf{X} is a known matrix, $\boldsymbol{\beta}$ is a vector of unknown fixed effect parameters, and \mathfrak{S} denotes the set of all possible vectors $\boldsymbol{\sigma}$ and $\mathfrak{N} = \{\boldsymbol{\Sigma}(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \mathfrak{S}\}$. We are interested in testing $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, for \mathbf{L}' an $(\ell \times p)$ fixed matrix.

We assume that the model satisfies assumptions A1-A8 where

(A1) \mathfrak{V} contains an r -dimensional rectangle.

(A2) $\Sigma(\sigma)$ is p.d. for all σ .

(A3) β and σ are functionally independent.

(A4) The model has orthogonal block structure (OBS). That is, \mathfrak{N} is commutative (i.e. $\Sigma(\sigma)\Sigma(\sigma^*) = \Sigma(\sigma^*)\Sigma(\sigma) \forall \sigma, \sigma^* \in \mathfrak{V}$). The orthogonal block structure is equivalent to the spectral decomposition:

$$\Sigma(\sigma) = \sum_{k=1}^q \alpha_k(\sigma) \mathbf{E}_k, \text{ where } \alpha_k(\sigma) \text{ is a linear function of } \sigma \left(\text{i.e. } \alpha_k(\sigma) = \sum_{i=1}^r \alpha_{k_i} \sigma_i \right), \text{ and}$$

\mathbf{E}_k 's are pairwise orthogonal o.p. matrices.

(A5) Zyskind's condition: A model is said to satisfy Zyskind's condition if

$$\mathbf{P}_X \Sigma = \Sigma \mathbf{P}_X \text{ for all } \Sigma.$$

(A6) Zyskind's condition for the submodel under the null hypothesis; $\Sigma \mathbf{P}_U = \mathbf{P}_U \Sigma$ for all Σ where $\mathfrak{R}(U) = \{\mathbf{X}\beta : \mathbf{L}'\beta = \mathbf{0}\}$.

(A7) $\mathfrak{R}(\mathbf{M}) \subset \mathfrak{R}(\mathbf{B}_s)$ for some s , where $\mathbf{M} = \mathbf{P}_X - \mathbf{P}_U$, and each \mathbf{E}_k can be expressed as $\mathbf{E}_k = \mathbf{B}_k + \mathbf{C}_k$ where \mathbf{B}_k and \mathbf{C}_k are symmetric idempotent matrices, with $\mathfrak{R}(\mathbf{C}_k) \subset \mathfrak{R}(\mathbf{X}')$, $\mathfrak{R}(\mathbf{B}_k) = \mathfrak{R}(\mathbf{E}_k \mathbf{X})$ and $\mathbf{B}_k \mathbf{C}_k = \mathbf{0}$.

(A8) $(\mathbf{I} - \mathbf{P}_X) \Sigma_i$'s are linearly independent.

(A9) $q = r$.

Remarks

(i) \mathbf{B}_k is an o.p.o on $\mathfrak{R}(\mathbf{E}_k \mathbf{X})$, $\mathbf{B}_k = \mathbf{E}_k \mathbf{P}_X$, and notice that $\mathfrak{R}(\mathbf{E}_k \mathbf{X}) \subset \mathfrak{R}(\mathbf{E}_k)$. Then, $\mathbf{C}_k = \mathbf{E}_k - \mathbf{B}_k = \mathbf{E}_k (\mathbf{I} - \mathbf{P}_X)$ is an o.p.o on $\mathfrak{R}(\mathbf{E}_k) \cap \mathfrak{R}(\mathbf{E}_k \mathbf{X})^\perp$ (Seely, 2002, problem, 2.F.3).

(ii) $\mathfrak{R}(\mathbf{E}_k) \cap \mathfrak{R}(\mathbf{E}_k \mathbf{X})^\perp \subset \mathfrak{R}(\mathbf{X}')$

Proof: Take any $\mathbf{x} \in \mathfrak{R}(\mathbf{E}_k) \cap \mathfrak{R}(\mathbf{E}_k \mathbf{X})^\perp$

$$\Rightarrow \mathbf{x} \in \mathfrak{R}(\mathbf{E}_k) \text{ and } \mathbf{x} \in \mathfrak{R}(\mathbf{E}_k \mathbf{X})^\perp = \mathfrak{R}(\mathbf{X}' \mathbf{E}_k)$$

$$\Rightarrow \mathbf{E}_k \mathbf{v} = \mathbf{x} \text{ for some } \mathbf{v}, \text{ and } \mathbf{X}' \mathbf{E}_k \mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \mathbf{X}' \mathbf{E}_k \mathbf{E}_k \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{X}' \mathbf{E}_k \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{X}' \mathbf{x} = \mathbf{0}, \text{ and this means that } \mathbf{x} \in \mathfrak{R}(\mathbf{X}').$$

(iii) From assumption(A7), $\Re(\mathbf{M}) \subset \Re(\mathbf{B}_s) = \Re(\mathbf{E}_s \mathbf{X})$, and since $\Re(\mathbf{E}_s \mathbf{X}) \subset \Re(\mathbf{E}_s)$, then $\Re(\mathbf{M}) \subset \Re(\mathbf{E}_s)$, and hence $\mathbf{E}_s \mathbf{M} = \mathbf{M}$.

(iv) Since $\mathbf{L}'\hat{\boldsymbol{\beta}}$ is estimable, then $\mathbf{L} = \mathbf{X}'\mathbf{A}$ for some \mathbf{A} . $r(\mathbf{L}) = r(\mathbf{A}) - \dim[\Re(\mathbf{A}) \cap \Re(\mathbf{X}')]]$, and $\Re(\mathbf{A}) \cap \Re(\mathbf{X}') = \Re(\mathbf{A}) \cap \Re(\mathbf{X})^\perp = \mathbf{0}$, because $\Re(\mathbf{A}) \subset \Re(\mathbf{X})$. Therefore, $r(\mathbf{L}) = r(\mathbf{A}) = r(\mathbf{M}) = \ell$.

(v) $\mathbf{C}_k \neq \mathbf{0} \forall k$.

Proof: Suppose that $\mathbf{C}_r = \mathbf{0}$, then $\boldsymbol{\Sigma} = \sum_{i=1}^r \sigma_i \boldsymbol{\Sigma}_i = \sum_{k=1}^{r-1} \alpha_k(\boldsymbol{\sigma}) \mathbf{E}_k$

$$\Rightarrow \sum_{i=1}^r \sigma_i (\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i = \sum_{k=1}^{r-1} \alpha_k(\boldsymbol{\sigma}) \mathbf{C}_k,$$

and this combined with assumption (A1) implies $(\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i$'s $\in \text{span}\{\mathbf{C}_1, \dots, \mathbf{C}_{r-1}\}$ which contradicts with assumption (A8) that $(\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i$'s are linearly independent.

(vi) $\mathbf{B} = \{\alpha_{ki}\}_{r \times r}$ is invertible.

Proof: Suppose $\mathbf{B}\boldsymbol{\sigma} = \mathbf{B}\boldsymbol{\sigma}^*$. Then $\alpha_k(\boldsymbol{\sigma}) = \alpha_k(\boldsymbol{\sigma}^*) \forall k \Rightarrow \sum_{k=1}^r \alpha_k(\boldsymbol{\sigma}) \mathbf{C}_k = \sum_{k=1}^r \alpha_k(\boldsymbol{\sigma}^*) \mathbf{C}_k$

$$\Rightarrow (\mathbf{I} - \mathbf{P}_X) \sum_{k=1}^r \alpha_k(\boldsymbol{\sigma}) \mathbf{E}_k = (\mathbf{I} - \mathbf{P}_X) \sum_{k=1}^r \alpha_k(\boldsymbol{\sigma}^*) \mathbf{E}_k \Rightarrow \sum_{i=1}^r \sigma_i (\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i = \sum_{i=1}^r \sigma_i^* (\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i,$$

and since $(\mathbf{I} - \mathbf{P}_X) \boldsymbol{\Sigma}_i$'s are linearly independent (assumption A8), then $\sigma_i = \sigma_i^* \forall i$.

Hence \mathbf{B} is injective on \mathfrak{I} , and by assumption (A1), it is injective on \mathbb{R}^r , and since it is square, then it is invertible.

(vii) The definition of OBS in (A4) was given by Birkes and Lee (2003) which is less restrictive than the definition given by VanLeeuwen *et al.* (1998). \square

Examples

1- Fixed effects linear models which include the balanced one-way Anova model are modified Rady models.

2- Balanced mixed classification nested models are modified Rady models.

3- Many balanced mixed classification models including balanced split-plot models are modified Rady models.

6.1.2 Useful Properties for the Modified Rady Model

A testing problem that satisfies all or some of the modified Rady testing problem assumptions has several properties that will be summarized in this section

Lemma 6.1.2.1 In a Rady model, $\hat{\Phi}_A = \hat{\Phi}$

Proof Consider $\mathbf{Q}_{ij} - \mathbf{P}_i \Phi \mathbf{P}_j$

$$= \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} - \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} ,$$

Note Since $\Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}'$ is a p.o. on $\mathfrak{R}(\Sigma^{-1} \mathbf{X})$ along $\mathfrak{R}(\mathbf{X}')$, Σ^{-1} is a nonsingular matrix (by assumption) and n.n.d. (Birkes, 2004), then it is a p.d. matrix (Seely, 2002, corollary 1.9.3)

$$\Rightarrow r(\mathbf{X}' \Sigma^{-1} \mathbf{X}) = r(\mathbf{X}') \quad (\text{Birkes, 2004}).$$

Since $r(\mathbf{X}' \Sigma^{-1} \mathbf{X}) = r(\mathbf{X}') = r(\mathbf{P}_X)$, then we have

$$\Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' = \Sigma^{-1} \mathbf{P}_X (\mathbf{P}_X \Sigma^{-1} \mathbf{P}_X)^+ \mathbf{P}_X \quad (\text{Seely, 2002, problem 1.10.B3})$$

$$\text{Also, } \mathbf{P}_X \Sigma^{-1} \mathbf{P}_X = \sum_{k=1}^q \mathbf{P}_X \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \mathbf{P}_X = \sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \mathbf{P}_X = \Sigma^{-1} \mathbf{P}_X$$

$$\Rightarrow (\mathbf{P}_X \Sigma^{-1} \mathbf{P}_X)^+ = \sum_{k=1}^q \alpha_k(\boldsymbol{\sigma}) \mathbf{E}_k \mathbf{P}_X = \Sigma \mathbf{P}_X,$$

$$\begin{aligned} \text{and hence } \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' &= \Sigma^{-1} \mathbf{P}_X (\mathbf{P}_X \Sigma^{-1} \mathbf{P}_X)^+ \mathbf{P}_X \\ &= \Sigma^{-1} \mathbf{P}_X \Sigma \mathbf{P}_X \mathbf{P}_X = \mathbf{P}_X \quad (\text{Zyskind's condition}) \end{aligned} \quad (6.1)$$

$$\begin{aligned} \mathbf{P}_X \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{P}_X &= \mathbf{P}_X \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{P}_X \\ &= \mathbf{P}_X \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^q \alpha_{ki} \mathbf{E}_k \right) \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^q \alpha_{kj} \mathbf{E}_k \right) \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \mathbf{P}_X \\ &= \sum_{k=1}^q \frac{\alpha_{ki} \alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^3} \mathbf{E}_k \mathbf{P}_X \end{aligned} \quad (6.2)$$

$$\text{Similarly, } \mathbf{P}_X \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{P}_X$$

$$\begin{aligned}
&= \mathbf{P}_X \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{P}_X \\
&= \mathbf{P}_X \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \mathbf{P}_X \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{P}_X \\
&= \mathbf{P}_X \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^q \alpha_{ki} \mathbf{E}_k \right) \mathbf{P}_X \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^q \alpha_{kj} \mathbf{E}_k \right) \left(\sum_{k=1}^q \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \mathbf{P}_X \\
&= \sum_{k=1}^q \frac{\alpha_{ki} \alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^3} \mathbf{E}_k \mathbf{P}_X
\end{aligned} \tag{6.3}$$

From expressions (6.2 and 6.3) we obtain

$$\begin{aligned}
&\mathbf{P}_X \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{P}_X - \mathbf{P}_X \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{P}_X = \mathbf{0} \\
&\Rightarrow \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} - \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_j} \mathbf{X} = \mathbf{0}
\end{aligned}$$

This shows that $\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j = \mathbf{0}$, and since $\mathbf{R}_{ij} = \mathbf{0}$, then $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$. \square

Corollary 6.1.2.2 In balanced mixed classification models, $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$.

Proof Balanced mixed classification models are Rady models (Rady, 1986), and hence by applying lemma 6.1.2.1, we have $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$. \square

Lemma 6.1.2.3 For a Rady testing problem, $A_1 = \ell A_2$.

Proof $\Omega = \{ \mathbf{X}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^p \} = \Re(\mathbf{X})$.

Since we are interested in testing $H_0: \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, then $\mathbf{L}'\boldsymbol{\beta}$ is assumed to be estimable $\Leftrightarrow \Re(\mathbf{L}) \subset \Re(\mathbf{X}') = \Re(\mathbf{X}'\mathbf{X}) \Leftrightarrow \mathbf{L} = \mathbf{X}'\mathbf{X}\mathbf{B}$ for some $\mathbf{B} \Leftrightarrow \mathbf{L} = \mathbf{X}'\mathbf{A}$ where $\mathbf{A} = \mathbf{X}\mathbf{B}$, $\Re(\mathbf{A}) \subset \Re(\mathbf{X})$, and hence

$$\Omega_H = \{ \mathbf{X}\boldsymbol{\beta} : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0} \} = \Re(\mathbf{X}) \cap \Re(\mathbf{A}') = \Re(\mathbf{U}).$$

Also, since $\Sigma \mathbf{P}_U = \mathbf{P}_U \Sigma$ for all Σ (Zyskind's condition for the submodel),
and $\Sigma \mathbf{P}_X = \mathbf{P}_X \Sigma$ for all Σ (Zyskind's condition),
then $\Sigma \mathbf{M} = \mathbf{M} \Sigma$ for all Σ where $\mathbf{M} = \mathbf{P}_X - \mathbf{P}_U$.

$$\begin{aligned}
\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}) &= -\text{tr}\left[\mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right] \\
&= -\text{tr}\left[\mathbf{X}'\mathbf{A}(\mathbf{A}'\mathbf{X}\mathbf{\Phi}\mathbf{X}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right] \\
&= -\text{tr}\left[\mathbf{P}_x\mathbf{A}(\mathbf{A}'\mathbf{X}\mathbf{\Phi}\mathbf{X}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{P}_x\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right] \quad (\text{from 6.1}). \\
&= -\text{tr}\left[\mathbf{A}(\mathbf{A}'\mathbf{X}\mathbf{\Phi}\mathbf{X}'\mathbf{A})^{-1}\mathbf{A}'\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right], \text{ because } \mathfrak{R}(\mathbf{A}) \subset \mathfrak{R}(\mathbf{X}) \\
&= -\text{tr}\left[\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}\mathbf{A})^{-1}\mathbf{A}'\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right], \quad \text{again notice that } \mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Phi}\mathbf{X}' = \mathbf{P}_x
\end{aligned}$$

Since $\mathfrak{R}(\mathbf{U}) \subset \mathfrak{R}(\mathbf{X})$, then \mathbf{M} is an o.p.o. on $\mathbf{\Omega} \cap \mathbf{\Omega}_H^\perp$ (Seely, 2002, problem 2.F.3),

$$\begin{aligned}
\mathbf{\Omega} \cap \mathbf{\Omega}_H^\perp &= \mathfrak{R}(\mathbf{X}) \cap [\mathfrak{R}(\mathbf{X}) \cap \mathfrak{R}(\mathbf{A}')]^\perp = \mathfrak{R}(\mathbf{X}) \cap [\mathfrak{R}(\mathbf{X})^\perp + \mathfrak{R}(\mathbf{A})] \\
&= \mathfrak{R}(\mathbf{X}) \cap \mathfrak{R}(\mathbf{X})^\perp + \mathfrak{R}(\mathbf{X}) \cap \mathfrak{R}(\mathbf{A}) = \mathfrak{R}(\mathbf{A})
\end{aligned}$$

So, we have $\mathfrak{R}(\mathbf{M}) = \mathfrak{R}(\mathbf{A})$.

Analogous to the argument given in the proof of lemma 6.2.1, we obtain

$$\mathbf{\Sigma}\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}\mathbf{A})^{-1}\mathbf{A}' = \mathbf{\Sigma}\mathbf{M}(\mathbf{M}\mathbf{\Sigma}\mathbf{M})^+\mathbf{M} = \mathbf{M}, \quad (6.4)$$

and hence,
$$\text{tr}(\mathbf{\Theta}\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}) = -\text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{M}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right).$$

$$\Rightarrow A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{M}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right) \text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{M}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_j}\right),$$

$$A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{M}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{M}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_j}\right)$$

A_1 and A_2 can also be expressed in \mathbf{E}_k 's terms as

$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \sum_{k=1}^q \frac{\alpha_{ki}}{\alpha_k(\boldsymbol{\sigma})} \text{tr}(\mathbf{E}_k\mathbf{M}) \sum_{k=1}^q \frac{\alpha_{kj}}{\alpha_k(\boldsymbol{\sigma})} \text{tr}(\mathbf{E}_k\mathbf{M}), \quad (6.5)$$

And

$$A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \sum_{k=1}^q \frac{\alpha_{ki}\alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^2} \text{tr}(\mathbf{E}_k\mathbf{M}) \quad (6.6)$$

Since $\mathbf{E}_k \mathbf{M} = \mathbf{M} \mathbf{E}_k$ (Zyskind's condition for the submodel), then

$\mathfrak{R}(\mathbf{E}_k \mathbf{M}) = \mathfrak{R}(\mathbf{M} \mathbf{E}_k) \subset \mathfrak{R}(\mathbf{M})$, and hence $\mathbf{E}_k \mathbf{M}$ is an o.p.o (Seely, 2002, problem 2.F.1).

$\Rightarrow \text{tr}(\mathbf{E}_k \mathbf{M}) = r(\mathbf{E}_k \mathbf{M})$ (Harville, 1997, corollary 10.2.2 and also Seely's notes).

$$= r(\mathbf{M}) - \dim[\mathfrak{R}(\mathbf{M}) \cap \mathfrak{R}(\mathbf{E}_k)] \quad (\text{Seely, 2002, proposition 1.8.2}).$$

When $k = s$, then $\dim[\mathfrak{R}(\mathbf{M}) \cap \mathfrak{R}(\mathbf{E}_s)] = 0$ (assumption A7),

and hence $r(\mathbf{E}_k \mathbf{M}) = r(\mathbf{M}) = \ell$ (from the remark above).

When $k \neq s$, then $\mathfrak{R}(\mathbf{M}) \subset \mathfrak{R}(\mathbf{E}_s) \Rightarrow \mathbf{E}_k \mathfrak{R}(\mathbf{M}) \subset \mathbf{E}_k \mathfrak{R}(\mathbf{E}_s) \Rightarrow \mathfrak{R}(\mathbf{E}_k \mathbf{M}) \subset \mathfrak{R}(\mathbf{E}_k \mathbf{E}_s) = \mathbf{0}$,

and then we have $\mathbf{E}_k \mathbf{M} = \mathbf{0}$.

Hence, expressions (6.5) and (6.6) can be simplified as

$$A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\alpha_{si} \alpha_{sj}}{[\alpha_s(\boldsymbol{\sigma})]^2} \ell^2,$$

and

$$A_2 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\alpha_{si} \alpha_{sj}}{[\alpha_s(\boldsymbol{\sigma})]^2} \ell. \quad \square \quad (6.7)$$

Lemma 6.1.2.4 For a modified Rady model, the approximation and the exact approaches to derive $\mathbf{W} = \text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}]$ are identical.

Proof *The Exact Approach*

$$f(\mathbf{z}) = (2\pi)^{-q/2} |\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}' (\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{z}\right], \quad \mathbf{z} \sim N(\mathbf{0}, \mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})$$

$$\Rightarrow \ell_R(\boldsymbol{\sigma}) = \text{constant} - \frac{1}{2} \log |\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K}| - \frac{1}{2} \mathbf{z}' (\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{z}$$

$$= \text{constant} - \frac{1}{2} \log |\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K}| - \frac{1}{2} \mathbf{y}' \mathbf{K} (\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{K}' \mathbf{y},$$

Since $\mathbf{B} = \{\alpha_{ki}\}_{r \times r}$ is invertible (from the remarks above) which is equivalent to

$\boldsymbol{\alpha} = \{\alpha_k(\boldsymbol{\sigma})\}_{r \times 1}$ a one to one function.

Now, we reparametrize by making $\gamma_k = \alpha_k(\boldsymbol{\sigma})$, and hence we have

$$\begin{aligned} \frac{\partial \ell_R(\boldsymbol{\gamma})}{\partial \gamma_k} &= -\frac{1}{2} \text{tr} \left[(\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{K}' \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_k} \mathbf{K} \right] + \frac{1}{2} \mathbf{y}' \mathbf{K} (\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{K}' \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_k} \mathbf{K} (\mathbf{K}' \boldsymbol{\Sigma} \mathbf{K})^{-1} \mathbf{K}' \mathbf{y} \\ &= -\frac{1}{2} \text{tr} \left(\mathbf{G} \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_k} \right) + \frac{1}{2} \mathbf{y}' \mathbf{G} \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_k} \mathbf{G} \mathbf{y}, \end{aligned}$$

where $\mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = \mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{K}' = \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{P}_\mathbf{X})$ (from 6.1)

$$\begin{aligned} \Rightarrow \quad \frac{\partial \ell_R(\boldsymbol{\gamma})}{\partial \gamma_k} &= -\frac{1}{2} \text{tr}(\mathbf{G}\mathbf{E}_k) + \frac{1}{2} \mathbf{y}'\mathbf{G}\mathbf{E}_k\mathbf{y}, \quad \text{since } \boldsymbol{\Sigma} = \sum_{k=1}^r \gamma_k \mathbf{E}_k \\ &= -\frac{1}{2} \text{tr}\left(\frac{1}{\gamma_k} \mathbf{C}_k\right) + \frac{1}{2} \mathbf{y}' \frac{\mathbf{C}_k}{\gamma_k^2} \mathbf{y}, \quad \text{since } \mathbf{G} = \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{P}_\mathbf{X}) = \sum_{k=1}^r \frac{1}{\gamma_k} \mathbf{C}_k \end{aligned}$$

$$\text{Equating } \frac{\partial \ell_R(\boldsymbol{\gamma})}{\partial \gamma_k} \text{ with zero,} \quad \alpha_k(\hat{\boldsymbol{\sigma}})_{\text{REML}} = \hat{\gamma}_k = \frac{\mathbf{y}'\mathbf{C}_k\mathbf{y}}{t_k}, \quad \text{for } k=1, \dots, r \quad (6.8)$$

$$\mathbf{B}\hat{\boldsymbol{\sigma}} = \begin{bmatrix} \alpha_1(\hat{\boldsymbol{\sigma}}) \\ \vdots \\ \alpha_r(\hat{\boldsymbol{\sigma}}) \end{bmatrix} \Rightarrow \hat{\boldsymbol{\sigma}}_{\text{REML}} = \mathbf{B}^{-1} \begin{bmatrix} \alpha_1(\hat{\boldsymbol{\sigma}}) \\ \vdots \\ \alpha_r(\hat{\boldsymbol{\sigma}}) \end{bmatrix} = \mathbf{B}^{-1}\mathbf{s}, \quad \text{where } \mathbf{s} = \begin{bmatrix} \vdots \\ \frac{\mathbf{y}'\mathbf{C}_k\mathbf{y}}{t_k} \\ \vdots \end{bmatrix}$$

The above expression is an explicit form for $\hat{\boldsymbol{\sigma}}_{\text{REML}}$.

$$\text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}] = \text{Var}[\mathbf{B}^{-1}\mathbf{s}] = \mathbf{B}^{-1} \text{Var}[\mathbf{s}](\mathbf{B}^{-1})'.$$

Observe that for any two different elements of \mathbf{s} ,

$$\text{Cov}[\mathbf{y}'\mathbf{C}_i\mathbf{y}, \mathbf{y}'\mathbf{C}_j\mathbf{y}] = 2\text{tr}(\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_j\boldsymbol{\Sigma}) + 4(\mathbf{X}\boldsymbol{\beta})'\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_j\mathbf{X}\boldsymbol{\beta} = 0 \quad (\text{Schott, 2005, theorem 10.22}).$$

And

$$\text{Cov}[\mathbf{y}'\mathbf{C}_i\mathbf{y}, \mathbf{y}'\mathbf{C}_i\mathbf{y}] = 2\text{tr}(\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_i\boldsymbol{\Sigma}) + 4(\mathbf{X}\boldsymbol{\beta})'\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_i\mathbf{X}\boldsymbol{\beta} = 2\text{tr}(\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_i\boldsymbol{\Sigma}),$$

$$\begin{aligned} \mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_i\boldsymbol{\Sigma} &= (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{E}_i \left(\sum_{k=1}^q \alpha_k(\boldsymbol{\sigma})\mathbf{E}_k \right) (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{E}_i \left(\sum_{k=1}^q \alpha_k(\boldsymbol{\sigma})\mathbf{E}_k \right) \\ &= (\mathbf{I} - \mathbf{P}_\mathbf{X})[\alpha_i(\boldsymbol{\sigma})]^2 \mathbf{E}_i, \quad \text{because } \mathbf{E}_k\text{'s are orthogonal o.p. matrices} \\ &= [\alpha_i(\boldsymbol{\sigma})]^2 \mathbf{C}_i \end{aligned}$$

$$\text{So, } \text{Var}[\mathbf{s}] = \text{diag} \left\{ \frac{2[\alpha_i(\boldsymbol{\sigma})]^2}{t_i} \right\}_{r \times r} = 2\mathbf{D},$$

$$\text{and hence} \quad \text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}] = 2\mathbf{B}^{-1}\mathbf{D}(\mathbf{B}^{-1})'. \quad (6.9)$$

The Approximation Approach

$$\mathbf{W} = \text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}] \approx 2 \left[\left\{ \text{tr} \left(\mathbf{G} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \mathbf{G} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j} \right)_{i,j=1}^r \right\}_m \right]^{-1} \quad (\text{Searle et al., 1992, P.253}),$$

$$\begin{aligned} \text{tr} \left(\mathbf{G} \frac{\partial \Sigma}{\partial \sigma_i} \mathbf{G} \frac{\partial \Sigma}{\partial \sigma_j} \right) &= \text{tr} \left[\Sigma^{-1} (\mathbf{I} - \mathbf{P}_X) \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} (\mathbf{I} - \mathbf{P}_X) \frac{\partial \Sigma}{\partial \sigma_j} \right] \\ &= \text{tr} \left[\sum_{k=1}^r \frac{\alpha_{ki} \alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^2} \mathbf{E}_k (\mathbf{I} - \mathbf{P}_X) \right] = \sum_{k=1}^r \frac{\alpha_{ki} \alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^2} t_k, \quad \text{where } \text{tr}(\mathbf{C}_k) = t_k \geq 1. \end{aligned}$$

Define $\boldsymbol{\Psi} = \left\{ \sum_{k=1}^r \alpha_{ki} \frac{t_k}{[\alpha_k(\boldsymbol{\sigma})]^2} \alpha_{kj} \right\}_{i,j=1}^r = \mathbf{B}' \mathbf{D}^{-1} \mathbf{B}$, where \mathbf{B} and \mathbf{D} are defined as above.

$$\Rightarrow \boldsymbol{\Psi}^{-1} = \mathbf{B}^{-1} \mathbf{D} (\mathbf{B}')^{-1},$$

$$\text{hence } \mathbf{W} = \text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}] \approx 2 \boldsymbol{\Psi}^{-1} = 2 \mathbf{B}^{-1} \mathbf{D} (\mathbf{B}')^{-1}. \quad (6.10)$$

From expressions (6.9 and 6.10), we see that both approaches give same result. \square

Corollary 6.1.2.5 For a modified Rady testing problem, $A_1 = \frac{2\ell^2}{t_s}$ and $A_2 = \frac{2\ell}{t_s}$.

Proof From expression (6.7), $A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\alpha_{si} \alpha_{sj}}{[\alpha_s(\boldsymbol{\sigma})]^2} \ell^2 = \frac{\ell^2}{[\alpha_s(\boldsymbol{\sigma})]^2} \sum_{i=1}^r \sum_{j=1}^r \alpha_{si} w_{ij} \alpha_{sj}$.

Notice that $\sum_{i=1}^r \sum_{j=1}^r \alpha_{si} w_{ij} \alpha_{sj}$ is an entry of \mathbf{BWB}' ; call it $(\mathbf{BWB}')_{ss}$, where \mathbf{B} as in above.

From expressions (6.9) or (6.10), we have $\mathbf{BWB}' = \mathbf{B}[\mathbf{B}^{-1} \mathbf{D} (\mathbf{B}')^{-1}] \mathbf{B}' = \mathbf{D}$,

$$\text{and hence } A_1 = \frac{\ell^2}{[\alpha_s(\boldsymbol{\sigma})]^2} \times \frac{2[\alpha_s(\boldsymbol{\sigma})]^2}{t_s} = \frac{2\ell^2}{t_s}, \quad \text{and } A_2 = \frac{2\ell}{t_s} \quad \square$$

Lemma 6.1.2.6 (Rady, 1986)

Under the null hypothesis, the Anova test statistic $F_R \sim F(\ell, t_s)$, where $F_R = \frac{t_s \mathbf{y}' \mathbf{M} \mathbf{y}}{\ell \mathbf{y}' \mathbf{C}_s \mathbf{y}}$.

Lemma 6.1.2.7 The K-R test statistic $F = F_R$.

$$\text{Proof} \quad F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}}_A \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}},$$

and since $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$ (lemma 6.1.2.1), then

$$F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}}$$

$$\begin{aligned} (i) \quad (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} &= \left[\mathbf{L}' (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{L} \right]^{-1} \\ &= \left[\mathbf{A}' \mathbf{X} (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{A} \right]^{-1} \quad (\text{notice that } \mathbf{L} = \mathbf{X}' \mathbf{A}) \\ &= \left[\mathbf{A}' \boldsymbol{\Sigma} \mathbf{P}_x \mathbf{A} \right]^{-1}, \quad \text{utilizing expression (6.1)} \\ &= \left[\mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \right]^{-1}, \quad \text{because } \Re(\mathbf{A}) \subset \Re(\mathbf{X}), \end{aligned}$$

$$(ii) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \quad (\text{utilizing expression 6.1}).$$

From (i) and (ii),

$$\begin{aligned} F &= \frac{1}{\ell} \mathbf{y}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{A} \left[\mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \right]^{-1} \mathbf{A}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \\ &= \frac{1}{\ell} \mathbf{y}' \mathbf{P}_x \mathbf{A} \left[\mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \right]^{-1} \mathbf{A}' \mathbf{P}_x \mathbf{y} \\ &= \frac{1}{\ell} \mathbf{y}' \mathbf{A} \left[\mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \right]^{-1} \mathbf{A}' \mathbf{y} \quad (\text{because } \Re(\mathbf{A}) \subset \Re(\mathbf{X})) \\ &= \frac{1}{\ell} \mathbf{y}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{M} \mathbf{y} \quad (\text{utilizing expression 6.4}), \end{aligned}$$

$$\text{but } \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{M} = \sum_{k=1}^q \frac{1}{\alpha_k(\hat{\boldsymbol{\sigma}})} \mathbf{E}_k \mathbf{M} = \frac{1}{\alpha_s(\hat{\boldsymbol{\sigma}})} \mathbf{M} \quad (\text{from the remarks of this section}),$$

$$\text{where } \alpha_s(\hat{\boldsymbol{\sigma}}) = \frac{\mathbf{y}' \mathbf{C}_s \mathbf{y}}{t_s} \quad (\text{from expression 6.8}).$$

$$\text{Therefore,} \quad F = \frac{t_s \mathbf{y}' \mathbf{M} \mathbf{y}}{\ell \mathbf{y}' \mathbf{C}_s \mathbf{y}} = F_R \quad \square$$

6.1.3 Estimating the Denominator Degrees of Freedom and the Scale Factor

From corollary, 6.1.2.5, $A_1 = \frac{2\ell^2}{t_s}$, and $A_2 = \frac{2\ell}{t_s}$, and then $\frac{A_1}{A_2} = \ell$.

By utilizing theorems (5.3.1 and 5.3.2) and corollary (5.3.3), K-R and our proposed

approaches are identical. $m^* = \frac{2\ell}{A_2} = t_s$, and $\lambda^* = 1$.

The estimates of the denominator degrees of freedom and the scale factor match the exact

values for the Anova test in lemma 6.1.2.6. So, the K-R approach produces the exact values for the modified Rady model which is more general than the balanced one-way Anova model which is one of the special cases considered in modifying the K-R approach.

6.2 A General Multivariate Linear Model

This model was studied by Mardia *et al.* (1992) where they show that for this model, we have an exact F test to test a linear hypothesis of the fixed effects.

6.2.1 The Model Setting

Consider the model defined by

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U},$$

where $\mathbf{Y}(n \times p)$ is an observed matrix of p response variables on each of n individuals, $\mathbf{X}(n \times q)$ is the design matrix, $\mathbf{B}(q \times p)$ is a matrix of unknown parameters, and \mathbf{U} is a matrix of unobserved random disturbances whose rows for given \mathbf{X} are uncorrelated. We also assume that \mathbf{U} is a data matrix from $N_p(0, \Sigma_p)$.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}'_{(1)} \\ \vdots \\ \boldsymbol{\beta}'_{(q)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_p & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma_p \end{bmatrix} = \mathbf{I}_n \otimes \Sigma_p$$

\mathbf{x}_i 's are $q \times 1$ vectors, and \mathbf{y}_i 's and $\boldsymbol{\beta}_i$'s are $p \times 1$ vectors.

Suppose that we are interested in testing: $H_0 : \mathbf{c}'\mathbf{B}\mathbf{M} = \mathbf{0}$,

where $\mathbf{c}(1 \times q)$, and $\mathbf{M}(p \times r)$ are given matrices, and \mathbf{M} have a rank r .

Note A special case of this general model is the Hotelling T^2 one sample test, in which $\mathbf{X} = \mathbf{1}_n$, $\mathbf{B} = \boldsymbol{\mu}'$, $\mathbf{c} = 1$, $\mathbf{M} = \mathbf{I}_p$. So, $p = r$, and $q = 1$. Also, the Hotelling two samples is a special case of this model.

Other Setting

To compute the Wald statistic for our problem so we are able to compute the estimates of the Kenward and Roger method, we rearrange \mathbf{Y} as a vector.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(q)} \end{bmatrix}$$

\mathbf{y} and $\boldsymbol{\beta}$ are $np \times 1$ and $pq \times 1$ vectors respectively.

Notice that the design matrix for the new setting will be $\mathbf{X}^* = \mathbf{X} \otimes \mathbf{I}_p$.

The testing problem $H_0: \mathbf{c}'\mathbf{B}\mathbf{M} = \mathbf{0}$,

$$\text{which is } \begin{bmatrix} c_1 & \cdots & c_q \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}'_{(1)} \\ \vdots \\ \boldsymbol{\beta}'_{(q)} \end{bmatrix} \mathbf{M} = \begin{bmatrix} c_1 & \cdots & c_q \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}'_{(1)}\mathbf{M} \\ \vdots \\ \boldsymbol{\beta}'_{(q)}\mathbf{M} \end{bmatrix} = \mathbf{0}.$$

$$\Leftrightarrow \sum_{j=1}^q c_j \boldsymbol{\beta}'_{(j)} \mathbf{M} = \mathbf{0} \Leftrightarrow \sum_{j=1}^q c_j \mathbf{M}' \boldsymbol{\beta}_{(j)} = \mathbf{0} \Leftrightarrow \begin{bmatrix} c_1 \mathbf{M}' & \cdots & c_q \mathbf{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(q)} \end{bmatrix} = \mathbf{0},$$

which is can be written as $H_0: \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{L} = \mathbf{c} \otimes \mathbf{M}$.

$$\begin{aligned} \text{So under this setting, } \Phi &= \left[(\mathbf{X} \otimes \mathbf{I}_p)' (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_p^{-1}) (\mathbf{X} \otimes \mathbf{I}_p) \right]^{-1} \\ &= \left(\mathbf{X}'\mathbf{X} \otimes \boldsymbol{\Sigma}_p^{-1} \right)^{-1} = (\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p, \\ \boldsymbol{\beta} &= \left[(\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p \right] (\mathbf{X}' \otimes \mathbf{I}_p) (\mathbf{I}_n \otimes \boldsymbol{\Sigma}_p^{-1}) \mathbf{y} \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I}_p \right] \mathbf{y}, \end{aligned}$$

$$\text{and } (\mathbf{L}'\Phi\mathbf{L})^{-1} = \left[(\mathbf{c}' \otimes \mathbf{M}') ((\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p) (\mathbf{c} \otimes \mathbf{M}) \right]^{-1} = \frac{(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}.$$

6.2.2 Useful Properties for the Model

In this section, we derive several lemmas to summarize the useful properties for this model under the specified testing problem.

Lemma 6.2.2.1 (Mardia et al., 1992, chapter 6)

Let $F_M = \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{M}(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{(n-q)\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$, where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimator

for Σ_p . Under the null hypothesis, $F_M \sim [r/(n-q-r+1)]F_{r,n-q-r+1}$.

Note Similar to lemma 4.2.3, $\hat{\Sigma}_p = \frac{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'}{n-q}$

Special Notation Since $\sigma = \{\sigma_{if}\}_{\frac{p(p+1)}{2}} (i \leq f)$, then like in the Hotelling T^2 case, we use

the notation $\mathbf{P}_{if}, \mathbf{P}_{jg}, \mathbf{Q}_{if,jg}, \mathbf{R}_{if,jg}$, and $w_{if,jg}$.

Lemma 6.2.2.2 $\hat{\Phi}_A = \hat{\Phi}$.

Proof $\mathbf{P}_{if} = (\mathbf{X}' \otimes \mathbf{I}_p) \left(\mathbf{I}_n \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \right) (\mathbf{X} \otimes \mathbf{I}_p) = \mathbf{X}' \mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}}$

$$\begin{aligned} \Rightarrow \mathbf{P}_{if} \Phi \mathbf{P}_{jg} &= \left(\mathbf{X}' \mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \right) \left[(\mathbf{X}' \mathbf{X})^{-1} \otimes \Sigma_p \right] \left(\mathbf{X}' \mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}} \right) \\ &= \mathbf{X}' \mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \Sigma_p \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}}, \end{aligned}$$

$$\begin{aligned} \mathbf{Q}_{if,jg} &= (\mathbf{X}' \otimes \mathbf{I}_p) \left(\mathbf{I}_n \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \right) (\mathbf{I}_n \otimes \Sigma_p) \left(\mathbf{I}_n \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}} \right) (\mathbf{X} \otimes \mathbf{I}_p) \\ &= \mathbf{X}' \mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \Sigma_p \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{jg}} \end{aligned}$$

$$\Rightarrow \mathbf{Q}_{if,jg} - \mathbf{P}_{if} \Phi \mathbf{P}_{jg} = \mathbf{0}, \text{ and since } \mathbf{R}_{if,jg} = \mathbf{0}, \text{ then we conclude that } \hat{\Phi}_A = \hat{\Phi}. \quad \square$$

Lemma 6.2.2.3 The K-R test statistic $F = \frac{r}{n-q} F_M$.

Proof $F = \frac{1}{\ell} \hat{\beta}' \mathbf{L} (\mathbf{L}' \hat{\Phi}_A \mathbf{L})^{-1} \mathbf{L}' \hat{\beta}$

$$= \frac{1}{\ell} \hat{\beta}' \mathbf{L} (\mathbf{L}' \hat{\Phi} \mathbf{L})^{-1} \mathbf{L}' \hat{\beta} \quad (\text{lemma 6.2.2.2})$$

$$= \frac{1}{\ell} \left(\left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I}_p \right] \mathbf{y} \right)' (\mathbf{c} \otimes \mathbf{M}) \frac{(\mathbf{M}' \hat{\Sigma}_p \mathbf{M})^{-1}}{\mathbf{c}' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{c}} (\mathbf{c}' \otimes \mathbf{M}') \left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I}_p \right] \mathbf{y}$$

$$= \frac{1}{r\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \mathbf{y}' \left\{ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}(\mathbf{M}'\hat{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \right\} \mathbf{y}.$$

In the expression above, we used $\ell = r$, and this is because $\dim(\mathbf{L}') = r \times pq$.

In addition,

$$\mathbf{y}' \left\{ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}(\mathbf{M}'\hat{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \right\} \mathbf{y} = \mathbf{y}'(\mathbf{d}\mathbf{d}' \otimes \mathbf{A})\mathbf{y} = (\text{vec } \mathbf{Y})'(\mathbf{d}\mathbf{d}' \otimes \mathbf{A})\text{vec } \mathbf{Y},$$

$$\text{and } \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{M}(\mathbf{M}'\hat{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = \mathbf{d}'\mathbf{Y}\mathbf{A}\mathbf{Y}'\mathbf{d},$$

where $\mathbf{d} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$, $\mathbf{A} = \mathbf{M}(\mathbf{M}'\hat{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'$, $\text{vec } \mathbf{Y}$ is the np dimensional column vector that we obtain when we stack the columns of \mathbf{Y} .

Since $(\text{vec } \mathbf{Y})'(\mathbf{d}\mathbf{d}' \otimes \mathbf{A})\text{vec } \mathbf{Y} = \mathbf{d}'\mathbf{Y}\mathbf{A}\mathbf{Y}'\mathbf{d}$ (Harville, 1997, theorem 16.2.2), the proof is completed. \square

Corollary 6.2.2.4 Under the null hypothesis, $\frac{n-q-r+1}{n-q}F \sim F(r, n-q-r+1)$.

Proof Direct from lemmas 6.2.2.1 and 6.2.2.3.

$$\textbf{Lemma 6.2.2.5} \quad A_1 = \frac{2\ell}{n-q}, \quad \text{and} \quad A_2 = \frac{\ell(\ell+1)}{n-q}.$$

Proof

$$A_1 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\Theta \Phi \mathbf{P}_{if} \Phi) \text{tr}(\Theta \Phi \mathbf{P}_{jg} \Phi),$$

$$\text{and} \quad A_2 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr}(\Theta \Phi \mathbf{P}_{if} \Phi \Theta \Phi \mathbf{P}_{jg} \Phi).$$

$$\Theta = \mathbf{L}(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}' = (\mathbf{c} \otimes \mathbf{M}) \frac{(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} (\mathbf{c}' \otimes \mathbf{M}') = \frac{\mathbf{c}\mathbf{c}' \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}'}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

$$\text{tr}(\Theta \Phi \mathbf{P}_{if} \Phi)$$

$$\begin{aligned} &= \frac{1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \text{tr} \left\{ \left[\mathbf{c}\mathbf{c}' \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}' \right] \left[(\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma_p \right] \left[\mathbf{X}'\mathbf{X} \otimes \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \right] \left[(\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma_p \right] \right\} \\ &= \frac{1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \text{tr} \left[\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}'\Sigma_p \frac{\partial \Sigma_p^{-1}}{\partial \sigma_{if}} \Sigma_p \right] \end{aligned}$$

$$= -\text{tr} \left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \right]$$

Accordingly, A_1 can be expressed as

$$A_1 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr} \left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \right] \text{tr} \left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}} \right],$$

and similarly,

$$A_2 = \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p w_{if,jg} \text{tr} \left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}} \right].$$

Let $\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}' = \mathbf{T}$. Since $\mathbf{T} = \{t_{ij}\}_{p \times p}$ is a symmetric matrix, then

$$\text{tr} \left(\mathbf{T} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \right) = \begin{cases} t_{ii} & \text{for } i = f \\ 2t_{if} & \text{for } i \neq f \end{cases}$$

$$\text{Or, } \text{tr} \left(\mathbf{T} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \right) = 2^{1-\delta_{if}} t_{if}, \quad \text{where } \delta_{if} = \begin{cases} 1 & \text{for } i = f \\ 0 & \text{for } i \neq f \end{cases} \quad (6.11)$$

$$\text{It's easy to check that } \text{tr} \left(\mathbf{T} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}} \mathbf{T} \frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}} \right) = 2^{1-\delta_{if}-\delta_{jg}} (t_{ig}t_{jf} + t_{jg}t_{if}) \quad (6.12)$$

Moreover, analogous to the proof given for expression (4.6) in the Hotelling T^2 case,

$$w_{if,jg} = \text{Cov}[\hat{\sigma}_{if}, \hat{\sigma}_{jg}] = \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)} \quad (6.13)$$

Utilizing expressions (6.11) and (6.13),

$$\begin{aligned} A_1 &= \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{j=1}^p \sum_{\substack{g=1, \\ j \leq g}}^p \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)} 2^{2-\delta_{if}-\delta_{jg}} t_{if}t_{jg} \\ &= \sum_{i=1}^p \sum_{f=1}^p \sum_{j=1}^p \sum_{g=1}^p \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)} t_{if}t_{jg} \end{aligned}$$

Note In the expression above for A_1 , we divided by $2^{2-\delta_{if}-\delta_{jg}}$, and this is true because

the summations repeat $w_{if,jg} 2^{2-\delta_{if}-\delta_{jg}}$ times.

$$\Rightarrow A_1 = \frac{1}{n-q} \left\{ \sum_{f=1}^p \sum_{g=1}^p \left(\sum_{i=1}^p \sigma_{ig}t_{if} \right) \left(\sum_{j=1}^p \sigma_{jf}t_{jg} \right) + \sum_{j=1}^p \sum_{i=1}^p \left(\sum_{g=1}^p \sigma_{fg}t_{jg} \right) \left(\sum_{f=1}^p \sigma_{ij}t_{if} \right) \right\}$$

Notice that
$$\sum_{f=1}^p \sum_{g=1}^p \left(\sum_{i=1}^p \sigma_{ig} t_{if} \right) \left(\sum_{j=1}^p \sigma_{jf} t_{jg} \right) = \text{tr}(\mathbf{T}\Sigma_p \mathbf{T}\Sigma_p),$$

and since $\mathbf{T}\Sigma_p = \mathbf{M}(\mathbf{M}'\Sigma_p \mathbf{M})^{-1} \mathbf{M}'\Sigma_p$ is o.p. on $\Re(\mathbf{M})$, then $\mathbf{T}\Sigma_p \mathbf{T}\Sigma_p = \mathbf{T}\Sigma_p$,
and $\text{tr}(\mathbf{T}\Sigma_p) = r(\mathbf{T}\Sigma_p) = r(\mathbf{M}) = r$ (Seely, 2002, chapter 2).

Hence
$$A_1 = \frac{1}{n-q}(r+r) = \frac{2r}{n-q}. \quad (6.14)$$

Similarly, and by utilizing expressions (6.12) and (6.13),

$$\begin{aligned} A_2 &= \sum_{i=1}^p \sum_{\substack{f=1, \\ i \leq f}}^p \sum_{\substack{j=1, \\ j \leq g}}^p \sum_{g=1}^p \frac{(\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij})}{(n-q)} (t_{ig} t_{jf} + t_{fg} t_{ij}) 2^{1-\delta_{if}-\delta_{ij}} \\ &= \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \frac{(\sigma_{ig} \sigma_{jf} + \sigma_{fg} \sigma_{ij})}{2(n-q)} (t_{ig} t_{jf} + t_{fg} t_{ij}) \\ &= \frac{1}{2(n-q)} \left\{ \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{ig} t_{ig} \sigma_{jf} t_{jf} + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{ig} t_{fg} \sigma_{jf} t_{ij} \right. \\ &\quad \left. + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{fg} t_{ig} \sigma_{ij} t_{jf} + \sum_{i=1}^p \sum_{j=1}^p \sum_{g=1}^p \sum_{f=1}^p \sigma_{fg} t_{fg} \sigma_{ij} t_{ij} \right\} \\ &= \frac{1}{2(n-q)} \left\{ \sum_{g=1}^p \sum_{i=1}^p \sigma_{ig} t_{ig} \sum_{f=1}^p \sum_{j=1}^p \sigma_{jf} t_{jf} + \sum_{i=1}^p \sum_{f=1}^p \left(\sum_{g=1}^p \sigma_{ig} t_{fg} \right) \left(\sum_{j=1}^p \sigma_{jf} t_{ij} \right) \right. \\ &\quad \left. + \sum_{i=1}^p \sum_{f=1}^p \left(\sum_{g=1}^p \sigma_{fg} t_{ig} \right) \left(\sum_{j=1}^p \sigma_{ij} t_{jf} \right) + \sum_{g=1}^p \sum_{f=1}^p \sigma_{fg} t_{fg} \sum_{j=1}^p \sum_{i=1}^p \sigma_{ij} t_{ij} \right\} \end{aligned}$$

Since $\sum_{f=1}^p \sum_{j=1}^p \sigma_{jf} t_{jf} = \text{tr}(\mathbf{T}\Sigma_p) = r(\mathbf{T}\Sigma_p) = r(\mathbf{M}) = r$

and $\sum_{i=1}^p \sum_{f=1}^p \left(\sum_{g=1}^p \sigma_{fg} t_{ig} \right) \left(\sum_{j=1}^p \sigma_{ij} t_{jf} \right) = r$ (from above),

therefore
$$A_2 = \frac{1}{2(n-q)}(r^2 + r + r + r^2) = \frac{r(r+1)}{n-q}. \quad (6.15)$$

6.2.3 Estimating the Denominator Degrees of Freedom and the Scale

$$A_1 = \frac{2r}{n-q}, \quad \text{and} \quad A_2 = \frac{r(r+1)}{n-q}, \quad \text{and then} \quad \frac{A_1}{A_2} = \frac{2}{r+1} \quad (\text{lemma 6.2.2.5}).$$

By utilizing theorems (5.3.1, 5.3.2) and corollary (5.3.3), K-R and the proposed approaches are identical,

$$\text{where} \quad m^* = \frac{r(r+1)}{A_2} - (r-1) = n - q - r + 1 \quad (\text{notice that } \ell = r),$$

$$\text{and} \quad \lambda^* = 1 - \frac{r-1}{r(r+1)} A_2 = 1 - \frac{r-1}{n-q} = \frac{n-q-r+1}{n-q}.$$

The estimates of the denominator degrees of freedom and the scale match the values for the exact multivariate test in corollary 6.2.2.4. So, the K-R approach produces the exact values for the general multivariate linear model considered in this section which is more general than the Hotelling T^2 model.

6.3 A Balanced Multivariate Model with a Random Group Effect

This model was studied by Birkes (2006) where he showed that there exists an exact F test to test a linear hypothesis of the fixed effects under a specific condition as we will see shortly. In this section, we prove that the K-R and the proposed approaches as well give the same exact values for the denominator degrees of freedom and the scale.

6.3.1 Model Setting

Suppose an experiment yields measurements on p characteristics of a sample of subjects, and suppose the subjects are in t groups each of size m . The measurements of the k -th characteristic of the j -th subject in the i -th group is denoted by y_{ijk} ($i = 1, \dots, t$, $j = 1, \dots, m$, $k = 1, \dots, p$). The experiment is designed so that the j -th subject in each group is associated with q covariates x_{jl} ($l = 1, \dots, q$), not depending on i . The design is balanced in so far as the groups all have the same size m and the same set of covariates. Each measurement y_{ijk} includes a random group effect a_{ik} and a random subject error e_{ijk} .

Notation

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_m \end{bmatrix}, \quad \mathbf{x}_j = \begin{bmatrix} x_{j1} \\ \vdots \\ x_{jq} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_t \end{bmatrix}, \quad \mathbf{Y}_i = \begin{bmatrix} \mathbf{y}'_{i1} \\ \vdots \\ \mathbf{y}'_{im} \end{bmatrix}, \quad \mathbf{y}_{ij} = \begin{bmatrix} y_{ij1} \\ \vdots \\ y_{ijp} \end{bmatrix},$$

$$\mathbf{B} = [\boldsymbol{\beta}^{(1)} \quad \dots \quad \boldsymbol{\beta}^{(p)}], \quad \boldsymbol{\beta}^{(k)} = \begin{bmatrix} \beta_{k1} \\ \vdots \\ \beta_{kp} \end{bmatrix}, \quad \mathbf{a}_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{ip} \end{bmatrix}, \quad \mathbf{e}_{ij} = \begin{bmatrix} e_{ij1} \\ \vdots \\ e_{ijp} \end{bmatrix}$$

Model

$y_{ijk} = \boldsymbol{\beta}^{(k)'} \mathbf{x}_j + a_{ik} + e_{ijk}$, hence $\mathbf{y}_{ij} = \mathbf{B}' \mathbf{x}_j + \mathbf{a}_i + \mathbf{e}_{ij}$. We assume the β_{kl} are fixed unknown parameters, the \mathbf{a}_i are i.i.d. $N_p(\mathbf{0}, \boldsymbol{\Lambda})$, the \mathbf{e}_{ij} are i.i.d. $N_p(\mathbf{0}, \boldsymbol{\Sigma}_p)$, and the \mathbf{a}_i are independent of the \mathbf{e}_{ij} . We also require $\mathbf{1}_m \in \mathfrak{R}(\mathbf{X})$.

Hypothesis

$H_0 : \mathbf{c}' \mathbf{B} \mathbf{M} = \mathbf{0}$, where \mathbf{c} is $q \times 1$ and \mathbf{M} is $p \times r$. We require $\mathbf{1}_m' \mathbf{X} (\mathbf{X}' \mathbf{X}) \mathbf{c} = 0$.

Other Setting

To compute the Wald statistic for our problem and in order to compare the statistic mentioned above and the Wald statistic, we rearrange \mathbf{Y} as a vector.

$$\mathbf{y} = \text{vec}(\mathbf{Y}'), \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\tau}_1 \\ \vdots \\ \boldsymbol{\tau}_q \end{bmatrix} \quad \text{where } \boldsymbol{\tau}_i = \begin{bmatrix} \beta_{1i} \\ \vdots \\ \beta_{pi} \end{bmatrix}$$

\mathbf{y} and $\boldsymbol{\beta}$ are $mtp \times 1$ and $pq \times 1$ vectors respectively.

$$\text{Notice that } \mathbf{B} = [\boldsymbol{\beta}^{(1)} \quad \dots \quad \boldsymbol{\beta}^{(p)}] = \begin{bmatrix} \beta_{11} & \dots & \beta_{p1} \\ \vdots & \vdots & \vdots \\ \beta_{1q} & \dots & \beta_{pq} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\tau}'_1 \\ \vdots \\ \boldsymbol{\tau}'_q \end{bmatrix},$$

the design matrix for the new setting will be $\mathbf{X}^* = \mathbf{1}_t \otimes (\mathbf{X} \otimes \mathbf{I}_p)$,

and $\boldsymbol{\Sigma} = \mathbf{I}_t \otimes \boldsymbol{\Lambda}$,

where $\boldsymbol{\Lambda} = \mathbf{I}_m \otimes \boldsymbol{\Sigma}_p + m \mathbf{P}_{\mathbf{1}_m} \otimes \boldsymbol{\Lambda}$

Special Notation As in sections 4.2 and 6.2, we use the notation \mathbf{P}_{if} , $\mathbf{Q}_{if,jg}$, $\mathbf{R}_{if,jg}$, and

$w_{if,jg}$.

6.3.2 Useful properties for the Model

In this section, we derive several lemmas that show useful properties about the model and the testing problem.

Lemma 6.3.2.1 $\hat{\Phi}_A = \hat{\Phi}$.

Proof $\Delta = \mathbf{I}_m \otimes \Sigma_p + m\mathbf{P}_{\mathbf{I}_m} \otimes \Lambda$

$$= (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \Sigma_p + \mathbf{P}_{\mathbf{I}_m} \otimes (\Sigma_p + m\Lambda),$$

$$\Delta^{-1} = (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \Sigma_p^{-1} + \mathbf{P}_{\mathbf{I}_m} \otimes (\Sigma_p + m\Lambda)^{-1} \quad (6.16)$$

$$\Rightarrow (\mathbf{X}' \otimes \mathbf{I}_p) \Delta^{-1} (\mathbf{X} \otimes \mathbf{I}_p) = (\mathbf{X}' \otimes \mathbf{I}_p) \left\{ (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \Sigma_p^{-1} + \mathbf{P}_{\mathbf{I}_m} \otimes (\Sigma_p + m\Lambda)^{-1} \right\} (\mathbf{X} \otimes \mathbf{I}_p)$$

$$= \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} \otimes \Sigma_p^{-1} + \mathbf{X}' \mathbf{P}_{\mathbf{I}_m} \mathbf{X} \otimes (\Sigma_p + m\Lambda)^{-1}$$

Hence,

$$\Phi = [(\mathbf{I}_t \otimes \mathbf{X} \otimes \mathbf{I}_p)' (\mathbf{I}_t \otimes \Delta^{-1}) (\mathbf{I}_t \otimes \mathbf{X} \otimes \mathbf{I}_p)]^{-1}$$

$$= \frac{[(\mathbf{X}' \otimes \mathbf{I}_p) \Delta^{-1} (\mathbf{X} \otimes \mathbf{I}_p)]^{-1}}{t}$$

$$= \frac{[\mathbf{X}' (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} \otimes \Sigma_p^{-1} + \mathbf{X}' \mathbf{P}_{\mathbf{I}_m} \mathbf{X} \otimes (\Sigma_p + m\Lambda)^{-1}]^{-1}}{t}$$

$$= \frac{(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \otimes \Sigma_p + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_{\mathbf{I}_m} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \otimes (\Sigma_p + m\Lambda)}{t} \quad (6.17)$$

$$\mathbf{P}_{if} = \mathbf{I}_t \otimes (\mathbf{X}' \otimes \mathbf{I}_p) \left(\mathbf{I}_t \otimes \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} \right) \mathbf{I}_t \otimes (\mathbf{X} \otimes \mathbf{I}_p)$$

$$= t(\mathbf{X}' \otimes \mathbf{I}_p) \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} (\mathbf{X} \otimes \mathbf{I}_p)$$

Hence, $\mathbf{P}_{if} \Phi \mathbf{P}_{jg} = t^2 (\mathbf{X}' \otimes \mathbf{I}_p) \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} (\mathbf{X} \otimes \mathbf{I}_p) \Phi (\mathbf{X}' \otimes \mathbf{I}_p) \frac{\partial \Delta^{-1}}{\partial \sigma_{jg}} (\mathbf{X} \otimes \mathbf{I}_p).$

$$\mathbf{Q}_{if,jg} = \mathbf{I}_t' \otimes (\mathbf{X}' \otimes \mathbf{I}_p) \left(\mathbf{I}_t \otimes \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} \right) (\mathbf{I}_t \otimes \Lambda) \left(\mathbf{I}_t \otimes \frac{\partial \Delta^{-1}}{\partial \sigma_{jg}} \right) \mathbf{I}_t \otimes (\mathbf{X} \otimes \mathbf{I}_p)$$

$$= t(\mathbf{X}' \otimes \mathbf{I}_p) \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} \Lambda \frac{\partial \Delta^{-1}}{\partial \sigma_{jg}} (\mathbf{X} \otimes \mathbf{I}_p).$$

$$\Rightarrow \mathbf{Q}_{if,jg} - \mathbf{P}_{if} \Phi \mathbf{P}_{jg} = t(\mathbf{X}' \otimes \mathbf{I}_p) \frac{\partial \Delta^{-1}}{\partial \sigma_{if}} \left\{ \Lambda - t(\mathbf{X} \otimes \mathbf{I}_p) \Phi (\mathbf{X}' \otimes \mathbf{I}_p) \right\} \frac{\partial \Delta^{-1}}{\partial \sigma_{jg}} (\mathbf{X} \otimes \mathbf{I}_p)$$

Consider $t(\mathbf{X} \otimes \mathbf{I}_p) \Phi (\mathbf{X}' \otimes \mathbf{I}_p)$

$$\begin{aligned}
&= (\mathbf{X} \otimes \mathbf{I}_p) \left\{ (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p \right. \\
&\quad \left. + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_{\mathbf{I}_m} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \right\} (\mathbf{X}' \otimes \mathbf{I}_p). \\
&= \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes \boldsymbol{\Sigma}_p + \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{P}_{\mathbf{I}_m} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \\
&= \mathbf{P}_\mathbf{X} (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{P}_\mathbf{X} \otimes \boldsymbol{\Sigma}_p + \mathbf{P}_\mathbf{X} \mathbf{P}_{\mathbf{I}_m} \mathbf{P}_\mathbf{X} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \\
&= (\mathbf{P}_\mathbf{X} - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p + \mathbf{P}_{\mathbf{I}_m} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \quad (\text{observe that } \mathbf{1}_m \in \mathfrak{R}(\mathbf{X})) \\
&\Rightarrow \boldsymbol{\Lambda} - t(\mathbf{X} \otimes \mathbf{I}_p) \boldsymbol{\Phi} (\mathbf{X}' \otimes \mathbf{I}_p) = (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p - (\mathbf{P}_\mathbf{X} - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p = (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \otimes \boldsymbol{\Sigma}_p
\end{aligned}$$

Moreover, $\left\{ \boldsymbol{\Lambda} - t(\mathbf{X} \otimes \mathbf{I}_p) \boldsymbol{\Phi} (\mathbf{X}' \otimes \mathbf{I}_p) \right\} \frac{\partial \boldsymbol{\Lambda}^{-1}}{\partial \sigma_{jg}} (\mathbf{X} \otimes \mathbf{I}_p)$

$$\begin{aligned}
&= \left\{ (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \otimes \boldsymbol{\Sigma}_p \right\} \left\{ (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} + \mathbf{P}_{\mathbf{I}_m} \otimes \frac{\partial (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})^{-1}}{\partial \sigma_{jg}} \right\} (\mathbf{X} \otimes \mathbf{I}_p). \\
&= (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \mathbf{X} \otimes \boldsymbol{\Sigma}_p \frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} + (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \mathbf{P}_{\mathbf{I}_m} \mathbf{X} \otimes \boldsymbol{\Sigma}_p \frac{\partial (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})^{-1}}{\partial \sigma_{jg}} = \mathbf{0}
\end{aligned}$$

This implies $\mathbf{Q}_{if,jg} - \mathbf{P}_{if} \boldsymbol{\Phi} \mathbf{P}_{jg} = \mathbf{0}$, and since $\mathbf{R}_{if,jg} = \mathbf{0}$, then $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$. \square

Lemma 6.3.2.2 (Birkes, 2006)

Let $F_d = \frac{(tm-t-q-r+2)t}{r} \cdot \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \bar{\mathbf{Y}} \cdot \mathbf{M} [\mathbf{M}' \mathbf{Y}' \mathbf{Q} \mathbf{Y} \mathbf{M}]^{-1} \mathbf{M}' \bar{\mathbf{Y}} \cdot \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}}$

where $\bar{\mathbf{Y}} = \sum_{i=1}^t \mathbf{Y}_i / t$ and $\mathbf{Q} = \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m})$.

Under H_0 , $F_d \sim F_{r, tm-t-q-r+2}$.

Lemma 6.3.2.3 $F = \frac{t}{r} \cdot \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \bar{\mathbf{Y}} \cdot \mathbf{M} (\mathbf{M}' \hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \mathbf{M}' \bar{\mathbf{Y}} \cdot \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}},$

where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimate of $\boldsymbol{\Sigma}_p$.

Proof Since $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$ (lemma 6.3.2.1), then under the null hypothesis,

$$F = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}}$$

$$\begin{aligned}
(i) \quad \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{\Phi}}(\mathbf{1}'_t \otimes \mathbf{X}' \otimes \mathbf{I}_p)(\mathbf{I}_t \otimes \hat{\boldsymbol{\Lambda}}^{-1})\mathbf{y} \\
&= \hat{\boldsymbol{\Phi}}\left\{\mathbf{1}'_t \otimes (\mathbf{X}' \otimes \mathbf{I}_p)\hat{\boldsymbol{\Lambda}}^{-1}\right\}\mathbf{y} = \left\{\mathbf{1}'_t \otimes \hat{\boldsymbol{\Phi}}(\mathbf{X}' \otimes \mathbf{I}_p)\hat{\boldsymbol{\Lambda}}^{-1}\right\}\mathbf{y},
\end{aligned}$$

From (6.17), $\boldsymbol{\Phi}(\mathbf{X}' \otimes \mathbf{I}_p)$

$$\begin{aligned}
&= \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\right\} \frac{\mathbf{X}' \otimes \mathbf{I}_p}{t} \\
&= \frac{1}{t} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m})\mathbf{P}_X \otimes \boldsymbol{\Sigma}_p + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m}\mathbf{P}_X \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\right\} \\
&= \frac{1}{t} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P}_X - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\right\},
\end{aligned}$$

and from (6.16), we have

$$\begin{aligned}
\boldsymbol{\Phi}(\mathbf{X}' \otimes \mathbf{I}_p)\boldsymbol{\Lambda}^{-1} &= \frac{1}{t} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P}_X - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\right\} \times \\
&\quad \left\{(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \boldsymbol{\Sigma}_p^{-1} + \mathbf{P}_{\mathbf{I}_m} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})^{-1}\right\} \\
&= \frac{1}{t} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P}_X - \mathbf{P}_{\mathbf{I}_m})(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \mathbf{I}_p + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P}_X - \mathbf{P}_{\mathbf{I}_m})\mathbf{P}_{\mathbf{I}_m} \otimes \boldsymbol{\Sigma}_p (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})^{-1}\right. \\
&\quad \left.+ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m}(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\boldsymbol{\Sigma}_p^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m} \otimes \mathbf{I}_p\right\} \\
&= \frac{1}{t} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I}_p\right\}
\end{aligned}$$

$$\text{So,} \quad \hat{\boldsymbol{\beta}} = \frac{1}{t} \left\{\mathbf{1}'_t \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I}_p\right\}\mathbf{y},$$

$$\mathbf{L}'\hat{\boldsymbol{\beta}} = \frac{1}{t}(\mathbf{c}' \otimes \mathbf{M}')\left\{\mathbf{1}'_t \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I}_p\right\}\mathbf{y} = \frac{1}{t}\left\{\mathbf{1}'_t \otimes \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}'\right\}\mathbf{y}$$

(ii) $\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}$

$$\begin{aligned}
&= \frac{1}{t}(\mathbf{c}' \otimes \mathbf{M}')\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_p\right. \\
&\quad \left.+ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\right\}(\mathbf{c} \otimes \mathbf{M}) \\
&= \frac{1}{t}\left\{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \otimes \mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M}\right. \\
&\quad \left.+ \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{I}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \otimes \mathbf{M}'(\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})\mathbf{M}\right\}
\end{aligned}$$

$$= \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{t} \mathbf{M}'\boldsymbol{\Sigma}_p \mathbf{M} \quad (\text{notice that } \mathbf{1}_m' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = 0) \quad (6.18)$$

Accordingly, from (i) and (ii),

$$\begin{aligned} F &= \frac{1}{rt} \cdot \frac{\mathbf{y}' \{ \mathbf{1}_t \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \otimes \mathbf{M} \} (\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \{ \mathbf{1}_t' \otimes \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}' \} \mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \\ &= \frac{1}{rt} \cdot \frac{\mathbf{y}' \{ [\mathbf{1}_t \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}] [\mathbf{1}_t' \otimes \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \otimes \mathbf{M} (\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \mathbf{M}' \} \mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \\ &= \frac{1}{rt} \cdot \frac{\mathbf{y}' \{ \mathbf{d}\mathbf{d}' \otimes \mathbf{A} \} \mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}, \text{ where } \mathbf{d} = \mathbf{1}_t \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}, \text{ and } \mathbf{A} = \mathbf{M} (\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \mathbf{M}'. \end{aligned}$$

$$\begin{aligned} \text{Since } \mathbf{y}' \{ \mathbf{d}\mathbf{d}' \otimes \mathbf{A} \} \mathbf{y} &= (\text{vec } \mathbf{Y}')' \{ \mathbf{d}\mathbf{d}' \otimes \mathbf{A} \} \text{vec } \mathbf{Y}' \\ &= \mathbf{d}' \mathbf{Y} \mathbf{A} \mathbf{Y}' \mathbf{d} \quad (\text{Harville, 1997, theorem 16.2.2}), \end{aligned}$$

$$\text{then } F = \frac{1}{rt} \cdot \frac{\{ \mathbf{1}_t' \otimes \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \} \mathbf{Y} \mathbf{M} (\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \mathbf{M}' \mathbf{Y}' \{ \mathbf{1}_t \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$$

Moreover, since

$$\{ \mathbf{1}_t' \otimes \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \} \mathbf{Y} = \sum_{i=1}^t \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbf{Y}_i = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \sum_{i=1}^t \mathbf{Y}_i = t \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \bar{\mathbf{Y}}.$$

$$\text{then } F = \frac{t}{r} \cdot \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \bar{\mathbf{Y}} \mathbf{M} (\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p \mathbf{M})^{-1} \mathbf{M}' \bar{\mathbf{Y}}' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimate of $\boldsymbol{\Sigma}_p$. \square

Lemma 6.3.2.4 The REML estimator for $\boldsymbol{\Sigma}_p$ is $\frac{\mathbf{Y}'\mathbf{Q}\mathbf{Y}}{mt - q - t + 1}$,

where $\mathbf{Q} = \mathbf{P}_{\mathbf{1}_t} \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{X}}) + (\mathbf{I}_t - \mathbf{P}_{\mathbf{1}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{1}_m})$.

Proof Choose \mathbf{T} such that $\mathbf{T}\mathbf{T}' = \mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}}$, and $\mathbf{T}'\mathbf{T} = \mathbf{I}_{mt-q}$.

Notice that $\Re(\mathbf{T}') = \Re(\mathbf{T}'\mathbf{T}) \Rightarrow \mathbf{r}(\mathbf{T}') = \mathbf{r}(\mathbf{T}'\mathbf{T}) = mt - q$,

and $\mathbf{T}'(\mathbf{1}_t \otimes \mathbf{X}) = \mathbf{0}$

(6.19)

Define $\mathbf{K} = \mathbf{T} \otimes \mathbf{I}_p$.

Notice that $\mathbf{K}'(\mathbf{1}_t \otimes \mathbf{X} \otimes \mathbf{I}_p) = (\mathbf{T}' \otimes \mathbf{I}_p)(\mathbf{1}_t \otimes \mathbf{X} \otimes \mathbf{I}_p) = \mathbf{T}'(\mathbf{1}_t \otimes \mathbf{X}) \otimes \mathbf{I}_p = \mathbf{0}$

$$\ell_R = \text{Constant} - \frac{1}{2} \log |\mathbf{K}'(\mathbf{I}_t \otimes \Delta) \mathbf{K}| - \frac{1}{2} \mathbf{y}' \mathbf{K} [\mathbf{K}'(\mathbf{I}_t \otimes \Delta) \mathbf{K}]^{-1} \mathbf{K}' \mathbf{y}.$$

$$\text{Let } \mathbf{D} = \mathbf{K}'(\mathbf{I}_t \otimes \Delta) \mathbf{K} = \mathbf{K}'(\mathbf{I}_t \otimes \Delta) \mathbf{K} \mathbf{K}' \mathbf{K}.$$

$$\begin{aligned} (\mathbf{I}_t \otimes \Delta) \mathbf{K} \mathbf{K}' &= (\mathbf{I}_t \otimes \Delta) (\mathbf{I}_{mpt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} \otimes \mathbf{I}_p) \\ &= \mathbf{I}_t \otimes \Delta - \mathbf{P}_{\mathbf{I}_t} \otimes [\Delta (\mathbf{P}_{\mathbf{X}} \otimes \mathbf{I}_p)], \end{aligned}$$

$$\begin{aligned} \text{where } \mathbf{I}_t \otimes \Delta &= \mathbf{I}_t \otimes (\mathbf{I}_m \otimes \Sigma_p + \mathbf{P}_{\mathbf{I}_m} \otimes m\Lambda) \\ &= \mathbf{I}_{mt} \otimes \Sigma_p + \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} \otimes m\Lambda. \end{aligned}$$

$$\begin{aligned} \text{and } \mathbf{P}_{\mathbf{I}_t} \otimes [\Delta (\mathbf{P}_{\mathbf{X}} \otimes \mathbf{I}_p)] &= \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{P}_{\mathbf{X}} \otimes \Sigma_p + \mathbf{P}_{\mathbf{I}_m} \otimes m\Lambda) \\ &= \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} \otimes \Sigma_p + \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m} \otimes m\Lambda. \end{aligned}$$

$$\text{Hence, } (\mathbf{I}_t \otimes \Delta) \mathbf{K} \mathbf{K}' = (\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}}) \otimes \Sigma_p + (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes m\Lambda.$$

$$\Rightarrow \mathbf{D} = \mathbf{K}' [(\mathbf{I}_t \otimes \Delta) \mathbf{K} \mathbf{K}'] \mathbf{K}$$

$$\begin{aligned} &= (\mathbf{T}' \otimes \mathbf{I}_p) \left\{ (\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}}) \otimes \Sigma_p + (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes m\Lambda \right\} (\mathbf{T} \otimes \mathbf{I}_p) \\ &= \mathbf{T}' \mathbf{T} \otimes \Sigma_p - [\mathbf{T}' (\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}}) \mathbf{T}] \otimes \Sigma_p + [\mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes m\Lambda - [\mathbf{T}' (\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes m\Lambda. \end{aligned}$$

$$\text{Since from expression (6.19), we have } \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{X}) = \mathbf{0} \text{ then, } [\mathbf{T}' (\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}}) \mathbf{T}] \otimes \Sigma_p = \mathbf{0}.$$

$$\text{Also, since } \mathbf{I}_m \in \mathfrak{R}(\mathbf{X}), \text{ then } [\mathbf{T}' (\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes m\Lambda = \mathbf{0}.$$

$$\text{Hence, } \ell_R = \text{Constant} - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{y}' \mathbf{K} \mathbf{D}^{-1} \mathbf{K}' \mathbf{y} \text{ where}$$

$$\mathbf{D} = \mathbf{I}_{mt-q} \otimes \Sigma_p + [\mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes m\Lambda.$$

We may re-express \mathbf{D} as

$$\mathbf{D} = [\mathbf{I}_{mt-q} - \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes \Sigma_p + [\mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes (\Sigma_p + m\Lambda), \quad (6.20)$$

$$\mathbf{D}^{-1} = [\mathbf{I}_{mt-q} - \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes \Sigma_p^{-1} + [\mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}] \otimes (\Sigma_p + m\Lambda)^{-1}$$

Notice that $\mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}$ is a p.o., and this is true because

$$\begin{aligned} \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} &= \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) (\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}}) (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \\ &= \mathbf{T}' \left\{ (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \right\} \mathbf{T} = \mathbf{T}' (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \end{aligned}$$

$$= \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T} \quad \left\{ \text{notice that } \mathbf{T}'(\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) = \mathbf{0} \right\}.$$

Moreover,

$$\begin{aligned} \text{r}[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] &= \text{tr}[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] = \text{trace}[\mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})] \\ &= \text{tr}[(\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}})(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})] \\ &= \text{tr}(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) - \text{trace}(\mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) = t - 1 \end{aligned} \quad (6.21)$$

$$\begin{aligned} \mathbf{K}\mathbf{D}^{-1}\mathbf{K}' &= (\mathbf{T} \otimes \mathbf{I}_p) \left\{ [\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] \otimes \Sigma_p^{-1} \right. \\ &\quad \left. + [\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] \otimes (\Sigma_p + m\Lambda)^{-1} \right\} (\mathbf{T}' \otimes \mathbf{I}_p) \\ &= [\mathbf{T}\mathbf{T}' - \mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}\mathbf{T}'] \otimes \Sigma_p^{-1} + [\mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}\mathbf{T}'] \otimes (\Sigma_p + m\Lambda)^{-1}, \end{aligned}$$

$$\begin{aligned} \text{where } [\mathbf{T}\mathbf{T}' - \mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}\mathbf{T}'] \otimes \Sigma_p^{-1} \\ &= [\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} - (\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}})(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})(\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}})] \otimes \Sigma_p^{-1} \\ &= [\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}] \otimes \Sigma_p^{-1}, \end{aligned}$$

$$\begin{aligned} \text{and } [\mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}\mathbf{T}'] \otimes (\Sigma_p + m\Lambda)^{-1} \\ &= (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes (\Sigma_p + m\Lambda)^{-1} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbf{K}\mathbf{D}^{-1}\mathbf{K}' &= [\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} + \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}] \otimes \Sigma_p^{-1} + (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes (\Sigma_p + m\Lambda)^{-1} \\ \text{Notice that } &\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m} \\ &= \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{I}_m - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{X}} + \mathbf{I}_{mt} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} + \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{I}_m \\ &= \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{X}}) + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \\ \Rightarrow \mathbf{K}\mathbf{D}^{-1}\mathbf{K}' &= \mathbf{Q} \otimes \Sigma_p^{-1} + (\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes (\Sigma_p + m\Lambda)^{-1}, \end{aligned} \quad (6.22)$$

$$\text{where } \mathbf{Q} = \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{X}}) + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}).$$

Since $\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}$, $\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}$, Σ_p , and $\Sigma_p + m\Lambda$ are n.n.d. matrices,

so is $[\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] \otimes \Sigma_p$ and $[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}] \otimes (\Sigma_p + m\Lambda)$, (Harville, 1997, P.369)

Hence, $\left[\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes \boldsymbol{\Sigma}_p = \mathbf{G}\mathbf{G}'$,

and $\left[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) = \mathbf{H}\mathbf{H}'$,

for some full column rank matrices \mathbf{G} and \mathbf{H} (Seely, 2002, Problem 2.D.2).

Notice that since

$$\left\{ \left[\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes \boldsymbol{\Sigma} \right\} \left\{ \left[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \right\} = \mathbf{0},$$

then $\mathbf{G}\mathbf{G}'\mathbf{H}\mathbf{H}' = \mathbf{0} \Rightarrow \mathbf{G}'\mathbf{G}\mathbf{G}'\mathbf{H}\mathbf{H}'\mathbf{H} = \mathbf{0} \Rightarrow \mathbf{G}'\mathbf{H} = \mathbf{0}$.

Hence, $\left| \left[\mathbf{I}_{mt-q} - \mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes \boldsymbol{\Sigma} + \left[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}) \right| = |\mathbf{G}\mathbf{G}' + \mathbf{H}\mathbf{H}'|$

$$= \left| \begin{bmatrix} \mathbf{G} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{G}' \\ \mathbf{H}' \end{bmatrix} \right| = \left| \begin{bmatrix} \mathbf{G}' \\ \mathbf{H}' \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{H} \end{bmatrix} \right| = \begin{vmatrix} \mathbf{G}'\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}'\mathbf{H} \end{vmatrix} = |\mathbf{G}'\mathbf{G}| |\mathbf{H}'\mathbf{H}|$$

Recall that $\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T}$ is a projection operation, and $\text{r} \left[\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m}) \mathbf{T} \right] = t-1$,

and hence, $|\mathbf{G}'\mathbf{G}| = |\boldsymbol{\Sigma}_p|^{mt-q-t+1}$, and $|\mathbf{H}'\mathbf{H}| = |\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}|^{t-1}$.

In addition, by using expression (6.22), we obtain

$$\mathbf{y}'\mathbf{K}\mathbf{D}^{-1}\mathbf{K}'\mathbf{y} = \mathbf{y}'(\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1})\mathbf{y} + \mathbf{y}' \left[(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m} - \mathbf{P}_{\mathbf{I}_t} \otimes \mathbf{P}_{\mathbf{I}_m}) \otimes (\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})^{-1} \right] \mathbf{y}.$$

Accordingly, $\ell_R = \text{Constant} - \frac{1}{2} \log |\boldsymbol{\Sigma}_p|^{mt-q-t+1} - \frac{1}{2} \mathbf{y}'(\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1})\mathbf{y} + f(\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda})$.

Since $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Lambda}$ are unstructured, then $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_p + m\boldsymbol{\Lambda}$ may be considered as different random variables. This means that to maximize ℓ_R with respect to $\boldsymbol{\Sigma}_p$, it suffices to maximize ℓ_R^* with respect to $\boldsymbol{\Sigma}_p$,

where $\ell_R^* = -\frac{1}{2}(mt-q-t+1) \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} \mathbf{y}'(\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1})\mathbf{y}$.

$$\mathbf{y}'(\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1})\mathbf{y} = (\text{vec} \mathbf{Y}')' (\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1}) \text{vec} \mathbf{Y}'$$

$$= \text{tr}(\mathbf{Y}\boldsymbol{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}) \quad (\text{Harville, 1997, theorem 16.2.2}).$$

$$= \text{tr}(\boldsymbol{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}\mathbf{Y})$$

So, $\ell_R^* = -\frac{1}{2}(mt-q-t+1) \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}\mathbf{Y})$,

and hence $\hat{\boldsymbol{\Sigma}}_p = \frac{\mathbf{Y}'\mathbf{Q}\mathbf{Y}}{mt-q-t+1}$ (Anderson, 2003, lemma 3.2.2). \square

Note We have two expressions for \mathbf{Q} :

$$\mathbf{Q} = \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_X) + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}),$$

and $\mathbf{Q} = \mathbf{T}\mathbf{T}' - \mathbf{T}\mathbf{T}'(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{I}_m})\mathbf{T}\mathbf{T}'$

\mathbf{Q} is a p.o. and $r(\mathbf{Q}) = mt - q - t + 1$

Corollary 6.3.2.5 $\frac{mt - t - q - r + 2}{mt - q - t + 1} F = F_d.$

Proof By combining the results of lemmas (6.3.23) and (6.3.2.4), we obtain

$$F = \frac{t(mt - q - t + 1)}{r} \cdot \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{Y}}\mathbf{M}(\mathbf{M}'\mathbf{Y}'\mathbf{Q}\mathbf{Y}\mathbf{M})^{-1}\mathbf{M}'\bar{\mathbf{Y}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

and therefore $\frac{mt - t - q - r + 2}{mt - q - t + 1} F = F_d. \quad \square$

Corollary 6.3.2.6 Under the null hypothesis,

$$\frac{mt - t - q - r + 2}{mt - q - t + 1} F \sim F(r, tm - t - q - r + 2)$$

Proof Direct from lemma 6.3.2.2 and corollary 6.3.2.5. \square

Lemma 6.3.2.7 For this model, the approximation and the exact approaches to derive

$\mathbf{W} = \text{Var}[\hat{\boldsymbol{\sigma}}_{\text{REML}}]$ are identical.

Proof Approximation approach

$$\ell_R^* = -\frac{1}{2}(mt - q - t + 1) \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} \mathbf{y}'(\mathbf{Q} \otimes \boldsymbol{\Sigma}_p^{-1}) \mathbf{y},$$

$$\frac{\partial \ell_R^*}{\partial \sigma_{if}} = -\frac{1}{2}(mt - q - t + 1) \text{tr} \left(\boldsymbol{\Sigma}_p^{-1} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right) - \frac{1}{2} \mathbf{y}' \left(\mathbf{Q} \otimes \frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if}} \right) \mathbf{y},$$

$$\frac{\partial^2 \ell_R^*}{\partial \sigma_{if} \partial \sigma_{jg}} = -\frac{1}{2}(mt - q - t + 1) \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right) - \frac{1}{2} \mathbf{y}' \left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) \mathbf{y}$$

$$\Rightarrow -E \left[\frac{\partial^2 \ell_R^*}{\partial \sigma_{if} \partial \sigma_{jg}} \right] = \frac{1}{2}(mt - q - t + 1) \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right)$$

$$+ \frac{1}{2} \left\{ E(\mathbf{y}') \left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) E(\mathbf{y}) + \text{tr} \left[\left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) (\mathbf{I}_t \otimes \boldsymbol{\Lambda}) \right] \right\}$$

$$\begin{aligned} \text{where } E(\mathbf{y}') \left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) E(\mathbf{y}) &= \boldsymbol{\beta}' (\mathbf{1}'_t \otimes \mathbf{X}' \otimes \mathbf{I}_p) \left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) (\mathbf{1}_t \otimes \mathbf{X} \otimes \mathbf{I}_p) \boldsymbol{\beta} \\ &= \boldsymbol{\beta}' \left[\mathbf{1}' \mathbf{P}_{\mathbf{I}_t} \mathbf{1}_t \otimes \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \mathbf{X} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} + \mathbf{1}' (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \mathbf{1}_t \otimes \mathbf{X}' (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \mathbf{X} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right] \boldsymbol{\beta} = 0 \\ \Rightarrow -E \left[\frac{\partial^2 \ell_R^*}{\partial \sigma_{if} \partial \sigma_{jg}} \right] &= \frac{1}{2} (mt - q - t + 1) \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right) + \frac{1}{2} \text{tr} \left[\left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) (\mathbf{I}_t \otimes \boldsymbol{\Lambda}) \right]. \end{aligned}$$

Moreover,

$$\begin{aligned} \left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) (\mathbf{I}_t \otimes \boldsymbol{\Lambda}) &= \\ &= \left[\mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right] \times \\ &\quad (\mathbf{I}_t \otimes \mathbf{I}_m \otimes \boldsymbol{\Sigma}_p + \mathbf{I}_t \otimes m \mathbf{P}_{\mathbf{I}_m} \otimes \boldsymbol{\Lambda}) \\ &= \mathbf{P}_{\mathbf{I}_t} \otimes (\mathbf{I}_m - \mathbf{P}_\mathbf{X}) \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p + (\mathbf{I}_t - \mathbf{P}_{\mathbf{I}_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{\mathbf{I}_m}) \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p \\ &= \mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p, \end{aligned}$$

and hence,

$$\begin{aligned} \text{tr} \left[\left(\mathbf{Q} \otimes \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \right) (\mathbf{I}_t \otimes \boldsymbol{\Lambda}) \right] &= \text{tr}(\mathbf{Q}) \text{tr} \left(\frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p \right) = (mt - q - t + 1) \text{tr} \left(\frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p \right). \\ \Rightarrow -E \left[\frac{\partial^2 \ell_R^*}{\partial \sigma_{if} \partial \sigma_{jg}} \right] &= \frac{1}{2} (mt - q - t + 1) \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right) + \frac{1}{2} (mt - q - t + 1) \text{tr} \left(\frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p \right). \end{aligned}$$

$$\text{Notice that since } \frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p = - \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} + \boldsymbol{\Sigma}_p^{-1} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \boldsymbol{\Sigma}_p \right),$$

$$\text{then, } \text{tr} \left(\frac{\partial^2 \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{if} \partial \sigma_{jg}} \boldsymbol{\Sigma}_p \right) = -2 \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right).$$

$$\Rightarrow -E \left[\frac{\partial^2 \ell_R^*}{\partial \sigma_{if} \partial \sigma_{jg}} \right] = -\frac{1}{2} (mt - q - t + 1) \text{tr} \left(\frac{\partial \boldsymbol{\Sigma}_p^{-1}}{\partial \sigma_{jg}} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right) = \frac{1}{2} (mt - q - t + 1) \text{tr} \left(\boldsymbol{\Sigma}_p^{-1} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{jg}} \boldsymbol{\Sigma}_p^{-1} \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right)$$

Since from expression (4.7), $\text{tr} \left(\Sigma_p^{-1} \frac{\partial \Sigma_p}{\partial \sigma_{jg}} \Sigma_p^{-1} \frac{\partial \Sigma_p}{\partial \sigma_{if}} \right) = 2^{1-\delta_{if}-\delta_{jg}} (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij})$,

$$\text{then } \mathbf{W} \approx \frac{2}{mt-q-t+1} \left[\left\{ 2^{1-\delta_{if}-\delta_{jg}} (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij}) \right\}_{i,f,j,g=1}^p \right]^{-1}.$$

With observing that $\sum_{i=1}^p \sigma_{ij} \sigma^{if} = \begin{cases} 1 & \text{for } j = f \\ 0 & \text{for } j \neq f \end{cases}$, it can be seen that

$$\left[\left\{ 2^{1-\delta_{if}-\delta_{jg}} (\sigma^{ig} \sigma^{jf} + \sigma^{fg} \sigma^{ij}) \right\}_{i,f,j,g=1}^p \right]^{-1} = \frac{1}{2} \left[(\sigma_{ig} \sigma_{if} + \sigma_{fg} \sigma_{ij})_{i,f,j,g=1}^p \right],$$

and hence,

$$\mathbf{W} \approx \frac{1}{mt-q-t+1} \left[(\sigma_{ig} \sigma_{if} + \sigma_{fg} \sigma_{ij})_{i,f,j,g=1}^p \right] \quad (6.23)$$

Exact Approach

$$\text{Since } \mathbf{Y}_i = \begin{bmatrix} \mathbf{y}'_{i1} \\ \vdots \\ \mathbf{y}'_{im} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \mathbf{B} + \mathbf{a}'_i + \mathbf{e}'_{i1} \\ \vdots \\ \mathbf{x}'_m \mathbf{B} + \mathbf{a}'_i + \mathbf{e}'_{im} \end{bmatrix},$$

then we may express \mathbf{Y}_i as $\mathbf{Y}_i = \mathbf{X} \mathbf{B} + \mathbf{1}_m \mathbf{a}'_i + \mathbf{E}_i$ where $\mathbf{E}_i = [\mathbf{e}_{i1}, \dots, \mathbf{e}_{im}]'$

$$\text{and } \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_t \end{bmatrix} = \mathbf{1}_t \otimes \mathbf{X} \mathbf{B} + \mathbf{G} + \mathbf{E} \quad \text{where } \mathbf{E} = \begin{bmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_t \end{bmatrix} \quad \text{and } \mathbf{G} = \begin{bmatrix} \mathbf{1}_m \mathbf{a}'_1 \\ \vdots \\ \mathbf{1}_m \mathbf{a}'_t \end{bmatrix}$$

$$\Rightarrow \mathbf{QY} = \mathbf{Q}(\mathbf{1}_t \otimes \mathbf{XB}) + \mathbf{QG} + \mathbf{QE}$$

Observe that $\mathbf{Q}(\mathbf{1}_t \otimes \mathbf{XB}) = \mathbf{Q}(\mathbf{1}_t \otimes \mathbf{X})(\mathbf{I}_t \otimes \mathbf{B})$,

$$\text{but } \mathbf{Q}(\mathbf{1}_t \otimes \mathbf{X}) = \mathbf{TT}'(\mathbf{1}_t \otimes \mathbf{X}) - \mathbf{TT}'(\mathbf{I}_t \otimes \mathbf{P}_{1_m})\mathbf{TT}'(\mathbf{1}_t \otimes \mathbf{X}) = \mathbf{0},$$

and hence, $\mathbf{Q}(\mathbf{1}_t \otimes \mathbf{XB}) = \mathbf{0}$.

$$\begin{aligned} \text{Also, } \mathbf{QG} &= \left[\mathbf{P}_{1_t} \otimes (\mathbf{I}_m - \mathbf{P}_X) + (\mathbf{I}_t - \mathbf{P}_{1_t}) \otimes (\mathbf{I}_m - \mathbf{P}_{1_m}) \right] \mathbf{G} \\ &= \mathbf{Q} \left[\mathbf{I}_t \otimes (\mathbf{I}_m - \mathbf{P}_{1_m}) \right] \mathbf{G} = \mathbf{0} \end{aligned}$$

We conclude that $\mathbf{QG} = \mathbf{0}$, and so we have $\mathbf{QY} = \mathbf{QE}$.

Since \mathbf{Q} is idempotent then, $\mathbf{Y}'\mathbf{QY} = \mathbf{E}'\mathbf{QE}$

In addition, since \mathbf{E} is a data matrix from $N_p(0, \Sigma_p)$, and \mathbf{Q} is idempotent with

$$r(\mathbf{Q}) = mt - q - t + 1,$$

then $\mathbf{E}'\mathbf{Q}\mathbf{E} = \mathbf{Y}'\mathbf{Q}\mathbf{Y} \sim W_p(\Sigma_p, mt - q - t + 1)$ (Mardia, 1992, theorem 3.4.4).

Analogous to the argument given for Hotelling T^2 case in section 4.2.4 that leads to expression (4.6), we have

$$w_{if,jg} = \frac{\sigma_{ig}\sigma_{if} + \sigma_{fg}\sigma_{ij}}{mt - q - t + 1} \quad (6.24)$$

From expressions (6.23) and (6.24), we can see that the approximation and the exact methods give the same estimate for $w_{if,jg}$. \square

Lemma 6.3.2.8 $A_1 = \frac{2r}{mt - q - t + 1}$ and $A_2 = \frac{r(r+1)}{mt - q - t + 1}$.

Proof From expression (6.18), we found that $\mathbf{L}'\Phi\mathbf{L} = \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{t}\mathbf{M}'\Sigma_p\mathbf{M}$.

$$\begin{aligned} \Rightarrow \Theta &= \mathbf{L}(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}' = \frac{t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}(\mathbf{c} \otimes \mathbf{M})(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}(\mathbf{c}' \otimes \mathbf{M}') \\ &= \frac{t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\mathbf{c}\mathbf{c}' \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}', \end{aligned}$$

and from expression (6.17), we obtain

$$\frac{\partial \Phi}{\partial \sigma_{if}} = -\Phi \mathbf{P}_{if} \Phi = -\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{1_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}{t} \otimes \frac{\partial \Sigma_p}{\partial \sigma_{if}}.$$

Hence, $\text{tr}(\Theta\Phi\mathbf{P}_{if}\Phi)$

$$\begin{aligned} &= \frac{-t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \text{tr} \left\{ \left[\mathbf{c}\mathbf{c}' \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}' \right] \left[\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{1_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}{t} \otimes \frac{\partial \Sigma_p}{\partial \sigma_{if}} \right] \right\} \\ &= \frac{-1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \text{tr} \left[\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{1_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \otimes \mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \Sigma_p}{\partial \sigma_{if}} \right] \\ &= \frac{-\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_m - \mathbf{P}_{1_m})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \text{tr} \left[\mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \Sigma_p}{\partial \sigma_{if}} \right] \\ &= -\text{tr} \left[\mathbf{M}(\mathbf{M}'\Sigma_p\mathbf{M})^{-1}\mathbf{M}' \frac{\partial \Sigma_p}{\partial \sigma_{if}} \right]. \quad (\text{notice that } \mathbf{P}_{1_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = \mathbf{0}) \end{aligned}$$

Accordingly, A_1 and A_2 can be expressed as

$$A_1 = \sum_{g=1}^p \sum_{f=1}^p \sum_{i=1}^p \sum_{j=1}^p w_{if,jg} \operatorname{tr} \left[\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1} \mathbf{M}' \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \right] \operatorname{tr} \left[\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1} \mathbf{M}' \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{jg}} \right],$$

$$A_2 = \sum_{g=1}^p \sum_{f=1}^p \sum_{i=1}^p \sum_{j=1}^p w_{if,jg} \operatorname{tr} \left[\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1} \mathbf{M}' \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{if}} \mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1} \mathbf{M}' \frac{\partial \boldsymbol{\Sigma}_p}{\partial \sigma_{jg}} \right].$$

Analogous to the proof of lemma 6.2.2.5, we obtain

$$A_1 = \frac{2r}{mt - q - t + 1},$$

$$A_2 = \frac{r(r+1)}{mt - q - t + 1}. \quad \square$$

6.3.3 Estimating the Denominator Degrees of Freedom and the Scale Factor

$$A_1 = \frac{2r}{mt - q - t + 1}, \quad \text{and} \quad A_2 = \frac{r(r+1)}{mt - q - t + 1}, \quad \text{and then} \quad \frac{A_1}{A_2} = \frac{2}{r+1} \quad (\text{lemma 6.3.2.8}).$$

By utilizing theorems (5.3.1 and 5.3.2) and corollary (5.3.3), K-R and the proposed approaches are identical,

$$\text{where} \quad m^* = \frac{r(r+1)}{A_2} - (r-1) = mt - q - t - r + 2,$$

$$\text{and} \quad \lambda^* = 1 - \frac{r-1}{r(r+1)} A_2 = 1 - \frac{r-1}{mt - q - t + 1} = \frac{mt - q - t - r + 2}{mt - q - t + 1}.$$

The estimates of the denominator degrees of freedom and the scale factor match the values for the exact multivariate test in corollary 6.3.2.6.

7. THE SATTERTHWAITE APPROXIMATIONS

Another method to approximate F tests is the Satterthwaite method. Based on the original Satterthwaite's approximation (1941), the method was developed by Giesbrecht and Burns (1985) and then by Fai and Cornelius (1996). In this chapter, we do not intend to investigate the theoretical derivation of the method. However, we present some useful theoretical results.

7.1 The Satterthwaite Method (SAS Institute Inc., 2002-2006)

In order to approximate the F test for the fixed effects, $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, the Satterthwaite method uses the Wald-type statistic

$$F_s = \frac{1}{\ell} \hat{\boldsymbol{\beta}}' \mathbf{L} (\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L})^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}}$$

(i) The multi-dimensional case ($\ell > 1$)

First, we perform the spectral decomposition; $\mathbf{L}' \hat{\boldsymbol{\Phi}} \mathbf{L} = \mathbf{P}' \mathbf{D} \mathbf{P}$ where \mathbf{P} is an orthogonal matrix of eigenvectors, and \mathbf{D} is a diagonal matrix of eigenvalues.

Let $\nu_m = \frac{2(d_m)^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m}$ where

d_m is the m th diagonal element of \mathbf{D} , and \mathbf{g}_m is the gradient of $\mathbf{a}_m \boldsymbol{\Phi} \mathbf{a}_m'$ with respect to $\boldsymbol{\sigma}$ evaluated at $\hat{\boldsymbol{\sigma}}$, where \mathbf{a}_m is the m th row of $\mathbf{P} \mathbf{L}'$.

Notice that

$$\mathbf{g}_m = \left[\dots \mathbf{a}_m \frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} \mathbf{a}_m' \dots \right] \Big|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}} = - \left[\dots \mathbf{a}_m \boldsymbol{\Phi} \mathbf{P}_i \boldsymbol{\Phi} \mathbf{a}_m' \dots \right] \Big|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}}$$

Then let

$$E = \sum_{m=1}^{\ell} \frac{\nu_m}{\nu_m - 2} I(\nu_m > 2)$$

so we eliminate the terms for which $\nu_m \leq 2$.

The degrees of freedom for F are then computed as

$$\nu = \begin{cases} \frac{2E}{E - \ell} & \text{for } E > \ell \\ 0 & \text{otherwise} \end{cases}$$

(ii) The one-dimensional case ($\ell = 1$)

In this case, F statistic is simplified as

$$F_s = \frac{(\mathbf{L}'\hat{\boldsymbol{\beta}})^2}{\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L}}.$$

Notice that we may use the t-statistic instead since the numerator degrees of freedom equals one. The denominator degrees of freedom is computed as

$$\nu = \frac{2(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^2}{\mathbf{g}'\mathbf{W}\mathbf{g}},$$

where \mathbf{g} is the gradient of $\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L}$ with respect to $\boldsymbol{\sigma}$. (we assume that $\mathbf{L}'\boldsymbol{\beta}$ is estimable).

7.2 The K-R, Satterthwaite, and Proposed Methods

In this section, we provide some useful lemmas that show the relationship among the Satterthwaite, K-R and the proposed methods.

Lemma 7.2.1 When $\ell = 1$, the Satterthwaite, the K-R and the proposed methods give the same estimate of the denominator degrees of freedom.

Proof The K-R and the proposed methods give the same estimate of the denominator degrees of freedom which is $\frac{2}{A_1}$ (corollary 5.3.5).

$$\begin{aligned} A_1 &= \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}) \text{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}), \quad \boldsymbol{\Theta} = \frac{\mathbf{L}\mathbf{L}'}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}} \\ &= \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}\left(\frac{\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}}\right) \text{tr}\left(\frac{\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}}\right) \\ &= \sum_{i=1}^r \sum_{j=1}^r w_{ij} \frac{\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}\mathbf{L}}{(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2} \\ &\Rightarrow \frac{2}{A_1} = \frac{2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2}{\sum_{i=1}^r \sum_{j=1}^r w_{ij} \mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}\mathbf{L}} = \frac{2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2}{\mathbf{g}'\mathbf{W}\mathbf{g}} \quad \text{where } \mathbf{g} = [\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}]_{r \times 1} \end{aligned}$$

By noticing that $\frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} = -\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}$, the proof is completed. \square

Note Even though the estimates of the denominator degrees of freedom are the same, the true levels do not have to be the same for the Satterthwaite method and the K-R and proposed methods. This is because the statistics used are not the same. The Satterthwaite statistic uses $\hat{\Phi}$ as the estimator of the variance-covariance matrix of the fixed effects estimator, whereas the K-R and the proposed methods statistic uses $\hat{\Phi}_A$. The following corollary clarifies this fact.

Corollary 7.2.2 When $\ell = 1$, and $\hat{\Phi}_A = \hat{\Phi}$, then the Satterthwaite, K-R and proposed methods are identical.

Proof From lemma 7.2.1, the Satterthwaite, K-R and proposed methods give the same estimate of the denominator degrees of freedom. Since the scale is 1 (corollary 5.3.5) and $\hat{\Phi}_A = \hat{\Phi}$, then we have the same statistics for all approaches. This shows that the four approaches are identical. \square

Corollary 7.2.3 In balanced mixed classification models, and when $\ell = 1$, the Satterthwaite, K-R and proposed methods are identical. The denominator degrees of freedom and the scale estimate are $\frac{2}{A_2}$ and 1 respectively.

Proof: Direct from corollaries 5.3.5, 6.1.1.2, and 7.2.2. \square

Lemma 7.2.4 When $\ell = 1$, then the Satterthwaite statistic $F_s \geq$ the K-R and the proposed methods statistic F for variance components models.

Proof It suffices to show that $\mathbf{L}'\hat{\Phi}_A\mathbf{L} \geq \mathbf{L}'\hat{\Phi}\mathbf{L}$.

$$\begin{aligned} \text{Since } \hat{\Phi}_A &= \hat{\Phi} + 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\Phi} \hat{\mathbf{P}}_j) \right\} \hat{\Phi} \\ \Rightarrow \hat{\Phi}_A - \hat{\Phi} &= 2\hat{\Phi} \left\{ \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} (\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i \hat{\Phi} \hat{\mathbf{P}}_j) \right\} \hat{\Phi} \end{aligned}$$

Notice that

$$\begin{aligned}
\mathbf{Q}_{ij} - \mathbf{P}_i \Phi \mathbf{P}_j &= \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{X} - \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{X} \\
&= \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \mathbf{G} \frac{\partial \Sigma}{\partial \sigma_j} \Sigma^{-1} \mathbf{X}, \quad \text{where } \mathbf{G} = \Sigma^{-1} - \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1}
\end{aligned}$$

and hence,

$$\begin{aligned}
\Phi_A - \Phi &= 2\Phi \left\{ \sum_{i=1}^r \sum_{j=1}^r w_{ij} \mathbf{B}_i \mathbf{G} \mathbf{B}_j' \right\} \Phi \quad \text{where } \mathbf{B}_i = \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_i} \\
&= 2\Phi \mathbf{B} \mathbf{M} \mathbf{B}' \Phi, \quad \text{where } \mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_r] \text{ and } \mathbf{M} = \mathbf{W} \otimes \mathbf{G}
\end{aligned}$$

Since \mathbf{W} and \mathbf{G} are both n.n.d, so is \mathbf{M} (Harville, 1997, P.367), and hence so is $\mathbf{B} \mathbf{M} \mathbf{B}'$.

So $\hat{\Phi}_A - \hat{\Phi}$ is n.n.d., and hence $\mathbf{L}'(\hat{\Phi}_A - \hat{\Phi})\mathbf{L} \geq 0$. \square

Definition The Loewner ordering of symmetric matrices (Pukelsheim, 1993, chapter 1):

For \mathbf{A} and \mathbf{B} symmetric matrices, we say $\mathbf{A} \geq \mathbf{B}$ when $\mathbf{A} - \mathbf{B}$ is n.n.d.

Lemma 7.2.5 For \mathbf{A} and \mathbf{B} p.d. matrices, and $\mathbf{A} \geq \mathbf{B}$, then $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$

Proof Since $\mathbf{A} \geq \mathbf{B}$, then $\mathbf{A} = \mathbf{B} + \mathbf{V} = \mathbf{B} + \mathbf{C}\mathbf{C}'$ for some n.n.d. matrix \mathbf{V}

$$\Rightarrow \mathbf{A}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{C} (\mathbf{C}' \mathbf{B}^{-1} \mathbf{C} + \mathbf{I})^{-1} \mathbf{C}' \mathbf{B}^{-1} \text{ (theorem 1.7, Schott, 2005)}$$

$$\Rightarrow \mathbf{B}^{-1} - \mathbf{A}^{-1} = \mathbf{B}^{-1} \mathbf{C} (\mathbf{C}' \mathbf{B}^{-1} \mathbf{C} + \mathbf{I})^{-1} \mathbf{C}' \mathbf{B}^{-1} \text{ which is n.n.d, and hence } \mathbf{B}^{-1} \geq \mathbf{A}^{-1}. \quad \square$$

Note Theorem 7.2.4 is applicable for any ℓ for the variance components models as long as the scale is less or equal to 1, and this is true because from lemma 7.2.5, we have

$$\mathbf{L}' \hat{\Phi}_A \mathbf{L} \geq \mathbf{L}' \hat{\Phi} \mathbf{L} \Rightarrow (\mathbf{L}' \hat{\Phi}_A \mathbf{L})^{-1} \leq (\mathbf{L}' \hat{\Phi} \mathbf{L})^{-1}, \text{ and hence } F \leq F_s. \quad \square$$

Consider the partition of the design matrix \mathbf{X} as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, \mathbf{X}_1 and \mathbf{X}_2 are matrices of dimensions $n \times p_1$ and $n \times p_2$ respectively where $p = p_1 + p_2$.

$$\mathbf{X}' \Sigma^{-1} \mathbf{X} = \begin{bmatrix} \mathbf{X}_1' \Sigma^{-1} \mathbf{X}_1 & \mathbf{X}_1' \Sigma^{-1} \mathbf{X}_2 \\ \mathbf{X}_2' \Sigma^{-1} \mathbf{X}_1 & \mathbf{X}_2' \Sigma^{-1} \mathbf{X}_2 \end{bmatrix}$$

Let \mathbf{X}_2 to be correspondent to the fixed effects that of our interest to be tested.

Consider the case where

(a) $\mathbf{X}'_1 \Sigma^{-1} \mathbf{X}_2 = \mathbf{0}$.

(b) $\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2 = f(\boldsymbol{\sigma}) \mathbf{A}$ where \mathbf{A} is a fixed matrix. \mathbf{A} is invertible when \mathbf{X} is a full column rank matrix.

Lemma 7.2.6 For a model that satisfies conditions (a) and (b) mentioned above, we have $A_1 = \ell A_2$.

Proof According to above, $\mathbf{X}' \Sigma^{-1} \mathbf{X} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}$ where $\mathbf{C}_2 = f(\boldsymbol{\sigma}) \mathbf{A}$

$$\Rightarrow \Phi = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} = \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix}$$

Also, since $\mathbf{L}' = \begin{bmatrix} \mathbf{0}_{\ell \times p_1} & \mathbf{B}'_{\ell \times p_2} \end{bmatrix}$, then

$$\mathbf{L}' \Phi \mathbf{L} = \begin{bmatrix} \mathbf{0} & \mathbf{B}' \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{B} \end{bmatrix} = \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B},$$

$$\Theta = \mathbf{L} (\mathbf{L}' \Phi \mathbf{L})^{-1} \mathbf{L}' = \begin{bmatrix} \mathbf{0}_{p_1 \times \ell} \\ \mathbf{B}_{p_2 \times \ell} \end{bmatrix} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \begin{bmatrix} \mathbf{0}_{\ell \times p_1} & \mathbf{B}'_{\ell \times p_2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \end{bmatrix},$$

$$\Phi \Theta \Phi = \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \end{bmatrix}$$

Moreover, since $\mathbf{P}_i = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \mathbf{X} = \begin{bmatrix} \frac{\partial \mathbf{C}_1}{\partial \sigma_i} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \end{bmatrix}$ then

$$\Phi \Theta \Phi \mathbf{P}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \end{bmatrix} \begin{bmatrix} \frac{\partial \mathbf{C}_1}{\partial \sigma_i} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \end{bmatrix}$$

$$\text{So, } \text{tr}(\Phi \Theta \Phi \mathbf{P}_i) = \text{tr} \left(\mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \right) \quad (7.1)$$

Since $\mathbf{C}_2 = f(\boldsymbol{\sigma}) \mathbf{A}$, then $\mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} = \frac{\mathbf{A}^{-1}}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \mathbf{A} = \frac{1}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \mathbf{I}$

So, expression (7.1) above can be simplified as

$$\begin{aligned} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) &= \frac{1}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \text{tr}(\mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}') \\ &= \frac{1}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \text{tr}(\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1}) = \frac{\ell}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \end{aligned} \quad (7.2)$$

$$\Rightarrow \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi) = \frac{\ell^2}{[f(\boldsymbol{\sigma})]^2} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \quad (7.3)$$

Also,

$$\begin{aligned} \Phi \Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_j} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_j} \end{bmatrix} \end{aligned}$$

So,

$$\begin{aligned} \text{tr}(\Phi \Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi) &= \text{tr} \left(\mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_i} \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \sigma_j} \right) \\ &= \frac{1}{[f(\boldsymbol{\sigma})]^2} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \text{tr}(\mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}') \\ &= \frac{1}{[f(\boldsymbol{\sigma})]^2} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \text{tr}(\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} (\mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B})^{-1}) \\ &= \frac{\ell}{[f(\boldsymbol{\sigma})]^2} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \end{aligned} \quad (7.4)$$

From expressions 7.3 and 7.4, we have

$$\text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi) = \ell \text{tr}(\Phi \Theta \Phi \mathbf{P}_i \Phi \Theta \Phi \mathbf{P}_j \Phi),$$

and hence

$$A_1 = \ell A_2 \quad \square$$

Lemma 7.2.7 For a model that satisfies conditions (a) and (b) mentioned above

- (i) The K-R and the proposed approaches are identical.
- (ii) The Satterthwaite, K-R and proposed approaches give the same estimate of the denominator degrees of freedom given that the estimate > 2 .

(iii) If $\hat{\Phi}_A = \hat{\Phi}$, and the estimate of the denominator degrees of freedom > 2 , then the Satterthwaite, K-R and proposed methods are identical.

Proof (i) Observe that $A_1 = \ell A_2$ (lemma 7.2.6), and hence the K-R, and the proposed methods are identical (corollary 5.3.3). They all give 1 and $\frac{2\ell}{A_2}$ as the estimates of the scale and the denominator degrees of freedom respectively (theorems 5.3.1 and 5.3.2).

(ii) Claim $\frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} = \frac{2\ell}{A_2}$ for any m .

Proof $A_1 = \sum_{i=1}^r \sum_{j=1}^r w_{ij} \text{tr}(\Theta \Phi \mathbf{P}_i \Phi) \text{tr}(\Theta \Phi \mathbf{P}_j \Phi) = \mathbf{a}' \mathbf{W} \mathbf{a}$ where $\mathbf{a} = [\text{tr}(\Theta \Phi \mathbf{P}_i \Phi)]_{r \times 1}$,

Since from expression (7.2), $\text{tr}(\Theta \Phi \mathbf{P}_i \Phi) = \frac{\ell}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}$, then $\mathbf{a} = \left[\frac{\ell}{f(\boldsymbol{\sigma})} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}$

And hence

$$\text{the R.H.S} = \frac{2\ell}{A_2} = \frac{2\ell^2}{A_1} = \frac{2\ell^2}{\mathbf{a}' \mathbf{W} \mathbf{a}} = \frac{2[f(\boldsymbol{\sigma})]^2}{\mathbf{g}' \mathbf{W} \mathbf{g}} \quad \text{where } \mathbf{g} = \left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}.$$

Since $\mathbf{g}_m = \left[\mathbf{a}_m' \frac{\partial \Phi}{\partial \sigma_i} \mathbf{a}_m' \right]_{r \times 1}$ where \mathbf{a}_m is the m th row of $\mathbf{P} \mathbf{L}'$, then $\mathbf{g}_m = \left[\mathbf{b}_m \mathbf{L}' \frac{\partial \Phi}{\partial \sigma_i} \mathbf{L} \mathbf{b}_m' \right]_{r \times 1}$

where \mathbf{b}_m is the m th row of \mathbf{P} .

$$\Rightarrow \mathbf{g}_m = \left[\mathbf{b}_m \mathbf{B}' \frac{\partial \mathbf{C}_2^{-1}}{\partial \sigma_i} \mathbf{B} \mathbf{b}_m' \right]_{r \times 1} = \left[\frac{-1}{[f(\boldsymbol{\sigma})]^2} \frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \mathbf{b}_m \mathbf{B}' \mathbf{A}^{-1} \mathbf{B} \mathbf{b}_m' \right]_{r \times 1} = - \frac{\mathbf{b}_m \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} \mathbf{b}_m'}{f(\boldsymbol{\sigma})} \left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}$$

$$\text{And the L.H.S} = \frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} = \frac{2d_m^2}{\left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}' \mathbf{W} \left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}} \left(\frac{f(\boldsymbol{\sigma})}{\mathbf{b}_m \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} \mathbf{b}_m'} \right)^2$$

Notice that $\mathbf{b}_m \mathbf{B}' \mathbf{C}_2^{-1} \mathbf{B} \mathbf{b}_m' = \mathbf{b}_m \mathbf{L}' \Phi \mathbf{L} \mathbf{b}_m' = d_m$.

$$\Rightarrow \text{the L.H.S} = \frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} = \frac{2[f(\boldsymbol{\sigma})]^2}{\left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}' \mathbf{W} \left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \right]_{r \times 1}} = \text{the R.H.S}$$

We note that the L.H.S does not depend on row m , and this means that our claim is true for any row m .

Since $\nu_m = \frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} = \frac{2\ell}{A_2} > 2$ for any m ,

then $E = \frac{2\ell d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} \left/ \left(\frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} - 2 \right) \right.$, and $\nu = \frac{2E}{E - \ell} = \frac{2d_m^2}{\mathbf{g}_m' \mathbf{W} \mathbf{g}_m} =$ the L.H.S.

(iii) Direct from parts (ii) and (iii). \square

Comments

(i) When conditions (a) and (b) are satisfied, then the K-R and the proposed methods are identical. However, these conditions are not enough to make the Satterthwaite approach identical with the K-R and the proposed methods.

(ii) However, when the denominator degrees of freedom estimate is greater than two, conditions (a) and (b) are enough to make the Satterthwaite produce the same estimate of the denominator degrees of freedom produced by the K-R and the proposed methods.

(iii) When the K-R and the proposed methods' estimate of the denominator degrees of freedom is less or equal than two, the Satterthwaite estimate is zero.

(iv) Even though the Satterthwaite method produces the same estimate of the denominator degrees of freedom as the one produced by K-R and the proposed methods when the denominator degrees of freedom estimate is greater than two and conditions (a) and (b) are satisfied, we should notice that this does not mean that the Satterthwaite approach is identical to the K-R and proposed methods. This is true because the statistics used are different. The Satterthwaite statistic uses $\hat{\Phi}$ as an estimator of the variance-covariance matrix of the fixed effects estimator, whereas the K-R and the proposed methods statistic uses $\hat{\Phi}_A$.

(v) Conditions (a) and (b) are not enough to make $\hat{\Phi} = \hat{\Phi}_A$. For example, in BIB designs, conditions (a) and (b) are satisfied; however, $\hat{\Phi} \neq \hat{\Phi}_A$ (chapter 8).

8. SIMULATION STUDY FOR BLOCK DESIGNS

To compare the tests discussed in the previous chapters, we have conducted simulation studies for three different types of block designs: partially balanced incomplete block design (PBIB), balanced incomplete block design (BIB), and complete block design with missing data (two observations to be missing: one is from the first block under first treatment and the other one is from the second block under the second treatment). To test the fixed effects in these models, there are two approaches. The first approach is the intra-block approach where we consider the blocks fixed, so that only the within blocks information is utilized. Even though this approach leads to exact F tests, it does not lead to an optimal test. When the design is efficient and blocking is effective, this approach is close to optimal. We use the second approach where the blocks are considered random, and information from both within and between blocks is utilized. With the second approach, the model does not satisfy Zyskind's condition, so it is not a Rady model, and hence we do not have an exact F-test to test the fixed effect.

8.1 Preparing Formulas for Computations

Model for a Block Design

$$y_{ij} = \mu + \alpha_i + b_j + e_{ij}$$

for $i = 1, \dots, t$, $j = 1, \dots, s$, where μ is the general mean, α_i is the treatments effects, b_j is the blocks effects, $b_j \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, and b_j 's and e_{ij} 's are all independent.

The model can be expressed as

$$\mathbf{y} = \mathbf{1}_n \mu + \mathbf{A}\mathbf{a} + \mathbf{B}\mathbf{b} + \mathbf{e},$$

where $\mathbf{b} = [b_1, \dots, b_s]'$, $\mathbf{a} = [\alpha_1, \dots, \alpha_t]'$,

$$E(\mathbf{y}) = \mathbf{1}_n \mu + \mathbf{A}\mathbf{a}, \text{ and } \text{Var}(\mathbf{y}) = \mathbf{\Sigma} = \sigma_e^2 \mathbf{I}_n + \sigma_b^2 \mathbf{B}\mathbf{B}' = \sum_{i=1}^2 \sigma_i^2 \mathbf{D}_i \text{ where } \mathbf{D}_1 = \mathbf{I}_n \text{ and } \mathbf{D}_2 = \mathbf{B}\mathbf{B}'$$

Suppose that we are interested in testing: $H: \alpha_1 = \dots = \alpha_t \Leftrightarrow \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$

$$\text{where } \mathbf{L}' = \begin{bmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{\beta}' = [\mu \quad \alpha_1 \quad \cdots \quad \alpha_t]$$

The design matrix; $\mathbf{X} = [\mathbf{1}_n, \mathbf{A}]$ is not of full column rank. We can proceed with this parameterization using the g-inverse. However, to simplify our computations, we will reparameterize the model so \mathbf{X} will be a full column rank matrix.

Reparameterize the model in the way that $\boldsymbol{\beta}^{*'} = [\mu^* \quad \alpha_1^* \quad \cdots \quad \alpha_{t-1}^*]$ with $\sum_{i=1}^t \alpha_i^* = 0$.

$$\text{and } \mathbf{L}^{*'} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

To simplify our notation, from now on, consider $\mathbf{X} = \mathbf{X}^*, \mathbf{L} = \mathbf{L}^*$, and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

The REML Estimates

The REML equations are:

$$\frac{2\partial\ell(\boldsymbol{\sigma})}{\partial\sigma_e^2} = -\text{tr}(\mathbf{G}) + \mathbf{y}'\mathbf{G}\mathbf{G}\mathbf{y}, \quad \frac{2\partial\ell(\boldsymbol{\sigma})}{\partial\sigma_b^2} = -\text{tr}(\mathbf{G}\mathbf{D}_2) + \mathbf{y}'\mathbf{G}\mathbf{D}_2\mathbf{G}\mathbf{y},$$

where $\mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = \mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{K}'$.

A useful form to find the REML estimates by iteration is

$$\left\{ \text{tr}(\mathbf{D}_i\mathbf{G}\mathbf{D}_j\mathbf{G}) \right\}_{i,j=1}^r \boldsymbol{\sigma} = \left\{ \mathbf{y}'\mathbf{G}\mathbf{D}_i\mathbf{G}\mathbf{y} \right\}_{i=1}^r,$$

where $\boldsymbol{\sigma}(r \times 1) = (\sigma_1, \dots, \sigma_r)'$ (Searle *et al.*, 1992, chapter 6).

To estimate w_{ij} , we may use the inverse of the expected information matrix where,

$$-E \left[\frac{\partial^2 \ell}{\partial\sigma_i^2 \partial\sigma_j^2} \right] = \frac{1}{2} \text{tr}(\mathbf{G}\mathbf{D}_j\mathbf{G}\mathbf{D}_i) \quad (\text{Searle } et al., 1992, \text{ chapter 6})$$

$$\text{and hence } -E \left[\frac{\partial^2 \ell}{\partial(\sigma_e^2)^2} \right] = \frac{1}{2} \text{tr}(\mathbf{G}^2), \quad -E \left[\frac{\partial^2 \ell}{\partial(\sigma_b^2)^2} \right] = \frac{1}{2} \text{tr}(\mathbf{G}\mathbf{D}_2\mathbf{G}\mathbf{D}_2),$$

and
$$-E \left[\frac{\partial^2 \ell}{\partial \sigma_e^2 \partial \sigma_b^2} \right] = \frac{1}{2} \text{tr}(\mathbf{G} \mathbf{D}_2 \mathbf{G})$$

The entries of the inverse of the information matrix are estimated by using the REML estimates of σ_e^2 and σ_b^2 .

How we get matrix \mathbf{K}

$\mathbf{K}' = \mathbf{C}'[\mathbf{I}_n - \mathbf{P}_X]$ for any \mathbf{C}' (Searle *et al.*, 1992, chapter 6)

Notice that $\mathbf{K}'\mathbf{X} = \mathbf{C}'[\mathbf{I}_n - \mathbf{P}_X]\mathbf{X} = \mathbf{C}'[\mathbf{X} - \mathbf{X}] = \mathbf{0}$.

However, it's desirable to have \mathbf{K}' with a full row rank so $\mathbf{K}'\Sigma\mathbf{K}$ is invertible.

For our example, we obtain $\mathbf{I}_n - \mathbf{P}_X$, and then we can pick the independent rows of it.

Computing \mathbf{P}_i , \mathbf{Q}_{ij} and \mathbf{R}_{ij}

$$\mathbf{P}_1 = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_e^2} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_e^2} \Sigma^{-1} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \Sigma^{-1} \mathbf{X},$$

$$\mathbf{P}_2 = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_b^2} \Sigma^{-1} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X}$$

$$\mathbf{Q}_{11} = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_e^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_e^2} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_e^2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_e^2} \Sigma^{-1} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \Sigma^{-1} \Sigma^{-1} \mathbf{X},$$

$$\mathbf{Q}_{12} = \mathbf{Q}_{21} = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_e^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_e^2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_b^2} \Sigma^{-1} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X}$$

$$\mathbf{Q}_{22} = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X}.$$

$\mathbf{R}_{ij} = \mathbf{0}$, for all $i, j = 1, 2$ (clear).

The above quantities are to be estimated by using the REML estimates of σ .

Block designs with equal block sizes

When the design has equal block sizes (e.g., BIBD, and PBIBD), then $\mathbf{D}_2 = \mathbf{B}\mathbf{B}' = \mathbf{I}_s \otimes \mathbf{J}_k$ where k is the block size. Also, $\Sigma = \mathbf{I}_s \otimes \Sigma_k$, where $\Sigma_k = \sigma_e^2 \mathbf{I}_k + \sigma_b^2 \mathbf{J}_k$. $\Sigma^{-1} = \mathbf{I}_s \otimes \Sigma_k^{-1}$,

where $\Sigma_k^{-1} = (\sigma_e^2 \mathbf{I}_k + \sigma_b^2 \mathbf{J}_k)^{-1} = \frac{1}{\sigma_e^2} \left[\mathbf{I}_k - \frac{\sigma_b^2}{\sigma_e^2 + k\sigma_b^2} \mathbf{J}_k \right]$ (Schott, 2005, theorem 1.7)

$$\mathbf{P}_2 = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_b^2} \Sigma^{-1} \mathbf{X} = -\mathbf{X}' \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X} = -\frac{\mathbf{X}' \mathbf{D}_2 \mathbf{X}}{(\sigma_e^2 + k\sigma_b^2)^2}$$

$$\text{Notice that } \mathbf{D}_2 \Sigma^{-1} = (\mathbf{I}_s \otimes \mathbf{J}_k)(\mathbf{I}_s \otimes \Sigma_k^{-1}) = \mathbf{I}_s \otimes \mathbf{J}_k \Sigma_k^{-1} = \frac{\mathbf{I}_s \otimes \mathbf{J}_k}{(\sigma_e^2 + k\sigma_b^2)} = \frac{\mathbf{D}_2}{(\sigma_e^2 + k\sigma_b^2)}.$$

$$\begin{aligned} \mathbf{Q}_{12} = \mathbf{Q}_{21} &= \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_e^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_e^2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_b^2} \Sigma^{-1} \mathbf{X} \\ &= \mathbf{X}' \Sigma^{-1} \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X} = \frac{\mathbf{X}' \mathbf{D}_2 \mathbf{X}}{(\sigma_e^2 + k\sigma_b^2)^3} \end{aligned}$$

$$\mathbf{Q}_{22} = \mathbf{X}' \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_b^2} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{D}_2 \Sigma^{-1} \mathbf{X} = \frac{\mathbf{X}' \mathbf{D}_2 \mathbf{D}_2 \mathbf{X}}{(\sigma_e^2 + k\sigma_b^2)^3} = \frac{k\mathbf{X}' \mathbf{D}_2 \mathbf{X}}{(\sigma_e^2 + k\sigma_b^2)^3} = k\mathbf{Q}_{12}.$$

Lemma For the mixed linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, $\mathbf{u} \sim N(\mathbf{0}, \Sigma)$, $\Sigma = \sum_{i=1}^r \sigma_i \mathbf{V}_i$, and the

null hypothesis is $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$.

(i) In simulation of the K-R and Satterthwaite statistics (F_{KR} and F_S) under H_0 , it suffices to take the fixed effects, $\boldsymbol{\beta} = \mathbf{0}$.

(ii) $F_{\text{KR}}(\mathbf{u}) = F_{\text{KR}}(a\mathbf{u})$ and $F_S(\mathbf{u}) = F_S(a\mathbf{u})$ for any constant $a > 0$.

Proof (i) $F_{\text{KR}} = \frac{1}{\ell} (\mathbf{L}'\hat{\boldsymbol{\beta}})' (\mathbf{L}'\hat{\boldsymbol{\Phi}}_A \mathbf{L})^{-1} (\mathbf{L}'\hat{\boldsymbol{\beta}}).$

To simulate F_{KR} under H_0 , we generate $\mathbf{u} \sim N(\mathbf{0}, \Sigma)$, and then $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ where $\mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$.

However, we should notice that $\hat{\Sigma}$, $\hat{\boldsymbol{\Phi}}$, and $\hat{\boldsymbol{\Phi}}_A$ do not depend on $\mathbf{X}\boldsymbol{\beta}$. Indeed, they depend only on \mathbf{u} .

$$\text{Moreover, } \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Phi}} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{y} = \hat{\boldsymbol{\Phi}} \mathbf{X}' \hat{\Sigma}^{-1} (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + \hat{\boldsymbol{\Phi}} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{u}.$$

$$\Rightarrow \mathbf{L}'\hat{\boldsymbol{\beta}} = \mathbf{L}'\boldsymbol{\beta} + \mathbf{L}'\hat{\boldsymbol{\Phi}} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{u} = \mathbf{L}'\hat{\boldsymbol{\Phi}} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{u} = \mathbf{L}'\hat{\boldsymbol{\beta}}_0 \text{ (under } H_0) \text{ where } \hat{\boldsymbol{\beta}}_0 \text{ is } \hat{\boldsymbol{\beta}} \text{ calculated from } \mathbf{u} = \mathbf{y} \text{ with } \boldsymbol{\beta} = \mathbf{0}.$$

Similarly, it can be shown that in simulation of F_S under H_0 , it suffices to take $\boldsymbol{\beta} = \mathbf{0}$

(ii) Observe that $\mathbf{u} \sim N(\mathbf{0}, \Sigma) \Leftrightarrow a\mathbf{u} \sim N(\mathbf{0}, a^2 \Sigma)$, and hence $\hat{\Sigma}(a\mathbf{u}) = a^2 \hat{\Sigma}(\mathbf{u})$.

$$\hat{\boldsymbol{\Phi}}(a\mathbf{u}) = \left(\mathbf{X}' \hat{\Sigma}^{-1}(a\mathbf{u}) \mathbf{X} \right)^{-1} = a^2 \left(\mathbf{X}' \hat{\Sigma}^{-1}(\mathbf{u}) \mathbf{X} \right)^{-1} = a^2 \hat{\boldsymbol{\Phi}}(\mathbf{u}),$$

$$\hat{\boldsymbol{\beta}}(a\mathbf{u}) = \hat{\boldsymbol{\Phi}}(a\mathbf{u}) \mathbf{X}' \hat{\Sigma}^{-1}(a\mathbf{u}) a\mathbf{u} = a \hat{\boldsymbol{\Phi}}(\mathbf{u}) \mathbf{X}' \hat{\Sigma}^{-1}(\mathbf{u}) \mathbf{u} = a \hat{\boldsymbol{\beta}}(\mathbf{u}).$$

$$\hat{\mathbf{P}}_i(a\mathbf{u}) = -\mathbf{X}'\hat{\Sigma}^{-1}(a\mathbf{u})\frac{\partial\hat{\Sigma}(a\mathbf{u})}{\partial\sigma_i}\hat{\Sigma}^{-1}(a\mathbf{u})\mathbf{X} = -\frac{1}{a^4}\mathbf{X}'\hat{\Sigma}^{-1}(\mathbf{u})\frac{\partial\hat{\Sigma}(\mathbf{u})}{\partial\sigma_i}\hat{\Sigma}^{-1}(\mathbf{u})\mathbf{X} = \frac{1}{a^4}\mathbf{P}_i(\mathbf{u}),$$

$$\text{and } \mathbf{Q}_{ij}(a\mathbf{u}) = \mathbf{X}'\hat{\Sigma}^{-1}(a\mathbf{u})\frac{\partial\hat{\Sigma}(a\mathbf{u})}{\partial\sigma_i}\hat{\Sigma}^{-1}(a\mathbf{u})\frac{\partial\hat{\Sigma}(a\mathbf{u})}{\partial\sigma_j}\hat{\Sigma}^{-1}(a\mathbf{u})\mathbf{X}$$

$$= \frac{1}{a^6}\mathbf{X}'\hat{\Sigma}^{-1}(\mathbf{u})\frac{\partial\hat{\Sigma}(\mathbf{u})}{\partial\sigma_i}\hat{\Sigma}^{-1}(\mathbf{u})\frac{\partial\hat{\Sigma}(\mathbf{u})}{\partial\sigma_j}\hat{\Sigma}^{-1}(\mathbf{u})\mathbf{X} = \frac{1}{a^6}\mathbf{Q}_{ij}(\mathbf{u}).$$

$$\begin{aligned} \text{Also, since } \mathbf{G}(a\mathbf{u}) &= \hat{\Sigma}^{-1}(a\mathbf{u}) - \hat{\Sigma}^{-1}(a\mathbf{u})\left(\mathbf{X}'\hat{\Sigma}^{-1}(a\mathbf{u})\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\Sigma}^{-1}(a\mathbf{u}) \\ &= \frac{1}{a^2}\left[\hat{\Sigma}^{-1}(\mathbf{u}) - \hat{\Sigma}^{-1}(\mathbf{u})\left(\mathbf{X}'\hat{\Sigma}^{-1}(\mathbf{u})\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\Sigma}^{-1}(\mathbf{u})\right] = \frac{1}{a^2}\mathbf{G}(\mathbf{u}), \end{aligned}$$

$$\begin{aligned} \text{then } \mathbf{W}(a\mathbf{u}) &\approx 2\left[\left\{\text{tr}\left(\mathbf{G}(a\mathbf{u})\frac{\partial\hat{\Sigma}(a\mathbf{u})}{\partial\sigma_i}\mathbf{G}(a\mathbf{u})\frac{\partial\hat{\Sigma}(a\mathbf{u})}{\partial\sigma_j}\right)_{i,j=1}^r\right\}\right]^{-1} \quad (\text{section 6.1.2}) \\ &= 2\left[\left\{\text{tr}\left(\frac{1}{a^4}\mathbf{G}(\mathbf{u})\frac{\partial\hat{\Sigma}(\mathbf{u})}{\partial\sigma_i}\mathbf{G}(\mathbf{u})\frac{\partial\hat{\Sigma}(\mathbf{u})}{\partial\sigma_j}\right)_{i,j=1}^r\right\}\right]^{-1} = a^4\mathbf{W}(\mathbf{u}). \end{aligned}$$

Since $\mathbf{R}_{ij} = \mathbf{0}$, then we obtain $\hat{\Phi}_A(a\mathbf{u}) = a^2\hat{\Phi}_A(\mathbf{u})$, and hence $F_{\text{KR}}(\mathbf{u}) = F_{\text{KR}}(a\mathbf{u})$ and $F_S(\mathbf{u}) = F_S(a\mathbf{u})$. \square

8.2 Simulation Results

Seven approaches for testing fixed effects are considered in the simulation study: Kenward-Roger (after modification), Kenward-Roger before the modification, Kenward-Roger using the conventional variance-covariance of the fixed effects, Satterthwaite, Containment, and our proposed methods. Five settings for the ratio of σ_b and σ_e that are denoted by ρ have been used: 0.25, 0.5, 1, 2, 4. For each setting of ρ , 10,000 data sets have been simulated from the corresponding multivariate distribution with zero treatment effects. The level for each test, mean of scales, and mean of the denominator degrees of freedom were computed for the data sets at nominal 0.05. We use $\dagger\dagger$ to denote levels below 0.0400 or above 0.0600, and \dagger for levels between 0.0400 and 0.0449 or between 0.0551 and 0.0600. So, all unmarked levels are between 0.0450 and 0.0550. For each data

set the REML estimates of the variance components have been calculated by the iteration method mentioned above, and convergence was achieved for almost all data sets. The convergence rate is reported. Also, we present the percentage relative bias in the conventional and the adjusted estimator of variance-covariance matrix of the fixed effects estimator. In each type of block design, we consider some examples where we have different efficiency factors. The relative efficiency of a random block model to a complete block model for a given number t of treatments and a given number of n of observations can be expressed as

$$E = \frac{t/n}{\text{trace} \left[\mathbf{T}' (\mathbf{A}' \boldsymbol{\Sigma}_\rho^{-1} \mathbf{A})^{-1} \mathbf{T} \right] / (t-1)} \text{ where } \boldsymbol{\Sigma}_\rho = \rho^2 \mathbf{B} \mathbf{B}' + \mathbf{I}_n, \quad \mathbf{T}' \mathbf{T} = \mathbf{I}_{t-1}, \text{ and } \mathbf{T}' \mathbf{1}_t = \mathbf{0}.$$

The efficiency factor which is denoted by q is defined to be the lower bound of E , and it describes the design alone. For an equireplicate block design with equal block sizes,

$$q = \frac{t-1}{r \text{ trace} \left(\left[\mathbf{A}' (\mathbf{I} - \mathbf{P}_B) \mathbf{A} \right]^+ \right)} \text{ where } r \text{ is the replication of the treatments (Birkes, 2006).}$$

Observe that BIB and PBIB designs are equireplicate and have equal block sizes.

8.2.1 Partially Balanced Incomplete Block Design (PBIB)

I. The first PBIB design (PBIB1) is from Cochran and Cox (1957, P.456) where we have fifteen blocks, fifteen treatments, four treatments per block and $q = 0.7955$.

Table 1.1
PBIB1 ($t = s = 15$, $k = 4$, and $n = 60$)

Block	Treat.	Block	Treat.	Block	Treat.
1	15,9,1,13	6	12,4,3,1	11	9,7,10,3
2	5,7,8,1	7	12,14,15,8	12	8,6,2,9
3	10,1,14,2	8	6,3,14,5	13	5,9,11,12
4	15,11,2,3	9	5,4,2,13	14	7,13,14,11
5	6,15,4,7	10	10,12,13,6	15	10,4,8,11

Table 1.2
Simulated size of nominal 5% Wald F- tests for PBIB1

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2	Satter	Contain
0.25	0.0528	0.0390 ^{††}	0.0466	0.0478	0.0468	0.0697 ^{††}	0.0674 ^{††}
0.5	0.0536	0.0423 [†]	0.0475	0.0483	0.0480	0.0622 ^{††}	0.0558 [†]
1.0	0.0573 [†]	0.0486	0.0521	0.0521	0.0521	0.0560 [†]	0.0528
2.0	0.0548	0.0467	0.0480	0.0480	0.0480	0.0497	0.0483
4.0	0.0544	0.0481	0.0484	0.0484	0.0484	0.0492	0.0485

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$. S.E. of entries $\in (0.0019, 0.0025)$

Table 1.3
Mean of estimated denominator degrees of freedom for PBIB1

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2	Satter	Contain
0.25	48.8821	31.9638	39.3355	41.7298	40.2042	38.8929	31
0.5	47.4572	31.8812	38.4884	39.6942	38.9313	38.3272	31
1.0	43.1678	30.0484	34.6682	34.8640	34.7480	34.6475	31
2.0	40.5233	29.9220	32.1060	32.1205	32.1125	32.1052	31
4.0	39.7016	30.5696	31.2880	31.2889	31.2884	31.2830	31

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$.

Table 1.4
Mean of estimated scale for PBIB1

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2
0.25	0.9890	0.8787	0.9974	0.9954	0.9975
0.5	0.9916	0.9459	0.9991	0.9978	0.9989
1.0	0.9913	0.9829	0.9999	0.9996	0.9998
2.0	0.9902	0.9970	1.0000	1.0000	1.0000
4.0	0.9898	0.9996	1.0000	1.0000	1.0000

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$.

Table 1.5
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for PBIB1

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-6.5	1.6	0.9993	0.9604
0.5	-4.2	1.0	0.9999	0.8999
1.0	-0.6	1.5	1.0000	0.8378
2.0	-0.4	0.0	1.0000	0.8080
4.0	-0.6	-0.5	1.0000	0.7987

II. The second design (PBIB2) is from Green (1974, P.65) where we have forty-eight blocks, sixteen treatments, two treatments per block and $q = 0.4762$.

Table 1.6
PBIB2 design ($t = 16, s = 48, k = 2$, and $n = 96$)

blk	trt	blk	trt	blk	trt	blk	trt	blk	trt	blk	trt
1	1,2	9	5,8	17	10,12	25	1,5	33	2,14	41	7,15
2	1,3	10	6,7	18	11,12	26	1,9	34	6,10	42	11,15
3	1,4	11	6,8	19	13,14	27	1,13	35	6,14	43	4,8
4	2,3	12	7,8	20	13,15	28	5,9	36	10,14	44	4,12
5	2,4	13	9,10	21	13,16	29	5,13	37	3,7	45	4,16
6	3,4	14	9,11	22	14,15	30	9,13	38	3,11	46	8,12
7	5,6	15	9,12	23	14,16	31	2,6	39	3,15	47	8,16
8	5,7	16	10,11	24	15,16	32	2,10	40	7,11	48	12,16

Table 1.7
Simulated size of nominal 5% Wald F- tests for PBIB2

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2	Satter	Contain
0.25	0.0593 [†]	0.0532	0.0568 [†]	0.0585 [†]	0.0573 [†]	0.0730 ^{††}	0.0440 [†]
0.5	0.0585 [†]	0.0529	0.0561 [†]	0.0579 [†]	0.0565 [†]	0.0711 ^{††}	0.0453
1.0	0.0570 [†]	0.0504	0.0540	0.0549	0.0543	0.0644 ^{††}	0.0468
2.0	0.0557 [†]	0.0473	0.0508	0.0511	0.0509	0.0542	0.0480
4.0	0.0562 [†]	0.0470	0.0493	0.0494	0.0493	0.0507	0.0480

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$. S.E. of entries $\in (0.0021, 0.0026)$

Table 1.8
Mean of estimated denominator degrees of freedom for PBIB2

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2	Satter	Contain
0.25	83.6874	65.3837	72.6495	82.2865	74.7209	68.9973	33
0.5	79.8284	61.8836	69.0306	77.4455	70.9267	65.9502	33
1.0	65.6409	49.2069	55.8740	59.9145	56.9966	54.6475	33
2.0	49.9531	36.2898	41.2927	41.9262	41.5348	41.1791	33
4.0	43.6954	32.8601	35.2892	35.3400	35.3120	35.2835	33

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$.

Table 1.9
Mean of estimated scales for PBIB2

ρ	K-R*	K-R**	K-R	Prop.1	Prop.2
0.25	0.9953	0.9482	0.9981	0.9949	0.9973
0.5	0.9952	0.9511	0.9983	0.9952	0.9975
1.0	0.9948	0.9639	0.9991	0.9968	0.9984
2.0	0.9934	0.9859	0.9999	0.9991	0.9996
4.0	0.9917	0.9974	1.0000	0.9999	1.0000

* before modification, ** with using $\hat{\Phi}$ instead of $\hat{\Phi}_A$.

Table 1.10
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for PBIB2

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-7.0	-2.5	1.0000	0.9478
0.5	-7.0	-2.7	1.0000	0.8408
1.0	-5.9	-2.4	1.0000	0.6705
2.0	-3.3	-1.6	1.0000	0.5451
4.0	-2.6	-2.2	1.0000	0.4955

8.2.2 Balanced Incomplete Block Design (BIBD)

Since the design is balanced with respect to treatments, then it can be shown that the model satisfies conditions (a) and (b) for theorem lemma 7.2.6, and hence $A_1 = \ell A_2$.

The K-R and the proposed approaches are identical. The scale estimate is 1 and the

denominator degree of freedom is estimate is $\frac{2\ell}{A_2}$ (theorems 5.3.1 and 5.3.2, and corollary

5.3.3). The Satterthwaite method gives the same estimate of the denominator degrees of freedom as the K-R and the proposed approaches when the estimate > 2 (lemma 7.2.7).

Six BIB designs with different efficiency factors are adopted from Cochran and Cox

(1957, chapter 11). Notice that the efficiency factor simplified as $q = \frac{t\alpha}{rk} = \frac{1-1/k}{1-1/t}$ where

r is the replication of the treatments and α is the number of blocks in which each pair of treatments occur together (Kuehl, 2000, chapter 9).

I. In the first design (BIB1), where we have fifteen blocks, six treatments, and two treatments per block ($t = 6$, $s = 15$, $k = 2$, and $n = 30$), and the efficiency factor, $q = 0.6$

Table 2.1
BIB1 ($t = 6$, $s = 15$, $k = 2$, and $n = 30$)

Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.
1	1,2	4	1,3	7	1,4	10	1,5	13	1,6
2	3,4	5	2,5	8	2,6	11	2,4	14	2,3
3	5,6	6	4,6	9	3, 5	12	3,6	15	4,5

Table 2.2
Simulated size of nominal 5% Wald F- tests for BIB1

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0786 ^{††}	0.0676 ^{††}	0.0998 ^{††}	0.0619 ^{††}
0.5	0.0731 ^{††}	0.0635 ^{††}	0.0930 ^{††}	0.0589 [†]
1.0	0.0710 ^{††}	0.0566 [†]	0.0746 ^{††}	0.0534
2.0	0.0697 ^{††}	0.0551 [†]	0.0599 [†]	0.0539
4.0	0.0727 ^{††}	0.0542	0.0559 [†]	0.0540

* before modification. S.E. of entries $\in (0.0023, 0.003)$

Table 2.3
Mean of estimated denominator degrees of freedom for BIB1

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	29.0917	20.1712	20.1712	10
0.5	28.0031	19.0591	19.0591	10
1.0	24.6097	15.5784	15.5784	10
2.0	21.0182	11.8539	11.8539	10
4.0	19.7201	10.4807	10.4807	10

* before modification.

Table 2.4
Mean of estimated scales for BIB1

ρ	K-R*	K-R
0.25	0.9753	1.0000
0.5	0.9725	1.0000
1.0	0.9619	1.0000
2.0	0.9446	1.0000
4.0	0.9344	1.0000

* before modification.

Table 2.5
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for BIB1

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-16.0	-3.7	0.9968	0.9556
0.5	-13.5	-2.3	0.9978	0.8667
1.0	-8.9	-1.2	0.9994	0.7333
2.0	-2.0	-0.6	1.0000	0.6444
4.0	0.6	0.9	1.0000	0.6121

II. In the second design (BIB2), we have ten blocks, six treatments, and three treatments per block ($t = 6$, $s = 10$, $k = 3$, and $n = 30$), $q = 0.8$

Table 2.6
BIB2 ($t = 6$, $s = 10$, $k = 3$, and $n = 30$)

Block	Treat.	Block	Treat.
1	1,2,5	6	2,3,4
2	1,2,6	7	2,3,5
3	1,3,4	8	2,4,6
4	1,3,6	9	3,5,6
5	1,4,5	10	4,5,6

Table 2.7
Simulated size of nominal 5% Wald F- tests for BIB2

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0673 ^{††}	0.0577 [†]	0.0830 ^{††}	0.0710 ^{††}
0.5	0.0670 ^{††}	0.0584 [†]	0.0775 ^{††}	0.0664 ^{††}
1.0	0.0633 ^{††}	0.0524	0.0604 ^{††}	0.0552 [†]
2.0	0.0606 ^{††}	0.0491	0.0515	0.0497
4.0	0.0579 [†]	0.0474	0.0482	0.0475

* before modification. S.E. of entries $\in (0.0019, 0.0025)$

Table 2.8
Mean of estimated denominator degrees of freedom for BIB2

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	29.4561	20.5425	20.5425	15
0.5	28.5419	19.6163	19.6163	15
1.0	26.3744	17.4069	17.4069	15
2.0	24.7693	15.7620	15.7620	15
4.0	24.2247	15.2012	15.2012	15

* before modification.

Table 2.9
Mean of estimated scales for BIB2

ρ	K-R*	K-R
0.25	0.9761	1.0000
0.5	0.9749	1.0000
1.0	0.9701	1.0000
2.0	0.9653	1.0000
4.0	0.9632	1.0000

* before modification.

Table 2.10
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency for BIB2

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-12.0	-2.1	0.9977	0.9684
0.5	-7.4	0.2	0.9991	0.9143
1.0	-5.3	-2.1	1.0000	0.8500
2.0	-2.6	-2.0	1.0000	0.8154
4.0	-0.5	-0.5	1.0000	0.8041

III. In the third design (BIB3), we have ten blocks, six treatments, and three treatments per block ($t = 7$, $s = 7$, $k = 4$, and $n = 28$), $q = 0.8750$

Table 2.11
BIB3 ($t = 7$, $s = 7$, $k = 4$, and $n = 28$)

Block	Treat.
1	3,5,6,7
2	1,4,6,7
3	1,2,5,7
4	1,2,3,6
5	2,3,4,7
6	1,3,4,5
7	2,4,5,6

Table 2.12
Simulated size of nominal 5% Wald F- tests for BIB3

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0569 [†]	0.0472	0.0697 ^{††}	0.0769 ^{††}
0.5	0.0613 ^{††}	0.0508	0.0660 ^{††}	0.0644 ^{††}
1.0	0.0607 ^{††}	0.0494	0.0553 [†]	0.0524
2.0	0.0581 [†]	0.0450	0.0470	0.0454
4.0	0.0570 [†]	0.0455	0.0456	0.0456

* before modification. S.E. of entries $\in (0.0021, 0.0027)$

Table 2.13
Mean of estimated denominator degrees of freedom for BIB3

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	26.3479	17.6322	17.6218	15
0.5	26.2043	17.3494	17.3468	15
1.0	25.1420	16.2466	16.2464	15
2.0	24.3069	15.3868	15.3868	15
4.0	24.0273	15.0992	15.0992	15

* before modification.

Table 2.14
Mean of estimated scales for BIB3

ρ	K-R*	K-R
0.25	0.9549	1.0000
0.5	0.9653	1.0000
1.0	0.9667	1.0000
2.0	0.9643	1.0000
4.0	0.9632	1.0000

* before modification.

Table 2.15
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency for BIB3

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-10.8	3.9	1.0000	0.9750
0.5	-8.9	-0.5	1.0000	0.9357
1.0	-0.4	2.2	1.0000	0.9000
2.0	-1.3	-0.9	1.0000	0.8824
4.0	-2.0	-1.9	1.0000	0.8769

IV. In the fourth design (BIB4), we have thirty six blocks, nine treatments, and two treatments per block ($t = 9$, $s = 36$, $k = 2$, and $n = 72$), and $q = 0.5625$

Table 2.16
BIB4 ($t = 9$, $s = 36$, $k = 2$, and $n = 72$)

Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.
1	1,2	10	1,3	19	1,4	28	1,5
2	2,8	11	2,5	20	2,6	29	2,4
3	3,4	12	3,6	21	2,3	30	3,8
4	4,7	13	4,9	22	4,5	31	4,6
5	5,6	14	5,8	23	5,7	32	3,5
6	1,6	15	6,7	24	6,8	33	6,9
7	3,7	16	1,7	25	7,9	34	2,7
8	8,9	17	4,8	26	1,8	35	7,8
9	5,9	18	2,9	27	3,9	36	1,9

Table 2.17
Simulated size of nominal 5% Wald F- tests for BIB4

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0578 [†]	0.0566 [†]	0.0731 ^{††}	0.0520
0.5	0.0587 [†]	0.0562 [†]	0.0709 ^{††}	0.0521
1.0	0.0555 [†]	0.0524	0.0617 ^{††}	0.0501
2.0	0.0598 [†]	0.0536	0.0592 [†]	0.0525
4.0	0.0599 [†]	0.0547	0.0557 [†]	0.0546

* before modification. S.E. of entries $\in (0.0022, 0.0026)$

Table 2.18
Mean of estimated denominator degrees of freedom for BIB4

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	66.8712	58.3711	58.3711	28
0.5	63.3018	54.7945	54.7945	28
1.0	52.5089	43.9746	43.9746	28
2.0	41.8840	33.3083	33.3083	28
4.0	38.0104	29.4118	29.4118	28

* before modification.

Table 2.19
Mean of estimated scales for BIB4

ρ	K-R*	K-R
0.25	0.9966	1.0000
0.5	0.9961	1.0000
1.0	0.9942	1.0000
2.0	0.9906	1.0000
4.0	0.9883	1.0000

* before modification.

Table 2.20
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for BIB4

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-7.5	-2.0	1.0000	0.9514
0.5	-5.9	-0.8	1.0000	0.8542
1.0	-4.2	-0.7	1.0000	0.7083
2.0	-0.7	0.5	1.0000	0.6111
4.0	0.9	1.1	1.0000	0.5758

V. In the fifth design (BIB5), we have ten blocks, six treatments, and three treatments per block ($t = 9$, $s = 18$, $k = 4$, and $n = 72$), and $q = 0.8438$

Table 2.21
BIB5 design ($t = 9$, $s = 18$, $k = 4$, and $n = 72$)

Block	Treat.	Block	Treat.	Block	Treat.
1	1,4,6,7	7	2,3,6,7	13	1,2,4,9
2	2,6,8,9	8	2,4,5,8	14	1,5,6,9
3	1,3,8,9	9	3,5,7,9	15	1,3,6,8
4	1,2,3,4	10	1,2,5,7	16	4,6,7,8
5	1,5,7,8	11	2,3,5,6	17	3,4,5,8
6	4,5,6,9	12	3,4,7,9	18	2,7,8,9

Table 2.22
Simulated size of nominal 5% Wald F- tests for BIB5

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0546	0.0532	0.0612 ^{††}	0.0585 [†]
0.5	0.0534	0.0523	0.0577 [†]	0.0546
1.0	0.0518	0.0494	0.0525	0.0514
2.0	0.0498	0.0468	0.0476	0.0480
4.0	0.0505	0.0472	0.0475	0.0480

* before modification. S.E. of entries $\in (0.0021, 0.0024)$

Table 2.23
Mean of estimated denominator degrees of freedom for BIB5

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	66.4586	57.9582	57.9582	46
0.5	63.1346	54.6281	54.6281	46
1.0	58.2671	49.7502	49.7502	46
2.0	55.6517	47.1284	47.1284	46
4.0	54.8206	46.2951	46.2951	46

* before modification.

Table 2.24
Mean of estimated scales for BIB5

ρ	K-R*	K-R
0.25	0.9966	1.0000
0.5	0.9962	1.0000
1.0	0.9955	1.0000
2.0	0.9950	1.0000
4.0	0.9949	1.0000

* before modification.

Table 2.25
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for BIB5

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-3.0	0.8	1.0000	0.9688
0.5	-2.4	0.0	1.0000	0.9219
1.0	-1.7	-0.8	1.0000	0.8750
2.0	-0.9	-0.7	1.0000	0.8529
4.0	-0.6	-0.6	1.0000	0.8462

VI. In the sixth design (BIB6), where we have twelve blocks, nine treatments, and six treatments per block ($t = 9$, $s = 12$, $k = 6$, and $n = 72$), and $q = 0.9375$

Table 2.26
BIB6 ($t = 9$, $s = 12$, $k = 6$, and $n = 72$)

Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.
1	4,5,6,7,8,9	4	2,3,4,5,7,9	7	1,3,4,5,8,9	10	1,2,4,5,7,8
2	2,3,5,6,8,9	5	1,3,5,6,7,8	8	1,2,5,6,7,9	11	1,2,3,7,8,9
3	2,3,4,6,7,8	6	1,3,4,6,7,9	9	1,2,4,6,8,9	12	1,2,3,4,5,6

Table 2.27
Simulated size of nominal 5% Wald F- tests for BIB6

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	0.0496	0.0477	0.0543	0.0519
0.5	0.0503	0.0483	0.0518	0.0503
1.0	0.0517	0.0501	0.0512	0.0501
2.0	0.0527	0.0509	0.0511	0.0510
4.0	0.0526	0.0505	0.0505	0.0505

* before modification. S.E. of entries $\in (0.0021, 0.0023)$

Table 2.28
Mean of estimated denominator degrees of freedom for BIB6

ρ	K-R*	K-R&prop.	Satter	Contain
0.25	65.7861	57.2847	57.2847	52
0.5	63.6891	55.1839	55.1839	52
1.0	61.7011	53.1920	53.1920	52
2.0	60.8480	52.3371	52.3371	52
4.0	60.5985	52.0870	52.0870	52

* before modification.

Table 2.29
Mean of estimated scales for BIB6

ρ	K-R*	K-R
0.25	0.9965	1.0000
0.5	0.9963	1.0000
1.0	0.9960	1.0000
2.0	0.9959	1.0000
4.0	0.9959	1.0000

* before modification.

Table 2.30
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for BIB6

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-1.5	0.7	1.0000	0.9750
0.5	0.2	1.3	1.0000	0.9357
1.0	-0.2	0.1	1.0000	0.9000
2.0	-0.4	-0.4	1.0000	0.8824
4.0	-0.8	-0.8	1.0000	0.8769

8.2.3 Complete Block Designs with Missing Data

As we mentioned in the introduction of this chapter, two observations are in each of the following designs; one is from the first block under the first treatment, and the other is from the second block under the second treatments.

I. Consider a complete block design with five blocks and four treatments (RCB1), $q = 0.9477$.

Table 3.1
Simulated size of nominal 5% Wald F- tests for RCB1

ρ	K-R*	K-R	Prop.1	Prop.2	Satter	Contain
0.25	0.0696 ^{††}	0.0557 [†]	0.0557 [†]	0.0557 [†]	0.0591 [†]	0.0562 [†]
0.5	0.0691 ^{††}	0.0545	0.0546	0.0545	0.0575 [†]	0.0558 [†]
1.0	0.0631 ^{††}	0.0490	0.0490	0.0490	0.0508	0.0498
2.0	0.0622 ^{††}	0.0494	0.0494	0.0494	0.0498	0.0497
4.0	0.0622 ^{††}	0.0490	0.0490	0.0490	0.0492	0.0491

* before modification. S.E. of entries $\in (0.0022, 0.0025)$

Table 3.2

Mean of estimated denominator degrees of freedom for RCB1

ρ	K-R*	K-R	Prop.1	Prop.2	Satter	Contain
0.25	20.3740	10.7678	10.7747	10.7699	10.7542	10
0.5	20.1975	10.5812	10.5863	10.5828	10.5708	10
1.0	19.9325	10.3008	10.3029	10.3014	10.2960	10
2.0	19.7490	10.1050	10.1054	10.1051	10.1039	10
4.0	19.6793	10.0296	10.0297	10.0297	10.0295	10

* before modification.

Table 3.3
Mean of estimated scales for RCB1

ρ	K-R*	K-R	Prop.1	Prop.2
0.25	0.9348	0.9997	0.9996	0.9997
0.5	0.9332	0.9998	0.9997	0.9998
1.0	0.9307	0.9999	0.9999	0.9999
2.0	0.9288	1.0000	1.0000	1.0000
4.0	0.9281	1.0000	1.0000	1.0000

* before modification.

Table 3.4
Percentage relative bias in the variance estimates,
convergence rate (CR) and efficiency (E) for RCB1

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-3.0	-0.5	0.9939	0.9814
0.5	-2.3	-0.5	0.9972	0.9706
1.0	-1.0	-0.3	0.9992	0.9577
2.0	-0.6	0.0	0.9998	0.9508
4.0	-0.2	-0.2	1.0000	0.9485

II. Consider a complete block design with seven blocks and six treatments (RCB2),
 $q = 0.9839$

Table 3.5
Simulated size of nominal 5% Wald F- tests for RCB2

ρ	K-R*	K-R	Prop.1	Prop.2	Satter	Contain
0.25	0.0564 [†]	0.0521	0.0521	0.0521	0.0531	0.0525
0.5	0.0510	0.0466	0.0466	0.0466	0.0472	0.0472
1.0	0.0576 [†]	0.0524	0.0524	0.0524	0.0528	0.0526
2.0	0.0585 [†]	0.0525	0.0525	0.0525	0.0525	0.0525
4.0	0.0585 [†]	0.0527	0.0527	0.0527	0.0527	0.0527

* before modification. S.E. of entries $\in (0.0021, 0.0023)$

Table 3.6

Mean of estimated denominator degrees of freedom for RCB2

ρ	K-R*	K-R	Prop.1	Prop.2	Satter	Contain
0.25	37.3202	28.5050	28.5173	28.5076	28.4925	28
0.5	37.1337	28.3205	28.3266	28.3218	28.3143	28
1.0	36.9442	28.1321	28.1335	28.1324	28.1307	28
2.0	36.8508	28.0386	28.0387	28.0386	28.0384	28
4.0	36.8226	28.0102	28.0102	28.0102	28.0102	28

* before modification.

Table 3.7

Mean of estimated scales for RCB2

ρ	K-R*	K-R	Prop.1	Prop.2
0.25	0.9873	1.0000	0.9999	1.0000
0.5	0.9872	1.0000	1.0000	1.0000
1.0	0.9871	1.0000	1.0000	1.0000
2.0	0.9870	1.0000	1.0000	1.0000
4.0	0.9870	1.0000	1.0000	1.0000

* before modification.

Table 3.8

Percentage relative bias in the variance estimates, convergence rate (CR) and efficiency (E) for RCB2

ρ	$\hat{\Phi}$	$\hat{\Phi}_A$	CR	E
0.25	-0.4	0.0	0.9995	0.9922
0.5	-0.5	-0.3	0.9998	0.9887
1.0	-1.6	-1.6	1.0000	0.9857
2.0	0.9	0.9	1.0000	0.9844
4.0	0.7	0.7	1.0000	0.9840

8.3 Comments and Conclusion

The bias of the conventional estimator of the variance-covariance matrix of the fixed effects estimator is large and negative in most cases as expected especially with small values of ρ (e.g., table 2.5 and table 2.10). The bias was found to be effected by three factors: q , ρ , and n . When q , ρ , and/or n . get larger, the bias decreases. This bias is reduced to an acceptable level by using the adjusted estimator. As expected, when the between-block variance increased, the estimate of the denominator degrees of freedom

decreased (e.g., tables 1.3, 2.2, and 3.1). This is because that the contribution of the information from between blocks to estimate the fixed effects and their covariance matrix gets smaller. Notice that for all approaches, the denominator degrees of freedom estimates approach the Containment method's estimate as ρ increases. In other words, as ρ increases, the optimal approximate test approaches the Intra-block exact F-test where the between-block information is not utilized.

In general, the K-R before modification does not perform as well as K-R after modification. Before modification, the K-R method tends to be more liberal than K-R after modification (e.g., table 3.1 and 3.5). Even though the K-R approach using the conventional estimator of the variance-covariance matrix of the fixed effects estimator give reasonable true levels (table 1.2), the approach is not appropriate since the denominator degrees of freedom and the scale estimates were found to be negative for some data sets. For PBIB1 design, and when $\rho=0.25$, a data set has -120.8 as an estimate of the denominator degrees of freedom.

The K-R and the proposed methods perform well and similarly and the similarity gets stronger when ρ gets larger (for BIB designs, the K-R and proposed methods are identical). Their true levels tend to be more conservative with designs where the efficiency factor is higher, and this is related to the small remaining bias in the adjusted estimator of variance-covariance matrix of the fixed effects estimator in most cases. This bias is not eliminated in the BIB designs where the efficiency factor is small and this makes the K-R and the proposed methods' true levels more liberal (e.g., tables 2.5, 2.10 and 2.15). Indeed, they perform better than other approaches in most cases except a few cases that will be mentioned below.

The Satterthwaite approach was found to perform poorly for designs with small values of ρ, q and n where the true level tends to be too liberal (e.g., table 2.2). This unwanted liberality of the true level is related to the negative large percentage relative bias in the conventional estimator of the variance-covariance matrix of the fixed effects estimator for the cases where we have small values of ρ, q and n (e.g., table 2.5).

When ρ, q and/or n get larger, the Satterthwaite method tends to be more conservative.

This is due to the relatively small values of the bias in the conventional estimate of the estimator of the variance-covariance matrix of the fixed effects estimator (e.g., table 2.9). Indeed, when q and ρ get larger, the Satterthwaite method tends to perform as well as the K-R and the proposed methods regardless of the sample size (e.g., table 2.12).

For the cases where the Satterthwaite method performs poorly, the Containment method was found to perform better than the Satterthwaite method generally (e.g., tables 1.2 and 2.2). Moreover, for the cases where the efficiency factors are large, and ρ and/or n get larger, the performance of the Containment method improves in the way that it becomes as good as the K-R and the proposed methods (e.g., tables 2.12 and 2.22). For the cases where the efficiency factor is relatively small, we found the Containment method's performance tends to be better than all other approaches including the K-R and the proposed methods when ρ is relatively small (e.g., table 2.2).

9. CONCLUSION

9.1 Summary

The strength of Kenward and Roger's approach comes from combining three components where no other approach does: First, instead of approximating the Wald-type F statistic itself by an F distribution as in the Satterthwaite approximation, Kenward and Roger considered a scaled form of the statistic that makes the approximation more flexible since they approximate the denominator degrees of freedom and the scale as well. Second, Kenward and Roger modified the procedure in such a way that they get the right values for both the denominator degrees of freedom and the scale for the two special cases considered in chapter 3. Third, the adjusted estimator of the variance-covariance matrix of the fixed effects estimator, which is less biased than the conventional estimator, as we saw in chapter 2, is used in constructing the Wald-type F statistic.

In this thesis, by modifying certain steps in Kenward and Roger's derivation, we obtain two alternative methods which are similar to the K-R method but simpler in derivation and application. Whenever the ratio of the two quantities A_1 and A_2 derived in chapter 3 is the same as for the two special cases, the K-R and the proposed methods are identical. We also showed that the K-R and the proposed methods produce the right values for the denominator degrees of freedom and the scale factor, not only in the special cases, but also in three general models where we have exact F tests for fixed effects.

A Wald-type statistic was constructed using the conventional, rather than the adjusted, estimator of the variance-covariance matrix of the fixed effects estimator, and then modified to attain the right values in the two special cases. Because A_3 was found to be null in all cases with exact F tests considered in this thesis, the modification step was problematic and this is consistent with the remark made by Kenward and Roger (1997).

Another approach to obtain approximate F tests for fixed effects is the Satterthwaite method. When a one-dimensional hypothesis is of interest, then the K-R, the Satterthwaite, and the proposed methods produce the same estimate of the

denominator degrees of freedom, and the scale estimate is one. Even though the estimates are the same, the Satterthwaite and the K-R tests are not necessarily identical, and this is true because in the K-R statistic, we use the adjusted estimator of the variance-covariance matrix of the fixed effects estimator, whereas the conventional estimator is used in the Satterthwaite test. However, in some cases like balanced mixed classification models where the adjustment in the estimator of the variance-covariance matrix of the fixed effects estimator is null, the Satterthwaite approach is identical to the K-R and the proposed methods when testing a one-dimensional hypothesis. When we make inference about a multidimensional hypothesis, we still have some cases where the K-R and the Satterthwaite approaches produce the same estimate of the denominator degrees of freedom, or are even identical approaches as in lemma 7.2.7.

In chapter 8, we conducted a simulation study for three kinds of block designs; partially balanced block designs, balanced incomplete block designs, and complete block designs with missing data. The K-R, the proposed, the Satterthwaite, and the Containment methods were compared in the study. Three factors were considered in the study: the variance components ratio, the efficiency factor, and the sample size. In most cases, we found the K-R and the proposed methods performed as well as or better than the other methods. For designs with small values for all the three factors, the Satterthwaite method performed poorly. The true level was found to be quite liberal due to the large bias in the conventional estimator of the variance-covariance matrix of the fixed effects estimator. In some designs where the efficiency factor and the ratio of the variance components are both relatively small, the K-R and the proposed methods were found to be liberal. For those cases, the Containment method was found to work as well as or better than all other approaches.

9.2 Future Research

In this thesis, some questions about the K-R and other methods used to make inference about fixed effects in mixed linear models were answered. However, many issues still need to be investigated. In chapter 3, we derived approximate expressions for

the expectation and variance of the Wald-type F statistic using the conventional estimator of the variance-covariance of the fixed effects estimator. However, as we saw in section 4.4, it was problematic to use the special cases to modify the approach since A_3 was null. As a matter of fact, it is also null in all three general models considered in chapter 6. A possible direction for further research is to investigate if there is any case where we have an exact F test and A_3 is not null, so that a better modification can be accomplished. Two different ratios A_1/A_2 were found in the two special cases and in the three general models considered in chapter 6. It is suggested to search if there is any case where we have an exact F test and the ratio A_1/A_2 is different than the two ratios we had. In case another ratio is found, it is suggested to use it in the modifying stage for the K-R and the proposed methods.

In balanced mixed classification models, and in the general models considered in sections 6.2 and 6.3, we noticed that the adjustment of the estimator of the variance-covariance matrix of the fixed effects estimator is null. It would be interesting to generalize these results.

When we make an inference about a multidimensional hypothesis of the fixed effects, lemma 7.2.7 provide conditions where the K-R and the Satterthwaite approaches produce the same estimate of the denominator degrees of freedom. It would be interesting to investigate the possibility of other conditions that lead to the same result.

Finally, based on the theoretical derivation of the methods proposed in chapter 5, and their simulation results for the block designs in chapter 8, these methods' performances are expected to be comparable to the K-R method. A natural proposal is to investigate the performance of these proposed methods in other different models.

BIBLIOGRAPHY

- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons.
- Bellavance, F., Tardif, S. and Stephens, M. (1996). Tests for the Analysis of Variance of Crossover Designs with Correlated Errors. *Biometrics* **52**, 607-612.
- Birkes, D. (2004). *Linear Model Theory Class Notes*, Unpublished. Oregon State University.
- Birkes, D. and Lee, Y. (2003). *Adjusted Orthogonal Block Structure in Mixed Linear Models*. Unpublished.
- Birkes, D. (2006). *Class Notes*. Unpublished. Oregon State University.
- Chen, X. and Wei, L. (2003). A Comparison of Recent Methods for the Analysis of Small-Sample Cross-Over Studies. *Statistics in Medicine* **22**, 2821-2833.
- Cochran, W. and Cox, G. (1975). *Experimental Designs*. New York: John Wiley & Sons.
- Fai, C and Cornelius, P. (1996). Approximate F Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analysis of Unbalanced Split-Plot Experiments. *Journal of Statistical Computation and Simulation* **54**, 363-378.
- Giesbrecht, F. and Burns, J. (1985). Two-Stage Anaysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics* **41**, 477-486.
- Gomes, E., Schaalje, G. and Fellingham, G. (2005). Performance of the Kenward-Roger Method when the Covariance Structure is Selected Using AIC and BIC. *Communications in Statistics-Simulation and Computation* **34**, 377-392.
- Graybill, F. and Wang, C. (1980). Confidence Intervals on Nonnegative Linear Cominations of Variances. *Journal of the American Statistical Association* **75**, 869-873.
- Green, P. (2001). On the Design of Choice Experiments Involving Multifactor Alternatives. *Journal of Consumer Research* **1**, 61-68.
- Harville, D. (1997). *Matrix Algebra from a Statistician's Prospective*. New York: Springer.
- Harville, D. and Jeske, D. (1992). Mean Squared Error of Estimation or Prediction Under a General Linear Model. *Journal of the American Sataistical Association* **87**, 724-731.
- Huynh, H. and Feldt, L. (1970). Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions. *Journal of the American Statistical Association* **65**, 1582-1589.

- Kackar, R. and Harville, D. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effect in Mixed Linear Models. *Journal of the American Staistical Association* **79**, 853-862.
- Kenward, M. and Roger, J. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics* **53**, 983-997.
- Krzanowski, W. (2000). *Principles of Multivariate Analysis*. Oxford, U.K.: Oxford University Press.
- Kuehl, R. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Pacific Grove: Duxbury Press.
- Mardia, K., Kent, J. and Bibby, J. (1992). *Multivariate Analysis*. New York: Academic Press.
- Pase, L. and Salvan, A. (1997). *Principles of Statistical Inference*. New Jersey: World Scientific.
- Prasad, N. and Rao, J. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association* **85**, 163-171.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: John Wiley & Sons.
- Rady, E. (1986). *Testing Fixed Effects in Mixed Linear Models*. Ph.D. thesis, Unpublished. Oregon State University.
- Rao, C. (2002). *Linear Statistical Inference and its Applications*. New York: John Wiley & Sons.
- SAS Institute, Inc. (2002-2006), SAS version 9.1.3, Online Help, Cary, NC: SAS Institute.
- Satterthwaite, F. (1941). Synthesis of Variance. *Psychometrika* **6**, 309-316.
- Satterthwaite, F. (1946). Approximate Distribution of Estimates of Variance Components. *Biometrics* **2**, 110-114.
- Savin, A., Wimmer, G. and Witkovsky, V. (2003). *Measurement Science Review* **3**, 53-56.
- Schaalje, G., McBride, J. and Fellingham, G. (2002). *Journal of Agricultural, Biological, and Enviromental Statistics* **7**, 512-524.
- Schott, J. (2005). *Matrix Analysis for Statistics*. New York: John Wiley & Sons.
- Searle, S., Casella, G. and McCullouch, C. (1992). *Variance Components*. New York: John Wiley & Sons.
- Seely, J. (2002). *Linear Model Theory Class Notes*, Unpublished. Oregon State University.

- Seely, J. and Rady, E. (1988). When Can Random Effects Be Treated As Fixed Effects For Computing A Test Statistic For A Linear Hypothesis? *Communications in Statistics- Theory and Methods* **17**, 1089-1109.
- Siefert, B. (1979). Optimal Testing for Fixed Effects in General Balanced Mixed Classification Models. *Math. Operationsforsch. Statist. Ser. Statistics* **10**, 237-256.
- Spike, J., Piepho, H. and Meyer, U. (2004). Approximating the Degrees of Freedom for Contrasts of Genotypes Laid Out as Subplots in an Alpha-Design in a Split-Plot Experiment. *Plan Breeding* **123**, 193-197.
- Utlaut, T., Birkes, D. and VanLeeuwen, D. (2001). Optimal Exact F Tests and Error-Orthogonal Designs. *Conference Proceeding of the 53rd ISI Session, Seoul 2001*.
- VanLeeuwen, D., Seely, J. and Birkes, D. (1998). Sufficient Conditions for Orthogonal Designs in Mixed Linear Models. *Journal of Statistical Planning and Inference* **73**, 373-389.
- VanLeeuwen, D., Birkes, D. and Seely, J. (1999). Balance and Orthogonality in Designs for Mixed Classification Models. *The Annals of Statistics* **27**, 1927-1947.