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

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

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,

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

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 $\mathbf{\Sigma}(n \times n)$ is an unknown variance-covariance matrix whose elements are assumed to be functions of r parameters, $\boldsymbol{\sigma}(r \times 1) = (\sigma_1, ..., \sigma_r)'$. The generalized least squares estimator of $\boldsymbol{\beta}$ is $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}$, and the matrix $\mathbf{\Phi} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$ 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 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{y}$. The center of our study is $\boldsymbol{\beta}$, and $\mathbf{\Sigma}$ is a nuisance to be addressed in the analysis. We are interested in testing $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, 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 $\mathbf{A}'$, $\Re(\mathbf{A})$, $\mathfrak{N}(\mathbf{A})$, $\mathrm{r}(\mathbf{A})$, $\mathrm{tr}(\mathbf{A})$, $|\mathbf{A}|$, 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. $\mathbf{P}_\mathbf{A}$ is o.p.o. on $\Re(\mathbf{A})$. In addition, we use the following

$$\mathbf{V} = \mathrm{Var}(\hat{\boldsymbol{\beta}})$$
$$\mathbf{W} = \mathrm{Var}(\hat{\boldsymbol{\sigma}}),\ w_{ij} = \mathrm{Cov}(\hat{\sigma}_i, \hat{\sigma}_j)$$
$$\mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$$
$$\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}'$$
$$\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$$
$$\mathbf{P}_i = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\boldsymbol{\Sigma}^{-1}\mathbf{X}$$
$$\mathbf{Q}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}$$
$$\mathbf{R}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}$$
$$\boldsymbol{\Lambda} = \mathrm{Var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$$
$$\tilde{\boldsymbol{\Lambda}} = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j)\boldsymbol{\Phi}$$
$$\hat{\boldsymbol{\Phi}}_\mathrm{A} = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{w}_{ij}(\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i\hat{\boldsymbol{\Phi}}\hat{\mathbf{P}}_j - \frac{1}{4}\hat{\mathbf{R}}_{ij})\right\}\hat{\boldsymbol{\Phi}}$$
$$A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})$$
$$A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}),$$
$$A_3 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}[\boldsymbol{\Theta\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\boldsymbol{\Phi}]$$

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

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

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

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

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

(A5) $\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{r}\mathrm{Cov}\left[(\frac{\partial\tilde{\boldsymbol{\beta}}}{\partial\sigma_i})(\frac{\partial\tilde{\boldsymbol{\beta}}}{\partial\sigma_j})', (\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\right] = O(n^{-5/2}).$

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

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

<u>*Remarks*</u>

(*i*) $\boldsymbol{\Sigma}_k$ is said to be bounded when $|\max(\text{elements of } \boldsymbol{\Sigma}_k)| \le b(\boldsymbol{\sigma})$ for some constant b .

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

(*ii*) Even though Kenward and Roger required the bias of the REML estimator to be ignored, we only require $\mathrm{E}[\hat{\boldsymbol{\sigma}}] = \boldsymbol{\sigma} + O(n^{-3/2})$ as stated as assumption (A3). In fact, for the model mentioned in section 2.1, $\mathrm{E}[\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}] = \mathbf{b}(\boldsymbol{\sigma}) + O(n^{-2})$, 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 $\mathrm{E}[\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}] = O(n^{-2})$ 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 $\mathrm{E}\left[(\frac{\partial\tilde{\boldsymbol{\beta}}}{\partial\sigma_i})(\frac{\partial\tilde{\boldsymbol{\beta}}}{\partial\sigma_j})' \mid \hat{\boldsymbol{\sigma}}\right]$ 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 $\mathbf{H}(\boldsymbol{\sigma}) = \mathrm{diag}_{1\le k\le m}(\mathbf{H}_k(\boldsymbol{\sigma}))$, 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 $1\le k\le m \Rightarrow \left|\max(\text{elements of } \mathbf{X}_k'\mathbf{H}_k(\boldsymbol{\sigma})\mathbf{X}_k)\right| \le a$ for $1\le k\le m$

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

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

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 set $\mathbb{C}$ 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 $\frac{1}{m}\sum_{k=1}^{m}\mathbf{X}_k'\boldsymbol{\Sigma}_1^{-1}\mathbf{X}_k \to \mathbf{A}$(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 $\mathrm{E}[\mathbf{X}_1'\boldsymbol{\Sigma}_1^{-1}\mathbf{X}_1]$.

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

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

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

However, in lemma (2.4.7), we have $\mathrm{E}[\mathbf{X}_k'\mathbf{B}_k(\boldsymbol{\sigma})\mathbf{y}_k] = \mathbf{0}$, 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, $\mathrm{E}[(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)(\hat{\sigma}_k - \sigma_k)] = O(n^{-2})$, 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. □

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

There are two main sources of bias in $\hat{\boldsymbol{\Phi}}$ when it is used as an estimator for Var($\hat{\boldsymbol{\beta}}$): $\hat{\boldsymbol{\Phi}}$ underestimates $\boldsymbol{\Phi}$, and $\boldsymbol{\Phi} = \mathrm{Var}(\tilde{\boldsymbol{\beta}})$ 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 Var($\hat{\boldsymbol{\beta}}$) can be partitioned as 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 Var($\hat{\boldsymbol{\beta}}$) which is denoted by $\hat{\boldsymbol{\Phi}}_A$. In this section, several lemmas are derived to lead to the expression for $\hat{\boldsymbol{\Phi}}_A$.

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

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

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

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

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

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

The REML estimator for $\boldsymbol{\sigma}$ 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

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

To prove the lemma, it suffices to show that

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

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

For part (a), it suffices to show that

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

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

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

Observe that $\boldsymbol{\Sigma}$ is nonsingular (assumption A2), and n.n.d (Birkes, 2004), then $\boldsymbol{\Sigma}$ is p.d. (Seely, 2002, corollary 1.9.3),

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

Also,we have $\Re(\mathbf{K}) = \mathfrak{N}(\mathbf{X}')$ (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 $\mathbf{K}'\mathbf{X}=\mathbf{0}$.

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

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

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

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

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

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

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

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

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

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

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

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, $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} \overset{d}{=} -(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \Rightarrow \mathrm{E}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathrm{E}[-(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]$

$\Rightarrow \mathrm{E}(\hat{\boldsymbol\beta}) - \boldsymbol\beta = -\mathrm{E}(\hat{\boldsymbol\beta}) + \boldsymbol\beta \Rightarrow 2\mathrm{E}(\hat{\boldsymbol\beta}) = 2\boldsymbol\beta \Rightarrow \mathrm{E}(\hat{\boldsymbol\beta}) = \boldsymbol\beta.$ □

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

Proof $\mathrm{Cov}\left[\tilde{\boldsymbol\beta}, (\mathbf{I} - \mathbf{P_X})\mathbf{y}\right] = \mathrm{Cov}\left[(\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}y}, (\mathbf{I} - \mathbf{P_X})\mathbf{y}\right]$

$$= (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}\Sigma}(\mathbf{I} - \mathbf{P_X}) = (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'}(\mathbf{I} - \mathbf{P_X})$$
$$= (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'} - (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'P_X}$$
$$= (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'} - (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'} = \mathbf{0}$$

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

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

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

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

$(\mathbf{I} - \mathbf{P_X})\mathbf{y}^{(1)} = (\mathbf{I} - \mathbf{P_X})\mathbf{y}^{(2)} \Rightarrow \boldsymbol\delta(\mathbf{y}^{(1)}) = \boldsymbol\delta(\mathbf{y}^{(2)})$

$$(\mathbf{I} - \mathbf{P_X})\mathbf{y}^{(1)} = (\mathbf{I} - \mathbf{P_X})\mathbf{y}^{(2)} \Rightarrow \mathbf{y}^{(1)} - \mathbf{P_X}\mathbf{y}^{(1)} = \mathbf{y}^{(2)} - \mathbf{P_X}\mathbf{y}^{(2)}$$
$$\Rightarrow \mathbf{y}^{(2)} = \mathbf{y}^{(1)} - \mathbf{P_X}\mathbf{y}^{(1)} + \mathbf{P_X}\mathbf{y}^{(2)} = \mathbf{y}^{(1)} + \mathbf{P_X}(\mathbf{y}^{(2)} - \mathbf{y}^{(1)})$$
$$\Rightarrow \boldsymbol\delta(\mathbf{y}^{(2)}) = \boldsymbol\delta(\mathbf{y}^{(1)} + \mathbf{Xb}) \quad \text{for some } \mathbf{b}$$
$$= \boldsymbol\delta(\mathbf{y}^{(1)}) \ \ (\boldsymbol\delta \text{ is translation invariant}).$$ □

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

Proof From lemma 2.4.4, $\tilde{\boldsymbol\beta}$ and $(\mathbf{I} - \mathbf{P_X})\mathbf{y}$ are independent, and from lemma 2.4.5, $\hat{\boldsymbol\beta} - \tilde{\boldsymbol\beta}$ is a function of $(\mathbf{I} - \mathbf{P_X})\mathbf{y}$. Then, $\tilde{\boldsymbol\beta}$ and $\hat{\boldsymbol\beta} - \tilde{\boldsymbol\beta}$ are independent.

Write $\hat{\boldsymbol\beta} = \tilde{\boldsymbol\beta} + \hat{\boldsymbol\beta} - \tilde{\boldsymbol\beta}$

$$\Rightarrow \mathrm{Var}(\hat{\boldsymbol\beta}) = \mathrm{Var}(\tilde{\boldsymbol\beta}) + \mathrm{Var}(\hat{\boldsymbol\beta} - \tilde{\boldsymbol\beta}) = \boldsymbol\Phi + \boldsymbol\Lambda.$$ □

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

Proof $\dfrac{\partial\tilde{\boldsymbol\beta}}{\partial\sigma_i} = \dfrac{\partial}{\partial\sigma_i}\left[(\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}y}\right]$

$$= \frac{\partial}{\partial\sigma_i}\left[(\mathbf{X'\Sigma^{-1}X})^{-1}\right]\mathbf{X'\Sigma^{-1}y} + (\mathbf{X'\Sigma^{-1}X})^{-1}\frac{\partial}{\partial\sigma_i}\left[\mathbf{X'\Sigma^{-1}y}\right]$$

$$= -\boldsymbol\Phi\mathbf{X'\Sigma^{-1}}\frac{\partial\boldsymbol\Sigma}{\partial\sigma_i}\mathbf{Gy} \quad \text{where } \mathbf{G} = \boldsymbol\Sigma^{-1} - \boldsymbol\Sigma^{-1}\mathbf{X}(\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}} \tag{2.1}$$

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

where $\mathrm{E}[\mathbf{Gy}] = \mathrm{E}[(\boldsymbol\Sigma^{-1} - \boldsymbol\Sigma^{-1}\mathbf{X}(\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}})\mathbf{y}]$

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

$\mathrm{E}\left[\dfrac{\partial\tilde{\boldsymbol\beta}}{\partial\sigma_i}(\hat\sigma_i - \sigma_i)\right] = -\boldsymbol\Phi\mathbf{X'\Sigma^{-1}}\dfrac{\partial\boldsymbol\Sigma}{\partial\sigma_i}\mathrm{E}[g(\mathbf{y})]$, where $g(\mathbf{y}) = \mathbf{Gy}(\hat\sigma_i(\mathbf{y}) - \sigma_i)$.

$g(-\mathbf{y}) = \mathbf{G}(-\mathbf{y})\left[\hat\sigma_i(-\mathbf{y}) - \sigma_i\right] = -\mathbf{Gy}\left[\hat\sigma_i(-\mathbf{y}) - \sigma_i\right] = -g(\mathbf{y})$, because $\hat\sigma_i$ is reflection equivariant. So, g is reflection-equaivariant.

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

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

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

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

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

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

$$= (\mathbf{Gy} + \mathbf{GX}\boldsymbol\beta)\left[\hat\sigma_i(\mathbf{y} + \mathbf{X}\boldsymbol\beta) - \sigma_i\right] = \mathbf{Gy}[\hat\sigma_i(\mathbf{y}) - \sigma_i] = g(\mathbf{y}),$$

and hence g is a translation invariant function.

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

$\Rightarrow 2\mathrm{E}[g(\mathbf{y})] = \mathbf{0} \Rightarrow \mathrm{E}[g(\mathbf{y})] = \mathbf{0}$, and hence $\mathrm{E}\left[(\dfrac{\partial\tilde{\boldsymbol\beta}}{\partial\sigma_i})(\hat\sigma_i - \sigma_i)\right] = \mathbf{0}$ for $i = 1,....,r$ □

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

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

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

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

Proof $(a)\ \dfrac{\partial \mathbf{\Phi}}{\partial \sigma_i} = -(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\dfrac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})}{\partial \sigma_i}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$

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

$$(b)\ \frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} = -\frac{\partial}{\partial \sigma_i}\left\{(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right\}$$

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

$$-(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{\partial^2 \mathbf{\Sigma}^{-1}}{\partial \sigma_i \partial \sigma_j}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$$

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

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

Observe that $\mathbf{X}'\dfrac{\partial^2 \mathbf{\Sigma}^{-1}}{\partial \sigma_i \partial \sigma_j}\mathbf{X} = \mathbf{X}'\dfrac{\partial}{\partial \sigma_i}\left(-\mathbf{\Sigma}^{-1}\dfrac{\partial \mathbf{\Sigma}}{\partial \sigma_j}\mathbf{\Sigma}^{-1}\right)\mathbf{X}$

$$= \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \sigma_i}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \sigma_j}\mathbf{\Sigma}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial^2 \mathbf{\Sigma}}{\partial \sigma_i \partial \sigma_j}\mathbf{\Sigma}^{-1}\mathbf{X} + \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \sigma_j}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}$$

$$= \mathbf{Q}_{ij} - \mathbf{R}_{ij} + \mathbf{Q}_{ji} \tag{2.3}$$

Combining expressions 2.3 and 2.3, we obtain

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

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

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

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

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

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

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

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

$w_{ij} = i^{ij} + a(\boldsymbol{\sigma})$, 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 $a(\boldsymbol{\sigma}) = O(n^{-2})$.

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

Since $\dfrac{\partial \mathbf{\Phi}}{\partial \sigma_i} = -\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}$, $\dfrac{\partial^2 \mathbf{\Phi}}{\partial \sigma_i \partial \sigma_j} = \mathbf{\Phi}(\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j + \mathbf{P}_j\mathbf{\Phi}\mathbf{P}_i - \mathbf{Q}_{ij} - \mathbf{Q}_{ji} + \mathbf{R}_{ij})\mathbf{\Phi}$(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 $\dfrac{\partial i^{ij}}{\partial \sigma_i} = O(n^{-1})$,

$\dfrac{\partial^2 i^{ij}}{\partial \sigma_i^2} = O(n^{-1})$, $\dfrac{\partial a(\boldsymbol{\sigma})}{\partial \sigma_i} = O(n^{-2})$, and $\dfrac{\partial^2 a(\boldsymbol{\sigma})}{\partial \sigma_i^2} = O(n^{-2})$hence results (j) and (l) are obtained. By computing the derivative of $\tilde{\mathbf{\Lambda}}$, 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 $\boldsymbol{\Lambda} = \text{Var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \tilde{\boldsymbol{\Lambda}} + O(n^{-5/2})$,

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

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

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

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

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

Applying lemma 2.4.7, we obtain

$$\boldsymbol{\Lambda} = \text{E}\left[\sum_{i=1}^{r}\sum_{j=1}^{r} (\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i})(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j})'(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\right] + O(n^{-5/2})$$

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

By assumption (A3), we have

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

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

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

$$= \boldsymbol{\Sigma}^{-1}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}][\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]$$

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

Then, using expression (2.1), we obtain

$$\text{Cov}\left[\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_i}, \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \sigma_j}\right] = \boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\boldsymbol{\Sigma}^{-1}[\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}$$

$$= \boldsymbol{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j)\boldsymbol{\Phi} \quad (2.6)$$

Combining expressions 2.4, 2.5 and 2.6 we have

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

Lemma 2.4.11 (*a*) $\text{E}[\hat{\boldsymbol{\Phi}}] = \boldsymbol{\Phi} - \tilde{\boldsymbol{\Lambda}} + \mathbf{R}^* + O(n^{-5/2})$

(*b*) $\text{E}[\hat{\boldsymbol{\Lambda}}] = \tilde{\boldsymbol{\Lambda}} + O(n^{-5/2})$

(*c*) $\text{E}[\hat{\mathbf{R}}^*] = \mathbf{R}^* + O(n^{-5/2})$

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

Proof (*a*) Using a Taylor series expansion about $\boldsymbol{\sigma}$,

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

By assumption (A3), we have

$$\text{E}[\hat{\boldsymbol{\Phi}}] = \boldsymbol{\Phi} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} \frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} w_{ij} + O(n^{-5/2})$$

$$= \boldsymbol{\Phi} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\Phi}(\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j + \mathbf{P}_j\boldsymbol{\Phi}\mathbf{P}_i - \mathbf{Q}_{ij} + \mathbf{R}_{ij} - \mathbf{Q}_{ji})\boldsymbol{\Phi} + O(n^{-5/2}) \text{ (lemma 2.4.8)} \quad (2.7)$$

$$= \boldsymbol{\Phi} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\Phi}(\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j - \mathbf{Q}_{ij})\boldsymbol{\Phi} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\Phi}(\mathbf{P}_j\boldsymbol{\Phi}\mathbf{P}_i - \mathbf{Q}_{ij})\boldsymbol{\Phi}$$

$$+ \frac{1}{2}\sum_{i=1}^{r}\sum_{i=1}^{r} w_{ij}\boldsymbol{\Phi}\mathbf{R}_{ij}\boldsymbol{\Phi} + O(n^{-5/2})$$

$$= \boldsymbol{\Phi} - \frac{1}{2}\tilde{\boldsymbol{\Lambda}} - \frac{1}{2}\tilde{\boldsymbol{\Lambda}} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\Phi}\mathbf{R}_{ij}\boldsymbol{\Phi} + O(n^{-5/2}) = \boldsymbol{\Phi} - \tilde{\boldsymbol{\Lambda}} + \mathbf{R}^* + O(n^{-5/2})$$

(*b*) Using a Taylor series expansion about $\boldsymbol{\sigma}$, we obtain

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

(*c*) Using a Taylor series expansion about $\boldsymbol{\sigma}$, we obtain

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

Proposition 2.4.12 $\text{E}[\hat{\boldsymbol{\Phi}}_A] = \text{Var}[\hat{\boldsymbol{\beta}}] + O(n^{-5/2})$, where $\hat{\mathbf{\Phi}}_{\mathrm{A}} = \hat{\mathbf{\Phi}} + 2\hat{\mathbf{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{w}_{ij}(\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i\hat{\mathbf{\Phi}}\hat{\mathbf{P}}_j - \frac{1}{4}\hat{\mathbf{R}}_{ij})\right\}\hat{\mathbf{\Phi}}.$

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

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

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

$= \mathrm{Var}(\hat{\boldsymbol{\beta}}) + O(n^{-5/2}).$ □

<u>*Comments*</u>

From proposition 2.4.12, we obtain

$$\mathrm{E}[\hat{\mathbf{\Phi}}_{\mathrm{A}}] = \mathrm{Var}(\hat{\boldsymbol{\beta}}) + O(n^{-5/2}),$$

and from lemma 2.4.10, we obtain

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

This shows that the adjusted estimator for $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ 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 $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ 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 $\lambda F \sim \mathrm{F}(\ell, m)$ 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 $\hat{\mathbf{\Phi}}_{\mathrm{A}}$ (this is what was done by Kenward and Roger, 1997), or through the conventional estimator $\hat{\mathbf{\Phi}}$. Both approaches are to be considered in this section.

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

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

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

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

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

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

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

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

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

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

$$= \mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})] + \mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})] .$$

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

$$(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1} = (\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}$$

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

and

$$(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) = \mathbf{I} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})$$

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

Since

$$\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] =$$

$$2\mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_j}\mathbf{L})\right] - \mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\mathbf{L})\right], \qquad (3.1)$$

then by assumption (A3), we have

$$\mathrm{E}[\mathrm{tr}\{(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\}] = \ell + \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_j}\mathbf{L})\right]$$

$$-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\mathbf{L})\right] + O(n^{-3/2})$$

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

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

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

Hence,

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

and,

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

$$= 1 + \frac{1}{\ell}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}) + \frac{2}{\ell}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathrm{tr}[\boldsymbol{\Theta}\boldsymbol{\Phi}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\boldsymbol{\Phi}] + O(n^{-3/2}).$$

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

where

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

and

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

3.1.1.2 Deriving an Approximate Expression for Var[F]

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

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. $\quad \mathrm{Var}[F\,|\,\hat{\boldsymbol{\sigma}}]] = \dfrac{1}{\ell^2}\mathrm{Var}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\mathbf{L}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\,|\,\hat{\boldsymbol{\sigma}}]$

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

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

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

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

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

and

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

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

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

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

(*i*) $[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2$

$$= (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})](\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})]$$
$$= [\mathbf{I} + (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})][\mathbf{I} + (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})]$$
$$= \mathbf{I} + 2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L}) + (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})$$
$$= \mathbf{I} + 2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L}) + O(n^{-2}) \tag{3.2}$$

(*ii*) $(\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L}) = [(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})](\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})]$

$$= (\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + 2(\mathbf{L}'\mathbf{\Lambda}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})$$

So, $\mathrm{tr}\left[(\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}\right]$

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

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

$$2A_2 - 2\sum_{i=1}^{r}\sum_{i=1}^{r}w_{ij}\,\mathrm{tr}[\mathbf{\Theta\Phi}(\mathbf{P}_i\mathbf{\Phi P}_j - \mathbf{Q}_{ij} + \frac{1}{2}\mathbf{R}_{ij})\mathbf{\Phi}] + O(n^{-2}).$$

(*iii*) $\dfrac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\dfrac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})$

$$= (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})]\times$$
$$(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_j}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L}) + (\mathbf{L}'\mathbf{\Lambda}\mathbf{L})]$$
$$= (\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_j}\mathbf{L})$$

$$+(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_j}\mathbf{L})(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})$$

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

And hence,

$$\mathrm{tr}\left(\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})\right)$$

$$=\mathrm{tr}[\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi\Theta\Phi P}_j\mathbf{\Phi})+O(n^{-2})$$

$$=A_2+O(n^{-2}) \tag{3.4}$$

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

$$\mathrm{E}[\mathrm{Var}[F\mid\hat{\boldsymbol{\sigma}}]]=$$

$$\frac{2}{\ell^2}\left\{\ell+2\mathrm{tr}(\mathbf{\Theta\Lambda})+3A_2-2\sum_{i=1}^{r}\sum_{i=1}^{r}w_{ij}\mathrm{tr}[\mathbf{\Theta\Phi}(\mathbf{P}_i\mathbf{\Phi P}_j-\mathbf{Q}_{ij}+\frac{1}{2}\mathbf{R}_{ij})\mathbf{\Phi}]\right\}+O(n^{-3/2}) \tag{3.5}$$

B. $\mathrm{Var}[\mathrm{E}[F\mid\hat{\boldsymbol{\sigma}}]]=\frac{1}{\ell^2}\mathrm{Var}[\mathrm{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]$

$$=\frac{1}{\ell^2}\left\{\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2-\left(\mathrm{E}\left\{\mathrm{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right\}\right)^2\right\}$$

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

$$\left(\mathrm{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right)^2=\mathrm{tr}\left[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\right.$$

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

$$=\left(\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2+2\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\mathrm{tr}\left[\frac{\partial(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]$$

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

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

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

$$\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2=\left(\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]$$

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

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

(i) $\quad(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})=\mathbf{I}+(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})$

$\Rightarrow\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]=\ell+\mathrm{tr}(\mathbf{\Theta\Lambda})$

$$\Rightarrow\left(\mathrm{tr}[(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right)^2=\ell^2+2\ell\mathrm{tr}(\mathbf{\Theta\Lambda})+O(n^{-2}) \tag{3.6}$$

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

$$=-(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L}-(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}'\frac{\partial\mathbf{\Phi}}{\partial\sigma_i}\mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})$$

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

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

and $\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]$

$$=\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})\right]+\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})\right]$$

$$=2\mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi\Theta\Phi P}_j\mathbf{\Phi})-\mathrm{tr}\left(\mathbf{\Theta}\frac{\partial^2\mathbf{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)+2\mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi\Theta\Phi P}_j\mathbf{\Phi\Theta\Lambda})$$

$$-\mathrm{tr}\left(\mathbf{\Theta}\frac{\partial^2\mathbf{\Phi}}{\partial\sigma_i\partial\sigma_j}\mathbf{\Theta\Lambda}\right)\text{ (from expression 3.1).} \tag{3.8}$$

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

$$\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right)^2 = \ell^2 + 2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}[\ell+\mathrm{tr}(\boldsymbol{\Theta\Lambda})]\left\{-\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)+2\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})\right.$$

$$\left.+2\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})-\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Theta\Lambda}\right)\right\}$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}[\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})+\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Lambda})][\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})+\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})]+O(n^{-3/2})$$

$$=\ell^2+2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\left\{2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})-\ell\,\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right.$$

$$+2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})-\ell\,\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\theta\Lambda}\right)+2\mathrm{tr}(\boldsymbol{\Theta\Lambda})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})$$

$$-\mathrm{tr}(\boldsymbol{\Theta\Lambda})\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)+2\mathrm{tr}(\boldsymbol{\Theta\Lambda})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})-\mathrm{tr}(\boldsymbol{\Theta\Lambda})\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Theta\Lambda}\right)$$

$$\left.+\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})+2\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})+\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Lambda})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi\Theta\Lambda})\right\}$$

$$+O(n^{-3/2})$$

$$=\ell^2+2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\left\{2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})-\ell\,\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right\}$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\left\{\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})\right\}+O(n^{-3/2})$$

$$=\ell^2+2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+2\ell A_2+A_1-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)+O(n^{-3/2}), \tag{3.9}$$

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

Also, from expression 3.2,

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

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

$\mathrm{Var}[\mathrm{E}[F|\hat{\boldsymbol{\sigma}}]]$

$$=\frac{1}{\ell^2}\left\{\ell^2+2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+2\ell A_2+A_1-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)-\ell^2-2\ell(A_2+2A_3)\right\}+O(n^{-3/2})$$

$$=\frac{1}{\ell^2}\left\{2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+A_1-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)-4\ell A_3\right\}+O(n^{-3/2})$$

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

$$\mathrm{Var}[F]=\frac{2}{\ell^2}\left\{\ell+2\mathrm{tr}(\boldsymbol{\Theta\Lambda})+3A_2-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right\}$$

$$+\frac{1}{\ell^2}\left\{2\ell\,\mathrm{tr}(\boldsymbol{\Theta\Lambda})+A_1-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)-4\ell A_3\right\}+O(n^{-3/2})$$

$$=\frac{2}{\ell}\left\{1+\frac{3A_2}{\ell}+\frac{A_1}{2\ell}-2A_3+\left(\frac{2}{\ell}+1\right)\mathrm{tr}(\boldsymbol{\Theta\Lambda})-\left(\frac{1}{\ell}+\frac{1}{2}\right)\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right\}+O(n^{-3/2})$$

$$=\frac{2}{\ell}\left\{1+\frac{1}{2\ell}\left[A_1+6A_2-4\ell A_3+2(\ell+2)\mathrm{tr}(\boldsymbol{\Theta\Lambda})-(\ell+2)\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right]\right\}+O(n^{-3/2})$$

$$=\frac{2}{\ell}\left\{1+\frac{1}{2\ell}\left[A_1+6A_2-4\ell A_3+(\ell+2)\left\{2\mathrm{tr}(\boldsymbol{\Theta\Lambda})-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)\right\}\right]\right\}+O(n^{-3/2})$$

$$=\frac{2}{\ell}\left\{1+\frac{1}{2\ell}\left[A_1+6A_2-4\ell A_3+(\ell+2)\left\{4\mathrm{tr}(\boldsymbol{\Theta\Lambda})-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{R}_{ij}\boldsymbol{\Phi})\right\}\right]\right\}+O(n^{-3/2}),$$

by noticing that

$$-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left(\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right)=$$

$$2\mathrm{tr}(\boldsymbol{\Theta\Lambda})-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{R}_{ij}\boldsymbol{\Phi})+O(n^{-5/2}) \quad \text{(lemmas 2.4.8 and 2.4.10)},$$

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

Proposition 3.1.1.2.1 $\mathrm{Var}[F]=\frac{2}{\ell}(1+B_1)+O(n^{-3/2})$, where $$B_1 = \frac{1}{2\ell}\left[A_1 + 6A_2 + 8A_3\right]$$

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

The Wald type pivot is $F = \frac{1}{\ell}(\hat{\boldsymbol\beta} - \boldsymbol\beta)'\mathbf{L}(\mathbf{L}'\hat{\mathbf{\Phi}}_A\mathbf{L})^{-1}\mathbf{L}'(\hat{\boldsymbol\beta} - \boldsymbol\beta)$. Similar to section 3.1.1, we derive expressions for $\mathrm{E}[F]$ and $\mathrm{Var}[F]$ 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)

$= \mathbf{I} - \left[\mathbf{I} - (\mathbf{I}+\mathbf{B})^{-1}\mathbf{B}\right]\mathbf{B}$, 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, $(\mathbf{I}+\mathbf{B})^{-1} = \mathbf{I} - \mathbf{B} + (\mathbf{I}+\mathbf{B})^{-1}\mathbf{B}\mathbf{B} = \mathbf{I} - \mathbf{B} + O(n^{-2})$. □

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

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

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

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

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

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

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

Notice that $\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]\right) = \ell + \mathrm{A}_2 + 2\mathrm{A}_3 + O(n^{-3/2})$, (3.12)

(proposition 3.1.1.1.1)

and $(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})$

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

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

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

From expressions (3.12 and 3.13), we obtain

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

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

Proof Direct from Lemmas 2.4.10 and 2.4.11 □

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

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

3.1.2.2 Deriving an Approximate Expression for Var[*F*]

$$\mathrm{Var}[F] = \mathrm{E}[\mathrm{Var}[F \mid \hat{\boldsymbol\sigma}]] + \mathrm{Var}[\mathrm{E}[F \mid \hat{\boldsymbol\sigma}]]$$

A. $\mathrm{Var}[F\,|\,\hat{\boldsymbol{\sigma}}]] = \dfrac{1}{\ell^2}\mathrm{Var}[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}\mathbf{L}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\,|\,\hat{\boldsymbol{\sigma}}]$

The expression above is a quadratic form and hence by assumption (A4),
$\mathrm{Var}[F\,|\,\hat{\boldsymbol{\sigma}}]] = \dfrac{2}{\ell^2}\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2]$ (Schott, 2005, theorem 10.22).

Recall that from expression (3.11), we had

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

and hence,

$$\mathrm{tr}\left([(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2\right) = \mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right]^2$$

$$-2\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})\right] + O_p(n^{-2})$$

$$= \mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\left((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right)(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\left((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right)\right]$$

$$-2\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\left((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right)(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\left((\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right)\right] + O_p(n^{-2})$$

$$= \mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] + 2\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right]$$

$$-2\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right] + O_p(n^{-2}).$$

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

$$\mathrm{tr}\left([(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V}\mathbf{L})]^2\right) =$$

$$\mathrm{tr}\left\{\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\right.$$

$$\left.\times\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\right\}$$

$$+2\mathrm{tr}\left\{\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\right.$$

$$\left.\times\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right]\right\}$$

$$-2\mathrm{tr}\left\{\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\right.$$

$$\times\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i^2}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]$$

$$\times\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})\right.$$

$$\left.\left.+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})\right]\right\}+O_p(n^{-2})$$

$$=\mathrm{tr}\left[\mathbf{I}+2\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]$$

$$+\mathrm{tr}\left[\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]$$

$$+2\mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Lambda}\mathbf{L})\right]-2\mathrm{tr}\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^{*}\mathbf{L})\right]+O_p(n^{-3/2}).$$

By taking the expectation and assumption (A3),

$$\mathrm{E}[\mathrm{Var}[F\,|\,\hat{\boldsymbol{\sigma}}]] = \frac{2}{\ell^2}\left\{\ell+\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left[\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}\right]\right.$$

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

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

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

Hence,

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

In addition, by lemmas 2.4.8 and 2.4.10, we have

$$2\mathrm{tr}[\boldsymbol{\Theta}\boldsymbol{\Lambda}]-2\mathrm{tr}[\boldsymbol{\Theta}\mathbf{A}^{*}]-\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left[\boldsymbol{\Theta}\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})}{\partial\sigma_i\partial\sigma_j}\right]$$

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

$$=2\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left[\boldsymbol{\Theta}\boldsymbol{\Phi}(\mathbf{Q}_{ij}-\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j)\boldsymbol{\Phi}\right]-4\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathrm{tr}\left[\boldsymbol{\Theta}\boldsymbol{\Phi}(\mathbf{Q}_{ij}-\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j-\frac{1}{4}\mathbf{R}_{ij})\boldsymbol{\Phi}\right]$$

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

$$\Rightarrow \mathrm{E}[\mathrm{Var}[F\,|\,\hat{\boldsymbol{\sigma}}]]=\frac{2}{\ell^2}\{\ell+3A_2\}+O(n^{-3/2}) \qquad (3.16)$$

B. $\mathrm{Var}[\mathrm{E}[F\,|\,\hat{\boldsymbol{\sigma}}]]=\dfrac{1}{\ell^2}\mathrm{Var}[\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})]$, by assumption (A4)

$$=\frac{1}{\ell^2}\left\{\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})]\right)^2-\left(\mathrm{E}\left\{\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})]\right\}\right)^2\right\},$$

(*i*) $\left(\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right)^2$

$$=\left\{\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})-(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})+O_p(n^{-2})\right]\right\}^2 \quad \text{(from 3.11)}$$

$$=\left\{\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]\right\}^2-2\,\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]$$
$$+O_p(n^{-2})$$

$$=\left\{\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]\right\}^2-2\,\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]$$
$$+O_p(n^{-2}) \qquad \text{(recall that } \mathbf{V}=\boldsymbol{\Phi}+\boldsymbol{\Lambda}).$$

From expression (3.9), $\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})]\right)^2$

$$=\ell^2+2\ell\,\mathrm{tr}[\boldsymbol{\Theta\Lambda}]+2\ell A_2+A_1-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\,\mathrm{tr}\left[\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right]+O(n^{-3/2}).$$

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

$$\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]=$$

$$\mathrm{tr}\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]$$

$$\times\mathrm{tr}\left\{\left[\mathbf{I}+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})\right]\right.$$

$$\times\left[(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})+\sum_{i=1}^{r}(\hat{\sigma}_i-\sigma_i)\frac{\partial(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})\right.$$

$$\left.\left.+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i-\sigma_i)(\hat{\sigma}_j-\sigma_j)\frac{\partial^2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}}{\partial\sigma_i\partial\sigma_j}(\mathbf{L}'\hat{\mathbf{A}}^*\mathbf{L})\right]\right\}+O_p(n^{-5/2}).$$

By assumption (A3), we obtain

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

From (*i*) and (*ii*),

$$\mathrm{E}\left(\mathrm{tr}[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})]\right)^2=\ell^2+2\ell\,\mathrm{tr}[\boldsymbol{\Theta\Lambda}]+2\ell A_2+A_1$$
$$-\sum_{i=1}^{r}\sum_{j=1}^{r}\ell w_{ij}\,\mathrm{tr}\left[\boldsymbol{\Theta}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}\right]-2\ell\,\mathrm{tr}[\boldsymbol{\Theta}\mathbf{A}^*]+O(n^{-3/2}),$$

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

then

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

Recall that from expression (3.14), we have

$$\mathrm{E}\left(\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]\right)=\ell\mathrm{E}[\mathrm{F}]=\ell+A_2+O(n^{-3/2})$$

$$\Rightarrow\left\{\mathrm{E}\left(\mathrm{tr}\left[(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{VL})\right]\right)\right\}^2=\ell^2+2\ell A_2+O(n^{-3/2}). \qquad (3.18)$$

From expressions (3.17 and 3.18), we obtain

$$\mathrm{Var}[\mathrm{E}[F\,|\,\hat{\boldsymbol{\sigma}}]]=\frac{1}{\ell^2}\left\{\ell^2+2\ell A_2+A_1-\ell^2-2\ell A_2\right\}+O(n^{-3/2})$$
$$=\frac{A_1}{\ell^2}+O(n^{-3/2}), \qquad (3.19)$$

Therefore from expressions (3.16 and 3.19),

Proposition 3.1.2.2.1 $\quad \mathrm{Var}[F]=\dfrac{2}{\ell}\left(1+B\right)+O(n^{-3/2}),$

where $\quad B=\dfrac{1}{2\ell}\left(A_1+6A_2\right).$

Summary

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

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

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

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

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

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

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

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

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

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

Or alternatively,

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

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

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

3.2 Estimating the Denominator Degrees of Freedom and the Scale Factor

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

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

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

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

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

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

By matching the first moments,

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

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

By matching the second moments,

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

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

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 $\hat{\boldsymbol{\sigma}}$ for $\boldsymbol{\sigma}$ 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 $i = 1,..,t,$ and $j = 1,....,m,$

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

Suppose that we are interested in testing $\mathrm{H}_0\!: \tau_1 = ... = \tau_t \Leftrightarrow \mathbf{L}'\boldsymbol{\beta} = \mathbf{0},$

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

$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, 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 $\boldsymbol{\beta}^{*\prime} = \begin{bmatrix} \mu^* & \tau_1^* & \cdots & \tau_{t-1}^* \end{bmatrix}$ with $\sum_{i=1}^{t} \tau_i^* = 0,$ and the hypothesis becomes

$$\mathrm{H}_0\!: \mathbf{L}^{*\prime}\boldsymbol{\beta}^* = \mathbf{0}, \quad \text{where} \quad \mathbf{L}^{*\prime} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

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 $\boldsymbol{\sigma}$

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

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

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

$$\mathbf{P}_1 = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2}\boldsymbol{\Sigma}^{-1}\mathbf{X} = -\frac{1}{(\sigma^2)^2}\mathbf{X}'\mathbf{X},$$

$$\mathbf{Q}_{11} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2}\boldsymbol{\Sigma}^{-1}\mathbf{X} = \frac{1}{(\sigma^2)^3}\mathbf{X}'\mathbf{X},$$

$$\text{and} \qquad \mathbf{R}_{11} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial (\sigma^2)^2}\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{0}, \text{ because } \frac{\partial^2 \boldsymbol{\Sigma}}{\partial (\sigma^2)^2} = \mathbf{0}$$

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

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

4.1.4 Computing A_1 and A_2

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

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

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

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

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 $\Re(\mathbf{L})$

$$\Rightarrow \mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_1\boldsymbol{\Phi}) = -\frac{1}{\sigma^2}\mathrm{tr}\left(\mathbf{L}\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\right)$$

$$= -\frac{1}{\sigma^2}\mathrm{r}\left(\mathbf{L}\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\right) \text{ (Harville, 1997, corollary 10.2.2)}$$

$$= -\frac{1}{\sigma^2}\mathrm{r}(\mathbf{L}) = -\frac{\ell}{\sigma^2} \quad (\ell = t-1).$$

Therefore,

$$A_1 = \frac{2(\sigma^2)^2}{n-t}\frac{\ell^2}{(\sigma^2)^2} = \frac{2\ell^2}{n-t},$$

$$A_2 = \frac{2(\sigma^2)^2}{n-t}\frac{\ell}{(\sigma^2)^2} = \frac{2\ell}{n-t}.$$

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,

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

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

$$\mathrm{r}(\mathbf{P}) = \mathrm{tr}(\mathbf{P}) = \mathrm{tr}\left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)$$

$$= \mathrm{tr}\left(\mathbf{L}\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\right)$$

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 $\Re(\mathbf{L})$, and hence $\mathrm{r}(\mathbf{P}) = \ell$,

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

Since $\Re(\mathbf{P}) \subset \Re(\mathbf{X})$, then under the null hypothesis, $F \sim \mathrm{F}(\ell, n-t)$ exactlly, where $n-t = \mathrm{r}\left(\mathbf{I} - \mathbf{P}_{\mathrm{X}}\right)$ (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.

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

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

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

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

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

$$= \frac{(n-t+2)^2}{n-t} = \frac{(n-t)^2 + 4(n-t) + 4}{n-t} = n-t+4+\frac{4}{n-t}$$

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 $\mathbf{T}^2$ model

4.2.1 The Model and Assumptions

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

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

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

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

Special Notation Since $\boldsymbol{\sigma}=\{\sigma_{if}\}_{\frac{p\times(p+1)}{2}}(i\le f)$, 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, $\hat{\mathbf{\Phi}}_{\mathrm{A}}=\hat{\mathbf{\Phi}}$.

Proof $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}=(\mathbf{1}_n\otimes\mathbf{I}_p)'(\mathbf{I}_p\otimes\mathbf{\Sigma}_p^{-1})(\mathbf{1}_n\otimes\mathbf{I}_p)$

$$=\mathbf{1}_n{}'\mathbf{I}_p\mathbf{1}_n\otimes\mathbf{I}_p\mathbf{\Sigma}_p^{-1}\mathbf{I}_p=n\mathbf{\Sigma}_p^{-1} \quad \text{(Harville,1997, chapter 16).}$$

$$\Rightarrow\mathbf{\Phi}=(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}=\left(n\mathbf{\Sigma}_p^{-1}\right)^{-1}=\frac{\mathbf{\Sigma}_p}{n}.$$

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

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

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

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

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

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

Theorem 4.2.3 The REML estimate of $\mathbf{\Sigma}_p$ is $\mathbf{S}=\dfrac{\mathbf{A}}{n-1}$,

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

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

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

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

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

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

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

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

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

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

In fact, Var[**z**] has the same construction as Var[**y**], but with $n-1$ blocks of $\mathbf{\Sigma}_p$ instead of n blocks. $\mathbf{z}_1,\ldots\ldots,\mathbf{z}_{n-1}$ are iid $\mathrm{N}_p(\mathbf{0},\mathbf{\Sigma}_p)$.

Lemma When $\mathbf{z}_1,\ldots\ldots,\mathbf{z}_{n-1}$ are iid $\mathrm{N}_p(\mathbf{0},\mathbf{\Sigma}_p)$, then $\dfrac{\sum_{i=1}^{n-1}\mathbf{z}_i\mathbf{z}_i'}{n-1}$ is the ML estimator of $\mathbf{\Sigma}_p$.

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

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

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

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

Remark

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

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

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

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

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

So, $\displaystyle \sum_{i=1}^{n-1} \mathbf{z}_i\ \mathbf{z}_i' = \sum_{i=1}^{n-1}\sum_{k=1}^{n}\sum_{j=1}^{n} m_{ji} m_{ki} \mathbf{y}_j \mathbf{y}_k' = \sum_{k=1}^{n}\sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_k' \left(\sum_{i=1}^{n-1} m_{ji} m_{ki}\right)$

$$= \sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' \left(\frac{n-1}{n}\right) + \sum_{\substack{j=1,\\ k\neq j}}^{n} \sum_{k=1}^{n} \mathbf{y}_j \mathbf{y}_k' \left(\frac{-1}{n}\right) \quad \text{(from the remark above)}$$

$$= \sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n}\sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n}\sum_{j=1}^{n} \mathbf{y}_j \sum_{k=1}^{n} \mathbf{y}_k'$$

$$= \sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n}\sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' - \frac{1}{n}\left[n\bar{\mathbf{y}} n\bar{\mathbf{y}}' - \sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j'\right] = \sum_{j=1}^{n} \mathbf{y}_j \mathbf{y}_j' - n\bar{\mathbf{y}}\ \bar{\mathbf{y}}'$$

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

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

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

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

4.2.4 Computing A_1 and A_2

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

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

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

$\mathrm{tr}(\mathbf{P}_{if}\mathbf{\Phi}) = -2^{1-\delta_{if}}\,\sigma^{if}$, and $\mathrm{tr}(\mathbf{P}_{jg}\mathbf{\Phi}) = -2^{1-\delta_{jg}}\,\sigma^{jg}$

$$\Rightarrow \mathrm{tr}(\mathbf{P}_{if}\mathbf{\Phi})\mathrm{tr}(\mathbf{P}_{jg}\mathbf{\Phi}) = 2^{2-\delta_{if}-\delta_{jg}}\,\sigma^{if}\sigma^{jg}, \tag{4.3}$$

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

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

Also, Anderson provided in section 7.3 that the characteristic function of $\mathrm{A}_{11}, \mathrm{A}_{22}, \ldots, \mathrm{A}_{pp}$, $2\mathrm{A}_{12}, 2\mathrm{A}_{13}, \ldots, 2\mathrm{A}_{p\text{-}1,p}$ as

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

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

Finding the moments

For the first moments,

$$\mathrm{E}[\mathrm{A}_{ij}] = \frac{1}{i} \times \left.\frac{\partial \Psi}{\partial \theta_{ij}}\right|_{\mathbf{\Theta}=\mathbf{0}}$$

$$= \frac{1}{i}(-\frac{1}{2}N)\left|\mathbf{I} - 2i\mathbf{\Theta\Sigma}_p\right|^{-\frac{1}{2}N-1}\left|\mathbf{I} - 2i\mathbf{\Theta\Sigma}_p\right| \mathrm{tr}\left[(\mathbf{I} - 2i\mathbf{\Theta\Sigma}_p)^{-1}\frac{\partial}{\partial\theta_{ij}}(\mathbf{I} - 2i\mathbf{\Theta\Sigma}_p)\right]\Bigg|_{\mathbf{\Theta}=\mathbf{0}}$$

$$= N\ \mathrm{tr}\left(\frac{\partial\mathbf{\Theta}}{\partial\theta_{ij}}\mathbf{\Sigma}_p\right) = \begin{cases} (n-1)\sigma_{ii} & \text{for } i = j \\ 2(n-1)\sigma_{ij} & \text{for } i \neq j \end{cases}, \quad \text{and hence } \mathrm{E}[\hat{\sigma}_{ij}] = \sigma_{ij}$$

For the second moments,

$$2^{2-\delta_{if}-\delta_{jg}}\,\mathrm{E}[\mathrm{A}_{if}\mathrm{A}_{jg}] = \frac{1}{i^2} \times \left.\frac{\partial^2 \Psi}{\partial\theta_{if}\partial\theta_{jg}}\right|_{\mathbf{\Theta}=\mathbf{0}}$$

$$=\frac{1}{i^2}\frac{\partial}{\partial\theta_{jg}}\left\{\frac{-N}{2}\left|\mathbf{I}-2i\boldsymbol{\Theta}\boldsymbol{\Sigma}_p\right|^{-N/2}\operatorname{tr}\left[(\mathbf{I}-2i\boldsymbol{\Theta}\boldsymbol{\Sigma}_p)^{-1}\left(-2i\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)\right]\right\}\Bigg|_{\boldsymbol{\Theta}=\mathbf{0}}$$

$$=N^2\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)+2N\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)$$

$$\Rightarrow 2^{2-\delta_{if}-\delta_{jg}}\operatorname{E}[\mathrm{A}_{if}\mathrm{A}_{jg}]=(n-1)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)+2(n-1)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)$$

$$\Rightarrow \operatorname{Cov}[\mathrm{A}_{if},\mathrm{A}_{jg}]=\operatorname{E}[\mathrm{A}_{if}\mathrm{A}_{jg}]-\operatorname{E}[\mathrm{A}_{if}]\operatorname{E}[\mathrm{A}_{jg}]$$

$$=\left(\frac{1}{2^{2-\delta_{if}-\delta_{jg}}}\right)\left\{(n-1)^2\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)+2(n-1)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)\right.$$

$$\left.-(n-1)^2\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\right)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)\right\}$$

$$=\left(\frac{2(n-1)}{2^{2-\delta_{if}-\delta_{jg}}}\right)\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)$$

$$\Rightarrow \operatorname{Cov}[\hat{\sigma}_{if},\hat{\sigma}_{jg}]=\frac{2\operatorname{tr}\left(\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Theta}}{\partial\theta_{jg}}\boldsymbol{\Sigma}_p\right)}{(n-1)2^{2-\delta_{if}-\delta_{jg}}} \tag{4.4}$$

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

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

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

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

$$=\sum_{i=1}^{p}\sum_{f=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\frac{\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij}}{(n-1)2^{2-\delta_{if}-\delta_{jg}}}2^{2-\delta_{if}-\delta_{jg}}\sigma^{if}\sigma^{jg}$$

<u>*Note*</u> 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.

$$\Rightarrow A_1=\frac{1}{n-1}\sum_{g=1}^{p}\sum_{f=1}^{p}\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma_{ig}\sigma^{if}\sigma_{jf}\sigma^{jg}+\sigma_{fg}\sigma^{jg}\sigma_{ij}\sigma^{if}$$

$$=\frac{1}{n-1}\left\{\sum_{f=1}^{p}\sum_{g=1}^{p}\left(\sum_{i=1}^{p}\sigma_{ig}\sigma^{if}\right)\left(\sum_{j=1}^{p}\sigma_{jf}\sigma^{jg}\right)+\sum_{j=1}^{p}\sum_{f=1}^{p}\left(\sum_{g=1}^{p}\sigma_{fg}\sigma^{jg}\right)\left(\sum_{i=1}^{p}\sigma_{ij}\sigma^{if}\right)\right\}$$

$$=\frac{1}{n-1}\left\{\sum_{f=1}^{p}\sum_{g=1}^{p}(\delta_{gf}\delta_{gf})+\sum_{j=1}^{p}\sum_{f=1}^{p}(\delta_{fj}\delta_{fj})\right\}$$

Therefore,

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

Also,

$$\operatorname{tr}(\mathbf{P}_{if}\boldsymbol{\Phi}\mathbf{P}_{jg}\boldsymbol{\Phi})=\operatorname{tr}\left(n\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{jg}}\right).$$

$$=(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij})2^{1-\delta_{if}-\delta_{ij}},\quad \text{by utilizing expresion (4.5) above} \tag{4.7}$$

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

$$A_2=\sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p}\frac{(\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij})}{(n-1)}(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij})2^{1-\delta_{if}-\delta_{ij}}$$

$$=\sum_{i=1}^{p}\sum_{f=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\frac{(\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij})}{(n-1)2^{2-\delta_{if}-\delta_{ij}}}(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij})2^{1-\delta_{if}-\delta_{ij}}$$

$$=\frac{1}{2(n-1)}\left\{\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{ig}\sigma^{ig}\sigma_{jf}\sigma^{jf}+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{ig}\sigma^{fg}\sigma_{jf}\sigma^{ij}\right.$$

$$\left.+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{fg}\sigma^{ig}\sigma_{ij}\sigma^{jf}+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{fg}\sigma^{fg}\sigma_{ij}\sigma^{ij}\right\}$$

$$=\frac{1}{2(n-1)}\left\{\sum_{g=1}^{p}\sum_{f=1}^{p}(1)(1)+\sum_{i=1}^{p}\sum_{f=1}^{p}\delta_{if}\delta_{if}+\sum_{i=1}^{p}\sum_{f=1}^{p}\delta_{if}\delta_{if}+\sum_{g=1}^{p}\sum_{j=1}^{p}(1)(1)\right\}$$

$$=\frac{1}{2(n-1)}(p^2+p+p+p^2)$$

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

4.2.5 Estimating the Denominator Degrees of Freedom and the Scale Factor

When testing $\mathrm{H}_0 : \boldsymbol{\mu} = \mathbf{0}$, 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 $\dfrac{n-p}{n-p+p-1} = \dfrac{n-p}{n-1}$.

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

$$m^* = 4 + \frac{\ell+2}{\ell\dfrac{V^*}{2E^{*2}}-1}, \tag{4.8}$$

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

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
$$E^* = \frac{1}{1-\dfrac{A_2}{\ell}} \tag{4.10}$$

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

$$\frac{1}{E^*\left[\dfrac{\ell(\ell+1)}{A_2}-(\ell+1)\right]} = \frac{A_2}{\ell(\ell+1)}$$

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

and hence
$$E^* = \frac{1}{1-\dfrac{A_2}{\ell}} \tag{4.11}$$

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

$$E^* = \frac{1}{1-\frac{A_2}{\ell}} \tag{4.12}$$

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

$$4+\frac{\ell+2}{\ell\frac{V^*}{2E^{*2}}-1}=\frac{\ell+6}{B} \tag{4.13}$$

$$\Rightarrow \frac{2(\ell+2)E^{*2}}{\ell V^*-2E^{*2}}=\frac{\ell+6}{B}-4,$$

$$V^*=\frac{2}{\ell}\left\{E^{*2}\left(1+\frac{\ell+2}{\frac{\ell+6}{B}-4}\right)\right\}.$$

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

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

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

and hence $V^*=\dfrac{2}{\ell}\left\{\dfrac{\ell+6+(\ell-2)B}{\left(1-\frac{2}{\ell+6}B\right)^2(\ell+6-4B)}\right\}$

$$=\frac{2}{\ell}\left\{\frac{1+\left(\frac{\ell-2}{\ell+6}\right)B}{\left(1-\frac{2}{\ell+6}B\right)^2\left(1-\frac{4B}{\ell+6}\right)}\right\}, \tag{4.14}$$

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

Similarly, applying expression (4.8) on the Hotelling T^2 model,

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

The expression above can be rewritten in B term as

$$4+\frac{\ell+2}{\ell\frac{V^*}{2E^{*2}}-1}=\frac{3\ell+4}{B}+1-\ell \tag{4.15}$$

$$\Rightarrow \frac{2(\ell+2)E^{*2}}{\ell V^*-2E^{*2}}=\frac{3\ell+4}{B}-3-\ell,$$

$$V^*=\frac{2}{\ell}\left\{E^{*2}\left(1+\frac{\ell+2}{\frac{3\ell+4}{B}-3-\ell}\right)\right\}$$

Moreover, Since for the Hotelling T^2model,

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

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

and hence $V^*=\dfrac{2}{\ell}\left\{\dfrac{3\ell+4-B}{\left(1-\frac{\ell+1}{3\ell+4}B\right)^2[3\ell+4-(\ell+3)B]}\right\}.$

$$=\frac{2}{\ell}\left\{\frac{1+\left(\frac{-1}{3\ell+4}\right)B}{\left(1-\frac{\ell+1}{3\ell+4}B\right)^2\left(1-\frac{\ell+3}{3\ell+4}B\right)}\right\}, \tag{4.16}$$

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

$$V^*=\frac{2}{\ell}\left\{\frac{1+d_1B}{(1-d_2B)^2(1-d_3B)}\right\},$$

where

$$d_1 = \frac{\ell-2}{\ell+6},\quad d_2 = \frac{2}{\ell+6},\quad d_3 = \frac{4}{\ell+6} \quad \text{for the balanced one-way Anova model,}$$

$$d_1 = \frac{-1}{3\ell+4},\quad d_2 = \frac{\ell+1}{3\ell+4},\quad d_3 = \frac{\ell+3}{3\ell+4} \quad \text{for Hotelling } \mathrm{T}^2 \text{ model}$$

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

$c(\ell-2)+d = \ell+6$ for the balanced one-way Anova model,

$-c+d = 3\ell+4$ for the Hotelling T^2 model.

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

the denominator of d_i's $= 3\ell + 2(1-$ 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.

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

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

$g = -1$ for the Hotelling T^2 model.

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

then $a\ell + b = \ell - 2,$

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

$$g = \frac{\ell+1}{\ell+2}\frac{\mathrm{A}_1}{\mathrm{A}_2} - \frac{\ell+4}{\ell+2}$$

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

<u>Summary</u>

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

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

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

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

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

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

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

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

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

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

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

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 $\hat{\mathbf{\Phi}}$ in section 3.1.1, we found that

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

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

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, $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, we use the same form of Wald-type F statistic used in K-R method, $F = \dfrac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}(\mathbf{L}'\hat{\mathbf{\Phi}}_A\mathbf{L})^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}$.

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

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

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

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

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 $\mathrm{E}[F]$ to be E^* as was derived in section 4.3; however, instead of constructing a modified estimator for $\mathrm{Var}[F]$, 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,

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

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

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

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

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

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

$$d_1 = -2d_0 + \ell - 2. \tag{5.3}$$

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.

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

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

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

Solving the two equations above,

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

And hence,

$$d_0 = -\frac{\ell(\ell+1)}{\ell+2} - 4 + \frac{\ell+1}{\ell+2}\left(\frac{A_1}{A_2}\right)$$

$$= \frac{(\ell+1)A_1 - (\ell^2 + 5\ell + 8)A_2}{(\ell+2)A_2} \tag{5.4}$$

Summary

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

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

such that

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

where

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

5.2 Second Proposed Modification

In order to approximate the F test for the fixed effects, $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, we use the same form of Wald-type F statistic used in K-R method, $F = \frac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}$.

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

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

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

and

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

And for the Hotelling T^2 model, we had

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

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

and

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

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

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

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

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

$$\zeta_2 = \frac{\ell + m_2 - 2}{m_2 - 4}. \tag{5.5}$$

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 $F \sim \mathrm{F}(\ell, n-t)$ exactly, then by (5.5),

$$\zeta_2 = \frac{\ell+n-t-2}{n-t-4}$$

$$= \frac{1 + \frac{\ell-2}{n-t}}{1 - \frac{4}{n-t}} \qquad \text{(dividing by } n-t\text{)}$$

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

$$\zeta_2 = \frac{n-2}{n-\ell-4}$$

$$= \frac{1 - \frac{1}{n-1}}{1 - \frac{\ell+3}{n-1}} \qquad \text{(dividing by } n-1\text{)}$$

For both models,

$$\zeta_2 = \frac{a_0 + a_1A_1 + a_2A_2}{b_0 + b_1A_1 + b_2A_2},$$

Notice that unlike the first proposed modification where we express h_1 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, $$\zeta_2 = \frac{a_0 + a_1\frac{2\ell^2}{n-t} + a_2\frac{2\ell}{n-t}}{b_0 + b_1\frac{2\ell^2}{n-t} + b_2\frac{2\ell}{n-t}} = \frac{1 + \frac{\ell-2}{n-t}}{1 - \frac{4}{n-t}}$$

$$\Rightarrow a_0 = c, \qquad 2\ell^2 a_1 + 2\ell a_2 = c(\ell-2) \tag{5.6}$$
$$b_0 = c, \qquad 2\ell^2 b_1 + 2\ell b_2 = c(-4) \tag{5.7}$$

for any $c \neq 0$

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

$$\Rightarrow a_0 = c, \qquad 2\ell a_1 + \ell(\ell+1)a_2 = -c \tag{5.8}$$
$$b_0 = c, \qquad 2\ell b_1 + \ell(\ell+1)b_2 = -c(\ell+3) \tag{5.9}$$

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

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

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

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

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

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

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

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 $\lambda_2 = \dfrac{m_2}{E^*(m_2 - 2)}$ 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 $A_1 = \dfrac{2}{\ell+1}A_2$, and the estimate values are $\dfrac{2\ell}{A_2}$, and $\dfrac{\ell(\ell+1)}{A_2} - (\ell-1)$ respectively.

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

K-R approach

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

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

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

then $$V^* = \frac{2}{\ell}\left\{\frac{1 + d_1 B}{(1-d_2B)^2(1-d_3B)}\right\} = \frac{2}{\ell}\left\{\frac{1 + \dfrac{\ell-2}{2\ell}A_2}{(1-A_2/\ell)^2(1-2A_2/\ell)}\right\}.$$

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

Therefore,

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

<u>*First proposed approach*</u>

$$d_0 = \frac{(\ell+1)A_1 - (\ell^2+5\ell+8)A_2}{(\ell+2)A_2} \ .$$

$$m_1 = 4 + d_0 + \frac{\ell - 2 - 2d_0}{B},$$

$$= 4 + d_0\left(1 - \frac{2}{B}\right) + \frac{\ell-2}{B}$$

$$= 4 + \frac{\ell(\ell+1)A_2 - (\ell^2+5\ell+8)A_2}{(\ell+2)A_2}\left(1 - \frac{4\ell}{(6+\ell)A_2}\right) + \frac{2\ell(\ell-2)}{(6+\ell)A_2}$$

$$= 4 + \frac{[\ell(\ell+1)A_2 - (\ell^2+5\ell+8)A_2][(6+\ell)A_2 - 4\ell] + 2\ell(\ell^2-4)A_2}{(6+\ell)(\ell+2)A_2^2}$$

$$= 4 + \frac{2\ell^3 + 16\ell^2 + 24\ell - (4\ell+8)(6+\ell)A_2}{(6+\ell)(\ell+2)A_2}$$

$$= 4 + \frac{2\ell(6+\ell)(2+\ell)}{(6+\ell)(\ell+2)A_2} - \frac{(4\ell+8)(6+\ell)A_2}{(6+\ell)(\ell+2)A_2} = \frac{2\ell}{A_2}$$

<u>*Second proposed approach*</u>

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

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

<u>*K-R approach*</u>

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

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

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

then $V^* = \dfrac{2}{\ell}\left\{\dfrac{1+d_1B}{(1-d_2B)^2(1-d_3B)}\right\}$

$$= \frac{2}{\ell}\left\{\frac{1 - \frac{1}{\ell(\ell+1)}A_2}{(1-A_2/\ell)^2(1 - \frac{\ell+3}{\ell(\ell+1)}A_2)}\right\},$$

$$\rho^* = \frac{V^*}{E^*} = \frac{1}{\ell}\left\{\frac{1 - \frac{1}{\ell(\ell+1)}A_2}{1 - \frac{\ell+3}{\ell(\ell+1)}A_2}\right\} = \frac{1}{\ell}\left\{\frac{\ell(\ell+1) - A_2}{\ell(\ell+1) - (\ell+3)A_2}\right\},$$

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

and therefore

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

<u>*First proposed approach*</u>

$$d_0 = \frac{(\ell+1)A_1 - (\ell^2+5\ell+8)A_2}{(\ell+2)A_2},$$

$$m_1 = 4 + d_0 + \frac{\ell - 2 - 2d_0}{B}$$

$$= 4 + d_0\left(1 - \frac{2}{B}\right) + \frac{\ell-2}{B}$$

$$= 4 + \frac{2A_2 - (\ell^2+5\ell+8)A_2}{(\ell+2)A_2}\left(1 - \frac{2\ell(\ell+1)}{(4+3\ell)A_2}\right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2}$$

$$= 4 + \frac{-(\ell^2+5\ell+6)A_2}{(\ell+2)A_2}\left(\frac{(4+3\ell)A_2 - 2\ell(\ell+1)}{(4+3\ell)A_2}\right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2}$$

$$= 4 - (\ell+3)\left(\frac{(4+3\ell)A_2 - 2\ell(\ell+1)}{(4+3\ell)A_2}\right) + \frac{\ell(\ell+1)(\ell-2)}{(4+3\ell)A_2}$$

$$= 4 + \frac{\ell(\ell+1)(\ell-2) - (\ell+3)[(4+3\ell)A_2 - 2\ell(\ell+1)]}{(4+3\ell)A_2}$$

$$= 4 + \frac{\ell(\ell+1)[2(\ell+3) + \ell - 2]}{(4+3\ell)A_2} - (\ell+3) = 1 - \ell + \frac{\ell(\ell+1)}{A_2}$$

<u>Second proposed approach</u>

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

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 $\mathrm{E}[F]$ 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 $A_1 = \ell A_2$, $m^* = \frac{2\ell}{A_2}$,

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

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

$$= \frac{\frac{(1-\ell)}{\ell(\ell+1)}A_2+1}{\left(\frac{1}{1-A_2/\ell}\right)\left(1-\frac{A_2}{\ell}\right)}, \qquad \text{multiplying by } \frac{A_2}{\ell(\ell+1)}$$

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

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 $A_1 = A_2$.

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

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

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

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

Also,

$$\mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi\Theta\Phi P}_j\mathbf{\Phi}) = \frac{1}{(\mathbf{L'\Phi L})^2}\mathrm{tr}(\mathbf{LL'\Phi P}_i\mathbf{\Phi LL'\Phi P}_j\mathbf{\Phi})$$

$$= \frac{\mathrm{tr}(\mathbf{L'\Phi P}_i\mathbf{\Phi LL'\Phi P}_j\mathbf{\Phi L})}{(\mathbf{L'\Phi L})^2} = \frac{\mathbf{L'\Phi P}_i\mathbf{\Phi LL'\Phi P}_j\mathbf{\Phi L}}{(\mathbf{L'\Phi L})^2}$$

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

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. □

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,

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

where $\mathbf{\Sigma} = \mathrm{Cov}[\mathbf{y}] = \sigma_1^2\mathbf{\Sigma}_1 + \cdots + \sigma_{r-1}^2\mathbf{\Sigma}_{r-1} + \sigma_r^2\mathbf{I}$, and $\mathbf{\Sigma}_i$ are n.n.d. matrices. $\mathbf{X}$ is a known matrix, $\boldsymbol{\beta}$ is a vector of unknown fixed effect parameters, and $\Im$ denotes the set of all possible vectors $\boldsymbol{\sigma}$ and $\aleph = \{\mathbf{\Sigma}(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \Im\}$. We are interested in testing $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, for $\mathbf{L}'$ an $(\ell \times p)$ fixed matrix .

We assume that the model satisfies assumptions A1-A8 where

(A1) $\Im$ contains an r-dimensional rectangle.

(A2) $\mathbf{\Sigma}(\boldsymbol{\sigma})$ is p.d. for all $\boldsymbol{\sigma}$.

(A3) $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are functionally independent.

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

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

(A6) Zyskind's condition for the submodel under the null hypothesis; $\mathbf{\Sigma P_U} = \mathbf{P_U \Sigma}$ for all $\mathbf{\Sigma}$ where $\Re(\mathbf{U}) = \{\mathbf{X}\boldsymbol{\beta} : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}\}$.

(A7) $\Re(\mathbf{M}) \subset \Re(\mathbf{B}_s)$ 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 $\Re(\mathbf{C}_k) \subset \mathfrak{N}(\mathbf{X}')$, $\Re(\mathbf{B}_k) = \Re(\mathbf{E}_k\mathbf{X})$ and $\mathbf{B}_k\mathbf{C}_k = \mathbf{0}$.

(A8) $(\mathbf{I} - \mathbf{P_X})\mathbf{\Sigma}_i$'s are linearly independent.

(A9) $q = r$.

<u>Remarks</u>

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

(ii) $\Re(\mathbf{E}_k) \bigcap \Re(\mathbf{E}_k\mathbf{X})^{\perp} \subset \mathfrak{N}(\mathbf{X}')$

Proof: Take any $\mathbf{x} \in \Re(\mathbf{E}_k) \bigcap \Re(\mathbf{E}_k\mathbf{X})^{\perp}$

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

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

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

(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 $\mathbf{A}$. $\mathrm{r}(\mathbf{L}) = \mathrm{r}(\mathbf{A}) - \dim[\Re(\mathbf{A}) \bigcap \mathfrak{N}(\mathbf{X}')]$, and $\Re(\mathbf{A}) \bigcap \mathfrak{N}(\mathbf{X}') = \Re(\mathbf{A}) \bigcap \Re(\mathbf{X})^{\perp} = \mathbf{0}$, because $\Re(\mathbf{A}) \subset \Re(\mathbf{X})$.

Therefore, $\mathrm{r}(\mathbf{L}) = \mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{M}) = \ell$.

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

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

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

and this combined with assumption (A1) implies $(\mathbf{I} - \mathbf{P_X})\mathbf{\Sigma}_i$'s $\in \mathrm{span}\{\mathbf{C}_1, \ldots, \mathbf{C}_{r-1}\}$ which contradicts with assumption (A8) that $(\mathbf{I} - \mathbf{P_X})\mathbf{\Sigma}_i$'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 $\alpha_k(\boldsymbol{\sigma}) = \alpha_k(\boldsymbol{\sigma}^*)\ \forall k \Rightarrow \sum_{k=1}^{r} \alpha_k(\boldsymbol{\sigma})\mathbf{C}_k = \sum_{k=1}^{r} \alpha_k(\boldsymbol{\sigma}^*)\mathbf{C}_k$

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

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

Hence $\mathbf{B}$ is injective on $\Im$, 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). □

<u>Examples</u>

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, $\hat{\mathbf{\Phi}}_A = \hat{\mathbf{\Phi}}$

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

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

<u>*Note*</u> Since $\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'$ is a p.o. on $\Re(\mathbf{\Sigma}^{-1}\mathbf{X})$ along $\mathfrak{N}(\mathbf{X}')$, $\mathbf{\Sigma}^{-1}$ 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)

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

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

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

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

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

and hence $\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}' = \mathbf{\Sigma}^{-1}\mathbf{P}_\mathbf{X}(\mathbf{P}_\mathbf{X}\mathbf{\Sigma}^{-1}\mathbf{P}_\mathbf{X})^{+}\mathbf{P}_\mathbf{X}$

$$= \mathbf{\Sigma}^{-1}\mathbf{P}_\mathbf{X}\mathbf{\Sigma}\mathbf{P}_\mathbf{X}\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X} \quad \text{(Zyskind's condition)} \tag{6.1}$$

$$\mathbf{P}_\mathbf{X} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{\Sigma} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_j} \mathbf{\Sigma}^{-1}\mathbf{P}_\mathbf{X}$$

$$= \mathbf{P}_\mathbf{X} \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathrm{E}_k \right) \left(\sum_{k=1}^{q} \alpha_{ki} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \alpha_{kj} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \mathbf{P}_\mathbf{X}$$

$$= \sum_{k=1}^{q} \frac{\alpha_{ki}\alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^3} \mathbf{E}_k \mathbf{P}_\mathbf{X} \tag{6.2}$$

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

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

$$= \mathbf{P}_\mathbf{X}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_i} \mathbf{P}_\mathbf{X}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_j} \mathbf{\Sigma}^{-1}\mathbf{P}_\mathbf{X}$$

$$= \mathbf{P}_\mathbf{X} \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \alpha_{ki} \mathbf{E}_k \right) \mathbf{P}_\mathbf{X} \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \alpha_{kj} \mathbf{E}_k \right) \left(\sum_{k=1}^{q} \frac{1}{\alpha_k(\boldsymbol{\sigma})} \mathbf{E}_k \right) \mathbf{P}_\mathbf{X}$$

$$= \sum_{k=1}^{q} \frac{\alpha_{ki}\alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^3} \mathbf{E}_k \mathbf{P}_\mathbf{X} \tag{6.3}$$

From expressions (6.2 and 6.3) we obtain

$$\mathbf{P}_\mathbf{X} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{\Sigma} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{X} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}' \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{P}_\mathbf{X} = \mathbf{0}$$

$$\Rightarrow \mathbf{X}' \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{\Sigma} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{X} - \mathbf{X}' \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}' \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{X} = \mathbf{0}$$

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

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

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

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

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

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

$$\mathbf{\Omega}_\mathrm{H} = \left\{ \mathbf{X}\boldsymbol{\beta} : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0} \right\} = \Re(\mathbf{X}) \cap \Re(\mathbf{A}') = \Re(\mathbf{U}).$$

Also, since $\mathbf{\Sigma}\mathbf{P}_\mathrm{U} = \mathbf{P}_\mathrm{U}\mathbf{\Sigma}$ for all $\mathbf{\Sigma}$ (Zyskind's condition for the submodel),

and $\mathbf{\Sigma}\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}\mathbf{\Sigma}$ for all $\mathbf{\Sigma}$ (Zyskind's condition),

then $\mathbf{\Sigma}\mathbf{M} = \mathbf{M}\mathbf{\Sigma}$ for all $\mathbf{\Sigma}$ where $\mathbf{M} = \mathbf{P}_\mathbf{X} - \mathbf{P}_\mathrm{U}$.

$$\mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi}) = -\mathrm{tr}\left[\mathbf{L}(\mathbf{L}'\mathbf{\Phi L})^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right]$$

$$= -\mathrm{tr}\left[\mathbf{X}'\mathbf{A}(\mathbf{A}'\mathbf{X\Phi X}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right]$$

$$= -\mathrm{tr}\left[\mathbf{P}_\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X\Phi X}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{P}_\mathbf{X}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right] \quad \text{(from 6.1).}$$

$$= -\mathrm{tr}\left[\mathbf{A}(\mathbf{A}'\mathbf{X\Phi X}'\mathbf{A})^{-1}\mathbf{A}'\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right] \quad \text{, because} \Re(\mathbf{A}) \subset \Re(\mathbf{X})$$

$$= -\mathrm{tr}\left[\mathbf{A}(\mathbf{A}'\mathbf{\Sigma A})^{-1}\mathbf{A}'\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\right], \quad \text{again notice that } \mathbf{\Sigma}^{-1}\mathbf{X\Phi X}' = \mathbf{P}_\mathbf{X}$$

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

$$\mathbf{\Omega} \cap \mathbf{\Omega}_\mathrm{H}^\perp = \Re(\mathbf{X}) \cap \left[\Re(\mathbf{X}) \cap \Re(\mathbf{A}')\right]^\perp = \Re(\mathbf{X}) \cap \left[\Re(\mathbf{X})^\perp + \Re(\mathbf{A})\right]$$

$$= \Re(\mathbf{X}) \cap \Re(\mathbf{X})^\perp + \Re(\mathbf{X}) \cap \Re(\mathbf{A}) = \Re(\mathbf{A})$$

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

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

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

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

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

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

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

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

And

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

Since $\mathbf{E}_k\mathbf{M} = \mathbf{M}\mathbf{E}_k$ (Zyskind's condition for the submodel), then $\Re(\mathbf{E}_k\mathbf{M}) = \Re(\mathbf{M}\mathbf{E}_k) \subset \Re(\mathbf{M})$, and hence $\mathbf{E}_k\mathbf{M}$ is an o.p.o (Seely, 2002, problem 2.F.1).
$\Rightarrow \mathrm{tr}(\mathbf{E}_k\mathbf{M}) = \mathrm{r}(\mathbf{E}_k\mathbf{M})$ (Harville, 1997, corollary 10.2.2 and also Seely's notes).

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

When $k = s$, then $\dim[\Re(\mathbf{M}) \cap \Re(\mathbf{E}_s)] = 0$ (assumption A7), and hence $\mathrm{r}(\mathbf{E}_k\mathbf{M}) = \mathrm{r}(\mathbf{M}) = \ell$ (from the remark above).
When $k \neq s$, then $\Re(\mathbf{M}) \subset \Re(\mathbf{E}_s) \Rightarrow \mathbf{E}_k\Re(\mathbf{M}) \subset \mathbf{E}_k\Re(\mathbf{E}_s) \Rightarrow \Re(\mathbf{E}_k\mathbf{M}) \subset \Re(\mathbf{E}_k\mathbf{E}_s) = \mathbf{0}$,

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

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

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

and

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

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

Proof *The Exact Approach*

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

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

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

Since $\mathbf{B} = \{\alpha_{ki}\}_{r\times r}$ is invertible (from the remarks above) which is equivalent to $\mathbf{\alpha} = \{\alpha_k(\mathbf{\sigma})\}_{r\times 1}$ a one to one function.
Now, we reparametrize by making $\gamma_k = \alpha_k(\mathbf{\sigma})$, and hence we have

$$\frac{\partial\ell_R(\mathbf{\gamma})}{\partial\gamma_k} = -\frac{1}{2}\mathrm{tr}\left[(\mathbf{K}'\mathbf{\Sigma K})^{-1}\mathbf{K}'\frac{\partial\mathbf{\Sigma}}{\partial\gamma_k}\mathbf{K}\right] + \frac{1}{2}\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{\Sigma K})^{-1}\mathbf{K}'\frac{\partial\mathbf{\Sigma}}{\partial\gamma_k}\mathbf{K}(\mathbf{K}'\mathbf{\Sigma K})^{-1}\mathbf{K}'\mathbf{y}$$

$$= -\frac{1}{2}\mathrm{tr}\left(\mathbf{G}\frac{\partial\mathbf{\Sigma}}{\partial\gamma_k}\right) + \frac{1}{2}\mathbf{y}'\mathbf{G}\frac{\partial\mathbf{\Sigma}}{\partial\gamma_k}\mathbf{G}\mathbf{y},$$

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

$$\Rightarrow \quad \frac{\partial \ell_R(\boldsymbol{\gamma})}{\partial \gamma_k} = -\frac{1}{2}\operatorname{tr}\left(\mathbf{G}\mathbf{E}_k\right) + \frac{1}{2}\mathbf{y}'\mathbf{G}\mathbf{E}_k\mathbf{G}\mathbf{y}, \quad \text{since } \boldsymbol{\Sigma} = \sum_{k=1}^{r}\gamma_k \mathbf{E}_k$$

$$= -\frac{1}{2}\operatorname{tr}\left(\frac{1}{\gamma_k}\mathbf{C}_k\right) + \frac{1}{2}\mathbf{y}'\frac{\mathbf{C}_k}{\gamma_k^2}\mathbf{y}, \quad \text{since } \mathbf{G} = \boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{P_X}) = \sum_{k=1}^{r}\frac{1}{\gamma_k}\mathbf{C}_k$$

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

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

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 $\mathbf{s}$,

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

And

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

$$\mathbf{C}_i\boldsymbol{\Sigma}\mathbf{C}_i\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{P_X})\mathbf{E}_i\left(\sum_{k=1}^{q}\alpha_k(\boldsymbol{\sigma})\mathbf{E}_k\right)(\mathbf{I} - \mathbf{P_X})\mathbf{E}_i\left(\sum_{k=1}^{q}\alpha_k(\boldsymbol{\sigma})\mathbf{E}_k\right)$$

$$= (\mathbf{I} - \mathbf{P_X})[\alpha_i(\boldsymbol{\sigma})]^2\mathbf{E}_i, \quad \text{because } \mathbf{E}_k\text{'s are orthogonal o.p. matrices}$$

$$= [\alpha_i(\boldsymbol{\sigma})]^2\mathbf{C}_i$$

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

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

__The Approximation Approach__

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

$$\operatorname{tr}\left(\mathbf{G}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\mathbf{G}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\right) = \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{P_X})\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{P_X})\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\right]$$

$$= \operatorname{tr}\left[\sum_{k=1}^{r}\frac{\alpha_{ki}\alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^2}\mathbf{E}_k(\mathbf{I} - \mathbf{P_X})\right] = \sum_{k=1}^{r}\frac{\alpha_{ki}\alpha_{kj}}{[\alpha_k(\boldsymbol{\sigma})]^2}t_k, \quad \text{where } \operatorname{tr}(\mathbf{C}_k) = t_k \geq 1.$$

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.

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

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

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

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 $\mathbf{BWB}' = \mathbf{B}[\mathbf{B}^{-1}\mathbf{D}(\mathbf{B}^{-1})']\mathbf{B}' = \mathbf{D}$,

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

Lemma 6.1.2.6 (Rady, 1986)

Under the null hypothesis, the Anova test statistic $F_R \sim \text{F}(\ell, t_s)$, 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 $F = F_R$.

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

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

(i) $(\mathbf{L}'\hat{\boldsymbol\Phi}\mathbf{L})^{-1} = \left[\mathbf{L}'(\mathbf{X}'\hat{\boldsymbol\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{L}\right]^{-1}$

$= \left[\mathbf{A}'\mathbf{X}(\mathbf{X}'\hat{\boldsymbol\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\right]^{-1}$ (notice that $\mathbf{L} = \mathbf{X}'\mathbf{A}$)

$= \left[\mathbf{A}'\boldsymbol\Sigma\mathbf{P}_{\mathbf{X}}\mathbf{A}\right]^{-1}$, utilizing expression (6.1)

$= \left[\mathbf{A}'\boldsymbol\Sigma\mathbf{A}\right]^{-1}$, because $\Re(\mathbf{A}) \subset \Re(\mathbf{X})$,

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

From (i) and (ii),

$$F = \frac{1}{\ell}\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\left[\mathbf{A}'\boldsymbol\Sigma\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \frac{1}{\ell}\mathbf{y}'\mathbf{P}_{\mathbf{X}}\mathbf{A}\left[\mathbf{A}'\boldsymbol\Sigma\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{P}_{\mathbf{X}}\mathbf{y}$$

$$= \frac{1}{\ell}\mathbf{y}'\mathbf{A}\left[\mathbf{A}'\boldsymbol\Sigma\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{y} \quad \text{(because } \Re(\mathbf{A}) \subset \Re(\mathbf{X})\text{)}$$

$$= \frac{1}{\ell}\mathbf{y}'\hat{\boldsymbol\Sigma}^{-1}\mathbf{M}\mathbf{y} \quad \text{(utilizing expression 6.4),}$$

but $\hat{\boldsymbol\Sigma}^{-1}\mathbf{M} = \sum_{k=1}^{q}\frac{1}{\alpha_k(\hat{\boldsymbol\sigma})}\mathbf{E}_k\mathbf{M} = \frac{1}{\alpha_s(\hat{\boldsymbol\sigma})}\mathbf{M}$ (from the remarks of this section),

where $\alpha_s(\hat{\boldsymbol\sigma}) = \dfrac{\mathbf{y}'\mathbf{C}_s\mathbf{y}}{t_s}$ (from expression 6.8).

Therefore, $\quad F = \dfrac{t_s\,\mathbf{y}'\mathbf{M}\mathbf{y}}{\ell\mathbf{y}'\mathbf{C}_s\mathbf{y}} = F_R \quad \square$

6.1.3 Estimating the Denominator Degrees of Freedom and the Scale Factor

From corollary, 6.1.2.5, $A_1 = \dfrac{2\ell^2}{t_s}$, and $A_2 = \dfrac{2\ell}{t_s}$, and then $\dfrac{A_1}{A_2} = \ell$.

By utilizing theorems (5.3.1 and 5.3.2) and corollary (5.3.3), K-R and our proposed approaches are identical. $m^* = \dfrac{2\ell}{A_2} = t_s$, 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

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U},$$

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 $\mathrm{N}_p(0, \boldsymbol\Sigma_p)$.

$$\mathbf{Y} = \begin{bmatrix}\mathbf{y}_1'\\ \vdots \\ \mathbf{y}_n'\end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix}\mathbf{x}_1'\\ \vdots \\ \mathbf{x}_n'\end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix}\boldsymbol\beta_{(1)}'\\ \vdots \\ \boldsymbol\beta_{(q)}'\end{bmatrix}, \quad \boldsymbol\Sigma = \begin{bmatrix}\boldsymbol\Sigma_p & \cdots & \mathbf{0}\\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol\Sigma_p\end{bmatrix} = \mathbf{I}_n \otimes \boldsymbol\Sigma_p$$

$\mathbf{x}_i$'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: $\mathrm{H}_0 : \mathbf{c}'\mathbf{B}\mathbf{M} = \mathbf{0}$,

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.

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{bmatrix},\qquad \boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_{(1)}\\ \vdots\\ \boldsymbol{\beta}_{(q)}\end{bmatrix}$$

$\mathbf{y}$ 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: $\mathbf{c}'\mathbf{BM}=\mathbf{0}$,

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 $\mathbf{L}=\mathbf{c}\otimes\mathbf{M}$.

So under this setting, $\boldsymbol{\Phi}=\left[(\mathbf{X}\otimes\mathbf{I}_p)'(\mathbf{I}_n\otimes\boldsymbol{\Sigma}_p^{-1})(\mathbf{X}\otimes\mathbf{I}_p)\right]^{-1}$

$$=\left(\mathbf{X}'\mathbf{X}\otimes\boldsymbol{\Sigma}_p^{-1}\right)^{-1}=(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p,$$

$$\boldsymbol{\beta}=\left[(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p\right](\mathbf{X}'\otimes\mathbf{I}_p)((\mathbf{I}_n\otimes\boldsymbol{\Sigma}_p^{-1})\mathbf{y}$$

$$=\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right]\mathbf{y},$$

and $$(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}=\left[(\mathbf{c}'\otimes\mathbf{M}')((\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p)(\mathbf{c}\otimes\mathbf{M})\right]^{-1}=\frac{(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}.$$

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 $F_{\mathrm{M}}=\dfrac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{YM}(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{(n-q)\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$, where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimator

for $\boldsymbol{\Sigma}_p$. Under the null hypothesis, $F_{\mathrm{M}}\ \sim\ [r/(n-q-r+1)]\mathrm{F}_{r,n-q-r+1}$.

Note Similar to lemma 4.2.3, $\hat{\boldsymbol{\Sigma}}_p=\dfrac{\sum_{i=1}^{n}(\mathbf{y}_i-\overline{\mathbf{y}})(\mathbf{y}_i-\overline{\mathbf{y}})'}{n-q}$

Special Notation Since $\boldsymbol{\sigma}=\{\sigma_{if}\}_{\frac{p\times(p+1)}{2}}(i\le f)$, 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 $\hat{\boldsymbol{\Phi}}_{\mathrm{A}}=\hat{\boldsymbol{\Phi}}$.

Proof $\mathbf{P}_{if}=(\mathbf{X}'\otimes\mathbf{I}_p)\left(\mathbf{I}_n\otimes\dfrac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\right)(\mathbf{X}\otimes\mathbf{I}_p)=\mathbf{X}'\mathbf{X}\otimes\dfrac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}$

$$\Rightarrow \mathbf{P}_{if}\boldsymbol{\Phi}\mathbf{P}_{jg}=\left(\mathbf{X}'\mathbf{X}\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\right)\left[(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p\right]\left(\mathbf{X}'\mathbf{X}\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\right)$$

$$=\mathbf{X}'\mathbf{X}\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}},$$

$$\mathbf{Q}_{if,jg}=(\mathbf{X}'\otimes\mathbf{I}_p)\left(\mathbf{I}_n\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\right)(\mathbf{I}_n\otimes\boldsymbol{\Sigma}_p)\left(\mathbf{I}_n\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\right)(\mathbf{X}\otimes\mathbf{I}_p)$$

$$=\mathbf{X}'\mathbf{X}\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}$$

$\Rightarrow \mathbf{Q}_{if,jg}-\mathbf{P}_{if}\boldsymbol{\Phi}\mathbf{P}_{jg}=\mathbf{0}$, and since $\mathbf{R}_{if,jg}=\mathbf{0}$, then we conclude that $\hat{\boldsymbol{\Phi}}_{\mathrm{A}}=\hat{\boldsymbol{\Phi}}$. □

Lemma 6.2.2.3 The K-R test statistic $F=\dfrac{r}{n-q}F_{\mathrm{M}}$.

Proof $F=\dfrac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L})^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}$

$$=\frac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}\quad\text{(lemma 6.2.2.2)}$$

$$=\frac{1}{\ell}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right]\mathbf{y}\right)'(\mathbf{c}\otimes\mathbf{M})\frac{(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M})^{-1}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}(\mathbf{c}'\otimes\mathbf{M}')\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right]\mathbf{y}$$

$$= \frac{1}{r\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\, \mathbf{y}'\left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}(\mathbf{M}'\hat{\mathbf{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'\right\}\mathbf{y}.$$

In the expression above, we used $\ell = r$, and this is because $\dim(\mathbf{L}') = r \times pq$.

In addition,

$\mathbf{y}'\left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{M}(\mathbf{M}'\hat{\mathbf{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'\right\}\mathbf{y} = \mathbf{y}'(\mathbf{d}\mathbf{d}' \otimes \mathbf{A})\mathbf{y} = (\text{vec}\,\mathbf{Y})'(\mathbf{d}\mathbf{d}' \otimes \mathbf{A})\text{vec}\,\mathbf{Y}$,

and $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{M}(\mathbf{M}'\hat{\mathbf{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = \mathbf{d}'\mathbf{Y}\mathbf{A}\mathbf{Y}'\mathbf{d}$,

where $\mathbf{d} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$, $\mathbf{A} = \mathbf{M}(\mathbf{M}'\hat{\mathbf{\Sigma}}_p\mathbf{M})^{-1}\mathbf{M}'$, 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. □

Corollary 6.2.2.4 Under the null hypothesis, $\dfrac{n-q-r+1}{n-q}F \sim \mathrm{F}(r, n-q-r+1)$.

Proof Direct from lemmas 6.2.2.1 and 6.2.2.3.

Lemma 6.2.2.5 $A_1 = \dfrac{2\ell}{n-q}$, and $A_2 = \dfrac{\ell(\ell+1)}{n-q}$.

Proof

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

and
$$A_2 = \sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p} w_{if,jg}\,\mathrm{tr}(\mathbf{\Theta\Phi P}_{if}\mathbf{\Phi\Theta\Phi P}_{jg}\mathbf{\Phi}).$$

$$\mathbf{\Theta} = \mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}' = (\mathbf{c}\otimes\mathbf{M})\frac{(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}(\mathbf{c}'\otimes\mathbf{M}') = \frac{\mathbf{c}\mathbf{c}'\otimes\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

$\mathrm{tr}(\mathbf{\Theta\Phi P}_{if}\mathbf{\Phi})$

$$= \frac{1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\,\mathrm{tr}\left\{\left[\mathbf{c}\mathbf{c}'\otimes\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\right]\left[(\mathbf{X}'\mathbf{X})^{-1}\otimes\mathbf{\Sigma}_p\right]\left[\mathbf{X}'\mathbf{X}\otimes\frac{\partial\mathbf{\Sigma}_p^{-1}}{\partial\sigma_{if}}\right]\left[(\mathbf{X}'\mathbf{X})^{-1}\otimes\mathbf{\Sigma}_p\right]\right\}$$

$$= \frac{1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\,\mathrm{tr}\left[\mathbf{c}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\otimes\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{\Sigma}_p\frac{\partial\mathbf{\Sigma}_p^{-1}}{\partial\sigma_{if}}\mathbf{\Sigma}_p\right]$$

$$= -\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\right]$$

Accordingly, A_1 can be expressed as

$$A_1 = \sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p} w_{if,jg}\,\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\right]\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{jg}}\right],$$

and similarly,

$$A_2 = \sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p} w_{if,jg}\,\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{jg}}\right].$$

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

$$\mathrm{tr}\left(\mathbf{T}\frac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\right) = \begin{cases} t_{ii} & \text{for } i = f \\ 2t_{if} & \text{for } i \ne f \end{cases}$$

Or, $\mathrm{tr}\left(\mathbf{T}\dfrac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\right) = 2^{1-\delta_{if}}t_{if}$, where $\delta_{if} = \begin{cases} 1 & \text{for } i = f \\ 0 & \text{for } i \ne f \end{cases}$ (6.11)

It's easy to check that $\mathrm{tr}\left(\mathbf{T}\dfrac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{if}}\mathbf{T}\dfrac{\partial\mathbf{\Sigma}_p}{\partial\sigma_{jg}}\right) = 2^{1-\delta_{if}-\delta_{jg}}(t_{ig}t_{jf} + t_{jg}t_{ij})$ (6.12)

Moreover, analogous to the proof given for expression (4.6) in the Hotelling T^2 case,

$$w_{if,jg} = \mathrm{Cov}[\hat{\sigma}_{if}, \hat{\sigma}_{jg}] = \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)} \tag{6.13}$$

Utilizing expressions (6.11) and (6.13),

$$A_1 = \sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p} \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)}2^{2-\delta_{if}-\delta_{jg}}t_{if}t_{jg}$$
$$= \sum_{i=1}^{p}\sum_{f=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p} \frac{\sigma_{ig}\sigma_{jf} + \sigma_{fg}\sigma_{ij}}{(n-q)}t_{if}t_{jg}$$

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 $\displaystyle\sum_{f=1}^{p}\sum_{g=1}^{p}\left(\sum_{i=1}^{p}\sigma_{ig}t_{if}\right)\left(\sum_{j=1}^{p}\sigma_{jf}t_{jg}\right)=\mathrm{tr}(\mathbf{T\Sigma}_p\mathbf{T\Sigma}_p)$,

and since $\mathbf{T\Sigma}_p=\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\mathbf{\Sigma}_p$ is o.p. on $\Re(\mathbf{M})$, then $\mathbf{T\Sigma}_p\mathbf{T\Sigma}_p=\mathbf{T\Sigma}_p$,

and $\mathrm{tr}(\mathbf{T\Sigma}_p)=\mathrm{r}(\mathbf{T\Sigma}_p)=\mathrm{r}(\mathbf{M})=r$ (Seely, 2002, chapter 2).

Hence
$$A_1=\frac{1}{n-q}(r+r)=\frac{2r}{n-q}. \tag{6.14}$$

Similarly, and by utilizing expressions (6.12) and (6.13),

$$A_2=\sum_{i=1}^{p}\sum_{\substack{f=1,\\ i\le f}}^{p}\sum_{j=1}^{p}\sum_{\substack{g=1,\\ j\le g}}^{p}\frac{(\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij})}{(n-q)}(t_{ig}t_{jf}+t_{fg}t_{ij})2^{1-\delta_{if}-\delta_{ij}}$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\frac{(\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij})}{2(n-q)}(t_{ig}t_{jf}+t_{fg}t_{ij})$$

$$=\frac{1}{2(n-q)}\left\{\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{ig}t_{ig}\sigma_{jf}t_{jf}+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{ig}t_{fg}\sigma_{jf}t_{ij}\right.$$

$$\left.+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{fg}t_{ig}\sigma_{ij}t_{jf}+\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{fg}t_{fg}\sigma_{ij}t_{ij}\right\}$$

$$=\frac{1}{2(n-q)}\left\{\sum_{g=1}^{p}\sum_{i=1}^{p}\sigma_{ig}t_{ig}\sum_{f=1}^{p}\sum_{j=1}^{p}\sigma_{jf}t_{jf}+\sum_{i=1}^{p}\sum_{f=1}^{p}\left(\sum_{g=1}^{p}\sigma_{ig}t_{fg}\right)\left(\sum_{j=1}^{p}\sigma_{jf}t_{ij}\right)\right.$$

$$\left.+\sum_{i=1}^{p}\sum_{f=1}^{p}\left(\sum_{g=1}^{p}\sigma_{fg}t_{ig}\right)\left(\sum_{j=1}^{p}\sigma_{ij}t_{jf}\right)+\sum_{g=1}^{p}\sum_{f=1}^{p}\sigma_{fg}t_{fg}\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma_{ij}t_{ij}\right\}$$

Since $\displaystyle\sum_{f=1}^{p}\sum_{j=1}^{p}\sigma_{jf}t_{jf}=\mathrm{tr}(\mathbf{T\Sigma}_p)=\mathrm{r}(\mathbf{T\Sigma}_p)=\mathrm{r}(\mathbf{M})=r$

and $\displaystyle\sum_{i=1}^{p}\sum_{f=1}^{p}\left(\sum_{g=1}^{p}\sigma_{fg}t_{ig}\right)\left(\sum_{j=1}^{p}\sigma_{ij}t_{jf}\right)=r$ (from above),

therefore
$$A_2=\frac{1}{2(n-q)}(r^2+r+r+r^2)=\frac{r(r+1)}{n-q}. \tag{6.15}$$

6.2.3 Estimating the Denominator Degrees of Freedom and the Scale

$$A_1=\frac{2r}{n-q},\quad\text{and } A_2=\frac{r(r+1)}{n-q},\quad\text{and then } \frac{A_1}{A_2}=\frac{2}{r+1}\quad\text{(lemma 6.2.2.5).}$$

By utilizing theorems (5.3.1, 5.3.2) and corollary (5.3.3), K-R and the proposed approaches are identical,

where $\displaystyle m^*=\frac{r(r+1)}{\mathrm{A}_2}-(r-1)=n-q-r+1$ (notice that $\ell=r$),

and $\displaystyle\lambda^*=1-\frac{r-1}{r(r+1)}A_2=1-\frac{r-1}{n-q}=\frac{n-q-r+1}{n-q}.$

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 that 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} $(i=1,...,t,$ $j=1,....,m,\ k=1,....,p)$. The experiment is designed so that the j-th subject in each group is associated with q covariates x_{jl} $(l=1,....,q)$, 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}.

<u>*Notation*</u>

$$\mathbf{X}=\begin{bmatrix}\mathbf{x}_1'\\ \vdots\\ \mathbf{x}_m'\end{bmatrix},\quad \mathbf{x}_j=\begin{bmatrix}x_{j1}\\ \vdots\\ x_{jq}\end{bmatrix},\quad \mathbf{Y}=\begin{bmatrix}\mathbf{Y}_1\\ \vdots\\ \mathbf{Y}_t\end{bmatrix},\quad \mathbf{Y}_i=\begin{bmatrix}\mathbf{y}_{i1}'\\ \vdots\\ \mathbf{y}_{im}'\end{bmatrix},\quad \mathbf{y}_{ij}=\begin{bmatrix}y_{ij1}\\ \vdots\\ y_{ijp}\end{bmatrix},$$

$$\mathbf{B}=\begin{bmatrix}\boldsymbol{\beta}^{(1)} & \cdots & \boldsymbol{\beta}^{(p)}\end{bmatrix},\quad \boldsymbol{\beta}^{(k)}=\begin{bmatrix}\beta_{k1}\\ \vdots \\ \beta_{kq}\end{bmatrix},\qquad \mathbf{a}_i=\begin{bmatrix}a_{i1}\\ \vdots \\ a_{ip}\end{bmatrix},\quad \mathbf{e}_{ij}=\begin{bmatrix}e_{ij1}\\ \vdots \\ e_{ijp}\end{bmatrix}$$

Model

$y_{ijk}=\boldsymbol{\beta}^{(k)\prime}\mathbf{x}_j+a_{ik}+e_{ijk}$, 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. $\mathrm{N}_p(\mathbf{0},\boldsymbol{\Lambda})$, the $\mathbf{e}_{ij}$ are i.i.d. $\mathrm{N}_p(\mathbf{0},\boldsymbol{\Sigma}_p)$, and the $\mathbf{a}_i$ are independent of the $\mathbf{e}_{ij}$. We also require $\mathbf{1}_m\in\Re(\mathbf{X})$.

Hypothesis

$\mathrm{H}_0:\mathbf{c}'\mathbf{BM}=\mathbf{0}$, 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.

$$\mathbf{y}=\operatorname{vec}(\mathbf{Y}'),\quad \boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\tau}_1\\ \vdots\\ \boldsymbol{\tau}_q\end{bmatrix}\ \text{ where }\boldsymbol{\tau}_i=\begin{bmatrix}\beta_{1i}\\ \vdots\\ \beta_{pi}\end{bmatrix}$$

$\mathbf{y}$ and $\boldsymbol{\beta}$ are $mtp\times 1$ and $pq\times 1$ vectors respectively.

Notice that $\mathbf{B}=\begin{bmatrix}\boldsymbol{\beta}^{(1)} & \cdots & \boldsymbol{\beta}^{(p)}\end{bmatrix}=\begin{bmatrix}\beta_{11} & \cdots & \beta_{p1}\\ \vdots & \vdots & \vdots\\ \beta_{1q} & \cdots & \beta_{pq}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\tau}_1'\\ \vdots\\ \boldsymbol{\tau}_q'\end{bmatrix}$,

the design matrix for the new setting will be $\mathbf{X}^*=\mathbf{1}_t\otimes(\mathbf{X}\otimes\mathbf{I}_p)$,

and $\qquad \boldsymbol{\Sigma}=\mathbf{I}_t\otimes\boldsymbol{\Delta}$,

where $\qquad \boldsymbol{\Delta}=\mathbf{I}_m\otimes\boldsymbol{\Sigma}_p+m\mathbf{P}_{\mathbf{1}_m}\otimes\boldsymbol{\Lambda}$

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 $\quad \hat{\boldsymbol{\Phi}}_{\mathrm{A}}=\hat{\boldsymbol{\Phi}}$.

Proof $\quad \boldsymbol{\Delta}=\mathbf{I}_m\otimes\boldsymbol{\Sigma}_p+m\mathbf{P}_{\mathbf{1}_m}\otimes\boldsymbol{\Lambda}$

$$=\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right),$$

$$\boldsymbol{\Delta}^{-1}=\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p^{-1}+\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}\tag{6.16}$$

$$\Rightarrow(\mathbf{X}'\otimes\mathbf{I}_p)\boldsymbol{\Delta}^{-1}(\mathbf{X}\otimes\mathbf{I}_p)=(\mathbf{X}'\otimes\mathbf{I}_p)\left\{\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p^{-1}+\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}\right\}(\mathbf{X}\otimes\mathbf{I}_p)$$

$$=\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}\otimes\boldsymbol{\Sigma}_p^{-1}+\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}$$

Hence,

$$\boldsymbol{\Phi}=\left[(\mathbf{1}_t\otimes\mathbf{X}\otimes\mathbf{I}_p)'(\mathbf{I}_t\otimes\boldsymbol{\Delta}^{-1})(\mathbf{1}_t\otimes\mathbf{X}\otimes\mathbf{I}_p)\right]^{-1}$$

$$=\frac{\left[(\mathbf{X}'\otimes\mathbf{I}_p)\boldsymbol{\Delta}^{-1}(\mathbf{X}\otimes\mathbf{I}_p)\right]^{-1}}{t}$$

$$=\frac{\left[\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}\otimes\boldsymbol{\Sigma}_p^{-1}+\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}\right]^{-1}}{t}$$

$$=\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)}{t}\tag{6.17}$$

$$\mathbf{P}_{if}=\mathbf{1}_t\otimes(\mathbf{X}'\otimes\mathbf{I}_p)\left(\mathbf{I}_t\otimes\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}\right)\mathbf{1}_t\otimes(\mathbf{X}\otimes\mathbf{I}_p)$$

$$=t(\mathbf{X}'\otimes\mathbf{I}_p)\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}(\mathbf{X}\otimes\mathbf{I}_p)$$

Hence, $\mathbf{P}_{if}\boldsymbol{\Phi}\mathbf{P}_{jg}=t^2(\mathbf{X}'\otimes\mathbf{I}_p)\dfrac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}(\mathbf{X}\otimes\mathbf{I}_p)\boldsymbol{\Phi}(\mathbf{X}'\otimes\mathbf{I}_p)\dfrac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{jg}}(\mathbf{X}\otimes\mathbf{I}_p)$.

$$\mathbf{Q}_{if,jg}=\mathbf{1}_t'\otimes(\mathbf{X}'\otimes\mathbf{I}_p)\left(\mathbf{I}_t\otimes\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}\right)(\mathbf{I}_t\otimes\boldsymbol{\Delta})\left(\mathbf{I}_t\otimes\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{jg}}\right)\mathbf{1}_t\otimes(\mathbf{X}\otimes\mathbf{I}_p)$$

$$=t(\mathbf{X}'\otimes\mathbf{I}_p)\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}\boldsymbol{\Delta}\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{jg}}(\mathbf{X}\otimes\mathbf{I}_p).$$

$$\Rightarrow\mathbf{Q}_{if,jg}-\mathbf{P}_{if}\boldsymbol{\Phi}\mathbf{P}_{jg}=t(\mathbf{X}'\otimes\mathbf{I}_p)\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{if}}\left\{\boldsymbol{\Delta}-t(\mathbf{X}\otimes\mathbf{I}_p)\boldsymbol{\Phi}(\mathbf{X}'\otimes\mathbf{I}_p)\right\}\frac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{jg}}(\mathbf{X}\otimes\mathbf{I}_p)$$

Consider $t(\mathbf{X}\otimes\mathbf{I}_p)\boldsymbol{\Phi}(\mathbf{X}'\otimes\mathbf{I}_p)$

$$= (\mathbf{X}\otimes\mathbf{I}_p)\Big\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p$$

$$+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\Big\}(\mathbf{X}'\otimes\mathbf{I}_p).$$

$$= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\boldsymbol{\Sigma}_p+\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)$$

$$= \mathbf{P}_{\mathbf{X}}\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{P}_{\mathbf{X}}\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{X}}\mathbf{P}_{\mathbf{1}_m}\mathbf{P}_{\mathbf{X}}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)$$

$$= \left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\quad\text{(observe that } \mathbf{1}_m\in\Re(\mathbf{X})\text{)}$$

$$\Rightarrow \boldsymbol{\Delta}-t(\mathbf{X}\otimes\mathbf{I}_p)\boldsymbol{\Phi}(\mathbf{X}'\otimes\mathbf{I}_p)=\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p-\left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p=\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)\otimes\boldsymbol{\Sigma}_p$$

Moreover, $\left\{\boldsymbol{\Delta}-t(\mathbf{X}\otimes\mathbf{I}_p)\boldsymbol{\Phi}(\mathbf{X}'\otimes\mathbf{I}_p)\right\}\dfrac{\partial\boldsymbol{\Delta}^{-1}}{\partial\sigma_{jg}}(\mathbf{X}\otimes\mathbf{I}_p)$

$$= \left\{(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}})\otimes\boldsymbol{\Sigma}_p\right\}\left\{\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}+\mathbf{P}_{\mathbf{1}_m}\otimes\frac{\partial\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}}{\partial\sigma_{jg}}\right\}(\mathbf{X}\otimes\mathbf{I}_p).$$

$$= (\mathbf{I}_m-\mathbf{P}_{\mathbf{X}})\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}\otimes\boldsymbol{\Sigma}_p\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}+(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}})\mathbf{P}_{\mathbf{1}_m}\mathbf{X}\otimes\boldsymbol{\Sigma}_p\frac{\partial\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}}{\partial\sigma_{jg}}=\mathbf{0}$$

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 $\hat{\boldsymbol{\Phi}}_{\mathrm{A}}=\hat{\boldsymbol{\Phi}}$. □

Lemma 6.3.2.2 (Birkes, 2006)

Let $F_{\mathrm{d}}=\dfrac{(tm-t-q-r+2)t}{r}.\dfrac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{Y}}_{\bullet}\mathbf{M}[\mathbf{M}'\mathbf{Y}'\mathbf{Q}\mathbf{Y}\mathbf{M}]^{-1}\mathbf{M}'\overline{\mathbf{Y}}_{\bullet}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$

where $\overline{\mathbf{Y}}_{\bullet}=\sum_{i=1}^{t}\mathbf{Y}_i/t$ and $\mathbf{Q}=\mathbf{P}_{\mathbf{1}_t}\otimes(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}})+(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t})\otimes(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m})$.

Under H_0, $F_{\mathrm{d}}\sim\mathrm{F}_{r,tm-t-q-r+2}$.

Lemma 6.3.2.3 $\quad F=\dfrac{t}{r}.\dfrac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{Y}}_{\bullet}\mathbf{M}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\mathbf{M}'\overline{\mathbf{Y}}_{\bullet}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$,

where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimate of $\boldsymbol{\Sigma}_p$.

Proof Since $\hat{\boldsymbol{\Phi}}_{\mathrm{A}}=\hat{\boldsymbol{\Phi}}$ (lemma 6.3.2.1), then under the null hypothesis,

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

(*i*) $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\Phi}}\left(\mathbf{1}_t'\otimes\mathbf{X}'\otimes\mathbf{I}_p\right)\left(\mathbf{I}_t\otimes\hat{\boldsymbol{\Delta}}^{-1}\right)\mathbf{y}$

$$=\hat{\boldsymbol{\Phi}}\left\{\mathbf{1}_t'\otimes\left(\mathbf{X}'\otimes\mathbf{I}_p\right)\hat{\boldsymbol{\Delta}}^{-1}\right\}\mathbf{y}=\left\{\mathbf{1}_t'\otimes\hat{\boldsymbol{\Phi}}\left(\mathbf{X}'\otimes\mathbf{I}_p\right)\hat{\boldsymbol{\Delta}}^{-1}\right\}\mathbf{y},$$

From (6.17), $\boldsymbol{\Phi}\left(\mathbf{X}'\otimes\mathbf{I}_p\right)$

$$=\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\right\}\frac{\mathbf{X}'\otimes\mathbf{I}_p}{t}$$

$$=\frac{1}{t}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{P}_{\mathbf{X}}\otimes\boldsymbol{\Sigma}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{P}_{\mathbf{X}}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\right\}$$

$$=\frac{1}{t}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\right\},$$

and from (6.16), we have

$$\boldsymbol{\Phi}\left(\mathbf{X}'\otimes\mathbf{I}_p\right)\boldsymbol{\Delta}^{-1}=\frac{1}{t}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\right\}\times$$

$$\left\{\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\boldsymbol{\Sigma}_p^{-1}+\mathbf{P}_{\mathbf{1}_m}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}\right\}$$

$$=\frac{1}{t}\Big\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\mathbf{I}_p+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{P}_{\mathbf{X}}-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{P}_{\mathbf{1}_m}\otimes\boldsymbol{\Sigma}_p\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}$$

$$+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\boldsymbol{\Sigma}_p^{-1}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\otimes\mathbf{I}_p\Big\}$$

$$=\frac{1}{t}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right\}$$

So, $\quad\hat{\boldsymbol{\beta}}=\dfrac{1}{t}\left\{\mathbf{1}_t'\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right\}\mathbf{y}$,

$$\mathbf{L}'\hat{\boldsymbol{\beta}}=\frac{1}{t}(\mathbf{c}'\otimes\mathbf{M}')\left\{\mathbf{1}_t'\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_p\right\}\mathbf{y}=\frac{1}{t}\left\{\mathbf{1}_t'\otimes\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{M}'\right\}\mathbf{y}$$

(*ii*) $\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}$

$$=\frac{1}{t}(\mathbf{c}'\otimes\mathbf{M}')\Big\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\boldsymbol{\Sigma}_p$$

$$+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\Big\}(\mathbf{c}\otimes\mathbf{M})$$

$$=\frac{1}{t}\Big\{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\otimes\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M}$$

$$+\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\otimes\mathbf{M}'\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)\mathbf{M}\Big\}$$

$$=\frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{t}\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M} \qquad (\text{ notice that } \mathbf{1}_m'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}=0) \tag{6.18}$$

Accordingly, from (*i*) and (*ii*),

$$F=\frac{1}{rt}.\frac{\mathbf{y}'\left\{\mathbf{1}_t\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\otimes\mathbf{M}\right\}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\left\{\mathbf{1}_t'\otimes\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{M}'\right\}\mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$$

$$=\frac{1}{rt}.\frac{\mathbf{y}'\left\{\left[\mathbf{1}_t\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\right]\left[\mathbf{1}_t'\otimes\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\otimes\mathbf{M}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\mathbf{M}'\right\}\mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$$

$$=\frac{1}{rt}.\frac{\mathbf{y}'\left\{\mathbf{d}\mathbf{d}'\otimes\mathbf{A}\right\}\mathbf{y}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\;, \text{ where } \mathbf{d}=\mathbf{1}_t\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}, \text{ and } \mathbf{A}=\mathbf{M}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\mathbf{M}'.$$

Since $\mathbf{y}'\left\{\mathbf{d}\mathbf{d}'\otimes\mathbf{A}\right\}\mathbf{y}=\left(\text{vec}\mathbf{Y}'\right)'\left\{\mathbf{d}\mathbf{d}'\otimes\mathbf{A}\right\}\text{vec}\mathbf{Y}'$

$=\mathbf{d}'\mathbf{Y}\mathbf{A}\mathbf{Y}'\mathbf{d}$ (Harville,1997, theorem 16.2.2),

then $$F=\frac{1}{rt}.\frac{\left\{\mathbf{1}_t'\otimes\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}\mathbf{Y}\mathbf{M}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\mathbf{M}'\mathbf{Y}'\left\{\mathbf{1}_t\otimes\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\right\}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$$

Moreover, since

$$\left\{\mathbf{1}_t'\otimes\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}\mathbf{Y}=\sum_{i=1}^{t}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i=\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sum_{i=1}^{t}\mathbf{Y}_i=t\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{Y}}.$$

then $$F=\frac{t}{r}.\frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{Y}}.\mathbf{M}\left(\mathbf{M}'\hat{\boldsymbol{\Sigma}}_p\mathbf{M}\right)^{-1}\mathbf{M}'\bar{\mathbf{Y}}.'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

where $\hat{\boldsymbol{\Sigma}}_p$ is the REML estimate of $\boldsymbol{\Sigma}_p$. □

Lemma 6.3.2.4 The REML estimator for $\boldsymbol{\Sigma}_p$ is $\dfrac{\mathbf{Y}'\mathbf{Q}\mathbf{Y}}{mt-q-t+1}$,

where $\mathbf{Q}=\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{I}_m-\mathbf{P}_\mathbf{X}\right)+\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)$.

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 $\Re(\mathbf{T}')=\Re(\mathbf{T}'\mathbf{T})\Rightarrow \text{r}(\mathbf{T}')=\text{r}(\mathbf{T}'\mathbf{T})=mt-q$,

and $\mathbf{T}'(\mathbf{1}_t\otimes\mathbf{X})=\mathbf{0}$ (6.19)

Define $\mathbf{K}=\mathbf{T}\otimes\mathbf{I}_p$.

Notice that $\mathbf{K}'(\mathbf{1}_t\otimes\mathbf{X}\otimes\mathbf{I}_p)=(\mathbf{T}'\otimes\mathbf{I}_p)(\mathbf{1}_t\otimes\mathbf{X}\otimes\mathbf{I}_p)=\mathbf{T}'(\mathbf{1}_t\otimes\mathbf{X})\otimes\mathbf{I}_p=\mathbf{0}$

$$\ell_R=\text{Constant}-\frac{1}{2}\log\left|\mathbf{K}'(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\right|-\frac{1}{2}\mathbf{y}'\mathbf{K}\left[\mathbf{K}'(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\right]^{-1}\mathbf{K}'\mathbf{y}.$$

Let $\mathbf{D}=\mathbf{K}'(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}=\mathbf{K}'(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\mathbf{K}'\mathbf{K}$.

$$(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\mathbf{K}'=(\mathbf{I}_t\otimes\boldsymbol{\Delta})\left(\mathbf{I}_{mpt}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\otimes\mathbf{I}_p\right)$$
$$=\mathbf{I}_t\otimes\boldsymbol{\Delta}-\mathbf{P}_{\mathbf{1}_t}\otimes\left[\boldsymbol{\Delta}\left(\mathbf{P}_\mathbf{X}\otimes\mathbf{I}_p\right)\right],$$

where $\mathbf{I}_t\otimes\boldsymbol{\Delta}=\mathbf{I}_t\otimes\left(\mathbf{I}_m\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{1}_m}\otimes m\boldsymbol{\Lambda}\right)$
$=\mathbf{I}_{mt}\otimes\boldsymbol{\Sigma}_p+\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\otimes m\boldsymbol{\Lambda}.$

and $\mathbf{P}_{\mathbf{1}_t}\otimes\left[\boldsymbol{\Delta}\left(\mathbf{P}_\mathbf{X}\otimes\mathbf{I}_p\right)\right]=\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{P}_\mathbf{X}\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{1}_m}\otimes m\boldsymbol{\Lambda}\right)$
$=\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\otimes\boldsymbol{\Sigma}_p+\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\otimes m\boldsymbol{\Lambda}.$

Hence, $(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\mathbf{K}'=\left(\mathbf{I}_{mt}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\right)\otimes\boldsymbol{\Sigma}_p+\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\otimes m\boldsymbol{\Lambda}.$

$$\Rightarrow\mathbf{D}=\mathbf{K}'\left[(\mathbf{I}_t\otimes\boldsymbol{\Delta})\mathbf{K}\mathbf{K}'\right]\mathbf{K}$$
$$=\left(\mathbf{T}'\otimes\mathbf{I}_p\right)\left\{\left(\mathbf{I}_{mt}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\right)\otimes\boldsymbol{\Sigma}_p+\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\otimes m\boldsymbol{\Lambda}\right\}\left(\mathbf{T}\otimes\mathbf{I}_p\right)$$
$$=\mathbf{T}'\mathbf{T}\otimes\boldsymbol{\Sigma}_p-\left[\mathbf{T}'\left(\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\right)\mathbf{T}\right]\otimes\boldsymbol{\Sigma}_p+\left[\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes m\boldsymbol{\Lambda}-\left[\mathbf{T}'\left(\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes m\boldsymbol{\Lambda}.$$

Since from expression (6.19), we have $\mathbf{T}'\left(\mathbf{1}_t\otimes\mathbf{X}\right)=\mathbf{0}$ then, $\left[\mathbf{T}'\left(\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\right)\mathbf{T}\right]\otimes\boldsymbol{\Sigma}_p=\mathbf{0}$.

Also, since $\mathbf{1}_m\in\Re(\mathbf{X})$, then $\left[\mathbf{T}'\left(\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes m\boldsymbol{\Lambda}=\mathbf{0}$.

Hence, $\ell_\text{R}=\text{Constant}-\dfrac{1}{2}\log\left|\mathbf{D}\right|-\dfrac{1}{2}\mathbf{y}'\mathbf{K}\mathbf{D}^{-1}\mathbf{K}'\mathbf{y}$ where

$\mathbf{D}=\mathbf{I}_{mt-q}\otimes\boldsymbol{\Sigma}_p+\left[\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes m\boldsymbol{\Lambda}.$

We may re-express $\mathbf{D}$ as

$$\mathbf{D}=\left[\mathbf{I}_{mt-q}-\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes\boldsymbol{\Sigma}_p+\left[\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right), \tag{6.20}$$
$$\mathbf{D}^{-1}=\left[\mathbf{I}_{mt-q}-\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes\boldsymbol{\Sigma}_p^{-1}+\left[\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right]\otimes\left(\boldsymbol{\Sigma}_p+m\boldsymbol{\Lambda}\right)^{-1}$$

Notice that $\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$ is a p.o., and this is true because

$$\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}=\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\left(\mathbf{I}_{mt}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_\mathbf{X}\right)\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$$
$$=\mathbf{T}'\left\{\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\right\}\mathbf{T}=\mathbf{T}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}-\mathbf{P}_{\mathbf{1}_t}\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$$

$= \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T} \quad \left\{\text{notice that } \mathbf{T}'\left(\mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) = \mathbf{0}\right\}.$

Moreover,

$$
\begin{aligned}
\mathrm{r}\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] &= \mathrm{tr}\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] = \mathrm{trace}\left[\mathbf{T}\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\right] \\
&= \mathrm{tr}\left[\left(\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}}\right)\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\right] \\
&= \mathrm{tr}\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right) - \mathrm{trace}\left(\mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) = t - 1 \qquad (6.21)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{K}\mathbf{D}^{-1}\mathbf{K}' &= \left(\mathbf{T} \otimes \mathbf{I}_p\right)\Big\{\left[\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \mathbf{\Sigma}_p^{-1} \\
&\qquad + \left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}\Big\}\left(\mathbf{T}' \otimes \mathbf{I}_p\right) \\
&= \left[\mathbf{T}\mathbf{T}' - \mathbf{T}\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\mathbf{T}'\right] \otimes \mathbf{\Sigma}_p^{-1} + \left[\mathbf{T}\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\mathbf{T}'\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1},
\end{aligned}
$$

where $\left[\mathbf{T}\mathbf{T}' - \mathbf{T}\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\mathbf{T}'\right] \otimes \mathbf{\Sigma}_p^{-1}$

$$
\begin{aligned}
&= \left[\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}} - \left(\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}}\right)\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\left(\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}}\right)\right] \otimes \mathbf{\Sigma}_p^{-1} \\
&= \left[\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right] \otimes \mathbf{\Sigma}_p^{-1},
\end{aligned}
$$

and $\left[\mathbf{T}\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\mathbf{T}'\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}$

$$
= \left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}
$$

Hence,

$\mathbf{K}\mathbf{D}^{-1}\mathbf{K}'$

$$
= \left[\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} + \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right] \otimes \mathbf{\Sigma}_p^{-1} + \left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}
$$

Notice that $\mathbf{I}_{mt} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}$

$$
\begin{aligned}
&= \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{I}_m - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{X}} + \mathbf{I}_{mt} - \mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} + \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{I}_m \\
&= \mathbf{P}_{\mathbf{1}_t} \otimes \left(\mathbf{I}_m - \mathbf{P}_{\mathbf{X}}\right) + \left(\mathbf{I}_t - \mathbf{P}_{\mathbf{1}_t}\right) \otimes \left(\mathbf{I}_m - \mathbf{P}_{\mathbf{1}_m}\right)
\end{aligned}
$$

$$
\Rightarrow \mathbf{K}\mathbf{D}^{-1}\mathbf{K}' = \mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1} + \left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}, \qquad (6.22)
$$

where $\mathbf{Q} = \mathbf{P}_{\mathbf{1}_t} \otimes \left(\mathbf{I}_m - \mathbf{P}_{\mathbf{X}}\right) + \left(\mathbf{I}_t - \mathbf{P}_{\mathbf{1}_t}\right) \otimes \left(\mathbf{I}_m - \mathbf{P}_{\mathbf{1}_m}\right)$.

Since $\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$, $\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$, $\mathbf{\Sigma}_p$, and $\mathbf{\Sigma}_p + m\mathbf{\Lambda}$ are n.n.d. matrices, so is $\left[\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \mathbf{\Sigma}_p$ and $\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)$, (Harville, 1997, P.369)

Hence, $\left[\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \mathbf{\Sigma}_p = \mathbf{G}\mathbf{G}'$,

and $\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right) = \mathbf{H}\mathbf{H}'$,

for some full column rank matrices $\mathbf{G}$ and $\mathbf{H}$ (Seely, 2002, Problem 2.D.2).

Notice that since

$$
\left\{\left[\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \mathbf{\Sigma}\right\}\left\{\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)\right\} = \mathbf{0},
$$

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, $\left|\left[\mathbf{I}_{mt-q} - \mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \mathbf{\Sigma} + \left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)\right| = \left|\mathbf{G}\mathbf{G}' + \mathbf{H}\mathbf{H}'\right|$

$$
= \left|\left[\mathbf{G}, \mathbf{H}\right]\begin{bmatrix}\mathbf{G}' \\ \mathbf{H}'\end{bmatrix}\right| = \left|\begin{bmatrix}\mathbf{G}' \\ \mathbf{H}'\end{bmatrix}\left[\mathbf{G}, \mathbf{H}\right]\right| = \begin{vmatrix}\mathbf{G}'\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}'\mathbf{H}\end{vmatrix} = \left|\mathbf{G}'\mathbf{G}\right|\left|\mathbf{H}'\mathbf{H}\right|
$$

Recall that $\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}$ is a projection operation, and $\mathrm{r}\left[\mathbf{T}'\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m}\right)\mathbf{T}\right] = t - 1$,

and hence, $\left|\mathbf{G}'\mathbf{G}\right| = \left|\mathbf{\Sigma}_p\right|^{mt-q-t+1}$, and $\left|\mathbf{H}'\mathbf{H}\right| = \left|\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right|^{t-1}$.

In addition, by using expression (6.22), we obtain

$$
\mathbf{y}'\mathbf{K}\mathbf{D}^{-1}\mathbf{K}'\mathbf{y} = \mathbf{y}'\left(\mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1}\right)\mathbf{y} + \mathbf{y}'\left[\left(\mathbf{I}_t \otimes \mathbf{P}_{\mathbf{1}_m} - \mathbf{P}_{\mathbf{1}_t} \otimes \mathbf{P}_{\mathbf{1}_m}\right) \otimes \left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right)^{-1}\right]\mathbf{y}.
$$

Accordingly, $\ell_{\mathrm{R}} = \text{Constant} - \frac{1}{2}\log\left|\mathbf{\Sigma}_p\right|^{mt-q-t+1} - \frac{1}{2}\mathbf{y}'\left(\mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1}\right)\mathbf{y} + f\left(\mathbf{\Sigma}_p + m\mathbf{\Lambda}\right).$

Since $\mathbf{\Sigma}_p$ and $\mathbf{\Lambda}$ are unstructured, then $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_p + m\mathbf{\Lambda}$ may be considered as different random variables. This means that to maximize ℓ_{R} with respect to $\mathbf{\Sigma}_p$, it sufficies to maximize ℓ_R^* with respect to $\mathbf{\Sigma}_p$,

where $\ell_R^* = -\frac{1}{2}(mt - q - t + 1)\log\left|\mathbf{\Sigma}_p\right| - \frac{1}{2}\mathbf{y}'\left(\mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1}\right)\mathbf{y}$.

$$
\begin{aligned}
\mathbf{y}'\left(\mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1}\right)\mathbf{y} &= \left(\mathrm{vec}\mathbf{Y}'\right)'\left(\mathbf{Q} \otimes \mathbf{\Sigma}_p^{-1}\right)\mathrm{vec}\mathbf{Y}' \\
&= \mathrm{tr}\left(\mathbf{Y}\mathbf{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}\right) \qquad \text{(Harville, 1997, theorem 16.2.2).} \\
&= \mathrm{tr}\left(\mathbf{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}\mathbf{Y}\right)
\end{aligned}
$$

So, $\ell_R^* = -\frac{1}{2}(mt - q - t + 1)\log\left|\mathbf{\Sigma}_p\right| - \frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_p^{-1}\mathbf{Y}'\mathbf{Q}\mathbf{Y}\right)$,

and hence $\hat{\mathbf{\Sigma}}_p = \dfrac{\mathbf{Y}'\mathbf{Q}\mathbf{Y}}{mt - q - t + 1}$ (Anderson, 2003, lemma 3.2.2). □

Note We have two expressions for $\mathbf{Q}$:

$$\mathbf{Q}=\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)+\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right),$$

and $\mathbf{Q}=\mathbf{TT}'-\mathbf{TT}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{TT}'$

$\mathbf{Q}$ is a p.o. and $\mathrm{r}(\mathbf{Q})=mt-q-t+1$

Corollary 6.3.2.5 $\dfrac{mt-t-q-r+2}{mt-q-t+1}F=F_{\mathrm{d}}.$

Proof By combining the results of lemmas (6.3.23) and (6.3.2.4), we obtain

$$F=\frac{t(mt-q-t+1)}{r}\cdot\frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{Y}}_{\cdot}\mathbf{M}\left(\mathbf{M}'\mathbf{Y}'\mathbf{QYM}\right)^{-1}\mathbf{M}'\overline{\mathbf{Y}}_{\cdot}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}},$$

and therefore $\dfrac{mt-t-q-r+2}{mt-q-t+1}F=F_{\mathrm{d}}.$ □

Corollary 6.3.2.6 Under the null hypothesis,

$$\frac{mt-t-q-r+2}{mt-q-t+1}F\sim \mathrm{F}(r,tm-t-q-r+2)$$

Proof Direct from lemma 6.3.2.2 and corollary 6.3.2.5. □

Lemma 6.3.2.7 For this model, the approximation and the exact approaches to derive $\mathbf{W}=\mathrm{Var}[\hat{\boldsymbol{\sigma}}_{\mathrm{REML}}]$ are identical.

Proof *Approximation approach*

$$\ell_R^*=-\frac{1}{2}(mt-q-t+1)\log\left|\boldsymbol{\Sigma}_p\right|-\frac{1}{2}\mathbf{y}'\left(\mathbf{Q}\otimes\boldsymbol{\Sigma}_p^{-1}\right)\mathbf{y},$$

$$\frac{\partial\ell_R^*}{\partial\sigma_{if}}=-\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)-\frac{1}{2}\mathbf{y}'\left(\mathbf{Q}\otimes\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}}\right)\mathbf{y},$$

$$\frac{\partial^2\ell_R^*}{\partial\sigma_{if}\partial\sigma_{jg}}=-\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)-\frac{1}{2}\mathbf{y}'\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\mathbf{y}$$

$$\Rightarrow-\mathrm{E}\left[\frac{\partial^2\ell_R^*}{\partial\sigma_{if}\partial\sigma_{jg}}\right]=\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)$$

$$+\frac{1}{2}\left\{\mathrm{E}(\mathbf{y}')\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\mathrm{E}(\mathbf{y})+\mathrm{tr}\left[\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\left(\mathbf{I}_t\otimes\boldsymbol{\Delta}\right)\right]\right\}$$

where $\mathrm{E}(\mathbf{y}')\left(\mathbf{Q}\otimes\dfrac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\mathrm{E}(\mathbf{y})=\boldsymbol{\beta}'\left(\mathbf{1}_t'\otimes\mathbf{X}'\otimes\mathbf{I}_p\right)\left(\mathbf{Q}\otimes\dfrac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\left(\mathbf{1}_t\otimes\mathbf{X}\otimes\mathbf{I}_p\right)\boldsymbol{\beta}$

$$=\boldsymbol{\beta}'\left[\mathbf{1}'\mathbf{P}_{\mathbf{1}_t}\mathbf{1}_t\otimes\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)\mathbf{X}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}+\mathbf{1}'\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\mathbf{1}_t\otimes\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)\mathbf{X}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right]\boldsymbol{\beta}=0$$

$$\Rightarrow-\mathrm{E}\left[\frac{\partial^2\ell_R^*}{\partial\sigma_{if}\partial\sigma_{jg}}\right]=\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)+\frac{1}{2}\mathrm{tr}\left[\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\left(\mathbf{I}_t\otimes\boldsymbol{\Delta}\right)\right].$$

Moreover,

$$\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\left(\mathbf{I}_t\otimes\boldsymbol{\Delta}\right)=$$

$$\left[\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}+\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right]\times$$

$$\left(\mathbf{I}_t\otimes\mathbf{I}_m\otimes\boldsymbol{\Sigma}_p+\mathbf{I}_t\otimes m\mathbf{P}_{\mathbf{1}_m}\otimes\boldsymbol{\Lambda}\right)$$

$$=\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p+\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p$$

$$=\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p,$$

and hence,

$$\mathrm{tr}\left[\left(\mathbf{Q}\otimes\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\right)\left(\mathbf{I}_t\otimes\boldsymbol{\Delta}\right)\right]=\mathrm{tr}(\mathbf{Q})\mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right)=(mt-q-t+1)\mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right).$$

$$\Rightarrow-\mathrm{E}\left[\frac{\partial^2\ell_R^*}{\partial\sigma_{if}\partial\sigma_{jg}}\right]=\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)+\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right).$$

Notice that since $\dfrac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p=-\left(\dfrac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\dfrac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}+\boldsymbol{\Sigma}_p^{-1}\dfrac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\dfrac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right),$

then, $\mathrm{tr}\left(\dfrac{\partial^2\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{if}\partial\sigma_{jg}}\boldsymbol{\Sigma}_p\right)=-2\mathrm{tr}\left(\dfrac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\dfrac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right).$

$$\Rightarrow-\mathrm{E}\left[\frac{\partial^2\ell_R^*}{\partial\sigma_{if}\partial\sigma_{jg}}\right]=-\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}_p^{-1}}{\partial\sigma_{jg}}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)=\frac{1}{2}(mt-q-t+1)\mathrm{tr}\left(\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{jg}}\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)$$

Since from expression (4.7), $\mathrm{tr}\left(\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{jg}}\boldsymbol{\Sigma}_p^{-1}\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right)=2^{1-\delta_{if}-\delta_{jg}}\left(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij}\right)$,

then $\mathbf{W}\approx\frac{2}{mt-q-t+1}\left[\left\{2^{1-\delta_{if}-\delta_{jg}}\left(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij}\right)\right\}_{i,f,j,g=1}^{p}\right]^{-1}$.

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

$\left[\left\{2^{1-\delta_{if}-\delta_{jg}}\left(\sigma^{ig}\sigma^{jf}+\sigma^{fg}\sigma^{ij}\right)\right\}_{i,f,j,g=1}^{p}\right]^{-1}=\frac{1}{2}\left[\left(\sigma_{ig}\sigma_{if}+\sigma_{fg}\sigma_{ij}\right)_{i,f,j,g=1}^{p}\right]$,

and hence,

$$\mathbf{W}\approx\frac{1}{mt-q-t+1}\left[\left(\sigma_{ig}\sigma_{if}+\sigma_{fg}\sigma_{ij}\right)_{i,f,j,g=1}^{p}\right] \tag{6.23}$$

<u>*Exact Approach*</u>

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 $\mathbf{Y}_i=\mathbf{XB}+\mathbf{1}_m\mathbf{a}_i'+\mathbf{E}_i$ where $\mathbf{E}_i=\left[\mathbf{e}_{i1},\ldots,\mathbf{e}_{im}\right]'$

and $\mathbf{Y}=\begin{bmatrix}\mathbf{Y}_1\\ \vdots\\ \mathbf{Y}_t\end{bmatrix}=\mathbf{1}_t\otimes\mathbf{XB}+\mathbf{G}+\mathbf{E}$ where $\mathbf{E}=\begin{bmatrix}\mathbf{E}_1\\ \vdots\\ \mathbf{E}_t\end{bmatrix}$ and $\mathbf{G}=\begin{bmatrix}\mathbf{1}_m\mathbf{a}_1'\\ \vdots\\ \mathbf{1}_m\mathbf{a}_t'\end{bmatrix}$

$\Rightarrow\mathbf{QY}=\mathbf{Q}\left(\mathbf{1}_t\otimes\mathbf{XB}\right)+\mathbf{QG}+\mathbf{QE}$

Observe that $\mathbf{Q}\left(\mathbf{1}_t\otimes\mathbf{XB}\right)=\mathbf{Q}\left(\mathbf{1}_t\otimes\mathbf{X}\right)\left(\mathbf{I}_t\otimes\mathbf{B}\right)$,

but $\mathbf{Q}\left(\mathbf{1}_t\otimes\mathbf{X}\right)=\mathbf{TT}'\left(\mathbf{1}_t\otimes\mathbf{X}\right)-\mathbf{TT}'\left(\mathbf{I}_t\otimes\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{TT}'\left(\mathbf{1}_t\otimes\mathbf{X}\right)=\mathbf{0}$,

and hence, $\mathbf{Q}\left(\mathbf{1}_t\otimes\mathbf{XB}\right)=\mathbf{0}$.

Also, $\mathbf{QG}=\left[\mathbf{P}_{\mathbf{1}_t}\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{X}}\right)+\left(\mathbf{I}_t-\mathbf{P}_{\mathbf{1}_t}\right)\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\right]\mathbf{G}$

$=\mathbf{Q}\left[\mathbf{I}_t\otimes\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\right]\mathbf{G}=\mathbf{0}$

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 $\mathrm{N}_p(0,\boldsymbol{\Sigma}_p)$, and $\mathbf{Q}$ is idempotent with $\mathrm{r}(\mathbf{Q})=mt-q-t+1$,

then $\mathbf{E}'\mathbf{QE}=\mathbf{Y}'\mathbf{QY}\sim\mathrm{W}_p(\boldsymbol{\Sigma}_p,mt-q-t+1)$ (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

$$w_{if,jg}=\frac{\sigma_{ig}\sigma_{jf}+\sigma_{fg}\sigma_{ij}}{mt-q-t+1} \tag{6.24}$$

From expressions (6.23) and (6.24), we can see that the approximation and the exact methods give the same estimate for $w_{if,jg}$. □

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 $\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}=\frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{t}\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M}$.

$$\Rightarrow\boldsymbol{\Theta}=\mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}'=\frac{t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}(\mathbf{c}\otimes\mathbf{M})(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}(\mathbf{c}'\otimes\mathbf{M}')$$
$$=\frac{t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\mathbf{cc}'\otimes\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}',$$

and from expression (6.17), we obtain

$$\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_{if}}=-\boldsymbol{\Phi}\mathbf{P}_{if}\boldsymbol{\Phi}=-\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}{t}\otimes\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}.$$

Hence, $\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_{if}\boldsymbol{\Phi})$

$$=\frac{-t}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\mathrm{tr}\left\{\left[\mathbf{cc}'\otimes\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\right]\left[\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}{t}\otimes\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right]\right\}$$

$$=\frac{-1}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\mathrm{tr}\left[\mathbf{cc}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\otimes\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right]$$

$$=\frac{-\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_m-\mathbf{P}_{\mathbf{1}_m}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right]$$

$$=-\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\boldsymbol{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial\boldsymbol{\Sigma}_p}{\partial\sigma_{if}}\right]. \qquad \text{(notice that } \mathbf{P}_{\mathbf{1}_m}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}=\mathbf{0}\text{)}$$

Accordingly, A_1 and A_2 can be expressed as

$$A_1 = \sum_{g=1}^{p}\sum_{f=1}^{p}\sum_{i=1}^{p}\sum_{j=1}^{p} w_{if,jg}\,\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}}\right]\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}}\right],$$

$$A_2 = \sum_{g=1}^{p}\sum_{f=1}^{p}\sum_{i=1}^{p}\sum_{j=1}^{p} w_{if,jg}\,\mathrm{tr}\left[\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{if}}\mathbf{M}(\mathbf{M}'\mathbf{\Sigma}_p\mathbf{M})^{-1}\mathbf{M}'\frac{\partial \mathbf{\Sigma}_p}{\partial \sigma_{jg}}\right].$$

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

$$A_1 = \frac{2r}{mt-q-t+1}, \quad \text{and } A_2 = \frac{r(r+1)}{mt-q-t+1}, \quad \text{and then } \frac{A_1}{A_2} = \frac{2}{r+1} \quad \text{(lemma 6.3.2.8).}$$

By utilizing theorems (5.3.1 and 5.3.2) and corollary (5.3.3), K-R and the proposed approaches are identical,

where $$m^* = \frac{r(r+1)}{\mathrm{A}_2} - (r-1) = mt-q-t-r+2 \ ,$$

and $$\lambda^* = 1 - \frac{r-1}{r(r+1)}A_2 = 1 - \frac{r-1}{mt-q-t+1} = \frac{mt-q-t-r+2}{mt-q-t+1}.$$

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, $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ in a model as described in section 2.1, the Satterthwaite method uses the Wald-type statistic

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

(i) The multi-dimensional case($\ell > 1$)

First, we perform the spectral decomposition; $\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L} = \mathbf{P}'\mathbf{D}\mathbf{P}$ where $\mathbf{P}$ is an orthogonal matrix of eigenvectors, and $\mathbf{D}$ is a diagonal matrix of eigenvalues.

Let $\nu_m = \dfrac{2(d_m)^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m}$ where

d_m is the mth diagonal element of $\mathbf{D}$, and $\mathbf{g}_m$ is the gradient of $\mathbf{a}_m\mathbf{\Phi}\mathbf{a}_m'$ with respect to $\boldsymbol{\sigma}$ evaluated at $\hat{\boldsymbol{\sigma}}$, where $\mathbf{a}_m$ is the mth row of $\mathbf{PL}'$.

Notice that

$$\mathbf{g}_m = \left[\ldots \mathbf{a}_m \frac{\partial \mathbf{\Phi}}{\partial \sigma_i^2}\mathbf{a}_m' \cdots\right]\Bigg|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}} = -\left[\cdots \mathbf{a}_m\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}\mathbf{a}_m' \cdots\right]\Big|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}}$$

Then let

$$\mathrm{E} = \sum_{m=1}^{\ell}\frac{\nu_m}{\nu_m - 2}I(\nu_m > 2)$$

so we eliminate the terms for which $\nu_m \le 2$.

The degrees of freedom for F are then computed as

$$\nu = \begin{cases} \dfrac{2\mathrm{E}}{\mathrm{E}-\ell} & \text{for } \mathrm{E}>\ell \\ 0 & \text{otherwise} \end{cases}$$

(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).

$$A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}), \qquad \boldsymbol{\Theta} = \frac{\mathbf{L}\mathbf{L}'}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}\left(\frac{\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}}\right)\mathrm{tr}\left(\frac{\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}}{\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}}\right)$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\frac{\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}\mathbf{L}}{(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2}$$

$$\Rightarrow \frac{2}{A_1} = \frac{2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2}{\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}\mathbf{L}} = \frac{2(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^2}{\mathbf{g}'\mathbf{W}\mathbf{g}} \quad \text{where } \mathbf{g} = [\mathbf{L}'\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{L}]_{r\times 1}$$

By noticing that $\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_i} = -\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}$, the proof is completed. □

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 $\hat{\boldsymbol{\Phi}}$ as the estimator of the variance-covariance matrix of the fixed effects estimator, whereas the K-R and the proposed methods statistic uses $\hat{\boldsymbol{\Phi}}_A$. The following corollary clarifies this fact.

Corollary 7.2.2 When $\ell = 1$, and $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$, 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 $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$, then we have the same statistics for all approaches. This shows that the four approaches are identical. □

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 form corollaries 5.3.5, 6.1.1.2, and 7.2.2. □

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 $\mathbf{L}'\hat{\boldsymbol{\Phi}}_A\mathbf{L} \geq \mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L}$.

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

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

Notice that

$$\mathbf{Q}_{ij} - \mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j = \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_j}\mathbf{\Sigma}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_j}\mathbf{\Sigma}^{-1}\mathbf{X}$$

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

and hence,

$$\mathbf{\Phi}_A - \mathbf{\Phi} = 2\mathbf{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\mathbf{B}_i\mathbf{G}\mathbf{B}_j'\right\}\mathbf{\Phi} \qquad \text{where } \mathbf{B}_i = \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\sigma_i}$$

$$= 2\mathbf{\Phi}\mathbf{B}\mathbf{M}\mathbf{B}'\mathbf{\Phi}, \qquad \text{where } \mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_r] \text{ and } \mathbf{M} = \mathbf{W}\otimes\mathbf{G}$$

Since $\mathbf{W}$ and $\mathbf{G}$ are both n.n.d, so is $\mathbf{M}$ (Harville, 1997, P.367), and hence so is $\mathbf{BMB}'$. So $\hat{\mathbf{\Phi}}_A - \hat{\mathbf{\Phi}}$ is n.n.d., and hence $\mathbf{L}'\left(\hat{\mathbf{\Phi}}_A - \hat{\mathbf{\Phi}}\right)\mathbf{L} \geq 0$. □

__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 $\mathbf{V}$

$$\Rightarrow \mathbf{A}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{C}\left(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} + \mathbf{I}\right)^{-1}\mathbf{C}'\mathbf{B}^{-1} \text{(theorem 1.7, Schott, 2005)}$$

$$\Rightarrow \mathbf{B}^{-1} - \mathbf{A}^{-1} = \mathbf{B}^{-1}\mathbf{C}\left(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} + \mathbf{I}\right)^{-1}\mathbf{C}'\mathbf{B}^{-1} \text{ which is n.n.d, and hence } \mathbf{B}^{-1} \geq \mathbf{A}^{-1}. \quad □$$

__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

$$\mathbf{L}'\hat{\mathbf{\Phi}}_A\mathbf{L} \geq \mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L} \Rightarrow \left(\mathbf{L}'\hat{\mathbf{\Phi}}_A\mathbf{L}\right)^{-1} \leq \left(\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L}\right)^{-1}, \text{ and hence } F \leq F_S. \quad □$$

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 $p = p_1 + p_2$.

$$\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_2 \end{bmatrix}$$

Let $\mathbf{X}_2$ to be correspondent to the fixed effects that of our interest to be tested.

Consider the case where

(a) $\mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_2 = \mathbf{0}$.

(b) $\mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_2 = f(\boldsymbol{\sigma})\mathbf{A}$ 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, $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}$ where $\mathbf{C}_2 = f(\boldsymbol{\sigma})\mathbf{A}$

$$\Rightarrow \mathbf{\Phi} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} = \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix}$$

Also, since $\mathbf{L}' = \begin{bmatrix} \mathbf{0}_{\ell\times p_1} & \mathbf{B}'_{\ell\times p_2} \end{bmatrix}$, then

$$\mathbf{L}'\mathbf{\Phi}\mathbf{L} = \begin{bmatrix} \mathbf{0} & \mathbf{B}' \end{bmatrix}\begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathbf{B} \end{bmatrix} = \mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B},$$

$$\mathbf{\Theta} = \mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}' = \begin{bmatrix} \mathbf{0}_{p_1\times\ell} \\ \mathbf{B}_{p_2\times\ell} \end{bmatrix}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\begin{bmatrix} \mathbf{0}_{\ell\times p_1} & \mathbf{B}'_{\ell\times p_2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}' \end{bmatrix},$$

$$\mathbf{\Phi\Theta\Phi} = \begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}' \end{bmatrix}\begin{bmatrix} \mathbf{C}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1} \end{bmatrix}$$

Moreover, since $\mathbf{P}_i = \mathbf{X}'\dfrac{\partial\mathbf{\Sigma}^{-1}}{\partial\sigma_i}\mathbf{X} = \begin{bmatrix} \dfrac{\partial\mathbf{C}_1}{\partial\sigma_i} & \mathbf{0} \\ \mathbf{0} & \dfrac{\partial\mathbf{C}_2}{\partial\sigma_i} \end{bmatrix}$ then

$$\mathbf{\Phi\Theta\Phi P}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1} \end{bmatrix}\begin{bmatrix} \dfrac{\partial\mathbf{C}_1}{\partial\sigma_i} & \mathbf{0} \\ \mathbf{0} & \dfrac{\partial\mathbf{C}_2}{\partial\sigma_i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\dfrac{\partial\mathbf{C}_2}{\partial\sigma_i} \end{bmatrix}$$

$$\text{So, } \mathrm{tr}(\mathbf{\Theta\Phi P}_i\mathbf{\Phi}) = \mathrm{tr}\left(\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\frac{\partial\mathbf{C}_2}{\partial\sigma_i}\right) \tag{7.1}$$

Since $\mathbf{C}_2 = f(\boldsymbol{\sigma})\mathbf{A}$, then $\mathbf{C}_2^{-1}\dfrac{\partial\mathbf{C}_2}{\partial\sigma_i} = \dfrac{\mathbf{A}^{-1}}{f(\boldsymbol{\sigma})}\dfrac{\partial f(\boldsymbol{\sigma})}{\partial\sigma_i}\mathbf{A} = \dfrac{1}{f(\boldsymbol{\sigma})}\dfrac{\partial f(\boldsymbol{\sigma})}{\partial\sigma_i}\mathbf{I}$

So, expression (7.1) above can be simplified as

$$\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi}) = \frac{1}{f(\boldsymbol{\sigma})}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\operatorname{tr}\left(\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\right)$$

$$= \frac{1}{f(\boldsymbol{\sigma})}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\operatorname{tr}\left(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\right) = \frac{\ell}{f(\boldsymbol{\sigma})}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i} \qquad (7.2)$$

$$\Rightarrow \operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}) = \frac{\ell^2}{[f(\boldsymbol{\sigma})]^2}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \qquad (7.3)$$

Also,

$$\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\dfrac{\partial \mathbf{C}_2}{\partial \sigma_i}\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\dfrac{\partial \mathbf{C}_2}{\partial \sigma_j}\end{bmatrix}$$

$$= \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\dfrac{\partial \mathbf{C}_2}{\partial \sigma_i}\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\dfrac{\partial \mathbf{C}_2}{\partial \sigma_j}\end{bmatrix}$$

So,

$$\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}) = \operatorname{tr}\left(\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\frac{\partial \mathbf{C}_2}{\partial \sigma_i}\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\frac{\partial \mathbf{C}_2}{\partial \sigma_j}\right)$$

$$= \frac{1}{[f(\boldsymbol{\sigma})]^2}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j}\operatorname{tr}\left(\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\right)$$

$$= \frac{1}{[f(\boldsymbol{\sigma})]^2}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j}\operatorname{tr}\left(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}(\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B})^{-1}\right)$$

$$= \frac{\ell}{[f(\boldsymbol{\sigma})]^2}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_j} \qquad (7.4)$$

From expressions 7.3 and 7.4, we have

$$\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}) = \ell\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}),$$

and hence $\qquad A_1 = \ell A_2 \quad \square$

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 $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$, 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 $\dfrac{2\ell}{A_2}$ as the estimates of the scale and the denominator degrees of freedom respectively (theorems 5.3.1 and 5.3.2).

(*ii*) <u>*Claim*</u> $\dfrac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} = \dfrac{2\ell}{A_2}$ for any *m*.

<u>*Proof*</u> $A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}) = \mathbf{a}'\mathbf{W}\mathbf{a}$ where $\mathbf{a} = [\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})]_{r\times 1}$,

Since form expression (7.2), $\operatorname{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi}) = \dfrac{\ell}{f(\boldsymbol{\sigma})}\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}$, then $\mathbf{a} = \left[\dfrac{\ell}{f(\boldsymbol{\sigma})}\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}$

And hence

the R.H.S $= \dfrac{2\ell}{A_2} = \dfrac{2\ell^2}{A_1} = \dfrac{2\ell^2}{\mathbf{a}'\mathbf{W}\mathbf{a}} = \dfrac{2[f(\boldsymbol{\sigma})]^2}{\mathbf{g}'\mathbf{W}\mathbf{g}}$ where $\mathbf{g} = \left[\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}$.

Since $\mathbf{g}_m = \left[\mathbf{a}_m\dfrac{\partial \boldsymbol{\Phi}}{\partial \sigma_i}\mathbf{a}_m'\right]_{r\times 1}$ where $\mathbf{a}_m$ is the *m*th row of $\mathbf{PL}'$, then $\mathbf{g}_m = \left[\mathbf{b}_m\mathbf{L}'\dfrac{\partial \boldsymbol{\Phi}}{\partial \sigma_i}\mathbf{L}\mathbf{b}_m'\right]_{r\times 1}$

where $\mathbf{b}_m$ is the *m*th row of $\mathbf{P}$.

$$\Rightarrow \mathbf{g}_m = \left[\mathbf{b}_m\mathbf{B}'\frac{\partial \mathbf{C}_2^{-1}}{\partial \sigma_i}\mathbf{B}\mathbf{b}_m'\right]_{r\times 1} = \left[\frac{-1}{[f(\boldsymbol{\sigma})]^2}\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\mathbf{b}_m\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_m'\right]_{r\times 1} = -\frac{\mathbf{b}_m\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}\mathbf{b}_m'}{f(\boldsymbol{\sigma})}\left[\frac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}$$

$$\text{And the L.H.S} = \frac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} = \frac{2d_m^2}{\left[\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}'\mathbf{W}\left[\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}}\left(\frac{f(\boldsymbol{\sigma})}{\mathbf{b}_m\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}\mathbf{b}_m'}\right)^2$$

Notice that $\mathbf{b}_m\mathbf{B}'\mathbf{C}_2^{-1}\mathbf{B}\mathbf{b}_m' = \mathbf{b}_m\mathbf{L}'\boldsymbol{\Phi}\mathbf{L}\mathbf{b}_m' = d_m$.

$$\Rightarrow \text{the L.H.S} = \frac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} = \frac{2[f(\boldsymbol{\sigma})]^2}{\left[\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}'\mathbf{W}\left[\dfrac{\partial f(\boldsymbol{\sigma})}{\partial \sigma_i}\right]_{r\times 1}} = \text{the R.H.S}$$

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 $\nu_m = \dfrac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} = \dfrac{2\ell}{\mathrm{A}_2} > 2$ for any m,

then $\mathrm{E} = \dfrac{2\ell d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} \Bigg/ \left(\dfrac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} - 2\right)$, and $\nu = \dfrac{2\mathrm{E}}{\mathrm{E}-\ell} = \dfrac{2d_m^2}{\mathbf{g}_m'\mathbf{W}\mathbf{g}_m} =$ the L.H.S.

(iii) Direct from parts (ii) and (iii). □

<u>*Comments*</u>

(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 $\hat{\mathbf{\Phi}}$ as an estimator of the variance-covariance matrix of the fixed effects estimator, whereas the K-R and the proposed methods statistic uses $\hat{\mathbf{\Phi}}_{\mathrm{A}}$.

(v) Conditions (a) and (b) are not enough to make $\hat{\mathbf{\Phi}} = \hat{\mathbf{\Phi}}_{\mathrm{A}}$. For example, in BIB designs, conditions (a) and (b) are satisfied; however, $\hat{\mathbf{\Phi}} \neq \hat{\mathbf{\Phi}}_{\mathrm{A}}$ (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 $i = 1,\ldots\ldots,t,\quad j = 1,\ldots\ldots,s,$ where μ is the general mean, α_i is the treatments effects, b_j is the blocks effects, $b_j \sim \mathrm{N}(0,\sigma_b^2)$, $e_{ij} \sim \mathrm{N}(0,\sigma_e^2)$, 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 $\mathbf{b} = [b_1,\cdots b_s]'$, $\mathbf{a} = [\alpha_1,\cdots,\alpha_t]'$,

$\mathrm{E}(\mathbf{y}) = \mathbf{1}_n\mu + \mathbf{A}\mathbf{a}$, and $\mathrm{Var}(\mathbf{y}) = \mathbf{\Sigma} = \sigma_e^2\mathbf{I}_n + \sigma_b^2\mathbf{B}\mathbf{B}' = \sum_{i=1}^{2}\sigma_i^2\mathbf{D}_i$ where $\mathbf{D}_1 = \mathbf{I}_n$ and $\mathbf{D}_2 = \mathbf{B}\mathbf{B}'$

Suppose that we are interested in testing: H: $\alpha_1 = \cdots = \alpha_t \Leftrightarrow \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$