Analysis of a mixture of fixed and random effects in a mixed model
by Miguel A Paz-Cuentas

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics
Montana State University

Abstract:
Mixed models arise in many experimental situations. In these experiments it is often of interest to test hypotheses about the fixed effects and to perform tests of significance and estimate variance components. Except in certain special cases, exact tests are not available for random effects in unbalanced mixed models. Ordinarily, researchers apply techniques developed for balanced mixed models or treat unknown variance components as known.

The design matrix for the particular mixed model of interest is obtained by concatenation of the design matrices for oneway fixed effects model and sets of oneway random effects models. The fixed and random components share a common variance. Standard methods of parameter estimation, such as, Minimum Norm Quadratic Unbiased Estimation (MINQUE), Maximum Likelihood (ML) estimation, and Restricted Maximum Likelihood (REML) estimation, were used. To solve the ML and REML equations, a Fisher scoring algorithm was used. Because of the complexity of the original equations, it was necessary to do some simplifications of the equations to gain efficiency in the computations.

The main objective is to develop procedures for making inferences about the means from the set of fixed effects and the expectations from the sets of random effects. From the simulation results, the F test with REML estimates appears to be promising for hypothesis testing under the particular mixed model of interest.

ANALYSIS OF A MIXTURE OF FIXED AND RANDOM EFFECTS IN A MIXED MODEL

by

MIGUEL A. PAZ-CUENTAS

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

January 1996

APPROVAL

of a thesis submitted by

MIGUEL A. PAZ-CUENTAS

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

2/2/96 — Robert J. Boik
Date — Robert J.Boik
Chairperson, Graduate Committee

Approved for the Major Department

2/2/96 — John Lund
Date — John Lund
Head, Mathematics

Approved for the College of Graduate Studies

4/1/96 — R. L. Brown
Date — Robert Brown
Graduate Dean

STATEMENT OF PERMISSION TO USE

ACKNOWLEDGEMENTS

I would like to thank my wife, Alicia, and my daughters Alicia, Mariceli and Evelyn, without whose help and support this thesis would never have been possible.

I would like to express my sincerest thanks to Professor Robert J. Boik for his invaluable assistance and guidance in the preparation of this dissertation. I am especially grateful for his patience and for his understanding of difficulties which were beyond anyone's control.

I would also like to thank Professors Pat Munholland and John Borkowski, who read the manuscript and made helpful suggestions.

I would like to thank Peter Burfening, Richard Lund, Jeff Banfield, and Rick Rossi for their help.

I would like to dedicate this to my mother Maria, my aunt Sofia, and my uncle Ricardo, who I promised a copy. Unless they are looking over my shoulder from some other dimension they will never be able to read my thesis.

Maria Cuentas

1917-1987

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Mixed models arise in many experimental situations. In these experiments it is often of interest to test hypotheses about the fixed effects and to perform tests of significance and estimate variance components. Except in certain special cases, exact tests are not available for random effects in unbalanced mixed models. Ordinarily, researchers apply techniques developed for balanced mixed models or treat unknown variance components as known.

The design matrix for the particular mixed model of interest is obtained by concatenation of the design matrices for oneway fixed effects model and sets of oneway random effects models. The fixed and random components share a common variance. Standard methods of parameter estimation, such as, Minimum Norm Quadratic Unbiased Estimation (MINQUE), Maximum Likelihood (ML) estimation, and Restricted Maximum Likelihood (REML) estimation, were used. To solve the ML and REML equations, a Fisher scoring algorithm was used. Because of the complexity of the original equations, it was necessary to do some simplifications of the equations to gain efficiency in the computations.

The main objective is to develop procedures for making inferences about the means from the set of fixed effects and the expectations from the sets of random effects. From the simulation results, the F test with REML estimates appears to be promising for hypothesis testing under the particular mixed model of interest.

CHAPTER 1

Introduction

Mixed models arise in many experimental situations. In these situations the usual interest is to perform tests of significance and estimate variance components, and to test fixed effects. To illustrate this situation, we consider the following two scenarios: (I) an academic scenario related to evaluation of students' performance. (II) an industrial scenario related to comparison of fabrics produced by different companies. EXAMPLE I. Suppose a mathematics department offers curriculum core courses which must be taken by a considerable number of students every semester. This situation requires the offering of many simultaneous sections of the same course. Although some sections are taught by professors, most of them are taught by teaching assistants (TAs) in masters and doctoral programs. Suppose that a supervisor of these courses is studying students' performance. Of particular interest are possible differences which say exist among students from sections taught by professors versus sections taught by TAs. Since TAs are not permanent employees, the expected performance of students across all sections taught by masters (MS) level TAs and the expected performance of students across all sections taught by doctoral (PhD) level TAs may be of interest. More specifically, suppose a course is offered with twelve sections, three of which are taught by professors, four by PhD level TAs, and five by Masters level TAs. The main goal is to compare five means measuring performance: three from sections taught by professors, one based on expected performance of students from four sections taught by PhD level TAs, and one from expected performance of students from five sections taught by Masters level TAs. When examining performance of students in sections taught by MS or

PhD level TAs there are two sources of variability: (i) variability in performance between these sections taught by MS level TAs and those sections taught by PhD level TAs, and (ii) variability in performance among students within sections.

<u>EXAMPLE II</u>. As another example, a vendor submits lots of fabric from national and foreign companies to a textile manufacturer which also produces this fabric. The manufacturer wants to know if the average breaking strength of the fabric produced by his/her factory is similar to breaking strength expected from fabric produced by national or foreign companies. Since there are many national and foreign companies, the interest is comparing three expectations: one from his/her factory, one from the set of national companies, and one from the set of foreign companies.

In both of these examples we find a common structure. In the academic example, the three sections taught by professors correspond to effects in a oneway fixed effects model, the set of four sections taught by PhD level TAs as a oneway random effects models, and the set of five sections taught by MS level TAs as a second oneway random effects model. In the fabric example, multiple lots of the manufacturers production may be modelled as a oneway fixed effects model, the set of national companies modelled as a oneway random effects model, and the set of foreign companies modelled as a second oneway random effects model. In each example, by concatenating the design matrices for fixed effects model and the random effects models we obtain a mixed model structure.

The analysis of the structure in each example highlights the problem considered in this thesis, that is, hypothesis testing related to the fixed effects in a particular mixed model, and estimation of parameters and variance components. The following are methods used in this thesis to obtain estimators: method of moments to obtain ANOVA estimates, Minimum Norm Quadratic Unbiased

Estimation (MINQUE), Maximum Likelihood (ML) estimation, and Restricted Maximum Likelihood (REML) estimation. To solve the ML and REML equations, a Fisher scoring algorithm is used. Because of the complexity of the original equations, it was necessary to do some simplifications to gain efficiency in the computations.

For hypothesis testing, the tests used are the likelihood ratio test, Wald's test with ML and REML estimates either with or without Kackar and Harville's (1988) correction, and two approximate F tests.

In Chapter 2 a description of the mixed model is presented. The methods used to obtain estimates are described in Chapter 3. The use of Fisher scoring is also discussed. The hypothesis testing procedures are described in Chapter 4. The results of simulations comparing tests of location parameters are presented in Chapter 5.

CHAPTER 2

Mixed Models

The general mixed model

Before describing the particular mixed model to be studied in this thesis, notation for mixed models will be introduced. The mixed model is a linear model in which some effects are fixed and some are random. The model can be written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{2.1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of data, $\boldsymbol{\beta}$ is an $p \times 1$ vector of fixed effects parameters, $\mathbf{X}$ is a known $N \times p$ design matrix corresponding to the fixed effects, $\mathbf{u}$ is an $m \times 1$ vector of random effects, $\mathbf{Z}$ is a known $N \times m$ design matrix corresponding to the random effects, and $\mathbf{e}$ is an $N \times 1$ vector of random errors.

The design matrix $\mathbf{Z}$ corresponding to the random effects can be partitioned as

$$\mathbf{Z} = [\mathbf{Z}_1 \quad \mathbf{Z}_2 \cdots \mathbf{Z}_k], \tag{2.2}$$

where $\mathbf{Z}_q$ is the design matrix for the q^{th} random factor or interaction. The vector of random effects $\mathbf{u}$ can be partitioned conformably with $\mathbf{Z}$ as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix}, \tag{2.3}$$

with m_q being the number of elements in $\mathbf{u}_q$. In the customary mixed model, the random effects have the properties

$$\mathrm{E}(\mathbf{e}) = \mathbf{0}_{N\times 1}, \tag{2.4}$$

$$\text{var}(\mathbf{e}) = \sigma_0^2 \mathbf{I}_N, \tag{2.5}$$

$$\text{E}(\mathbf{u}) = \mathbf{0}_{m\times 1}, \tag{2.6}$$

$$\text{var}(\mathbf{u}_q) = \sigma_q^2 \mathbf{I}_{m_q} \quad q = 1, 2, \ldots, k, \tag{2.7}$$

$$\text{cov}(\mathbf{u}_q, \mathbf{u}_{q'}) = \mathbf{0} \quad \forall q \neq q', \text{and} \tag{2.8}$$

$$\text{cov}(\mathbf{u}, \mathbf{e}) = \mathbf{0}. \tag{2.9}$$

Utilizing (2.3) and (2.7)–(2.8), the variance structure of $\mathbf{u}$ is

$$\mathbf{D} = \bigoplus_{q=1}^{k} \sigma_q^2 \mathbf{I}_{m_q}, \tag{2.10}$$

where $\bigoplus$ denotes direct sum of matrices (Searle, Casella and McCulloch 1992). The random component, $\mathbf{Zu}$ in (2.1), can be written as

$$\mathbf{Zu} = [\mathbf{Z}_1 \cdots \mathbf{Z}_k] \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \sum_{q=1}^{k} \mathbf{Z}_q \mathbf{u}_q \tag{2.11}$$

and substituting in (2.1) gives

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{q=1}^{k} \mathbf{Z}_q \mathbf{u}_q + \mathbf{e}. \tag{2.12}$$

Hence, from (2.1) to (2.12) the variance of $\mathbf{y}$ is

$$\boldsymbol{\Sigma} = \mathbf{ZDZ}' + \sigma_0^2 \mathbf{I}_N = \sum_{q=1}^{k} \sigma_q^2 \mathbf{Z}_q \mathbf{Z}_q' + \sigma_0^2 \mathbf{I}_N \tag{2.13}$$

Since $\mathbf{e}$ in (2.12) is a vector of random variables, just as each $\mathbf{u}_q$, we can define $\mathbf{e}$ as another $\mathbf{u}$-vector, say $\mathbf{u}_0$, and incorporate it into (2.11); Let

$$\mathbf{u}_0 \equiv \mathbf{e} \quad \mathbf{Z}_0 \equiv \mathbf{I}_N \quad \sigma_0^2 \equiv \sigma_e^2.$$

Then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{q=0}^{k} \mathbf{Z}_q \mathbf{u}_q \tag{2.14}$$

and

$$\boldsymbol{\Sigma} = \sum_{q=0}^{k} \mathbf{Z}_q \mathbf{Z}_q' \sigma_q^2. \tag{2.15}$$

Description of the mixed model of interest

In the examples in chapter 1, the fixed effects may be represented by $\mathbf{X}_0\boldsymbol{\mu}_0$, where $\mathbf{X}_0$ is a known design matrix corresponding to the fixed effects and $\boldsymbol{\mu}_0$ is a vector of means corresponding to the fixed effects. The subsets of random effects may be represented by $\mathbf{Z}\boldsymbol{\mu}_1$, where $\mathbf{Z}$ is a known design matrix corresponding to the subsets of random effects and $\boldsymbol{\mu}_1$ is a random vector of means corresponding to the subsets of random effects.

A matrix formulation of the model that shall be used is

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\mu}_0 + \mathbf{Z}\boldsymbol{\mu}_1 + \mathbf{e}, \tag{2.16}$$

where $\mathbf{y}$ is an $N \times 1$ vector of data, $\boldsymbol{\mu}_0$ is an $m_0 \times 1$ vector of fixed effects parameters, $\mathbf{X}_0$ is a known $N \times m_0$ design matrix for a oneway fixed effects model, $\boldsymbol{\mu}_1$ is an $m \times 1$ vector of random effects. The $N \times m$ matrix $\mathbf{Z}$ is assumed known and is the design matrix corresponding to the random effects, and $\mathbf{e}$ is the $N \times 1$ vector of random errors.

The matrix $\mathbf{X}_0$ in (2.16) can be written as

$$\mathbf{X}_0 = \begin{bmatrix} \mathbf{X}_{0u} \\ \mathbf{0}_{n \times m_0} \end{bmatrix}, \tag{2.17}$$

where u stands for the upper portion of the matrix,

$$\mathbf{X}_{0u} = \bigoplus_{i=1}^{m_0} \mathbf{1}_{n_{0i}}, \tag{2.18}$$

where $\mathbf{0}_{n\times m_0}$ is an $n \times m_0$ matrix of zeros, n is the total sample size for the random effects model, and n_{0i} is the sample size for the i^{th} fixed effect.

The random components $\mathbf{Z}\boldsymbol{\mu}_1$ in (2.16) can be partitioned as

$$\mathbf{Z}\boldsymbol{\mu}_1 = [\mathbf{Z}_1 \cdots \mathbf{Z}_k] \begin{bmatrix} \boldsymbol{\mu}_{11} \\ \vdots \\ \boldsymbol{\mu}_{1k} \end{bmatrix} = \sum_{q=1}^{k} \mathbf{Z}_q \boldsymbol{\mu}_{1q}, \tag{2.19}$$

where $\mathbf{Z}$ is a concatenation of k design matrices from oneway random effects models, every $\mathbf{Z}_q$ is a design matrix corresponding to a random effects model, and $\boldsymbol{\mu}_{1q}$ is the vector of random effects for factor q. The number of levels of the q^{th} random factor, and hence the order of $\boldsymbol{\mu}_{1q}$, is denoted by m_q, and

$$\boldsymbol{\mu}_{1q} \sim (\mathbf{1}_{m_q}\mu_q, \sigma_q^2 \mathbf{I}_{m_q}). \tag{2.20}$$

The design matrix $\mathbf{Z}$ can be written as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{0}_{n_0 \times m} \\ \mathbf{Z}_l \end{bmatrix}, \tag{2.21}$$

where $\mathbf{0}_{n_0\times m}$ is an $n_0 \times m$ matrix of zeros, n_0 will denote the total sample size for the fixed effects model, l stands for the lower portion of the matrix, and

$$\mathbf{Z}_l = \bigoplus_{q=1}^{k} \bigoplus_{i=1}^{m_q} \mathbf{1}_{n_{iq}}. \tag{2.22}$$

Hence, the random effects can be modeled as

$$\boldsymbol{\mu}_1 = W\boldsymbol{\mu} + \mathbf{u}, \tag{2.23}$$

where $\boldsymbol{\mu}$ is a $k \times 1$ vector of means, $\mathbf{W}$ is an $m \times k$ design matrix given by

$$\mathbf{W} = \bigoplus_{q=1}^{k} \mathbf{1}_{m_q}, \tag{2.24}$$

and

$$\mathbf{u} \sim (\mathbf{0}, \mathbf{D}), \tag{2.25}$$

for $\mathbf{D}$ in (2.10).

Using (2.23), the model in (2.16) can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{X_0}\boldsymbol{\mu}_0 + \mathbf{Z}(\mathbf{W}\boldsymbol{\mu} + \mathbf{u}) + \mathbf{e} \\ &= \mathbf{X_0}\boldsymbol{\mu}_0 + \mathbf{ZW}\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{e}, \end{aligned} \tag{2.26}$$

where $\mathbf{ZW}$ is a design matrix for fixed effects. Thus

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e} \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \end{aligned} \tag{2.27}$$

where

$$\mathbf{X} = [\mathbf{X}_0 \ \ \mathbf{ZW}], \tag{2.28}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu} \end{bmatrix}, \tag{2.29}$$

and

$$\boldsymbol{\eta} = \mathbf{Zu} + \mathbf{e}. \tag{2.30}$$

The variance of $\boldsymbol{\eta}$ is

$$\mathbf{Var}(\boldsymbol{\eta}) = \boldsymbol{\Sigma}, \tag{2.31}$$

where $\boldsymbol{\Sigma}$ is given in (2.13).

Theorem 2.1 Expected Mean Squares (*EMS*) for several sources of variation are given in tables B.1. – B.3. (Appendix B).

Proof: The results are obtained by repeated use of

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \mathrm{E}(\mathbf{y})'\mathbf{A}\mathrm{E}(\mathbf{y}). \tag{2.32}$$

The following, is a proof for the Expected Mean Squares among fixed effects $\mathrm{E}(MS_1)$ for case I, where $m_q = \frac{m}{k}$, $niq = \frac{N}{M}$, and $M = m_0 + m$.

The MS_1 can be written as

$$MS_1 = \frac{SS_1}{df_1}, \tag{2.33}$$

where SS_1 represents the sum of squares among fixed effects for case I, and the corresponding degrees of freedom are $df_1 = m_0 - 1$.

Taking expectations to both sides of equation (2.33), we obtain

$$\mathrm{E}(MS_1) = \frac{\mathrm{E}(SS_1)}{df_1}. \tag{2.34}$$

The SS_1 can be written as

$$SS_1 = \mathbf{y}'(\mathbf{P}_{x_0} - \mathbf{P}_{x_{01}})\mathbf{y}, \tag{2.35}$$

where

$$\mathbf{P}_{x_0} = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0', \tag{2.36}$$

$$\mathbf{P}_{x_{01}} = \mathbf{X}_{01}(\mathbf{X}_{01}'\mathbf{X}_{01})^{-1}\mathbf{X}_{01}', \tag{2.37}$$

$\mathbf{X}_0$ is given in (2.17), and

$$\begin{aligned} \mathbf{X}_{01} &= \mathbf{X}_0\mathbf{1}_{m_0} \\ &= \begin{bmatrix} \mathbf{1}_{n_0} \\ \mathbf{0}_{n\times 1} \end{bmatrix}. \end{aligned} \tag{2.38}$$

Using (2.32), $\mathrm{E}(SS_1)$ is

$$\mathrm{E}(SS_1) = \mathrm{tr}(\mathbf{P}_{x_0}\boldsymbol{\Sigma}) - \mathrm{tr}(\mathbf{P}_{x_{01}}\boldsymbol{\Sigma}) + \boldsymbol{\beta}'\mathbf{X}'\mathbf{P}_{x_0}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{P}_{x_{01}}\mathbf{X}\boldsymbol{\beta}. \tag{2.39}$$

Using equation (2.17), the matrix $\mathbf{P}_{x_0}$ in (2.36) can be written as

$$\mathbf{P}_{x_0} = \begin{bmatrix} \oplus_{i=1}^{m_0} \frac{1}{n_{i0}} \mathbf{1}_{n_{i0}} \mathbf{1}'_{n_{i0}} & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \mathbf{0}_{n \times n} \end{bmatrix}. \tag{2.40}$$

Using equation (2.38), the matrix $\mathbf{P}_{x_{01}}$ in (2.37) can be written as

$$\mathbf{P}_{x_{01}} = \begin{bmatrix} \frac{1}{n_0} \mathbf{1}_{n_0} \mathbf{1}'_{n_0} & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \mathbf{0}_{n \times n} \end{bmatrix}. \tag{2.41}$$

The matrix $\mathbf{\Sigma}$ in (2.31) can be partitioned conformably with the matrices $\mathbf{P}_{x_0}$ in (2.40), and $\mathbf{P}_{x_{01}}$ in (2.41), as

$$\mathbf{\Sigma} = \begin{bmatrix} \oplus_{i=1}^{n_0} \sigma_0^2 & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \oplus_{q=1}^{k} \oplus_{i=1}^{m_q} \mathbf{1}_{n_{iq}} \mathbf{1}'_{n_{iq}} \sigma_q^2 + \oplus_{i=1}^{n} \sigma_0^2 \end{bmatrix}. \tag{2.42}$$

Using matrices $\mathbf{P}_{x_0}$ in (2.40), $\mathbf{P}_{x_{01}}$ in (2.41), and $\mathbf{\Sigma}$ in (2.42). $\mathrm{E}(SS_1)$ becomes

$$\begin{aligned} \mathrm{E}(SS_1) &= m_0\sigma_0^2 - \sigma_0^2 + \sum_{i=1}^{m_0} n_{i0}\mu_{i0}^2 - \frac{1}{n_0}\left(\sum_{i=1}^{m_0} n_{i0}\mu_{i0}\right)^2 \\ &= (m_0 - 1)\sigma_0^2 + \sum_{i=1}^{m_0} n_{i0}(\mu_{i0} - \bar{\mu}_{\cdot 0})^2 \\ &= (m_0 - 1)\sigma_0^2 + \frac{N}{M}\sum_{i=1}^{m_0}(\mu_{i0} - \bar{\mu}_{\cdot 0})^2, \end{aligned} \tag{2.43}$$

because $n_{i0} = \frac{N}{M}$.

Therefore the expected mean squares for fixed effects $\mathrm{E}(MS_1)$ in (2.34) becomes

$$\begin{aligned} \mathrm{E}(MS_1) &= \frac{\mathrm{E}(SS_1)}{m_0 - 1} \\ &= \sigma_0^2 + \frac{N}{M(m_0 - 1)}\sum_{i=1}^{m_0}(\mu_{i0} - \bar{\mu}_{\cdot 0})^2. \end{aligned} \tag{2.44}$$

The proofs for all other expected mean squares are identical and will not be presented.

□

Table 1: Data

Fixed Effects		Random Subset 1		Random Subset 2		
level 1	level 2	level 1	level 2	level 1	level 2	level 3
86	81	89	75	91	82	88
90	88	85	77	86	85	72
97	99	87	85	87		82
	87		86			72

Table 2: Notation

Fixed Effects		Random Subset 1		Random Subset 2		
level 1	level 2	level 1	level 2	level 1	level 2	level 3
y_{110}	y_{210}	y_{111}	y_{211}	y_{112}	y_{212}	y_{312}
y_{120}	y_{220}	y_{121}	y_{221}	y_{122}	y_{222}	y_{322}
y_{130}	y_{230}	y_{131}	y_{231}	y_{132}		y_{332}
	y_{240}		y_{241}			y_{342}

Example

Assume some data (Tables 1 and 2) for the model of interest (2.26):

$$y_{ijq} = \text{response in the } i^{th} \text{ effect, } j^{th} \text{ replication, and } q^{th}\text{subset}$$

$$i = 1, 2, ..., m_q$$

$$j = 1, 2, ..., n_{iq}$$

$$q = 0, 1, 2, ..., k$$

$$m_q = \text{number of means in the } q^{th} \text{ subset of random effects}$$

$$n_{iq} = \text{number of observations in the } i^{th} \text{ effect, and } q^{th}\text{subset}$$

$$k = \text{number of random subsets}$$

$$M = \text{total number of fixed and random effects}$$

m_0 = total number of fixed effects

m = total number of random effects

m_q = number of means in the q^{th} subset of random effects

N = total number of observations

n = total number of observations in the random effects

n_{iq} = number of observations in the i^{th} effect, and q^{th}subset

n_q = number of observations in the q^{th}subset

$$
\begin{aligned}
M &= m_0 + m \\
m &= \sum_{q=1}^{k} m_q \\
N &= n_0 + n \\
n_0 &= \sum_{i=1}^{m_0} n_{0i} \\
n_q &= \sum_{i=1}^{m_q} n_{iq} \\
n &= \sum_{q=0}^{k} \sum_{i=1}^{m_q} n_{iq} \\
p &= m_0 + k
\end{aligned}
$$

β_{0i} = parameter corresponding to the i^{th} fixed effect

β_q = parameter corresponding to the q^{th} subset of random effects

u_{iq} = random component corresponding to the i^{th} random effect, and q^{th}subset

e_{ijq} = random error in the i^{th} effect, j^{th} replication, and q^{th}subset.

The model for the data in Table 1 is

$$
\begin{bmatrix} 86\\90\\97\\81\\88\\99\\87\\89\\85\\87\\75\\77\\85\\86\\91\\86\\87\\82\\85\\88\\72\\82\\72 \end{bmatrix}
=
\left[\begin{pmatrix} 1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\0&0 \end{pmatrix}
\;\middle|\;
\begin{pmatrix} 0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&1&0\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1 \end{pmatrix}
\begin{pmatrix} 1&0\\1&0\\0&1\\0&1\\0&1 \end{pmatrix}\right]
\begin{bmatrix} \beta_{01}\\ \beta_{02}\\ \beta_1\\ \beta_2 \end{bmatrix}
+
\begin{bmatrix} 0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&1&0\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1 \end{bmatrix}
\begin{bmatrix} u_{11}\\u_{21}\\u_{12}\\u_{22}\\u_{32} \end{bmatrix}
$$

$$
+\begin{bmatrix} e_{110}\\ e_{120}\\ \vdots\\ e_{332}\\ e_{342} \end{bmatrix}
$$

$$
\begin{aligned}
\mathbf{y} &= [\ \mathbf{X}_0 \mid \mathbf{ZW}\ ]\,\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}\\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}
\end{aligned}
$$

CHAPTER 3

Variance Components Estimation

Introduction

General procedures used for parameter estimation are described in Searle, Casella and McCulloch (1992). The methods developed in this thesis are based on those general procedures. The final equations are simplified for computational purposes.

Maximum Likelihood Estimation (ML Estimation)

Likelihood Function

Assuming

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{3.1}$$

the likelihood function is

$$L = L(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}) = \frac{e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}}{(2\pi)^{\frac{1}{2}N}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}. \tag{3.2}$$

ML Estimation Equations

Maximizing L in (3.2) can be achieved by maximizing the logarithm of L which shall be denoted by ℓ:

$$\ell = \log L = -\frac{1}{2}N\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{3.3}$$

To maximize ℓ, differentiate (3.3), first with respect to $\boldsymbol{\beta}$, which yields

$$\ell_\beta = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{3.4}$$

Second, differentiate (3.3), with respect to σ_q^2 to obtain

$$\ell_{\sigma_q^2} = \frac{\partial \ell}{\partial \sigma_q^2} = -\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q') + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{3.5}$$

for $q = 0, 1, \ldots, k$.

A general principle for maximizing ℓ with respect to $\boldsymbol{\beta}$ and the σ_q^2 is to equate (3.4) and (3.5) to zero and solve the resulting equations for $\boldsymbol{\beta}$ and the σ_q^2 s. The ML estimates are obtained by solving the following equations:

$$\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}, \tag{3.6}$$

and

$$\text{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_q\mathbf{Z}_q') = \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\widehat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right), \tag{3.7}$$

for $q = 0, \ldots, k$.

An equivalent expression for the ML equation (3.7) can be written as

$$\sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \left\{\text{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_q\mathbf{Z}_q')\right\} = \sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \left\{\mathbf{y}'\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q'\widehat{\mathbf{P}}\mathbf{y}\right\}, \tag{3.8}$$

where $\mathbf{e}_q^k$ is the q^{th} column of $\mathbf{I}_k$, and $\otimes$ denotes Kronecker product (Searle, Casella and McCulloch, 1992),

$$\widehat{\boldsymbol{\Sigma}} = \sum_{q=0}^{k} \mathbf{Z}_q\mathbf{Z}_q'\widehat{\sigma}_q^2, \tag{3.9}$$

and

$$\widehat{\mathbf{P}} = \widehat{\boldsymbol{\Sigma}}^{-1} - \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}')^{-}\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}. \tag{3.10}$$

Before deriving alternative expressions for these equations, note that (except in what turns out to be limited cases with balanced data (Searle, Casella and McCulloch, 1992)) closed form expressions for the solutions of ML equations (3.6) and (3.8) cannot be obtained. Therefore, solutions to these equations have to be obtained numerically.

Asymptotic Dispersion Matrices for ML Estimators

The model in (2.27), with the replacement of $\mathbf{Zu}$ by the equivalent expression from (2.11), can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \cdots + \mathbf{Z}_k\mathbf{u}_k + \mathbf{e}. \tag{3.11}$$

Letting $\mathbf{G}_q = \mathbf{Z}_q\mathbf{Z}'_q$, and $\mathbf{G}_0 = \mathbf{I}_N$, the following assumptions (Miller, 1977) are made about the model in (3.11).

ASSUMPTION 3.1. The random vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{e}$ are mutually independent, with

$$\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma_0^2\mathbf{I}_N), \text{ and} \tag{3.12}$$

$$\mathbf{u}_q \sim \mathrm{N}(\mathbf{0}, \sigma_q^2\mathbf{I}_{m_q}), \tag{3.13}$$

for $q = 0, 1, \ldots, k$.

ASSUMPTION 3.2. The matrix $\mathbf{X}$ has full rank p.

ASSUMPTION 3.3. $\mathrm{N} \geq p + k + 1$.

ASSUMPTION 3.4. The partitioned matrix $[\mathbf{X} : \mathbf{Z}_q]$ has rank greater than p, for each q, $q = 1, 2, \ldots, k$.

ASSUMPTION 3.5. The matrices $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_k$ in (3.11) are linearly independent.

ASSUMPTION 3.6. The matrix $\mathbf{Z}_q$ consists only of zeros and ones and there is exactly one 1 in each row and at least one 1 in each column, $q = 1, 2, \ldots, k$.

ASSUMPTION 3.7. Each m_q, $q = 1, 2, \ldots, k$, tend to infinity. By implication, N also tends to infinity.

Under these assumptions, a useful property of the ML estimator $\widehat{\boldsymbol{\theta}}$ of

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\sigma}^2 \end{bmatrix}, \tag{3.14}$$

where

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \vdots \\ \sigma_k^2 \end{bmatrix},$$

is that as $N \to \infty$ and $m_q \to \infty$ for all q,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{L} \mathrm{N}(\mathbf{0}, N[\mathbf{I}(\boldsymbol{\theta})]^{-1}),$$

where $\mathbf{I}(\boldsymbol{\theta})$ is known as the information matrix, and defined as

$$\mathbf{I}(\boldsymbol{\theta}) \equiv -\mathbf{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right). \tag{3.15}$$

Accordingly

$$\mathrm{var}(\widehat{\boldsymbol{\theta}}) \simeq [\mathbf{I}(\boldsymbol{\theta})]^{-1}. \tag{3.16}$$

Note, the result in (3.15) depends on $m_q \to \infty$. The case where $m_q \not\to \infty$ is treated later.

In (3.4) we used the symbol ℓ_β for $\partial \ell / \partial \boldsymbol{\beta}$. This is extended to using $\ell_{\beta\beta}$ for $\partial^2 \ell / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'$ and $\ell_{\beta\sigma^2}$ for $\partial^2 \ell / \partial \boldsymbol{\beta} \partial \boldsymbol{\sigma}^{2\prime}$. Then,

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\sigma}^2 \end{bmatrix} = -\mathrm{E}\begin{bmatrix} \ell_{\beta\beta} & \ell_{\beta\sigma^2} \\ \ell_{\sigma^2\beta} & \ell_{\sigma^2\sigma^2} \end{bmatrix}, \tag{3.17}$$

where the components of the matrix of second derivatives, known as the Hessian (**H**), are

$$\ell_{\beta\beta} = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}, \tag{3.18}$$

$$\ell_{\beta\sigma_q^2} = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta), \tag{3.19}$$

for $q = 0, \ldots, k$, and

$$\begin{aligned} \ell_{\sigma_q^2\sigma_{q'}^2} &= \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q') \\ &- (\mathbf{y} - \mathbf{X}\beta)'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta), \end{aligned} \tag{3.20}$$

for $q = 0, \ldots, k$; $q' = 0, \ldots, k$.

Therefore the information matrix is

$$\mathbf{I}\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{G}_q\boldsymbol{\Sigma}^{-1}\mathbf{G}_q\right) \end{bmatrix}, \tag{3.21}$$

where $\mathbf{G}_q = \mathbf{Z}_q\mathbf{Z}'_q$. Asymptotically the variance matrix of the estimators is approximately the inverse of the information matrix giving

$$\operatorname{var}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\sigma}}^2 \end{bmatrix} \simeq \begin{bmatrix} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0} & 2\left[\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{G}_q\boldsymbol{\Sigma}^{-1}\mathbf{G}_q\right)\right]^{-1} \end{bmatrix}. \tag{3.22}$$

It is of interest to know what asymptotic properties exist if $N \to \infty$, but $m_q \not\to \infty$ for some q.

Lemma 3.1 $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ is a block diagonal matrix.

Proof:

The matrix $\boldsymbol{\Sigma}$ in (2.31) can be written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \bigoplus_{i=1}^{n_0}\sigma_0^2 & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\sigma_q^2 + \bigoplus_{i=1}^{n}\sigma_0^2 \end{bmatrix},$$

then

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \bigoplus_{i=1}^{n_0}\frac{1}{\sigma_0^2} & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \left[\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\sigma_q^2 + \bigoplus_{i=1}^{n}\sigma_0^2\right]^{-1} \end{bmatrix}.$$

The matrix $\mathbf{X}$ in (2.28) can be written as

$$\mathbf{X} = \begin{bmatrix} \bigoplus_{i=1}^{m_0}\mathbf{1}_{n_{i0}} & \mathbf{0}_{n_0 \times k} \\ \mathbf{0}_{n \times m_0} & \bigoplus_{q=1}^{k}\mathbf{1}_{n_q} \end{bmatrix},$$

so

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} = \begin{bmatrix} \bigoplus_{i=1}^{m_0} n_{i0}\frac{1}{\sigma_0^2} & \mathbf{0}_{m_0 \times k} \\ \mathbf{0}_{k \times m_0} & \bigoplus_{q=1}^{k}\mathbf{1}'_{n_q}\left[\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\sigma_q^2 + \bigoplus_{i=1}^{n}\sigma_0^2\right]^{-1}\bigoplus_{q=1}^{k}\mathbf{1}_{n_q} \end{bmatrix}.$$

Therefore, $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ is a block diagonal matrix. □

Theorem 3.2 Let $\mathbf{I}(\boldsymbol{\theta})$ be the information matrix from the mixed model defined in (3.21) partitioned as

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}(\boldsymbol{\theta})_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{33} & \mathbf{I}(\boldsymbol{\theta})_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{43} & \mathbf{I}(\boldsymbol{\theta})_{44} \end{bmatrix},$$

where

$$\begin{bmatrix} \mathbf{I}(\boldsymbol{\theta})_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{22} \end{bmatrix} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X},$$

$$\begin{bmatrix} \mathbf{I}(\boldsymbol{\theta})_{33} & \mathbf{I}(\boldsymbol{\theta})_{34} \\ \mathbf{I}(\boldsymbol{\theta})_{43} & \mathbf{I}(\boldsymbol{\theta})_{44} \end{bmatrix} = \frac{1}{2}\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\right),$$

and the zeros are conformable partitions of the matrices of zeros in (3.21). The dimensions of the matrices are: $\mathbf{I}(\boldsymbol{\theta})_{11}$ is an $m_0 \times m_0$ matrix, $\mathbf{I}(\boldsymbol{\theta})_{22}$ is a $k \times k$ matrix, $\mathbf{I}(\boldsymbol{\theta})_{33}$ is a 1×1 matrix, $\mathbf{I}(\boldsymbol{\theta})_{34}$ is a $1 \times k$ matrix, $\mathbf{I}(\boldsymbol{\theta})_{43}$ is a $k \times 1$ matrix, $\mathbf{I}(\boldsymbol{\theta})_{44}$ is a $k \times k$ matrix, and the zeros are the corresponding matrices of zeros. For a fixed m_q, $q = 1, 2, \ldots, k$, if $N \to \infty$, then

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where, $\mathbf{K}_{11}$ is an $m_0 \times m_0$ diagonal matrix of constants whose values are in the interval $[0, \frac{1}{\sigma_0^2}]$, and $\mathbf{K}_{33} = \frac{1}{2(\sigma_0^2)^2}$, a scalar.

Proof:

The information matrix $\mathbf{I}(\boldsymbol{\theta})$ in (3.21) can be written as

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}(\boldsymbol{\theta})_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{33} & \mathbf{I}(\boldsymbol{\theta})_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}(\boldsymbol{\theta})_{43} & \mathbf{I}(\boldsymbol{\theta})_{44} \end{bmatrix},$$

where

$$\mathbf{I}(\boldsymbol{\theta})_{11} = \bigoplus_{i=1}^{m_0} n_{i0}\frac{1}{\sigma_0^2},$$

$$\mathbf{I}(\boldsymbol{\theta})_{22} = \bigoplus_{q=1}^{k} \mathbf{1}'_{n_q}\left[\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q} \mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\sigma_q^2 + \bigoplus_{i=1}^{n}\sigma_0^2\right]^{-1}\bigoplus_{q=1}^{k}\mathbf{1}_{n_q}$$

$$= \frac{1}{\sigma_0^2}\bigoplus_{q=1}^{k}\left\{\sum_{i=1}^{m_q}\left(\frac{1}{\frac{1}{n_{iq}}+\frac{\sigma_q^2}{\sigma_0^2}}\right)\right\},$$

$$\mathbf{I}(\boldsymbol{\theta})_{33} = \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\right)$$

$$= \frac{1}{2}\left\{\frac{N-m}{(\sigma_0^2)^2} + \sum_{q=1}^{k}\sum_{i=1}^{m_q}\left(\frac{1}{\sigma_0^2 + n_{iq}\sigma_q^2}\right)^2\right\},$$

$$\mathbf{I}(\boldsymbol{\theta})_{34} = \frac{1}{2}\sum_{q=1}^{k}\mathbf{e}_q^{k\prime}\otimes\text{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}'_q\right)$$

$$= \frac{1}{2}\sum_{q=1}^{k}\mathbf{e}_q^{k\prime}\otimes\left\{\sum_{i=1}^{m_q}\frac{1}{n_{iq}}\left(\frac{1}{\frac{\sigma_0^2}{n_{iq}}+\sigma_q^2}\right)^2\right\},$$

$$\mathbf{I}(\boldsymbol{\theta})_{43} = [\mathbf{I}(\boldsymbol{\theta})_{34}]',$$

$$\mathbf{I}(\boldsymbol{\theta})_{44} = \frac{1}{2}\sum_{q=1}^{k}\sum_{q=1}^{k}\mathbf{e}_q^{k}\otimes\mathbf{e}_q^{k\prime}\otimes\text{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}'_q\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}'_q\right)$$

$$= \frac{1}{2}\bigoplus_{q=1}^{k}\left\{\sum_{i=1}^{m_q}\left(\frac{1}{\frac{\sigma_0^2}{n_{iq}}+\sigma_q^2}\right)^2\right\}.$$

Additional details on the first and second derivatives are given in Theorem 3.5.

Taking limits yields

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{11} = \mathbf{K}_{11},$$

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{22} = \mathbf{0},$$

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{33} = \mathbf{K}_{33},$$

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{34} = \mathbf{0},$$

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{43} = \mathbf{0}, \text{and}$$

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{I}(\boldsymbol{\theta})_{44} = \mathbf{0},$$

where, $\mathbf{K}_{11}$ is an $m_0 \times m_0$ diagonal matrix of constants whose i^{th} diagonal element is

$$\frac{1}{\sigma_0^2}\lim_{N\to\infty}\frac{n_{i0}}{N}, \qquad \frac{n_{i0}}{N} \neq 0,$$

and $\mathbf{K}_{33} = \frac{1}{2(\sigma_0^2)^2}$ is a scalar.

□

From Theorem 3.2 the conclusion is that $\widehat{\boldsymbol{\mu}}_0$ and $\hat{\sigma}_0^2$ are consistent estimators of $\boldsymbol{\mu}_0$ in (2.29) and σ_0^2, the first component of $\boldsymbol{\sigma}^2$ in (3.14) respectively. As $N \to \infty$, the amount of information regarding $\boldsymbol{\mu}$ in (2.29) and $\sigma_q^2, q = 1, 2, \ldots, k$ does not increase to infinity if m_q does not increase to infinity. Specifically, it can be shown that

$$\lim_{N\to\infty}\mathbf{I}(\boldsymbol{\theta})_{22} = \bigoplus_{q=1}^{k}\frac{m_q}{\sigma_q^2}, \text{and}$$

$$\lim_{N\to\infty} \mathbf{I}(\boldsymbol{\theta})_{44} = \bigoplus_{q=1}^{k} \frac{m_q}{2\sigma_q^4}.$$

One implication is that $\widehat{\boldsymbol{\mu}}$ in (2.29) and $\widehat{\sigma}_q^2, q = 1, 2, \ldots, k$ the components of $\boldsymbol{\sigma}^2$ excluding σ_0^2 in (3.14) are not consistent estimators. Therefore $\left[\mathbf{I}(\widehat{\boldsymbol{\theta}})\right]^{-1}$ will provide poor estimators of $\text{var}(\widehat{\boldsymbol{\beta}})$ and $\text{var}(\widehat{\boldsymbol{\sigma}}^2)$ when m_q are small, even though N may be large.

Restricted Maximum Likelihood Estimation (REML Estimation)

Rather than using $\mathbf{y}$ directly, REML (Thompson, 1962) is based on linear combinations of elements of $\mathbf{y}$, chosen in such a way that those combinations do not contain any fixed effects, no matter what their value. These linear combinations are shown to be equivalent to residuals obtained after fitting the fixed effects. This results from starting with a set of values $\mathbf{k}^{*\prime}\mathbf{y}$ where vectors $\mathbf{k}^{*\prime}$ are chosen so that

$$\mathbf{k}^{*\prime}\mathbf{y} = \mathbf{k}^{*\prime}\mathbf{X}\boldsymbol{\beta} + \mathbf{k}^{*\prime}\mathbf{Z}\mathbf{u}, \tag{3.23}$$

contains no term in $\boldsymbol{\beta}$, so that

$$\mathbf{k}^{*\prime}\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \quad \forall \boldsymbol{\beta} \neq \mathbf{0}. \tag{3.24}$$

Hence

$$\mathbf{k}^{*\prime}\mathbf{X} = \mathbf{0}. \tag{3.25}$$

Therefore, the form of $\mathbf{k}^{*\prime}$ must be

$$\mathbf{k}^{*\prime} = \mathbf{c}'\mathbf{M}_x, \tag{3.26}$$

for any $\mathbf{c}'$ and where

$$\mathbf{M}_x = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'. \tag{3.27}$$

Since $\mathbf{X}$ is of order $N \times p$ and rank p equation (3.24) is satisfied by only $N-p$ linearly independent values of $\mathbf{k}^{*'}$. Thus, in using a set of linearly independent vectors $\mathbf{k}^{*'}$ as rows of $\mathbf{K}^{*'}$ we confine attention to $\mathbf{K}^{*'}\mathbf{y}$. This approach is discussed in Searle, Casella and McCulloch (1992).

REML Estimation Equations

With

$$\mathbf{y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}\right) \tag{3.28}$$

we have, for $\mathbf{K}^{*'}\mathbf{X} = \mathbf{0}$,

$$\mathbf{K}^{*'}\mathbf{y} \sim \mathrm{N}(\mathbf{0}, \mathbf{K}^{*'}\boldsymbol{\Sigma}\mathbf{K}^{*}). \tag{3.29}$$

The REML equations can therefore be derived from the ML equations of (3.8), by suitable replacements:

$$\begin{array}{ccl} \mathbf{y} & \text{by} & \mathbf{K}^{*'}\mathbf{y}, \\ \mathbf{X} & \text{by} & \mathbf{K}^{*'}\mathbf{X}, \\ \mathbf{Z} & \text{by} & \mathbf{K}^{*'}\mathbf{Z}, \text{and} \\ \boldsymbol{\Sigma} & \text{by} & \mathbf{K}^{*'}\boldsymbol{\Sigma}\mathbf{K}^{*}. \end{array}$$

The Information Matrix for REML is given by

$$\mathbf{I}(\boldsymbol{\sigma}^2) = \frac{1}{2}\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)'} \otimes \left\{\mathrm{tr}(\mathbf{P}\mathbf{G}_q\mathbf{P}\mathbf{G}_q)\right\},$$

where

$$\begin{aligned} \mathbf{P} &= \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X}')^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1} \\ &= \mathbf{K}^{*}(\mathbf{K}^{*'}\boldsymbol{\Sigma}\mathbf{K}^{*})^{-1}\mathbf{K}^{*'}. \end{aligned}$$

With

$$\begin{aligned}\widehat{\mathbf{P}} &= \widehat{\Sigma}^{-1} - \widehat{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\widehat{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\widehat{\Sigma}^{-1} \\ &= \mathbf{K}^{*}(\mathbf{K}^{*\prime}\widehat{\Sigma}\mathbf{K}^{*})^{-1}\mathbf{K}^{*\prime}, \end{aligned} \tag{3.30}$$

from Theorem in Appendix M.4f,(87) in Searle, Casella and McCulloch (1992), equation (3.8) becomes

$$\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \operatorname{tr}\left(\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q'\right) = \sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \left\{\mathbf{y}'\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q'\widehat{\mathbf{P}}\mathbf{y}\right\}. \tag{3.31}$$

An alternative form (Searle, Casella and McCulloch, 1992) of (3.31) can be developed by direct multiplication. It is easily established that $\mathbf{P}\Sigma\mathbf{P} = \mathbf{P}$. Hence in the left-hand side of (3.31) we can use the identity and put the equations in the form

$$\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \left\{\operatorname{tr}(\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q'\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q')\right\}\widehat{\sigma}^2 = \sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \left\{\mathbf{y}'\widehat{\mathbf{P}}\mathbf{Z}_q\mathbf{Z}_q'\widehat{\mathbf{P}}\mathbf{y}\right\}. \tag{3.32}$$

Minimum Norm Quadratic Unbiased Estimation

(MINQUE Estimation)

MINQUE estimation requires distributional properties concerning first and second moments in (2.4) – (2.9) [see Rao (1971a) and Rao (1971b)], but does not require the normality assumption that is the foundation of MLE and REML procedures. Also MINQUE estimation does not require iteration, just the solution to a system of linear equations. However, estimators obtained by MINQUE are functions of *a priori* values used in place of the variance components in the estimation procedure itself. Any MINQUE estimate is the same as a first iterate of REML, using *a priori* values needed for MINQUE as the starting values for REML iteration.

MINQUE Estimation Equations

The MINQUE estimation equations are

$$\sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k} \mathbf{e}_{q+1}^{(k+1)\prime} \otimes \left\{ \mathrm{tr}(\mathbf{P}_0 \mathbf{Z}_q \mathbf{Z}_q' \mathbf{P}_0 \mathbf{Z}_{q} \mathbf{Z}_{q}') \right\} \tilde{\boldsymbol{\sigma}}^2 = \sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \left\{ \mathbf{y}' \mathbf{P}_0 \mathbf{Z}_q \mathbf{Z}_q' \mathbf{P}_0 \mathbf{y} \right\}, \tag{3.33}$$

where

$$\tilde{\boldsymbol{\sigma}}^2 = \begin{bmatrix} \tilde{\sigma}_0^2 \\ \tilde{\sigma}_1^2 \\ \vdots \\ \tilde{\sigma}_k^2 \end{bmatrix},$$

$$\mathbf{P}_0 = \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}_0^{-1} \mathbf{X})^{-} \mathbf{X}' \boldsymbol{\Sigma}_0^{-1}, \tag{3.34}$$

and

$$\boldsymbol{\Sigma}_0 = \sum_{q=0}^{k} \mathbf{Z}_q \mathbf{Z}_q' \sigma_q^{2(0)}, \tag{3.35}$$

where the $\sigma_q^{2(0)}$s are *a priori* arbitrary values. The MINQUE estimation equations in (3.33) are similar to those of REML in (3.32). A MINQUE solution is equivalent to the result of the first iteration using the REML equations (3.32) provided that the initial values for a REML procedure are the same as the *a priori* values for a MINQUE procedure.

MINQUE0 Estimation

A particular easy form of MINQUE is the MINQUE0 form when the *a priori* values for each σ_q^2, $q = 0, \ldots, k$, are set equal to zero, except for $\sigma_0^{2(0)} = 1$.

MINQUE0 Estimation Equations

The MINQUE equations (3.33) reduce to

$$\sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k} \mathbf{e}_{q+1}^{(k+1)\prime} \otimes \left\{ \mathrm{tr}(\mathbf{M}_x \mathbf{Z}_q \mathbf{Z}_q' \mathbf{M}_x \mathbf{Z}_q \mathbf{Z}_q') \right\} \tilde{\boldsymbol{\sigma}}_0^2 = \sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \left\{ \mathbf{y}' \mathbf{M}_x \mathbf{Z}_q \mathbf{Z}_q' \mathbf{M}_x \mathbf{y} \right\}, \tag{3.36}$$

where

$$\tilde{\boldsymbol{\sigma}}_0^2 = \begin{bmatrix} \tilde{\sigma}_{00}^2 \\ \tilde{\sigma}_{01}^2 \\ \vdots \\ \tilde{\sigma}_{0k}^2 \end{bmatrix}$$

and

$$\mathbf{M}_x = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'. \tag{3.37}$$

Equation (3.36) is solved directly for $\tilde{\boldsymbol{\sigma}}_0^2$.

Solving the Likelihood Equations

The solutions to the ML equations in (3.6) and (3.8), or to the REML equations in (3.32) do not have a close form. After considering some iterative procedures, the Fisher Scoring method was adopted.

Fisher Scoring Method

This method is based on the derivatives of the log-likelihood function in (3.3), and uses the expected value of the Hessian matrix in (3.17). The log-likelihood function must be maximized within the parameter space (Harville, 1977).

This iterative method of maximizing the log-likelihood function begins with an initial estimate $\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$ in (3.14) and then proceeds by calculating new estimates, $\boldsymbol{\theta}^{(t+1)}, t = 0, 1, 2, \ldots$. Iteration continues until the estimates converge. The iteration for this method is as follows.

$$\begin{aligned} \boldsymbol{\theta}^{(t+1)} &= \boldsymbol{\theta}^{(t)} - \left[(\mathrm{E}[\mathbf{H}])^{(t)} \right]^{-1} \left. \frac{\partial \ell}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^{(t)}} \\ &= \boldsymbol{\theta}^{(t)} + \left[\mathbf{I}\left(\boldsymbol{\theta}^{(t)} \right) \right]^{-1} \left. \frac{\partial \ell}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^{(t)}} \end{aligned} \tag{3.38}$$

where $\mathbf{H}$ is the Hessian matrix given in (3.17) and $\mathbf{I}\left(\boldsymbol{\theta}^{(t)}\right)$ is the Information matrix evaluated at $\boldsymbol{\theta}^{(t)}$. The convergence criterion is based on the norm of

$$\left[\mathbf{I}\left(\boldsymbol{\theta}^{(t)}\right)\right]^{-1} \left.\frac{\partial \ell}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^{(t)}}.$$

Simplified Equations for the Model of Interest

The non-zero elements from the first and second derivatives in (3.4), (3.5), and (3.18) – (3.20) are given in Theorem 3.5. To obtain some of these simplifications the following lemmas are useful.

Lemma 3.3 $\mathbf{Z}_{q'}'\mathbf{Z}_q = \mathbf{0}$, for $q \neq q'$.

Proof:

Assume $q = 1$, $q' = 2$. The proof for all other pairs $q \neq q'$ is identical and will not be presented,

$$\mathbf{Z}_1 = \begin{bmatrix} \mathbf{0}_{n_0 \times m_1} \\ \bigoplus_{i=1}^{m_1} \mathbf{1}_{n_{i1}} \\ \mathbf{0}_{n_2 \times m_1} \\ \mathbf{0}_{n_3 \times m_1} \\ \vdots \\ \mathbf{0}_{n_k \times m_1} \end{bmatrix},$$

$$\mathbf{Z}_2 = \begin{bmatrix} \mathbf{0}_{n_0 \times m_2} \\ \mathbf{0}_{n_1 \times m_2} \\ \bigoplus_{i=1}^{m_2} \mathbf{1}_{n_{i2}} \\ \mathbf{0}_{n_3 \times m_2} \\ \vdots \\ \mathbf{0}_{n_k \times m_2} \end{bmatrix},$$

then

$$\begin{aligned} \mathbf{Z}_1'\mathbf{Z}_2 &= \left[\, \mathbf{0}_{m_1 \times n_0} \bigoplus_{i=1}^{m_1} \mathbf{1}_{n_{i1}}' \mathbf{0}_{m_1 \times n_2} \mathbf{0}_{m_1 \times n_3} \cdots \mathbf{0}_{m_1 \times n_k} \,\right] \begin{bmatrix} \mathbf{0}_{n_0 \times m_2} \\ \mathbf{0}_{n_1 \times m_2} \\ \bigoplus_{i=1}^{m_2} \mathbf{1}_{n_{i2}} \\ \mathbf{0}_{n_3 \times m_2} \\ \vdots \\ \mathbf{0}_{n_k \times m_2} \end{bmatrix} \\ &= \mathbf{0}_{m_1 \times m_2}. \end{aligned}$$

□

Lemma 3.4 $\mathbf{Z}_{q'}'\Sigma^{-1}\mathbf{Z}_q = \mathbf{0}$, for $q \neq q'$.

Proof: Assume $q = 1$, $q' = 2$. The proof for all other pairs $q \neq q'$ is identical and will not be presented. From equation (2.13)

$$\Sigma = \sigma_0^2 \mathbf{I}_N + \mathbf{ZDZ}',$$

and

$$\begin{aligned}\Sigma^{-1} &= \left[\sigma_0^2 \mathbf{I}_N + \mathbf{ZDZ}'\right]^{-1} \\ &= \frac{1}{\sigma_0^2}\left[\mathbf{I}_N - \mathbf{Z}\left(\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q}\frac{1}{\frac{\sigma_0^2}{\sigma_q^2} + n_{iq}}\right)\mathbf{Z}'\right], \qquad (3.39)\end{aligned}$$

using the Woodbury's binomial inverse theorem (Press, 1982). Then

$$\begin{aligned}\mathbf{Z}_{q'}'\Sigma^{-1}\mathbf{Z}_q &= \mathbf{Z}_{q'}'\left\{\frac{1}{\sigma_0^2}\left[\mathbf{I}_N - \mathbf{Z}\left(\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q}\frac{1}{\frac{\sigma_0^2}{\sigma_q^2} + n_{iq}}\right)\mathbf{Z}'\right]\right\}\mathbf{Z}_q \\ &= \frac{1}{\sigma_0^2}\mathbf{Z}_{q'}'\mathbf{Z}_q \\ &- \frac{1}{\sigma_0^2}\left[\bigoplus_{i=1}^{m_1}\frac{1}{\frac{\sigma_0^2}{n_{i1}\sigma_1^2} + 1}\right] \\ &\quad \left[\ \mathbf{I}_{m_1}\mathbf{0}_{m_1\times m_2}\mathbf{0}_{m_1\times m_3}\cdots\mathbf{0}_{m_1\times m_k}\ \right]\begin{bmatrix}\mathbf{0}_{m_1\times m_2}\\ \mathbf{I}_{m_2}\\ \mathbf{0}_{m_3\times m_2}\\ \vdots \\ \mathbf{0}_{m_k\times m_2}\end{bmatrix} \\ &\quad \left[\bigoplus_{i=1}^{m_2} n_{i2}\right].\end{aligned}$$

Therefore, by lemma 3.3,

$$\mathbf{Z}_{q'}'\Sigma^{-1}\mathbf{Z}_q = \mathbf{0}.$$

□

Theorem 3.5 For the model of interest in (2.27), the first and second derivatives in (3.4), (3.5), and (3.18) – (3.20) simplify to the following,

$$\ell_\beta = \frac{1}{\sigma_0^2}\begin{bmatrix} \sum_{i=1}^{m_0} \mathbf{e}_i^{m_0} \otimes (y_{i\cdot 0} - n_{i0}\beta_{i0}) \\ \sum_{q=1}^{k} \mathbf{e}_q^k \otimes \sum_{i=1}^{m_q}\left[1 - \frac{n_{iq}}{\frac{\sigma_0^2}{\sigma_q^2} + n_{iq}}\right][y_{i\cdot q} - n_{iq}\beta_q] \end{bmatrix}, \quad q = 1,\ldots,k, \qquad (3.40)$$

where β_{i0} is the i^{th} β_0 corresponding to the fixed effects, $i = 1,\ldots,m_0$, β_q is the q^{th} β corresponding to the random effects, $q = 1,\ldots,k$,

$$y_{i\cdot q} = \sum_{j=1}^{n_{iq}} y_{ijq},$$

$$\ell_{\sigma_q^2} = \frac{1}{2}\left(\sum_{i=1}^{m_q}\left[-\frac{1}{\frac{\sigma_0^2}{n_{iq}} + \sigma_q^2} + \left(\frac{e_{i\cdot q}}{\sigma_0^2 + n_{iq}\sigma_q^2}\right)^2\right]\right), \qquad (3.41)$$

where

$$e_{i\cdot q} = \sum_{j=1}^{n_{iq}} e_{ijq},$$

$$e_{ijq} = \begin{cases} y_{ij0} - \beta_{i0} & \text{if } q = 0 \\ y_{ijq} - \beta_q & \text{otherwise,} \end{cases}$$

$$\ell_{\beta\beta} = -\frac{1}{\sigma_0^2}\begin{bmatrix} \oplus_{i=1}^{m_0} n_{i0} & \mathbf{0}_{m_0\times k} \\ \mathbf{0}_{k\times m_0} & \oplus_{q=1}^{k}\left\{\sum_{i=1}^{m_q}\left(\frac{1}{\frac{1}{n_{iq}} + \frac{\sigma_q^2}{\sigma_0^2}}\right)\right\} \end{bmatrix}, \qquad (3.42)$$

$$\ell_{\beta\sigma_q^2} = -\frac{1}{\sigma_0^4}\begin{bmatrix} \mathbf{0}_{m_0\times 1} \\ \sum_{i=1}^{m_q} n_{iq} e_{i\cdot q}\left(1 - \frac{1}{\frac{\sigma_0^2}{n_{iq}\sigma_q^2} + 1}\right)^2 \\ \mathbf{0}_{(k-1)\times 1} \end{bmatrix}, \qquad (3.43)$$

$$\ell_{\sigma_q^2\sigma_q^2} = \frac{1}{2\sigma_0^4}\left\{\sum_{i=1}^{m_q}\left[niq\left(1-\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)\right]^2\right\}$$

$$- \frac{1}{\sigma_0^6}\left\{\sum_{i=1}^{m_q} niqe_{i\cdot q}^2\left[1-\frac{3niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}+3\left(\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^2-\left(\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^3\right]\right\}, \tag{3.44}$$

$$\ell_{\sigma_0^2} = -\frac{1}{2\sigma_0^2}\left(N-\sum_{q=1}^{k}\sum_{i=1}^{m_q}\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)$$

$$+ \frac{1}{2\sigma_0^4}\sum_{q=1}^{k}\sum_{i=1}^{m_q}\sum_{j=1}^{n_{iq}}\left(e_{ijq}-\frac{e_{i\cdot q}}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^2+\frac{1}{2\sigma_0^4}\sum_{i=1}^{m_0}\sum_{j=1}^{n_{i0}}\left(e_{ij0}\right)^2, \tag{3.45}$$

$$\ell_{\sigma_0^2\sigma_0^2} = \frac{1}{2\sigma_0^4}\left[N-m+\sum_{q=1}^{k}\sum_{i=1}^{m_q}\left(1-\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^2\right]$$

$$- \frac{1}{\sigma_0^6}\left[\sum_{q=1}^{k}\sum_{i=1}^{m_q}\sum_{j=1}^{n_{iq}}\left(e_{ijq}-\frac{e_{i\cdot q}}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^2\right]$$

$$+ \frac{1}{2\sigma_0^6}\left[\sum_{q=1}^{k}\sum_{i=1}^{m_q}e_{i\cdot q}^2\left(1-\frac{niq}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)^2\left(\frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+niq}\right)\right]$$

$$- \frac{1}{\sigma_0^6}\left[\sum_{i=1}^{m_0}\sum_{j=1}^{n_{i0}}\left(e_{ij0}\right)^2\right], \tag{3.46}$$

$$\ell_{\beta\sigma_0^2} = -\frac{1}{\sigma_0^4}\left[\begin{array}{c} \sum_{i=1}^{m_0} \mathbf{e}_i^{m_0} \otimes ei\cdot 0 \\ \sum_{q=1}^{k} \mathbf{e}_q^k \otimes \sum_{i=1}^{n_{iq}} ei\cdot q \left(1 - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2} + niq}\right)\left(\frac{1}{1+\frac{\sigma_q^2 n_{iq}}{\sigma_0^2}}\right) \end{array}\right]. \tag{3.47}$$

Proof:

Equations (3.40) – (3.47) are the result of applying standard matrix algebra results. The following is a proof for the first derivative ℓ_β in (3.40).

Using matrices $\mathbf{X}_0$ in (2.17), $\mathbf{Z}$ in (2.21), and $\mathbf{W}$ in (2.24), matrix $\mathbf{X}$ in (2.28) can be partitioned conformably with the random and fixed effects as

$$\mathbf{X} = \begin{bmatrix} \oplus_{i=1}^{m_0} \mathbf{1}_{n_{i0}} & \mathbf{0}_{n_0 \times k} \\ \mathbf{0}_{n \times m_0} & \oplus_{q=1}^{k} \mathbf{1}_{n_q} \end{bmatrix}. \tag{3.48}$$

Using matrices $\mathbf{Z}$ in (2.21) and $\mathbf{Z}_l$ in (2.22), the matrix $\boldsymbol{\Sigma}^{-1}$ in (3.39) can be partitioned conformably with matrix $\mathbf{X}$ as

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \frac{1}{\sigma_0^2}\left[\mathbf{I}_N - \mathbf{Z}\left(\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q} \frac{1}{\frac{\sigma_0^2}{\sigma_q^2} + niq}\right)\mathbf{Z}'\right] \\
&= \frac{1}{\sigma_0^2}\left[\begin{pmatrix} \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \mathbf{I}_n \end{pmatrix} - \begin{pmatrix} \mathbf{0}_{n_0 \times m} \\ \mathbf{Z}_l \end{pmatrix}\left(\bigoplus_{q=1}^{k}\bigoplus_{i=1}^{m_q} \frac{1}{\frac{\sigma_0^2}{\sigma_q^2} + niq}\right)\begin{pmatrix} \mathbf{0}_{m \times n_0} & \mathbf{Z}_l \end{pmatrix}\right] \\
&= \frac{1}{\sigma_0^2}\begin{bmatrix} \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times n} \\ \mathbf{0}_{n \times n_0} & \oplus_{q=1}^{k}\oplus_{i=1}^{m_q}\left(\mathbf{I}_{n_{iq}} - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+n_{iq}}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\right) \end{bmatrix}.
\end{aligned} \tag{3.49}$$

Equation (3.4) can be written as

$$\begin{aligned}
\ell_\beta &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned} \tag{3.50}$$

where

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \oplus_{i=1}^{m_0}\mathbf{1}'_{n_{i0}} & \mathbf{0}_{m_0\times n} \\ \mathbf{0}_{k\times n_0} & \oplus_{q=1}^{k}\mathbf{1}'_{n_q} \end{bmatrix} \frac{1}{\sigma_0^2} \begin{bmatrix} \mathbf{I}_{n_0} & \mathbf{0}_{n_0\times n} \\ \mathbf{0}_{n\times n_0} & \oplus_{q=1}^{k}\oplus_{i=1}^{m_q}\left(\mathbf{I}_{n_{iq}} - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+n_{iq}}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\right) \end{bmatrix}$$

$$= \frac{1}{\sigma_0^2}\begin{bmatrix} \oplus_{i=1}^{m_0}\mathbf{1}'_{n_{i0}} & \mathbf{0}_{m_0\times n} \\ \mathbf{0}_{k\times n_0} & \left(\oplus_{q=1}^{k}\mathbf{1}'_{n_q}\right)\oplus_{q=1}^{k}\oplus_{i=1}^{m_q}\left(\mathbf{I}_{n_{iq}} - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+n_{iq}}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\right) \end{bmatrix}, \quad (3.51)$$

and

$$\mathbf{y}-\mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \end{bmatrix} - \begin{bmatrix} \oplus_{i=1}^{m_0}\mathbf{1}_{n_{i0}} & \mathbf{0}_{n_0\times k} \\ \mathbf{0}_{n\times m_0} & \oplus_{q=1}^{k}\mathbf{1}_{n_q} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{y}_0 - (\oplus_{i=1}^{m_0}\mathbf{1}_{n_{i0}})\boldsymbol{\beta}_0 \\ \mathbf{y}_1 - (\oplus_{q=1}^{k}\mathbf{1}_{n_q})\boldsymbol{\beta} \end{bmatrix}. \quad (3.52)$$

Therefore ℓ_β in (3.50) can be written as

$$\ell_\beta = \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

$$= \frac{1}{\sigma_0^2}\begin{bmatrix} \oplus_{i=1}^{m_0}\mathbf{1}'_{n_{i0}} & \mathbf{0}_{m_0\times n} \\ \mathbf{0}_{k\times n_0} & (\oplus_{q=1}^{k}\mathbf{1}'_{n_q})\oplus_{q=1}^{k}\oplus_{i=1}^{m_q}\left(\mathbf{I}_{n_{iq}} - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+n_{iq}}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\right) \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y}_0 - (\oplus_{i=1}^{m_0}\mathbf{1}_{n_{i0}})\boldsymbol{\beta}_0 \\ \mathbf{y}_1 - (\oplus_{q=1}^{k}\mathbf{1}_{n_q})\boldsymbol{\beta} \end{bmatrix}$$

$$= \frac{1}{\sigma_0^2}\begin{bmatrix} (\oplus_{i=1}^{m_0}\mathbf{1}'_{n_{i0}})\mathbf{y}_0 - (\oplus_{i=1}^{m_0}\mathbf{1}'_{n_{i0}}\mathbf{1}_{n_{i0}})\boldsymbol{\beta}_0 \\ \left(\oplus_{q=1}^{k}\mathbf{1}'_{n_q} - \oplus_{q=1}^{k}\mathbf{1}'_{n_q}\oplus_{i=1}^{m_q}\left(\mathbf{I}_{n_{iq}} - \frac{1}{\frac{\sigma_0^2}{\sigma_q^2}+n_{iq}}\mathbf{1}_{n_{iq}}\mathbf{1}'_{n_{iq}}\right)\right)(\mathbf{y}_1 - (\oplus_{q=1}^{k}\mathbf{1}_{n_q})\boldsymbol{\beta}) \end{bmatrix}. \quad (3.53)$$

For the upper portion of the matrix in (3.53), where $q=0$, assume $i=1$. This portion becomes

$$[\mathbf{1}'_{n_{10}} \quad \mathbf{0}_{1\times(n_0-n_{10})}]\mathbf{y}_0 - [\mathbf{1}'_{n_{10}} \quad \mathbf{0}_{1\times(n_0-n_{10})}]\begin{bmatrix} \mathbf{1}_{n_{10}} \\ \mathbf{0}_{(n_0-n_{10})\times 1} \end{bmatrix}\beta_{10}$$

$$= y_{1\cdot 0} - n_{10}\beta_{10}. \quad (3.54)$$

The proof for any other value of i is identical and will not be presented.

For the lower portion of the matrix in (3.53), assume $q = 1$. This portion becomes

$$\left([\mathbf{1}'_{n_1} \quad \mathbf{0}_{1\times(n-n_1)}] - [\mathbf{1}'_{n_1} \quad \mathbf{0}_{1\times(n-n_1)}] \bigoplus_{i=1}^{m_1} \left(\mathbf{I}_{n_{i1}} - \frac{1}{\frac{\sigma_0^2}{\sigma_1^2} + n_{i1}} \mathbf{1}_{n_{i1}} \mathbf{1}'_{n_{i1}} \right) \right)$$

$$\left(\mathbf{y}_1 - \begin{bmatrix} \mathbf{1}_{n_1} \\ \mathbf{0}_{(n-n_1)\times 1} \end{bmatrix} \beta_1 \right)$$

$$\begin{aligned}
&= [\mathbf{1}'_{n_1} \quad \mathbf{0}_{1\times(n-n_1)}] \left(\mathbf{y}_1 - \begin{bmatrix} \mathbf{1}_{n_1} \\ \mathbf{0}_{(n-n_1)\times 1} \end{bmatrix} \beta_1 \right) \\
&- [\mathbf{1}'_{n_1} \quad \mathbf{0}_{1\times(n-n_1)}] \bigoplus_{i=1}^{m_1} \left(\mathbf{I}_{n_{i1}} - \frac{1}{\frac{\sigma_0^2}{\sigma_1^2} + n_{i1}} \mathbf{1}_{n_{i1}} \mathbf{1}'_{n_{i1}} \right) \left(\mathbf{y}_1 - \begin{bmatrix} \mathbf{1}_{n_1} \\ \mathbf{0}_{(n-n_1)\times 1} \end{bmatrix} \beta_1 \right) \\
&= \sum_{i=1}^{m_1} (y_{i\cdot 1} - n_{i1}\beta_1) - \sum_{i=1}^{m_1} \left[\frac{n_{i1}}{\frac{\sigma_0^2}{\sigma_1^2} + n_{i1}} \right] [y_{i\cdot 1} - n_{i1}\beta_1] \\
&= \sum_{i=1}^{m_1} \left[1 - \frac{n_{i1}}{\frac{\sigma_0^2}{\sigma_1^2} + n_{i1}} \right] [y_{i\cdot 1} - n_{i1}\beta_1]. \qquad (3.55)
\end{aligned}$$

The proof for any other value of q is identical and will not be presented.

□

Corollary 3.6 The solution to $\ell_\beta = \mathbf{0}$ for ℓ_β in (3.40) is

A) For any n_{iq}, $i = 1, \ldots, m_q$, and $q = 0, 1, \ldots, k$

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\mu}}_0 \\ \widehat{\boldsymbol{\mu}} \end{bmatrix}$$

(3.56)

$$= \begin{bmatrix} (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{y}_0 \\ \sum_{q=1}^{k}\mathbf{e}_q^k \otimes \left[\dfrac{\sum_{i=1}^{m_q}\sum_{j=1}^{n_{iq}} y_{ijq}\left(\dfrac{1}{1+\frac{n_{iq}\sigma_q^2}{\sigma_0^2}}\right)}{\sum_{i=1}^{m_q}\left(\dfrac{n_{iq}}{1+\frac{n_{iq}\sigma_q^2}{\sigma_0^2}}\right)} \right] \end{bmatrix}.$$

where $\mathbf{X}_0$ is given in (2.17), and $\mathbf{y}_0$ is the corresponding vector $\mathbf{y}$ conformably partitioned according to equation (2.28).

B) When data are balanced (i. e., $n_{iq}=\frac{N}{M}$, for $i = 1,\ldots,m_q$, and $q = 0,1,\ldots,k$), then

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \begin{bmatrix} \overline{\mathbf{y}}_0 \\ \overline{\mathbf{y}} \end{bmatrix}, \end{aligned} \tag{3.57}$$

where $\widehat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$, $\overline{\mathbf{y}}_0$ is a vector of ordinary sample means for the m_0 levels of fixed effects, and $\overline{\mathbf{y}}$ is a vector of ordinary sample means for the k subsets of random effects.

Corollary 3.7 When data are balanced ($n_{iq}=\frac{N}{M}$, for $i = 1,\ldots,m_q$, and $q = 0,1,\ldots,k$), the MLE equations for σ_0^2 and σ_q^2 for $q = 1,\ldots,k$ are

$$\frac{1}{2}\left(\sum_{i=1}^{m_q}\left[-\frac{1}{\frac{M\sigma_0^2}{N}+\sigma_q^2}+\left(\frac{e_{i\cdot q}}{\sigma_0^2+\frac{N\sigma_q^2}{M}}\right)^2\right]\right) = 0, \text{and} \tag{3.58}$$

$$\frac{1}{2\sigma_0^4}\sum_{q=1}^{k}\sum_{i=1}^{m_q}\sum_{j=1}^{n_{iq}}\left(e_{ijq}-\frac{e_{i\cdot q}}{\frac{\sigma_0^2}{\sigma_q^2}+\frac{N}{M}}\right)^2+\frac{1}{2\sigma_0^4}\sum_{i=1}^{m_0}\sum_{j=1}^{n_{i0}}(e_{ij0})^2$$
$$-\frac{1}{2\sigma_0^2}\left(N-\sum_{q=1}^{k}\sum_{i=1}^{m_q}\frac{\frac{N}{M}}{\frac{\sigma_0^2}{\sigma_q^2}+\frac{N}{M}}\right) = 0. \tag{3.59}$$

The solutions to these equations are

$$\widehat{\sigma}_0^2 = \frac{\sum_{q=1}^{k}\sum_{i=1}^{m_q}\sum_{j=1}^{n_{iq}} e_{ijq}^2 + \sum_{i=1}^{m_q}\sum_{j=1}^{n_{i0}} e_{ij0}^2 - \frac{M}{N}\sum_{q=1}^{k}\sum_{i=1}^{m_q} e_{i\cdot q}^2}{N-m} \tag{3.60}$$

$$= MS_7 \frac{N-M}{N-m},$$

for MS_7 in Table B.2 (Appendix B), and

$$\widehat{\sigma}_q^2 = \frac{M^2}{m_q N^2}\sum_{i=1}^{m_q} e_{i\cdot q}^2 - \frac{M}{N}\widehat{\sigma}_0^2 \tag{3.61}$$

$$= MS_4 \frac{M}{N} - MS_7 \frac{M(N-M)}{N(N-m)},$$

for MS_4 and MS_7 in Table B.2 (Appendix B).

Corollary 3.8 The Fisher Scoring Equations (3.38) for the t^{th} iteration are the following

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left[(\mathrm{E}[\mathbf{H}])^{(t)}\right]^{-1} \left.\frac{\partial \ell}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}^{(t)}} \tag{3.62}$$

$$= \boldsymbol{\theta}^{(t)} - \left[\left(\mathrm{E}\begin{bmatrix} \ell_{\beta\beta} & \ell_{\beta\sigma^2} \\ \ell_{\sigma^2\beta} & \ell_{\sigma^2\sigma^2} \end{bmatrix}\right)^{(t)}\right]^{-1} \begin{bmatrix} \ell_{\beta} \\ \ell_{\sigma^2} \end{bmatrix}^{(t)},$$

where expressions for the first and second derivatives are given in Theorem 3.5.

Example (continued)

Using the matrices from example in Chapter 2, and appropriate equations from this chapter, the results obtained are the following:

MINQUE0 estimates

Solving equation (3.36), the results obtained for $\tilde{\boldsymbol{\sigma}}_0^2$ are:

$$\tilde{\boldsymbol{\sigma}}_0^2 = \begin{bmatrix} \tilde{\sigma}_0^2 \\ \tilde{\sigma}_1^2 \\ \tilde{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 31.9325 \\ 10.2176 \\ 20.4699 \end{bmatrix} \tag{3.63}$$

and, the ordinary least squares solutions for $\widehat{\boldsymbol{\beta}}$ are

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_{01} \\ \widehat{\beta}_{02} \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 91.0000 \\ 88.7500 \\ 83.4286 \\ 82.7778 \end{bmatrix}. \tag{3.64}$$

REML estimates

Solving equation (3.32) using Fisher scoring method (3.38) with MINQUE0 estimates as the initial values, the results are

$$\widehat{\boldsymbol{\sigma}}^2 = \begin{bmatrix} \widehat{\sigma}_0^2 \\ \widehat{\sigma}_1^2 \\ \widehat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 33.2270 \\ 9.8400 \\ 14.9354 \end{bmatrix}. \tag{3.65}$$

Using the REML estimates, the vector $\boldsymbol{\beta}$ can be estimated by

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \\ &= \begin{bmatrix} 91.0000 \\ 88.7500 \\ 83.6535 \\ 83.1278 \end{bmatrix}. \end{aligned} \tag{3.66}$$

MLE

Solving equations (3.6) and (3.8) using Fisher scoring method (3.38) with REML estimates as the initial values, the results are

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_{01} \\ \widehat{\beta}_{02} \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 91.0000 \\ 88.7500 \\ 83.4759 \\ 83.0591 \end{bmatrix} \tag{3.67}$$

and

$$\widehat{\boldsymbol{\sigma}}^2 = \begin{bmatrix} \widehat{\sigma}_0^2 \\ \widehat{\sigma}_1^2 \\ \widehat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 29.4456 \\ 1.0180 \\ 8.0130 \end{bmatrix} \tag{3.68}$$

CHAPTER 4

Hypothesis Testing

Introduction

In experimental designs with mixed model structure, it is often of interest to test hypotheses about the fixed effects. Except in certain special cases, exact tests are not available for unbalanced mixed models. Commonly used techniques developed for balanced mixed models are applied or unknown variance components are treated as known. A research goal is the development of a procedure to analyze experiments with the mixed model structure in equation (2.16). The hypotheses of interest concern expected means from a set of fixed effects, $\boldsymbol{\mu}_0$, and expectations from sets of random effects, $\boldsymbol{\mu}_1$.

Hypotheses

The general linear null and alternative hypotheses of interest are

$$\mathrm{H}_0\colon \mathbf{C}'\boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad \mathrm{H}_a\colon \mathbf{C}'\boldsymbol{\beta} \neq \mathbf{0}, \tag{4.1}$$

where $\mathbf{C}$ is a $p \times r$ matrix of rank $r \leq p$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects and expected random effects, and $\mathbf{0}$ is an $r \times 1$ vector of zeros. For the hypotheses of interest, the tests used are the Likelihood Ratio test, Wald's test with ML and REML estimates either with or without Kackar and Harville correction (Kackar and Harville, 1984), and two approximate F tests.

Likelihood Ratio Test

For the Likelihood function in (3.2), the hypothesis of interest in (4.1), and the MLEs from equations (3.6) and (3.8), the likelihood ratio test statistic (Searle, 1971) is

$$\begin{aligned}\Lambda &= \frac{\sup_{\theta\epsilon\Omega_0} L_{\boldsymbol{\theta}}}{\sup_{\theta\epsilon\Omega} L_{\boldsymbol{\theta}}} \\ &= \frac{L_{\widehat{\boldsymbol{\theta}}_0}}{L_{\widehat{\boldsymbol{\theta}}}},\end{aligned} \tag{4.2}$$

where $\boldsymbol{\theta}$ is defined in (3.14), $L_{\widehat{\boldsymbol{\theta}}_0}$ is the likelihood function in (3.2) maximized under the null hypothesis, $L_{\widehat{\boldsymbol{\theta}}}$ is the likelihood function in (3.2) evaluated at the MLEs.

Under assumptions 3.1 to 3.7 regarding the model in (3.11), the test statistic $\boldsymbol{\Lambda}$ asymptotically follows

$$-2\log_e \boldsymbol{\Lambda} \sim \chi^2_{(r)}, \tag{4.3}$$

where $\chi^2_{(r)}$ is a standard notation for a chi-squared distribution with r degrees of freedom. If at least one m_q is small, then the approximation in (4.3) may be poor.

The test statistic $-2\log_e \boldsymbol{\Lambda}$ in (4.3), when the null hypothesis is not true follows an approximate non-central χ^2 distribution

$$-2\log_e \boldsymbol{\Lambda} \dot{\sim} \chi^2_{(r,\lambda)},$$

where λ may be represented by

$$\lambda = \frac{\boldsymbol{\beta}'\mathbf{C}\left[\mathbf{C}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}\right]^{-1}\mathbf{C}'\boldsymbol{\beta}}{2}.$$

and $\mathbf{C}$ is given in (4.1), $\boldsymbol{\beta}$ is given in (2.29), and $\boldsymbol{\Sigma}$ is given in (2.13).

Wald's Test with MLE Estimates

Under the same conditions described for the likelihood ratio test, the Wald's test statistic, described by Rao (1973), is

$$W = (\mathbf{C}'\widehat{\boldsymbol{\beta}})'[\mathbf{C}'(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})\mathbf{C}]^{-1}(\mathbf{C}'\widehat{\boldsymbol{\beta}}). \tag{4.4}$$

where $\mathbf{C}$ is given in (4.1), $\widehat{\boldsymbol{\beta}}$ is the vector of MLEs, and $\boldsymbol{\Sigma}$ is given in (2.13). $\widehat{\boldsymbol{\Sigma}}$ is estimated using the MLEs.

Under assumptions 3.1 to 3.7 about the model in (3.11), the Wald's test statistic W asymptotically follows

$$W \sim \chi^2_{(r)}, \tag{4.5}$$

The test statistic W in (4.4), when the null hypothesis is not true follows an approximate non-central χ^2 distribution

$$W \dot{\sim} \chi^2_{(r,\lambda)},$$

where λ may be represented by

$$\lambda = \frac{\boldsymbol{\beta}'\mathbf{C}\left[\mathbf{C}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}\right]^{-1}\mathbf{C}'\boldsymbol{\beta}}{2}.$$

and $\mathbf{C}$ is given in (4.1), $\boldsymbol{\beta}$ is given in (2.29), and $\boldsymbol{\Sigma}$ is given in (2.13).

Wald's Test with MLE Estimates

Kackar and Harville's Approximation

The traditional Wald's test procedure uses an estimate of $\text{var}(\mathbf{C}'\widehat{\boldsymbol{\beta}})$ that in some cases seriously underestimates the true variance. Kackar and Harville (1984) suggest that a more conservative estimate, namely $\widehat{\mathrm{var}}_{\mathrm{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}})$, be used instead. The procedure to obtain the new estimate is as follows:

Considering the model in (2.27), let

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \vdots \\ \sigma_k^2 \end{bmatrix}, \tag{4.6}$$

and a set of linear combinations of $\boldsymbol{\beta}$

$$\boldsymbol{\Psi} = \mathbf{C}'\boldsymbol{\beta}. \tag{4.7}$$

In this section a more convenient notation about $\boldsymbol{\Sigma}$ will be used, $\boldsymbol{\Sigma}_{\sigma^2}$ instead of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}_{\sigma^2}$ instead of $\widehat{\boldsymbol{\Sigma}}$. Thus

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_{\sigma^2} &= \mathbf{C}'\widehat{\boldsymbol{\beta}}_{\sigma^2} \\ &= \mathbf{C}'\left(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{y}, \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_{\sigma^2} &= \mathbf{C}'\widehat{\boldsymbol{\beta}}_{\sigma^2} \\ &= \mathbf{C}'\left(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{y}. \end{aligned} \tag{4.9}$$

Note that

$$\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \boldsymbol{\Psi} = \left[\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right] + \left[\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \boldsymbol{\Psi}\right]. \tag{4.10}$$

Kackar and Harville (1984) show that $\left[\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right]$ and $\left[\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \boldsymbol{\Psi}\right]$ are independent. They also show that $\widehat{\boldsymbol{\Psi}}_{\sigma^2}$ is unbiased for $\boldsymbol{\Psi}$. Accordingly,

$$\begin{aligned} \mathrm{var}\left[\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \boldsymbol{\Psi}\right)\right] &= \mathrm{var}\left[\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right)\right] + \mathrm{var}\left[\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \boldsymbol{\Psi}\right)\right] \\ &= \mathrm{var}\left[\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right)\right] + N\mathbf{C}'\left(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X}\right)^{-1}\mathbf{C}. \end{aligned} \tag{4.11}$$

Using first order Taylor series expansion, the approximate variance of $\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right)$ is

$$\text{var}\left[\sqrt{N}\left(\widehat{\boldsymbol{\Psi}}_{\sigma^2} - \widehat{\boldsymbol{\Psi}}_{\sigma^2}\right)\right] \simeq N\mathbf{C}'\mathbf{E}^*\left[\text{var}(\hat{\boldsymbol{\sigma}}^2) \otimes \boldsymbol{\Sigma}_{\sigma^2}\right]\mathbf{E}^{*\prime}\mathbf{C}, \tag{4.12}$$

where $\mathbf{C}$ is given in (4.1),

$$\mathbf{E}^* = [\mathbf{E}_0^* \quad \mathbf{E}_1^* \cdots \mathbf{E}_k^*], \tag{4.13}$$

and

$$\begin{aligned}\mathbf{E}_q^* &= \frac{\partial\left(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}}{\partial\sigma_q^2} \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}^{-1}[\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} - \mathbf{I}_N],\end{aligned} \tag{4.14}$$

for $q = 0, 1, \ldots, k$. The corresponding estimate of $\mathbf{E}_q^*$ is

$$\widehat{\mathbf{E}}_q^* = (\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{Z}_q\mathbf{Z}_q'\boldsymbol{\Sigma}_{\sigma^2}^{-1}[\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1} - \mathbf{I}_N], \tag{4.15}$$

therefore, the expression for the new estimate is

$$\widehat{\text{var}}_{\text{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}}) = \mathbf{C}'(\mathbf{X}'\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{X})^{-1}\mathbf{C} + \mathbf{C}'\widehat{\mathbf{E}}^*(\widehat{\mathbf{V}} \otimes \boldsymbol{\Sigma}_{\sigma^2})\widehat{\mathbf{E}}^{*\prime}\mathbf{C}, \tag{4.16}$$

where $\mathbf{X}$ is given in (2.28), $\boldsymbol{\Sigma}$ is given in (2.13), $\boldsymbol{\Sigma}_{\sigma^2}$ is obtained using the MLEs,

$$\begin{aligned}\widehat{\mathbf{V}} &= \widehat{\text{var}}(\hat{\boldsymbol{\sigma}}^2) \\ &= 2\left[\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \text{tr}\left(\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{G}_q\boldsymbol{\Sigma}_{\sigma^2}^{-1}\mathbf{G}_q\right)\right]^{-1},\end{aligned} \tag{4.17}$$

where $\mathbf{G}_q = \mathbf{Z}_q\mathbf{Z}_q'$, and

$$\hat{\boldsymbol{\sigma}}^2 = \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_k^2 \end{bmatrix}. \tag{4.18}$$

The components of $\widehat{\mathbf{V}}$ are obtained from the inverse of the information matrix, equation (3.22).

Walds test statistic becomes

$$W_{\mathrm{kh}} = (\mathbf{C}'\widehat{\boldsymbol{\beta}})'[\widehat{\mathrm{var}}_{\mathrm{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}})]^{-1}(\mathbf{C}'\widehat{\boldsymbol{\beta}}). \tag{4.19}$$

Corollary 4.1 From Corollary 3.6, when data are balanced,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

does not depend on unknown variance components. Thus

$$\begin{aligned} \mathbf{E}^* &= \frac{\partial\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'}{\partial \boldsymbol{\sigma}^2} \\ &= \mathbf{0}. \end{aligned}$$

Then, in equation (4.16)

$$\mathbf{C}'\widehat{\mathbf{E}}^*(\widehat{\mathbf{V}} \otimes \boldsymbol{\Sigma}_{\hat{\sigma}^2})\widehat{\mathbf{E}}^{*\prime}\mathbf{C} = \mathbf{0}. \tag{4.20}$$

Therefore, the expression for $\widehat{\mathrm{var}}_{\mathrm{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}})$ in (4.16), for balanced data becomes

$$\widehat{\mathrm{var}}_{\mathrm{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}}) = \mathbf{C}'(\mathbf{X}'\boldsymbol{\Sigma}_{\hat{\sigma}^2}^{-1}\mathbf{X})^{-1}\mathbf{C}. \tag{4.21}$$

That is, the Kackar and Harville's correction disappears when data are balanced.

The test statistic W_{kh} in (4.19), when the null hypothesis is not true follows an approximate non-central χ^2 distribution

$$W_{kh} \mathrel{\dot\sim} \chi^2_{(r,\lambda)},$$

where λ may be represented by

$$\lambda = \frac{\boldsymbol{\beta}'\mathbf{C}\left[\mathbf{C}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}\right]^{-1}\mathbf{C}'\boldsymbol{\beta}}{2}.$$

and $\mathbf{C}$ is given in (4.1), $\boldsymbol{\beta}$ is given in (2.29), and $\boldsymbol{\Sigma}$ is given in (2.13).

Wald's Test with REML Estimates

The same test statistic described in (4.4) will be used, except that instead of using the MLEs, the REML estimates will be used.

Wald's Test with REML Estimates

Kackar and Harville's Approximation

The procedure used to obtain $\widehat{\text{var}}_{\text{kh}}(\mathbf{C}'\widehat{\boldsymbol{\beta}})$ in (4.4) will be used, but the REML estimates will be used instead of using MLEs. The same replacement in equation (4.16) is needed to obtain the new W_{kh}, Wald's statistic with Kackar and Harville correction.

Note that $\widehat{\mathbf{V}}$ in (4.17) can be written as

$$\begin{aligned}\widehat{\mathbf{V}} &= \widehat{\text{var}}(\widehat{\boldsymbol{\sigma}}^2) \\ &= 2\left[\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1}\otimes\sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime}\otimes \text{tr}\left(\widehat{\mathbf{P}}\mathbf{G}_q\widehat{\mathbf{P}}\mathbf{G}_q\right)\right]^{-1}, \qquad (4.22)\end{aligned}$$

for $\widehat{\mathbf{P}}$ in (3.30).

Approximate F Test with MINQUE0 Estimates

This approximation was suggested by Kachman (1988). In model (2.27), the ordinary least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

The associated quadratic form for testing the hypothesis in (4.1) is

$$\begin{aligned}SS_H &= \widehat{\boldsymbol{\beta}}'\mathbf{C}\left[\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\right]^{-1}\mathbf{C}'\widehat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{A}\mathbf{y}, \qquad (4.23)\end{aligned}$$

where

$$\mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\left[\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\right]^{-1}\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{4.24}$$

The expectation of SS_H under the null hypothesis is

$$\begin{aligned} \mathrm{E}[SS_H] &= \mathrm{tr}\,(\mathbf{A}\boldsymbol{\Sigma}) \\ &= \mathcal{L}'\boldsymbol{\sigma}^2, \end{aligned} \tag{4.25}$$

where $\boldsymbol{\Sigma}$ is given by (2.13),

$$\mathcal{L} = \sum_{q=0}^{k} \mathbf{e}_{q+1}^{k+1} \otimes \mathrm{tr}\left(\mathbf{Z}_q'\mathbf{A}\mathbf{Z}_q\right), \tag{4.26}$$

and

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \vdots \\ \sigma_k^2 \end{bmatrix}. \tag{4.27}$$

To construct a test statistic, Kachman (1988) suggested estimating $\mathcal{L}'\widehat{\boldsymbol{\sigma}}^2$ by the MINQUE0 procedure with an additional constraint to force independence between SS_H and the estimator of $\mathcal{L}'\boldsymbol{\sigma}^2$.

Theorem 4.2 An unbiased estimator of $\mathcal{L}'\boldsymbol{\sigma}^2$ which is independent of SS_H is given by

$$\widehat{\mathcal{L}'\boldsymbol{\sigma}^2} = \mathbf{y}'\mathbf{B}\mathbf{y}, \tag{4.28}$$

where $\mathcal{L}$ is given by (4.26),

$$\begin{aligned} \mathbf{B} &= \mathbf{A}_0\mathbf{T}\mathbf{A}_0', \\ \mathbf{A}_0\mathbf{A}_0' &= \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \end{aligned}$$

note that $\mathbf{A}_0\mathbf{A}_0'$ is any factorization of $\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, such that $\mathbf{A}_0$ has full column rank, that is $\mathbf{A}_0\mathbf{A}_0' = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The factorization is not unique, so the structure of a particular $\mathbf{A}_0$ is of no importance,

$$\begin{aligned}
\mathbf{T} &= \mathbf{M}'_{\mathrm{B}^*}\sum_{q=0}^{k}\mathbf{T}_q\lambda_{1q}\mathbf{M}_{\mathrm{B}^*}, \\
\mathbf{M}_{\mathrm{B}^*} &= (\mathbf{I}_{N-p} - \mathbf{P}_{\mathrm{B}^*}), \\
\mathbf{P}_{\mathrm{B}^*} &= \mathbf{B}^*(\mathbf{B}^{*\prime}\mathbf{B}^*)^{-}\mathbf{B}^{*\prime}, \\
\mathbf{B}^* &= [\mathbf{b}_1^* \cdots \mathbf{b}_k^*], \\
\mathbf{b}_q^* &= \mathbf{A}'_0\mathbf{Z}_q\mathbf{n}_q, \\
\mathbf{n}_q &= \begin{bmatrix} n_{1q} \\ \vdots \\ n_{m_q q} \end{bmatrix}, \\
\mathbf{T}_q &= \mathbf{A}'_0\mathbf{Z}_q\mathbf{Z}'_q\mathbf{A}_0, \\
\boldsymbol{\lambda}_1 &= \left[\sum_{q=0}^{k}\mathbf{e}_{q+1}^{k+1} \otimes \sum_{q=0}^{k}\mathbf{e}_{q+1}^{(k+1)\prime} \otimes \{\mathrm{tr}(\mathbf{T}_q\mathbf{M}_{\mathrm{B}^*}\mathbf{T}_q\mathbf{M}'_{\mathrm{B}^*})\}\right]^{-1}\mathcal{L},
\end{aligned}$$

and $\boldsymbol{\lambda}_{1q}$ is the q^{th} element of $\boldsymbol{\lambda}_1$, for $q = 0, 1, \ldots, k$.

Proof:

In model (2.14), random vectors $\mathbf{u}_q$ are unknown. Rao's (1972) MINQUE procedure considers a "natural" unbiased estimator, $\widehat{\Phi}$, assuming that the vectors $\mathbf{u}_q$ were observable:

$$\begin{aligned}
\widehat{\Phi} &= \widehat{\mathcal{L}'\boldsymbol{\sigma}}^2 \\
&= \dot{\mathbf{u}}'\,\mathbf{M}\,\dot{\mathbf{u}},
\end{aligned}$$

where

$$\dot{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_0 \\ \vdots \\ \mathbf{u}_k \end{bmatrix},$$

$$\mathbf{M} = \bigoplus_{g=1}^{2} M_g,$$

$$\mathbf{M}_1 = \frac{\mathcal{L}_0}{N}\mathbf{I}_N,$$

$$\mathbf{M}_2 = \bigoplus_{q=1}^{k} \frac{\mathcal{L}_q}{m_q}\mathbf{I}_{m_q},$$

and $\mathcal{L}_q$ is the q^{th} element from the vector $\mathcal{L}$ given in (4.26), for $q = 0, 1, \ldots, k$. In contrast, the use of $\mathbf{y}'\mathbf{B}\mathbf{y}$ as an estimator is considered, because $\dot{\mathbf{u}}$ is not observable. Annihilation of the fixed effects from model (2.14) yields the invariant estimator

$$\begin{aligned}\widehat{\Phi}^* &= \mathbf{y}'\mathbf{A}_0\mathbf{T}\mathbf{A}_0'\mathbf{y} \\ &= \dot{\mathbf{u}}'\mathbf{M}^*\dot{\mathbf{u}},\end{aligned}$$

where

$$\mathbf{M}^* = \mathbf{Z}_{(0)}'\mathbf{A}_0\mathbf{T}\mathbf{A}_0'\mathbf{Z}_{(0)},$$

$$\mathbf{Z}_{(0)} = [\mathbf{Z}_0 \quad \mathbf{Z}], \text{ and}$$

$$\mathbf{Z}_0 = \mathbf{I}_N.$$

The main goal is to find $\mathbf{M}^*$ as close as possible to $\mathbf{M}$, subject to unbiased and independence constraints. Note that

$$\mathbf{B} = \mathbf{A}_0\mathbf{T}\mathbf{A}_0',$$

is a function of $\mathbf{T}$. The procedure involves the minimization of an Euclidean norm

$$\|\mathbf{W}_0^{\frac{1}{2}}(\mathbf{M}^* - \mathbf{M})\mathbf{W}_0^{\frac{1}{2}}\|^2 = \text{tr}\left(\left[\mathbf{W}_0^{\frac{1}{2}}(\mathbf{M}^* - \mathbf{M})\mathbf{W}_0^{\frac{1}{2}}\right]'\left[\mathbf{W}_0^{\frac{1}{2}}(\mathbf{M}^* - \mathbf{M})\mathbf{W}_0^{\frac{1}{2}}\right]\right), \quad (4.29)$$

where $\mathbf{W}_0$ is a matrix of weights that can be written as

$$\mathbf{W}_0 = \bigoplus_{i=1}^{2} \mathbf{W}_{0i}, \tag{4.30}$$

where

$$\mathbf{W}_{01} = \mathrm{w}_0 \mathbf{I}_N, \tag{4.31}$$

$$\mathbf{W}_{02} = \bigoplus_{q=1}^{k} \mathrm{w}_q \mathbf{I}_{m_q}, \tag{4.32}$$

and w_q, for $q = 0, 1, \ldots, k$ are non-zero constants. The values of w_q represent "guesses" for σ_q^2, $q = 0, 1, \ldots, k$.

The norm in (4.29) is minimized subject to the following three constraints:

1. Unbiased constraint

$$\mathrm{E}[\mathbf{y}'\mathbf{B}\mathbf{y}] = \mathcal{L}'\boldsymbol{\sigma}^2 \tag{4.33}$$

which implies,

$$\mathbf{L}'\mathrm{t} = \mathcal{L}, \tag{4.34}$$

where

$$\mathbf{L} = [\mathbf{L}_0 \quad \mathbf{L}_1 \cdots \mathbf{L}_k],$$

$$\mathbf{L}_q = \mathrm{vec}(\mathbf{T}_q),$$

the $\mathrm{vec}(\mathbf{T}_q)$ (vec for vector) is obtained by successively stacking the columns of $\mathbf{T}_q$ into a single vector,

$$\mathbf{T}_q = \mathbf{A}_0'\mathbf{Z}_q\mathbf{Z}_q'\mathbf{A}_0,$$

$$\mathbf{B} \;=\; \mathbf{A}_0\mathbf{T}\mathbf{A}_0',$$