A LEAST SQUARES APPROACH
TO QUADRATIC UNBIASED ESTIMATION
OF VARIANCE COMPONENTS
IN GENERAL MIXED MODEL

by
Cheng-Hsuan Yuan

Institute of Statistics
Mimeograph Series No. 1132
Raleigh, N.C. August 1977
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. PRELIMINARY FORMULATIONS</td>
<td>7</td>
</tr>
<tr>
<td>2.1 The General Mixed Model</td>
<td>7</td>
</tr>
<tr>
<td>2.2 The Inner Product Space A and the Linear Transformation H</td>
<td>9</td>
</tr>
<tr>
<td>3. THE LEAST SQUARES APPROACH TO QUADRATIC UNBIASED ESTIMATION OF VARIANCE COMPONENTS</td>
<td>13</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2 General Results</td>
<td>15</td>
</tr>
<tr>
<td>3.3 Quadratic Least Squares Estimation of Variance Components</td>
<td>19</td>
</tr>
<tr>
<td>3.4 Invariant Estimation of Variance Components</td>
<td>30</td>
</tr>
<tr>
<td>4. ANOTHER LOOK AT SOME OTHER ESTIMATION PROCEDURES</td>
<td>38</td>
</tr>
<tr>
<td>4.1 Seely's Work on Quadratic Unbiased Estimation in Mixed Models</td>
<td>38</td>
</tr>
<tr>
<td>4.2 Koch's Symmetric Sum Approach to the Estimation of Variance Components</td>
<td>42</td>
</tr>
<tr>
<td>4.3 A Least Squares Approach to the MINQUE</td>
<td>46</td>
</tr>
<tr>
<td>4.4 A Geometric Approach to Quadratic Least Squares Estimation and the MINQUE</td>
<td>52</td>
</tr>
<tr>
<td>5. CONCLUSIONS</td>
<td>57</td>
</tr>
<tr>
<td>6. LIST OF REFERENCES</td>
<td>60</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Estimation of variance components has been a long-standing problem associated with statistical inference in general linear models. In its early stage of development, the estimators were derived as linear combinations of sums of squares from analysis of variance tables for testing linear hypotheses. The well known methods proposed by Henderson (1953) and extended to more complex models by Searle (1968) still retained the form as linear combinations of sums of squares. In the case of balanced models, the estimators obtained from the analysis of variance have been proved to be uniformly minimum variance quadratic unbiased estimators (cf. Graybill, 1961). However, for the unbalanced models, the analysis of variance methods do not, in general, give unbiased estimates for all quadratic estimable functions. Searle (1968) pointed out Henderson's Method 1 does not give unbiased estimates in mixed models and Method 2 cannot be used if there is an interaction between fixed and random effects. Harville (1967) gave an example where Henderson's Method 3 does not give unbiased estimates of estimable variance components in a two-way classification random model. Limitation in the estimability of analysis of variance methods is due to the restriction that the estimators are linear combinations of sums of squares instead of all quadratic forms of observations.

Some significant developments took place in the last decade. Koch (1967) proposed a symmetric sum approach to the estimation of variance components. The underlying idea was to consider the linear combination of all squares and cross products of the observations, as Koch stated, "Any estimable parameter can be estimated by forming an appropriate
linear combination of certain symmetric sums of (linearly independent) random variables having the same expectation such that the resulting quadratic form is an unbiased estimate of it." He considered the observation vector $Y$ from a certain experiment and denoted

$$\text{E}(Y) = \mu, \quad \text{Var}(Y) = V$$

and

$$\text{E}(YY') = \mu\mu' + V = S.$$ 

Given the design of the experiment, $\mu$ and $V$ have a particular linear structure which can be utilized in constructing estimators. With this basic theory Koch then demonstrated the symmetric sums approach in a class of random models including some nested models, cross classification models and an experiment of mixed type. Even though the general form of the symmetric sum estimators was not formulated, the general feature of the method in different models is clear. The estimating procedure consists of calculating the averages of the entries in the matrix $YY'$ having the same expectation ($i.e.$, the identical entries in $S$), equating the averages to their expectations and solving the set of equations. For the models Koch used to demonstrate his methods, the coefficient matrices of the equations are non-singular. Therefore the solutions are unbiased estimators of the parameters. The estimators were found to be not invariant under shifts in location of the data. Koch (1968) modified the symmetric sum method by using squared differences of observations instead of the squares and cross products of observations. Forthofer and Koch (1974) extended the method of mixed models and the cross products of differences of observations were also used as well as squared differences.
In his work on estimation in finite-dimensional linear spaces, Seely (1969, 1970a, 1970b) treated the quadratic unbiased estimation in mixed models as a special case of his general theory. He also considered the expectation of $YY'$ as Koch did. Based on the linear structure in

$$E(Y) = X\beta \quad \text{and} \quad \text{Var}(Y) = \sum_{i=1}^{m} V_i \sigma^2_i$$

he introduced the representation

$$E(YY') = H\theta$$

(1.1)

where $\theta$ is the $M \times 1$ vector of parameters and $H$ is a linear transformation from $\mathbb{R}^M$ into $\mathcal{A}$, the linear space of all $n \times n$ symmetric matrices. Seely first considered the general case of quadratic estimation of linear functions of $\theta$ in (1.1). He obtained necessary and sufficient conditions for the estimability of linear functions $Z'\theta$ and also derived equations for the estimation of $Z'\theta$. As will be shown in Chapter 4 of this dissertation, Seely's estimation equations lead to estimators identical to Koch's (1967) in the random models considered by Koch. Seely also derived a set of equations for estimation of variance components but the consistency of the equations was not verified. Above all, Seely's work presents an important motivation to the present work on the least squares approach to quadratic estimation of variance components.

Another significant development was Rao's (1970, 1971a, 1971b, 1972) theory of minimum norm quadratic unbiased estimation (known as the MINQUE). In the general mixed model

$$Y = X\beta + U_\varepsilon$$
suppose the random effects \( \varepsilon \) were observable. Then a natural estimator
of \( \sum_{i=1}^{m} p_i \sigma_i^2 \) would be \( \varepsilon' \Delta \varepsilon \) with \( \Delta \) a suitably defined diagonal matrix.

If \( Y'AY \) with \( AX = 0 \) is an unbiased estimator of \( \sum_{i=1}^{m} p_i \sigma_i^2 \), then the difference between \( Y'AY \) and \( \varepsilon' \Delta \varepsilon \) is \( \varepsilon'(U'AU - \Delta)\varepsilon \). The MINQUE, as Rao stated, is to make \( \varepsilon'(U'AU - \Delta)\varepsilon \) "small in some sense by minimizing \( \|U'AU - \Delta\|\)", or equivalently, by minimizing \( \|U'AU\| \). In case that condition \( AX = 0 \) is not consistent with unbiasedness, then the MINQUE will be obtained by minimizing \( \text{tr}(A(V + 2XX')AV) \) where \( V = UU' \). In either case, it will be shown in Chapter 4 that the MINQUE can be considered a special case of the least squares quadratic estimation of variance components.

The aim of the present work is two-fold:

1. To extend and apply the theory of least squares estimation to quadratic estimation of variance components in the general mixed model.

2. To demonstrate that the variance component estimation procedures proposed by some previous authors can be unified in the general quadratic least squares approach.

Since Gauss' work in the early years of the nineteenth century, the theory of least squares, through the effort of various authors, has been established with much generality in linear estimation. To the present author, it seems to be a natural extension to apply the least squares principles to the quadratic estimation in mixed models. In this dissertation we shall demonstrate this without making special assumptions on the distribution of the random variables, except for the existence
of the second degree moment. Quadratic estimates $Y'AY$ are linear functionals of $YY'$. Application of least squares estimation with the symmetric matrix $YY'$ as the dependent variable does not fit into the conventional form of least squares estimation where the dependent variable is in the form of a column vector. To facilitate the application of the principle of least squares estimation, Seely's representation $E(YY') = H\theta$ is adopted and linear transformations on the linear space $A$ of all n x n symmetric matrices are defined and derived. These will be presented as part of the results in Chapter 3. In Chapter 4 we shall verify that Koch's symmetric sums estimators and the MINQUE are both special cases of least squares estimators. Also, Seely's work will be further discussed.

At the completion of the present work, a recent article by Pukelsheim (1976) was brought to the author's attention. Pukelsheim considered the model $E(Y) = X\beta$, $\text{Var}(Y) = \sum_{i=1}^{k} V_i t_i$, and derived the model

$$E(MY \otimes MY) = D_M^t$$

(1.2)

where $M$ is the orthogonal projection matrix on the null space of $X'$, and $D_M$ is the known coefficient matrix of the vector $t$ of variance components. Among the results derived from this model, he pointed out that

$$\hat{t} = D_M^+ \cdot MY \otimes MY$$

(1.3)

where $D_M^+$ is the Moore-Penrose inverse of $D_M$, is the ordinary least squares estimate of $t$ and he also showed that, with a suitably chosen positive definite matrix $T$, Rao's MINQUE with invariance can be obtained from the normal equation

$$D'(M \otimes M \cdot T \otimes T \cdot M \otimes M)^+ D t = D'(M \otimes M \cdot T \otimes T \cdot M \otimes M)^+ Y \otimes Y$$

(1.4)
The above results coincide with part of the findings in the present work which is developed independently of Pukelsheim's work. In Chapters 3 and 4 in this dissertation, special remarks will be made where the equivalence to Pukelsheim's results is observed.
2. PRELIMINARY FORMULATIONS

2.1 The General Mixed Model

By the general mixed model we mean the nx1 vector $Y$ of observations with the linear structure

$$ Y = X\beta + \sum_{i=1}^{m} U_i \varepsilon_i $$

(2.1)

where $X$ is an npx matrix of known constants; $\beta$ is a px1 vector of unknown constants; $U_i$ are nxq_i matrices of known constants; $\varepsilon_i$ are q_ixl vectors of random variables with

$$ E(\varepsilon_i) = 0 \quad \text{and} \quad Var(\varepsilon_i) = C_i \sigma_i^2 $$

where $C_i$ are q_i xq_i matrices of known constants and

$$ E(\varepsilon_i \varepsilon_j') = 0, \quad \text{for } i \neq j. $$

With q_m = n and $U_m$ and $C_m$ non-singular, this model is a generalized representation of the so-called fixed, random and mixed models. When p = 1, (2.1) is a random model and when m = 1, it is a fixed model. From model (2.1), it follows immediately that

$$ E(Y) = X\beta = \sum_{i=1}^{p} X_i \beta_i $$

(2.2)

and

$$ Var(Y) = \sum_{i=1}^{m} U_i C_i U_i' \sigma_i^2 = \sum_{i=1}^{m} V_i \sigma_i^2 $$

where $\beta_i$ and $X_i$ are the ith entry in $\beta$ and the ith column of $X$, respectively, and $V_i = U_i C_i U_i'$. Denote $W = YY'$. Then
\[ E(W) = \sum_{i=1}^{p} \sum_{j=1}^{p} X_i X_j^\prime \beta_i \beta_j + \sum_{i=1}^{m} V_i^2. \]

Denote
\[
\begin{align*}
B_{ii} &= X_i X_i^\prime, \quad i=1, \ldots, p \\
B_{ij} &= X_i X_j^\prime + X_j X_i^\prime, \quad i \neq j \\
\theta_{ij} &= \beta_i \beta_j, \quad i,j=1, \ldots, p
\end{align*}
\]

and
\[ M = \frac{1}{2}p(p+1) + m. \]

Then \[ E(W) = \sum_{i=1}^{p} \sum_{j=1}^{p} B_{ij} \theta_{ij} + \sum_{i=1}^{m} V_i^2. \] (2.3)

With this fixed set \{B_{ij}, V_i\} of M symmetric matrices, Seely (1969, 1970b) introduced a linear transformation H from \( \mathbb{R}^M \) into the linear space \( \mathbb{A} \) of all \( n \times n \) symmetric matrices over the field of real numbers.

The linear transformation H is defined as
\[ H = \sum_{i=1}^{p} \sum_{j=1}^{p} B_{ij} \rho_{ij} + \sum_{i=1}^{m} V_i \rho_i \] (2.4)

for every Mx1 vector \( \sigma = (\rho_{11} \ldots \rho_{pp} \ldots \rho_{12} \ldots \rho_{pp}^2 \ldots \rho_{mm}^2)' \) in \( \mathbb{R}^M \), where \( V_i \) and \( B_{ij} \) are given in (2.2) and (2.3), respectively. With this definition, we write Seely's representations as
\[ E(W) = H \theta \] (2.5)

where \( \theta = (\theta_{11} \ldots \theta_{pp} \theta_{12} \ldots \theta_{pp}^2 \ldots \theta_{mm}^2) \) and \( \theta_{ij} \) are given in (2.3).

To facilitate the later development, it is helpful to make the partition
\[ H \theta = H_1 \theta_1 + H_2 \theta_2 \] (2.6)
where \( H_1 \theta_1 = \sum_{i=1}^{p} \sum_{j=1}^{p} B_{ij} \theta_{ij} \) and \( H_2 \theta_2 = \sum_{i=1}^{m} V_i \sigma_i^{2} \) \hspace{1cm} (2.7)

Accordingly, we partition the parameter space \( \Omega \) as the product space \( \Omega_1 \times \Omega_2 \), where \( \Omega_1 \) is the set of all \( \frac{1}{2}^{p(p+1)} \) parametric vectors \( \theta_1 \) and \( \Omega_2 \) is the set of all \( mx1 \) parametric vectors \( \theta_2 \). We adopt Seely's assumption on \( \Omega \) that \( \lambda' \theta = 0 \) for all \( \theta \) in \( \Omega \) implies \( \lambda = 0 \). Actually, this is the consequence of the conventional assumption that \( -\infty < \beta_i < +\infty \), \( 0 < \sigma_i^{2} < +\infty \) and that \( \beta_i, \beta_j, \sigma_k, \sigma_j^{2} \) are functionally independent whenever \( i \neq j \) and \( k \neq l \). (Any linear dependence among \( \beta_i \) can be removed by reparameterization.)

Because \( \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_{ij} \beta_i \beta_j = 0 \) for all \( \beta \) in \( R^p \) implies \( \lambda_{ij} = 0, 1 \leq i \leq j \leq p \)
and \( \sum_{i=1}^{m} \lambda_i \sigma_i^{2} = 0 \) for all \( \theta_2 \) in \( \Omega_2 \) implies \( \lambda_i = 0, 1 \leq i \leq m \).

2.2 The Inner Product Space \( A \) and the Linear Transformation \( H \)
We are interested in the quadratic forms \( Y'AY \) whose expectation is a certain linear functional \( \lambda' \theta \). Since we intend to deal with general model (2.1) or (2.5) and the parametric linear functionals \( \lambda' \theta \), we shall consider every \( nxn \) symmetric matrix as a potential candidate to be used in the quadratic form of \( Y \). Denote \( A \) to be the linear space of all \( nxn \) symmetric real matrices. Since \( W = YY' \) is in \( A \) for every \( Y \) in \( R^n \) and we want to express \( Y'AY \) as a linear functional of \( W \), we need to define an inner product \( ( , ) \) on \( A \) such that

\( (A,W) = Y'AY \) for every \( A \) in \( A \) and \( Y \) in \( R^n \).

Consequently the inner product is to be defined, for any pair \( A,B \) in \( A \), as the trace of the product matrix \( AB \), i.e.,

\( (A,B) = tr(AB) \) for every \( A,B \) in \( A \).
Also the inner product $<,>$ on $\mathbb{R}^n$ is defined to be

$$<a, b> = a^t b$$

for every $a, b$ in $\mathbb{R}^n$.

With the inner products on $A$ and $\mathbb{R}^M$ defined, the adjoint $H^*$ of $H$ is determined as the unique linear transformation from $A$ into $\mathbb{R}^M$ such that

$$(H_\rho, A) = <\rho, H^* A>$$

for every $A$ in $A$, $\rho$ in $\mathbb{R}^M$.

Since

$$(H_\rho, A) = \left( \sum_{i=1}^P \sum_{j=1}^P B_{ij} \rho_{ij} + \sum_{i=1}^m V_i \rho_i, A \right)$$

$$= \sum_{i=1}^P \sum_{j=1}^P (B_{ij}, A) \rho_{ij} + \sum_{i=1}^m (V_i, A) \rho_i$$

$$= (\rho_{11} \ldots \rho_{pp} \rho_{1 \ldots} \rho_m)$$

therefore, $H^*$ is defined by

$$H^* A = \begin{bmatrix} (B_{11}, A) \\ \vdots \\ (B_{pp}, A) \\ (V_1, A) \\ \vdots \\ (V_m, A) \end{bmatrix}.$$
Corresponding to the partition of \( H \) in (2.6), we partition \( H^* \) as

\[
H^*A = \begin{bmatrix}
H^*_1 \\
H^*_2
\end{bmatrix}
\]

where \( H^*_1 \) and \( H^*_2 \) are the adjoints of \( H_1 \) and \( H_2 \) respectively. \( H^*_1 A \) is the upper \( \frac{1}{2} p(p+1) \times 1 \) subvector in \( H^*A \) and \( H^*_2 A \) the lower \( mx1 \) subvector.

We denote \( R(H) \) to be the range space of \( H \), \( r(H) \) the rank of \( H \), \( N(H^*) \) the null space of \( H^* \) and \( n(H^*) \) the nullity of \( H^* \).

Based on the fact that \( (A, H_0) = \langle H^* A, \rho \rangle \) for every \( A \) in \( A \) and \( \rho \) in \( \mathbb{R}^M \), it can easily be proved that \( R(H) \) and \( N(H^*) \) are orthogonal complements in \( A \). It follows that for every \( A \) in \( A \), there exist unique \( A_1 \) in \( R(H) \) and \( A_0 \) in \( N(H^*) \) such that \( A = A_1 + A_0 \). Since \( A_1 \) is in \( R(H) \), \( A_1 = H_0 \) for some \( \rho \) in \( \mathbb{R}^M \). Therefore

\[
H^* A = H^* A_1 = H^* H_0.
\]

This implies \( R(H^*) = R(H^* H) \). It should be noted that the composite linear transformation \( H^* H \) maps from \( \mathbb{R}^M \) into \( \mathbb{R}^M \) and can be expressed as an \( n \times n \) matrix. Seely showed that

\[
H^* H = \begin{bmatrix}
(B_{11}, B_{11}) & \cdots & (B_{11}, B_{pp}) & (B_{11}, V_1) & \cdots & (B_{11}, V_m) \\
& \ddots & \vdots & \vdots & \ddots & \vdots \\
& & \ddots & \vdots & \ddots & \vdots \\
(B_{pp}, B_{11}) & \cdots & (B_{pp}, B_{pp}) & (B_{pp}, V_1) & \cdots & (B_{pp}, V_m) \\
& \ddots & \vdots & \vdots & \ddots & \vdots \\
& & \ddots & \vdots & \ddots & \vdots \\
(V_1, B_{11}) & \cdots & (V_1, B_{pp}) & (V_1, V_1) & \cdots & (V_1, V_m) \\
& \ddots & \vdots & \vdots & \ddots & \vdots \\
& & \ddots & \vdots & \ddots & \vdots \\
(V_m, B_{11}) & \cdots & (V_m, B_{pp}) & (V_m, V_1) & \cdots & (V_m, V_m)
\end{bmatrix}
\]
Further, since

\[ \rho' H^* H \rho = (H \rho, H \rho) \geq 0 \] for every \( \rho \) in \( \mathbb{R}^M \),

the matrix \( H^* H \) is non-negative definite and is positive definite if and only if the set \( \{ B_{ij}, V_i \} \) is linearly independent.
3. THE LEAST SQUARES APPROACH TO QUADRATIC UNBIASED ESTIMATION 
OF VARIANCE COMPONENTS

3.1 Introduction

As was stated in Chapter 1, one of the aims of the present work is 
to extend and apply the theory of least squares to quadratic estimation 
of variance components in general mixed models. In this section we shall 
outline some fundamental principles of least squares estimation and 
describe the basic analogy between linear and quadratic estimation in 
application of the theory of least squares.

Let us consider a finite dimensional real inner product space \( U \). 
Let \( S \) be a subspace of \( U \) and \( S^\perp \) the orthogonal complement of \( S \). The 
 orthogonal projection of a point \( u \) in \( U \) on \( S \) is the point \( \hat{u} \) in \( S \) such 
that \( u - \hat{u} \) is in \( S^\perp \). Since

\[
\| u - \hat{u} \| = \min_{s \in S} \| u - s \| ,
\]

\( \hat{u} \) is called the best approximation to \( u \) by elements in \( S \) (cf. p. 283, 

In the case of linear estimation, we consider the observation \( Y \) 
taken from an experiment to be an nx1 vector. The expectation \( \mu \) of \( Y \) 
is unknown but the model implies that \( \mu = X\beta \), meaning that \( \mu \) lies in 
the range space \( R(X) \) of an nxp matrix \( X \), a linear transformation from 
\( \mathbb{R}^p \) into \( \mathbb{R}^n \). The least squares estimate \( \hat{\mu} \) of \( \mu \) is the best approximation 
to \( Y \) by elements in \( R(X) \) in the sense that

\[
\| Y - \hat{\mu} \| = \min_{x \in R(X)} \| Y - x \| .
\]
Denote $P$ to be the orthogonal projection on $R(X)$, \textit{i.e.}, $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $P = P' = P^2$ and $Px = x$ for every $x$ in $R(X)$. Then $\hat{Y} = PY$ is the best approximation to $Y$ in $R(X)$. Also, $\hat{Y}$ has the same expectation as $Y$ and if $\ell'Y$ is an unbiased estimate for a certain parametric function, $\ell'\hat{Y}$ is the least squares estimate.

In quadratic unbiased estimation in general mixed models, the expectation of $W = YY'$ is $S = \mu\mu' + V$. The representation $S = H\theta$ reveals that $S$ lies in the range space of $H$. Therefore the quadratic least squares estimate of $S$ is the orthogonal projection $W_1$ of $W$ on $R(H)$, and if $(A,W)$ is unbiased for a certain parametric function, then the quadratic least squares estimate of that function is $(A,W_1)$.

In the following sections of this chapter we shall work in the following areas:

(1) The explicit form of the orthogonal projection on $R(H)$ will be presented in terms of $H$ and $H^*$. 

(2) Quadratic least squares estimators of variance components will be derived. A linear operator $T$ on $A$ such that the expectation of $W$ under the mapping of $T$ is not a function of $\theta_1$ will be introduced.

(3) Quadratic least squares estimators invariant with respect to the fixed effects $\beta$ will be obtained. A quadratic estimator $Y'AY$ is invariant with respect to $\beta$ if and only if $AX = 0$. A linear operator whose range space is $\{A \in A : AX = 0\}$ will be introduced.
3.2 General Results

In this section, we shall deal with the least squares quadratic unbiased estimation of the parametric functions of the form

$$<\ell, \theta> = \sum_{i=1}^{p} \sum_{j=1}^{p} \ell_{ij} \theta_{ij} + \sum_{i=1}^{m} \ell_{i1} \sigma_{0i}^2.$$  \hspace{1cm} (3.1)

We adopt the following definition of quadratic estimability which was given by Graybill and Hultquist (1960), Koch (1967) and Seely (1969).

**Definition 3.1.** A parametric function of the form (3.1) is said to be quadratic estimable if there is an nxm symmetric matrix A such that

$$E(A, W) = <\ell, \theta>$$

for all $\theta$.

In the following we first outline Seely's results and then undertake our own exploration from there on.

Based on his representation $E(W) = H\theta$, Seely (1969, 1970b) obtained the following:

(1) The parametric function $<\ell, \theta>$ is quadratic estimable if, and only if, there exist $\rho$ in $\mathbb{R}^M$ such that

$$H^*H\rho = \ell.$$  \hspace{1cm} (3.2)

(2) If $<\ell, \theta>$ is quadratic estimable and if $\hat{\theta}$ is a solution of the equation

$$H^*H\hat{\theta} = H^*W,$$  \hspace{1cm} (3.3)

then $<\ell, \hat{\theta}> = \sum_{i=1}^{p} \sum_{j=1}^{p} \ell_{ij} \hat{\theta}_{ij} + \sum_{i=1}^{m} \ell_{i1} \hat{\sigma}_{0i}^2$ is an unbiased estimator for $<\ell, \theta>$.
Based on these results given by Seely, we now proceed as follows. An immediate consequence from Definition 3.1 is the following lemma.

**Lemma 3.1.** A parametric function $\langle \lambda, \theta \rangle$ is quadratic estimable if, and only if, there exist $A$ in $\mathcal{A}$ such that

$$H^*A = \lambda .$$

(3.4)

**Proof.** Since

$$E(A, W) = (A, H\theta) = \langle H^*A, \theta \rangle$$

for every $\theta$ in $\Omega$,

that $\langle \lambda, \theta \rangle$ is quadratic estimable implies $H^*A = \lambda$ and vice versa.

Comparing (3.2) and (3.4) we have $A - H\rho$ is in $N(H^*)$. Therefore $H\rho$ is the orthogonal projection of $A$ on $R(H)$. In general, the orthogonal projection of $A$ in $\mathcal{A}$ on $R(H)$ is obtained from the equation

$$H^*H\rho = H^*A .$$

(3.5)

Observe that this equation is consistent since $R(H^*) = R(H^*H)$. Therefore the solution exists and is of the form

$$\rho = (H^*H)^{-1}H^*A ,$$

where $(H^*H)^{-1}$ is a generalized inverse of $(H^*H)^{-1}$. When $(H^*H)^{-1}$ is not unique (i.e., when $H^*H$ is singular), $\rho = (H^*H)^{-1}H^*A$ is not the unique solution of (3.5). But $H\rho = H(H^*H)^{-1}H^*A$, the orthogonal projection of $A$ on $R(H)$ is unique regardless of the choice of $(H^*H)^{-1}$. Therefore, the linear operator $H(H^*H)^{-1}H^*$ on $\mathcal{A}$ is the orthogonal projection on $R(H)$.

An alternative proof that $H(H^*H)^{-1}H^*$ is the orthogonal projection on $R(H)$ is given in the following.
**Theorem 3.1.** The linear operator $H(H^*H)^{-1}H^*$ on $A$ is unique, regardless of the choice of $(H^*H)^{-1}$, symmetric and idempotent, in short, $H(H^*H)^{-1}H^*$ is the orthogonal projection on $R(H)$.

**Proof.**

(1) **Uniqueness:** Let $\phi$ and $\psi$ be two distinct solutions of the equation (3.5). Then

$$H^*H\phi = H^*w = H^*H\psi$$

or $H\phi - H\psi$ is in $N(H^*)$. But $H\phi - H\psi$ is also in $R(H)$. Therefore, $H\phi - H\psi = 0$ and $H\phi = H(H^*H)^{-1}HA$ is unique for every $A$ in $A$.

(2) **Symmetry:** Since $H^*H$ is symmetric, we can choose $(H^*H)^{-1}$ to be symmetric. Since $H(H^*H)^{-1}H^*$ is unique, $H(H^*H)^{-1}H^*$ must be symmetric, i.e.,

$$(A, H(H^*H)^{-1}HB) = (H(H^*H)^{-1}HA, B)$$

for every $A, B$ in $A$.

(3) **Idempotency:** First observe that

$$H^* = H^*H(H^*H)^{-1}H^* .$$

This is true because from (3.5)

$$H^*A = H^*H\phi = H^*H(H^*H)^{-1}HA$$

for every $A$ in $A$. It follows that

$$(H(H^*H)^{-1}H^*)^2 = H(H^*H)^{-1}H^* ,$$

i.e., $H(H^*H)^{-1}H^*$ is idempotent.
That $\hat{H}(H^*H)^{-1}H^*$ is the orthogonal projection on $R(H)$ follows by Theorem 1, §75 of Halmos (1958).

Now let us go back to the equation (3.3). The above results have shown that $\hat{H}$ is the orthogonal projection of $W$ on $R(H)$. Therefore, for any $\hat{\theta}$ satisfying the equation (3.3), $\hat{H}\hat{\theta}$ is the quadratic least squares estimator of $H\theta$ in the sense that

$$\|W - H\hat{\theta}\| = \min_{\theta \in R^M} \|W - H\theta\|.$$ 

In other words, the equation (3.3) is equivalent to the equation

$$\frac{\partial}{\partial \theta} \|W - H\hat{\theta}\|^2 = 0. \tag{3.6}$$

However, if we want to minimize $\|W - H\hat{\theta}\|^2$ with the restriction that $\theta$ is in $\Omega$, a subset of $R^M$, then we have the following non-linear program

$$\min \{ \|W - H\hat{\theta}\|^2 \mid \hat{\theta} \in \Omega \}. \tag{3.7}$$

The solution of (3.7) will preserve the functional dependence among the parameters,

$$\hat{\theta}_{ij}^2 = \theta_{ii} \theta_{jj}, \tag{3.8}$$

in the estimates $\hat{\theta}_{ij}$ and assures positive value of $\hat{\theta}_{ii}^2$. In general, solution of (3.7) is neither quadratic in $Y$ nor unbiased for $\theta$. Therefore, solving the non-linear program (3.7) is not within the scope of the present work and we do not intend to pursue this problem any further.
3.3 Quadratic Least Squares Estimation of Variance Components

From now on we shall concentrate on the estimation of parametric functions of the form

$$<x_2, \theta> = \sum_{i=1}^{m} x_i^2 \sigma_i^2$$  \hspace{1cm} (3.9)

If this parametric function is estimable, then by Lemma 3.1, there exists \( A \) in \( \hat{A} \) such that

$$H_1^* A = 0 \quad \text{and} \quad H_2^* A = x_2. \quad \text{We therefore have the following.}$$

**Lemma 3.2.** The parametric function (3.9) is estimable if, and only if, there exists \( A \) in \( N(H_1^*) \) such that \( x_2 = H_2^* A \).

It should be noted that \( A \) in \( N(H_1^*) \) means

$$(B_{ij}, A) = 0, \quad 1 \leq i \leq j \leq p,$$

which, in turn, is equivalent to \( X'AX = 0 \). Therefore we have

$$N(H_1^*) = \{A \in \hat{A}: X'AX = 0\}. \quad \text{(3.10)}$$

Consider the linear operators on \( \hat{A} \) whose range space coincides with \( N(H_1^*) \). Some of these operators will be discussed in Chapter 4 when we discuss the estimation procedures of other authors. The one of special importance is introduced in the following.

**Lemma 3.3.** Let \( T \) be a linear operator on \( \hat{A} \) defined by

$$TA = A - PAP \quad \text{for every } A \in \hat{A}, \quad \text{(3.11)}$$

where \( P \) is the orthogonal projection matrix on the range space of \( X \).
Then,

(1) \( R(T) = \{ A \in A : X'AX = 0 \} \),

(2) \( T = T^2 = T^* \).

In other words, \( T \) is the orthogonal projection on \( N(H_1^*) \).

**Proof.**

(1) Since

\[ X'(A - PAP)X = X'AX - X'AX = 0 \]

therefore \( A \) in \( R(T) \) implies \( A \) in \( N(H_1^*) \) and this shows

\[ R(T) \subseteq N(H_1^*) \].

If \( A \) is in \( N(H_1^*) \), that is, \( X'AX = 0 \), then \( TA = A \) since

\( PAP = 0 \). Therefore, \( A \) is in \( R(T) \) and \( N(H_1^*) \supseteq R(T) \).

Hence \( R(T) = N(H_1^*) \).

(2) Since

\[ T(PAP) = PAP - PAP = 0 \]

\[ T^2(A) = T(A - PAP) = TA - T(PAP) = TA \].

Therefore \( T = T^2 \).

To prove \( T = T^* \), let \( A \) and \( B \) denote any two fixed members of \( A \). Then,

\[ (TA,B) = (A - PAP,B) \]

\[ = (A,B) - (PAP,B) \]

\[ = (A,B) - (A,PBP) \]
\[ = (A, B - PBP) = (A, TB). \]

Therefore \( T = T^* \).

First of all, the above proof serves the purpose of demonstrating the existence of a linear operator whose range space coincides with \( N(H^*_1) \). With this, we can restate Lemma 3.2 as follows.

**Lemma 3.4.** Let \( T \) be a linear operator given in (3.11). Then the parametric function \( \ell_2 = \langle \ell_2, \theta_2 \rangle \) is estimable if, and only if,

\[ \ell_2 = H^*_2TA \]

for some \( A \) in \( A \).

Consider the composite transformation \( TH \). Since \( R(T) = N(H^*_1) \) and \( N(H^*_1) \) and \( R(H_1) \) are orthogonal complements, it follows that \( TH_1 \) is a zero transformation, that is, \( TH_1 \rho_1 = 0 \) for every \( \rho_1 \) in \( R^{p(p+1)/2} \). Therefore, \( TH = TH_2 \rho_2 \), where \( \rho_2 \) is the subvector of \( \rho \) of order \( m \). Consequently, \( R(TH) = R(TH_2) \) and \( N(H^*T) = N(H^*_2T) \). Also, if \( A_0 \) is a member of \( N(H^*) \), then since \( N(H^*) \subset N(H^*_1) = R(T) \), \( TA_0 = A_0 \). Therefore, \( H^*TA_0 = H^*A_0 = 0 \) and we have \( N(H^*) \subset N(H^*T) = N(H^*_2T) \).

For every \( A \) in \( A \), put \( A = A_1 + A_2 \) with \( A_1 \) in \( R(H) \) and \( A_0 \) in \( N(H^*) \). We have

\[ H^*_2TA = H^*_2TA_1 + H^*_2TA_0 \]

\[ = H^*TH_2 \rho \]

\[ = H^*TH_2 \rho_2. \]

The above shows that \( R(H^*_2T) = R(H^*_2T) \). This, together with Lemma 3.4, proves the following.
Lemma 3.5. Let $T$ be the linear operator defined in (3.11). Then the parametric function $<\ell_2, \theta_2>$ is estimable if, and only if, there exists $\rho_2$ in $\mathbb{R}^m$ such that

$$H_2^T \rho_2 = \ell_2.$$

(3.12)

Notice that the transformation $H_2^T \rho_2$ is a linear operator on $\mathbb{R}^m$ and is expressible in the form of an $m \times m$ matrix as follows.

$$
H_2^T \rho_2 = \begin{bmatrix}
(V_1, V_1 - PV_1P) & \ldots & (V_1, V_m - PV_mP) \\
(V_2, V_1 - PV_1P) & \ldots & (V_2, V_m - PV_mP) \\
\vdots & \ddots & \vdots \\
(V_m, V_1 - PV_1P) & \ldots & (V_m, V_m - PV_mP)
\end{bmatrix}
$$

This provides us a useful device for examining the estimability of $<\ell_2, \theta_2>$ by the consistency of equation (3.12).

Let $<\ell_2, \theta_2>$ be quadratic estimable and $(A, W)$ be an unbiased estimator. From the above results, we must have $A$ in $R(T)$, i.e., $A = TA$. Therefore, $(A, W) = (TA, W) = (A, TW)$ since $T$ is symmetric. This implies that, in the quadratic unbiased estimation of variance components, we get identical estimators whether we use $W$ or $TW = W - PWP$. Since in general, $W$ is in $A$ and $TW$ is in $R(T)$, we can restrict our attention in $R(T)$ rather than the whole space $A$. Hence, we shall derive estimators from the transformed model

$$E(TW) = TH\theta = TH_2 \theta_2.$$

(3.13)
Since \( TW = W - PWP \) is the orthogonal projection of \( W \) on \( N(H_1^*) \),
PWP = W - TW is the orthogonal projection of \( W \) on \( R(H_1) \). In other words,

\[
PWP = H_1(H_1^*H_1)^{-1}H_1^*W .
\]  

(3.14)

The above expression should be the least squares estimator of \( H_0 \) if the model were \( E(W) = H_1 \theta_1 \). Of course, this is a biased estimator of \( H_1 \theta_1 \) but by analogy with linear estimation, \( TW = W - PWP \) can be considered as "The data adjusted for the fixed effect parameters in \( \theta_1 \)."

Let \( \langle \xi_2, \theta_2 \rangle \) be a quadratic estimable function, then by Lemma 3.5 we have

\[
\langle \xi_2, \theta_2 \rangle = \langle H_2^*TH_2 \rho_2, \theta_2 \rangle \\
= \langle \rho_2, H_2^*TH_2 \xi_2 \rangle \\
= E \langle \rho_2, H_2^*TW \rangle .
\]

This shows that every estimable function of the form \( \langle \rho_2, \theta_2 \rangle \) can be estimated by a linear functional of \( H_2^*TW \).

Consider the equation

\[
H_2^*TH_2 \hat{\theta}_2 = H_2^*TW .
\]

(3.15)

Since \( R(H_2^*TH_2) = R(H_2^*T) \), equation (3.15) is consistent. The use of solution \( \hat{\theta}_2 \) in the estimation of \( \theta_2 \) is shown in the following.

**Theorem 3.2.** Let \( \hat{\theta}_2 \) be a solution of (3.15). Then \( \langle \xi_2, \hat{\theta}_2 \rangle \) is unbiased for \( \langle \xi_2, \theta_2 \rangle \) if, and only if, \( \langle \xi_2, \theta_2 \rangle \) is estimable.

**Proof.** Let \( \hat{\theta}_2 \) be a solution of (3.15), i.e.,
\[ \hat{\theta}_2 = (H^*_2 TH_2) - H^*_2 TW \]

Where \((H^*_2 TH_2)^{-1}\) is a generalized inverse of \(H^*_2 TH_2\), an \(m \times m\) matrix. Then

\[
\begin{align*}
E<\hat{\lambda}_2, \hat{\theta}_2> &= E<H^*_2 TH_2 \rho, \hat{\theta}_2> \\
&= E<\rho, H^*_2 TH_2 \hat{\theta}_2> \\
&= E<\rho, H^*_2 TW> \\
&= \langle \rho, H^*_2 TH_2 \theta_2 \rangle \\
&= \langle H^*_2 TH_2 \rho, \theta_2 \rangle \\
&= \langle \lambda_2, \theta_2 \rangle.
\end{align*}
\]

Therefore \(\langle \lambda_2, \hat{\theta}_2 \rangle\) is unbiased for \(\langle \lambda_2, \theta_2 \rangle\).

**Theorem 3.3.** \(TH_2 \hat{\theta}_2\) is the orthogonal projection of \(TW\) on \(R(TH_2)\) if, and only if, \(\hat{\theta}_2\) satisfies (3.15).

**Proof.** To prove that

\[ TH_2 \hat{\theta}_2 = TH_2 (H^*_2 TH_2)^{-1} H^*_2 TW \]

is the orthogonal projection of \(TW\) on \(R(TH_2)\), we shall show that \(TH_2 \hat{\theta}_2\) is unique for all \(\hat{\theta}_2\) that satisfy (3.10) and that the linear operator \(TH_2 (H^*_2 TH_2)^{-1} H^*_2 T\) is symmetric and idempotent.

(1) **Uniqueness:** let \(\hat{\theta}_2\) and \(\hat{\phi}_2\) be two distinct solutions of equation (3.15). Then

\[ H^*_2 TH_2 \hat{\theta}_2 = H^*_2 TH_2 \hat{\phi}_2. \]
Hence, \( TH_2(\hat{\theta}_2 - \hat{\phi}_2) \) is in \( N(H^*T) \) but \( TH_2(\hat{\theta}_2 - \hat{\phi}_2) \) is also in \( R(TH_2) \). Since \( N(H^*T) \) and \( R(TH_2) \) are orthogonal complements, therefore, \( TH_2(\hat{\theta}_2 - \hat{\phi}_2) = 0 \) and \( TH_2\hat{\theta}_2 \) is unique.

(2) Symmetry: Let \( A \) and \( B \) be any two members in \( A \). Since

\[
((H^*TH_2)^-)^{-1}
\]

is also a generalized inverse of \( H^*TH_2 \) and \( TH_2(H^*TH_2)^-H^*T \) is unique, then

\[
(A, TH_2(H^*TH_2)^-H^*TB) = (TH_2(H^*TH_2)^-H^*TA, B)
\]

This proves symmetry.

(3) Idempotency: Observe that for every \( \rho_2 \) in \( \mathbb{R}^m \),

\[
\| TH_2\rho_2 - TH_2(H^*TH_2)^-H^*TH_2\rho_2 \|^2
\]

\[
= \text{tr}(TH_2\rho_2 - TH_2(H^*TH_2)^-H^*TH_2\rho_2)^2
\]

\[
= 0.
\]

It follows that

\[
TH_2(H^*TH_2)^-H^*TH_2 = TH_2.
\]

We then have

\[
(TH_2(H^*TH_2)^-H^*T)^2 = TH_2(H^*TH)^-H^*T
\]

or \( TH_2(H^*TH_2)^-H^*T \) is idempotent.

Conversely if \( TH_2\hat{\theta}_2 \) is the orthogonal projection of \( TW \) on \( R(TH_2) \), then

\[
\| TW - TH_2\hat{\theta}_2 \|^2 = \min_{\theta_2 \in \mathbb{R}^m} \| TW - TH_2\theta_2 \|^2.
\]
In other words, \( \hat{\theta}_2 \) satisfies

\[
\frac{3}{2} \left\|TW - TH_2 \hat{\theta}_2\right\|^2 = 0.
\]

But the above equation is equivalent to equation (3.15) and the proof is completed.

It should be pointed out that the equation (3.3) and (3.15) together are also consistent, that is, if \( \hat{\theta}' = (\hat{\theta}_1', \hat{\theta}_2') \) is a solution of (3.3), then the subvector \( \hat{\theta}_2' \) is also a solution of (3.15). This is proved in the following.

**Theorem 3.4.** Equations (3.3) and (3.15) are consistent.

**Proof.** Let \( \hat{\theta}' = (\hat{\theta}_1', \hat{\theta}_2') \) be a solution of (3.3). Recall that \( W - H\hat{\theta} \) is in \( N(H^*) \subset N(H^*) \). We have

\[
H^*TW = H^*TH\hat{\theta} + H^*T(W - H\hat{\theta})
\]

\[
= H^*TH\hat{\theta} = H^*TH_2 \hat{\theta}_2.
\]

This shows that \( \hat{\theta}_2' \) from (3.3) is also a solution of (3.15).

Recall that \( W_1 = H\hat{\theta} \) is the orthogonal projection of \( W \) on \( R(H) \) and \( TW_1 - TH_2 \hat{\theta}_2 \) is the orthogonal projection of \( W_1 \) on \( R(T) \). By Theorem 3.3, \( TH_2 \hat{\theta}_2 \) is the orthogonal projection of \( TW \) on \( R(TH_2) \). Notice that \( TH_2 \hat{\theta}_2 = TH_2(H^*TH_2)^{-1}H^*TW \) is also the orthogonal projection of \( W \) on \( R(TH_2) = R(TH) \). This shows that the orthogonal projection of \( W \) on \( R(TH) \) can be taken in two steps. The projection on \( R(H) \) followed by the projection on \( R(T) \) gives the same projection of \( W \) on \( R(TH) \).
Example 1 One-stage nested random model—Unbiased estimation.

\[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i=1,2, \ldots, t, \quad j=1,2, \ldots, n_i \]

\[ E(W) = J_n \mu^2 + V_1 \sigma_1^2 + I_n \sigma^2, \]

where \( n = \sum_{i=1}^{t} n_i \), \( J_n = 1_{n \times n} \), \( 1_n \) is the nxl vector of unity and

\[
V_1 = \begin{bmatrix}
J_{n_1} & 0 & \ldots & 0 \\
0 & J_{n_2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & J_{n_t}
\end{bmatrix}
\]

Since the set \( \{J_n, V_1, I_n\} \) is linearly independent, \( H^*H \) is positive definite. This means that \( \mu^2 \), \( \sigma_1^2 \) and \( \sigma^2 \) are quadratic estimable.

From the above model we have,

\[ P = \frac{1}{n} J_n, \quad PV_1 P = \frac{1}{n^2} \left( \sum_{i=1}^{t} n_i^2 \right) J_n, \]

\[ (V_1, TV_1) = (V_1, V_1) - (V_1, PV_1 P) \]

\[ = \sum_{i=1}^{t} n_i^2 - \frac{1}{n} \left( \sum_{i=1}^{t} n_i \right)^2, \]

\[ (V_1, TV_2) = (V_1, I_n) - (V_1, P) \]

\[ = n - \frac{1}{n} \sum_{i=1}^{t} n_i. \]
\[(V_2, TV_2) = (I_n, I_n) - (I_n, P) = n - 1,\]

\[(V_1, W) = \frac{1}{t} \sum_{i=1}^{t} Y_i^2, \]

\[(V_1, \text{PWP}) = (PV_1^P, W) = \frac{1}{n^2} \sum_{i=1}^{t} \frac{n_i^2}{n^2} Y_i^2, \]

therefore,

\[(V_1, TW) = (V_1, W) - (V_1, \text{PWP})\]

\[= \frac{1}{t} \sum_{i=1}^{t} Y_i^2 - \frac{1}{n^2} \sum_{i=1}^{t} \frac{n_i^2}{n^2} Y_i^2, \]

\[(V_2, TW) = (I, W) - (P, W)\]

\[= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} Y_{ij} - \frac{1}{n} n Y_i^2. \]

Now we have

\[
H_2^* TH_2 = \left[ \begin{array}{ccc}
\frac{1}{n^2} \left( n^2 - \sum_{i=1}^{t} \frac{n_i^2}{n^2} \right) & \frac{1}{n} \left( n^2 - \sum_{i=1}^{t} \frac{n_i^2}{n^2} \right) \\
\frac{1}{n} \left( n^2 - \sum_{i=1}^{t} \frac{n_i^2}{n^2} \right) & (n - 1)
\end{array} \right]
\]
\[ H_{2TW} = \begin{bmatrix} \frac{t}{n} \sum_{i=1}^{t} y_{i.}^2 - \frac{\sum_{i=1}^{n_1} y_{i.}^2}{n} y_{..} \end{bmatrix} \]

Here \( H_{2TH_2} \) is non-singular and its inverse is

\[
(H_{2TH_2})^{-1} = \frac{1}{t} \begin{bmatrix} \frac{n(n-1)}{n^2 - \sum_{i=1}^{n_1} y_{i.}^2} \ & -1 \\ \sum_{i=1}^{n_1} y_{i.}^2 - n & \frac{\sum_{i=1}^{n_1} n_i^2}{n} \end{bmatrix}
\]

\[
\hat{\sigma}_1^2 = \frac{1}{t} \left( \frac{n(n-1)}{n^2 - \sum_{i=1}^{n_1} y_{i.}^2} \right) \left( \sum_{i=1}^{n_1} y_{i.}^2 - \frac{\sum_{i=1}^{n_1} y_{i.}^2}{n} y_{..} \right)
\]

\[
- \left( \sum_{i=1}^{n_1} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{1}{n} y_{..}^2 \right)
\]

\[
\hat{\sigma}^2 = \left( \sum_{i=1}^{n_1} n_i^2 - n \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n_1} \left( \sum_{j=1}^{n_i} y_{ij}^2 - \frac{1}{n} y_{..}^2 \right) \right)
\]

\[
- \left( \sum_{i=1}^{n_1} y_{i.}^2 - \frac{1}{n^2} \sum_{i=1}^{n_1} n_i^2 y_{..}^2 \right)
\]
3.4 **Invariant Estimation of Variance Components**

In Model (2.1), if we replace $Y$ by $Y + X\beta_0$, then the model becomes

$$Y + X\beta_0 = X(\beta + \beta_0) + \sum_{i=1}^{m} U_i \varepsilon_i.$$  \hspace{1cm} (3.16)

By invariance in quadratic unbiased estimation of variance components we mean the estimates $(A,W) = Y'AY$ of $<\lambda_2, \theta_2>$ remain identical under the translation of the $\beta$-parameters in (3.16), i.e.,

$$Y'AY = (Y + X\beta_0)'A(Y + X\beta_0)$$ \hspace{1cm} (3.17)

for all $\beta_0$ in $\mathbb{R}^p$. This means that the value of the estimate is independent of the value of the $\beta$-parameters. The condition (3.17) is equivalent to

$$AX = 0.$$ 

Therefore, to say that $<\lambda_2, \theta_2>$ is quadratic estimable with invariance, or, in short, that $<\lambda_2, \theta_2>$ is invariant estimable we mean that there exists $A$ in $\mathbb{A}$ with $AX = 0$ such that $(A,W)$ is an unbiased estimate of $<\lambda_2, \theta_2>$. In Lemma 3.2, we have given a necessary and sufficient condition of the estimability of $<\lambda_2, \theta_2>$. Now, with the additional requirement of invariance we have the following:

**Lemma 3.6.** $<\lambda_2, \theta_2>$ is invariant estimable if, and only if, there exist some $A$ in the subspace $\{A \in \mathbb{A}: AX = 0\}$ such that

$$\lambda_2 = H\#A.$$ 

In the following lemma we shall give a linear operator $T_2$ on $\mathbb{A}$ whose range space is this subspace of $\mathbb{A}$. 
Lemma 3.7. Let $T_2$ be a linear operator on $A$ defined by

$$T_2 A = (I - P) A (I - P) \quad (3.18)$$

then $T_2$ is symmetric and idempotent and

$$R(T_2) = \{ A \in A : AX = 0 \} . \quad (3.19)$$

In short, $T_2$ is the orthogonal projection on the subspace $\{ A \in A : AX = 0 \}$ of $A$.

Proof. By the idempotency and symmetry of $(I - P)$, the orthogonal projection on the null space of $X'$, we have

$$T_2^2 A = T(TA) = (I - P)^2 A (I - P)^2$$

$$= (I - P) A (I - P) = T_2 A ,$$

and

$$(TA, B) = \text{tr}((I - P) A (I - P) B)$$

$$= \text{tr}(A (I - P) B (I - P))$$

$$= (A, TB) .$$

These prove the idempotency and symmetry of $T_2$.

To show that

$$R(T_2) = \{ A \in A : AX = 0 \} ,$$

first observe that if $AX = 0$, then

$$A = (I - P) A = A (I - P) = (I - P) A (I - P) ,$$

i.e., $A$ is in $R(T_2)$. Therefore

$$\{ A \in A : AX = 0 \} \subset R(T_2) .$$
Conversely,

\[(I - P)A(I - P)X = 0 \text{ for every } A \text{ in } A.\]

Therefore, if \(A \in \text{R}(T_2)\) then \(AX = 0\), i.e.,

\[R(T_2) \subseteq \{A \in A: AX = 0\}\]

and we have

\[R(T_2) = \{A \in A: AX = 0\}.\]

With the linear operator \(T_2\) introduced above we now restate Lemma 3.7 as follows:

**Lemma 3.8.** \(^1\) \(<\ell_2, \theta_2\>\) is invariant estimable if, and only if,

\[\ell_2 = H^{\ast}T_2A\]

for some \(A\) in \(A\).

**Lemma 3.9.** \(<\ell_2, \theta_2\>\) is invariant estimable if, and only if,

\[\ell_2 = H^{\ast}T_2H_2^\rho_2\]

(3.20)

for some \(\rho_2\) in \(R^m\).

**Proof.** Lemma 3.8 says that \(<\ell_2, \theta_2\>\) is invariant estimable if, and only if, \(\ell_2\) is in \(\text{R}(H^{\ast}T_2)\). For \(A\) in \(A\), let \(T_2H_2^\rho_2\) be its orthogonal projection on \(\text{R}(T_2H_2)\). Then

\[H^{\ast}T_2A = H^{\ast}T_2H_2^\rho_2\]

since \(T_2\) is symmetric and idempotent.

---

\(^1\)Pukelshein's results imply that the unbiased estimator of \(q't\) exists if \(q\) is in \(\text{R}(D_M')\). Lemma 3.8 presents equivalent results.
This shows that $R(\mathbf{HT}_2) = R(\mathbf{HT}_2^T \mathbf{H}_2)$. Therefore $\langle \lambda_2, \theta_2 \rangle$ is invariant estimable if and only if $\lambda_2$ is in $R(\mathbf{HT}_2^T \mathbf{H}_2)$, i.e., if and only if there exists $\rho_2$ in $\mathbb{R}^m$ such that $\lambda_2 = \mathbf{HT}_2^T \mathbf{H}_2 \rho_2$.

Notice that the linear operator $\mathbf{HT}_2^T \mathbf{H}_2$ on $\mathbb{R}^m$ can be written as an $m \times m$ symmetric matrix with its $(i,j)$th entry being

$$(V_i, T_2 V_j) = (V_i, (I - P)V_j(I - P)).$$

Lemma 2.9 provides a useful tool in examining the quadratic estimability with invariance of $\langle \lambda_2, \theta_2 \rangle$.

In general, $R(T_2)$ is a subspace of $R(T)$. Hence, $R(\mathbf{HT}_2^T \mathbf{H}_2) = R(\mathbf{HT}_2)$ is a subspace of $R(\mathbf{HT}_2^T \mathbf{H}_2) = R(\mathbf{HT})$. This means that a quadratic estimable function $\langle \lambda_2, \theta_2 \rangle$ is not necessarily invariant estimable.

If $\langle \lambda_2, \theta_2 \rangle$ is invariant estimable, and $(A, W)$ is an invariant unbiased estimator of $\langle \lambda_2, \theta_2 \rangle$, then we must have $A$ in $R(T_2)$. By symmetry and idempotency of $T_2$, $(A, W) = (A, T_2 W)$ for every $W$ in $A$. Hence, it suffices to derive invariant unbiased estimators from the model

$$E(T_2 W) = T_2 H_2 \theta_2. \quad (3.21)$$

Following the approach developed in the last section, we set up the equation

$$H_2^T T_2 \hat{\theta}_2 = H_2^T W. \quad (3.22)$$

---

2 The model (3.21) is equivalent to Pukelsheim's model $E(MY \Theta MY) D_M T$.

3 The ordinary least squares estimate $\hat{t} = D_M^+ \cdot Y \Theta Y$ given by Pukelsheim is a solution of Equation (3.22).
This equation is consistent since \( R(H^*T_2H_2) = R(H^*T_2) \) as shown in the proof of Lemma 3.9.

By analogy with Theorem 3.3, we have the following.

**Theorem 3.5.** \( T_2H_2\hat{\theta}_2 \) is the orthogonal projection of \( T_2W \) on \( R(TH_2) \) if and only if \( \hat{\theta} \) is a solution of equation (3.22).

**Theorem 3.6.** Let \( \hat{\theta}_2 \) be a solution of equation (3.22). Then \( \langle \ell_2, \hat{\theta}_2 \rangle \) is an invariant unbiased estimator of \( \langle \ell_2, \theta_2 \rangle \) if and only if \( \langle \ell_2, \theta_2 \rangle \) is invariant estimable.

**Proof.** Unbiasedness follows by analogy with the proof of Theorem 3.2. Invariance is verified by observing

\[
\langle \ell_2, \hat{\theta}_2 \rangle = \langle H^*T_2H_2\rho_2, \hat{\theta}_2 \rangle \\
= \langle \rho_2, H^*T_2H_2\hat{\theta}_2 \rangle \\
= \langle \rho_2, H^*T_2W \rangle \\
= \langle T_2H_2\rho_2, W \rangle ,
\]

where \( \rho_2 \) is a solution of equation (3.20).

Equations (3.22) and (3.15) do not have a common solution in general. A sufficient condition for the equations to have a common solution is given in the following.

**Theorem 3.7.** Equation (3.22) and (3.15) have a common solution if \( R(T_2H_2) \subseteq R(TH_2) \).

**Proof.** It can easily be shown that \( T_2TA = TT_2A = T_2A \) for every \( A \) in \( A \).

---

4 \( \langle \ell_2, \hat{\theta}_2 \rangle \) is equivalent to \( q'\hat{t} \) in Pukelsheim's notation.
For every $\hat{\theta}_2$ satisfying equation (3.15), we have

$$H^{*T}_2 T_2 W = H^{*T}_2 (W - T_2 \hat{\theta}_2) + H^{*T}_2 T_2 \hat{\theta}_2.$$ 

If $R(T_2 H_2) \subseteq R(\hat{T}_2)$, then $N(H^{*T}_2) \subseteq N(H^{*T}_2)$ and

$$H^{*T}_2 (W - T_2 \hat{\theta}_2) = 0$$

since $W - T_2 \hat{\theta}_2$ is in $N(H^{*T}_2)$.

Therefore, if $R(T_2 H_2) \subseteq R(\hat{T}_2)$, then

$$H^{*T}_2 T_2 \hat{\theta}_2 = H^{*T}_2 W$$

for every solution $\hat{\theta}_2$ of equation (3.15). In other words, a solution of (3.15) is also a solution of (3.22) if $R(T_2 H_2) \subseteq R(\hat{T}_2)$.

In the following example, the model is the same as that in Example 1.

The invariate estimators of $\sigma_1^2$ and $\sigma^2$ are obtained by solving equation (3.22). They are different from those given in Example 1.

Example 2 One-stage nested model—Invariant estimation.

The matrices $P$, $V_1$ and $V_2$ are the same as in Example 1. Then

$$(V_1, T_2 V_1) = (V_1, (I - P)V_1 (I - P))$$

$$= (V_1, V_1) - (V_1, PV_1 V_1 + V_1, P) + (V_1, PV_1 P)$$

$$= \frac{1}{n} \sum_{i=1}^{n} n_i^2 - \frac{2}{n} \sum_{i=1}^{n} n_i^3 + \frac{1}{n} \left( \sum_{i=1}^{n} n_i^2 \right)^2$$

$$(V_1, T_2 V_2) = (V_1, (I - P))$$

$$= n - \frac{1}{n} \sum_{i=1}^{n} n_i^2.$$
\[(V_2, T_2 V_2) = (I, (I - P)) = n - 1,\]

\[(V_1, T_2 W) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n_1} (Y_{ij} - \bar{Y}_{..}) \right)^2\]

\[(V_2, T_2 W) = (I, T_2 W) = ((I - P), W) \]

\[= \sum_{i=1}^{t} \sum_{j=1}^{n} Y_{ij}^2 - \frac{1}{n} \sum_{i=1}^{t} \sum_{j=1}^{n} Y_{ij}^2,\]

\[H^{T}_2 T_2 H_2 = \begin{bmatrix}
F & n - \frac{1}{n} \sum_{i=1}^{t} n_i^2 \\
\frac{1}{n} \sum_{i=1}^{t} n_i^2 & n - 1
\end{bmatrix}\]

where \(F = \sum_{i=1}^{t} n_i^2 - \frac{1}{n} \sum_{i=1}^{t} n_i^3 + \frac{1}{n^2} \left( \sum_{i=1}^{t} n_i^2 \right)^2,\)

\[\left(H^{T}_2 T_2 H_2\right)^{-1} = \frac{1}{D} \begin{bmatrix}
n - 1 & \frac{1}{n} \sum_{i=1}^{t} n_i^2 - n \\
\frac{1}{n} \sum_{i=1}^{t} n_i^2 - n & F
\end{bmatrix}\]

where \(D = (n + 1) \sum_{i=1}^{t} n_i^2 - \frac{2(n - 1)}{n} \sum_{i=1}^{t} n_i^3 + \frac{(n - 2)(n + 1)}{n^2} \left( \sum_{i=1}^{t} n_i^2 \right)^2 - n^2.\)
Then
\[
\hat{\sigma}_1^2 = \frac{1}{D} \left\{ (n - 1) (V_1, T_2 W) + \left( \frac{1}{n} \sum_{i=1}^{t} n_i^2 - n \right) (V_2, T_2 W) \right\}
\]
\[
= \frac{1}{D} \left\{ (n - 1) \sum_{i=1}^{t} (Y_i - \frac{n_i Y_{..}}{n})^2 \right. \\
+ \left. \left( \frac{1}{n} \sum_{i=1}^{t} n_i^2 - n \right) \left( \frac{1}{t} \sum_{j=1}^{t} \sum_{i=1}^{n_i} \frac{y_{ij}^2}{n_i^2} - \frac{1}{n} Y_{..}^2 \right) \right\},
\]

\[
\hat{\sigma}^2 = \frac{1}{D} \left\{ \left( \frac{1}{n} \sum_{i=1}^{t} n_i^2 - n \right) \sum_{i=1}^{t} (Y_i - \frac{n_i Y_{..}}{n})^2 \right. \\
+ \left. F \left( \sum_{i=1}^{t} \sum_{j=1}^{n_i} \frac{y_{ij}^2}{n_i} - \frac{1}{n} Y_{..}^2 \right) \right\}.
\]
4. ANOTHER LOOK AT SOME OTHER ESTIMATION PROCEDURES

4.1 Seely's Work on Quadratic Unbiased Estimation in Mixed Models

In his work on quadratic unbiased estimation in mixed models, Seely introduced the model

\[ E(W) = H\theta \]

induced from the general mixed linear model

\[ Y = X\beta + \sum_{i=1}^{m} U_i \varepsilon_i. \]

He showed that estimable functions of the form \( \langle \lambda, \theta \rangle \) can be estimated by \( \langle \lambda, \hat{\theta} \rangle \) if \( \hat{\theta} \) is a solution of the equation (3.3)

\[ H^*H\hat{\theta} = H^*W. \]

In the last chapter we proved that if \( \hat{\theta} \) is a solution of the above equation, then

1. \( \hat{\theta} \) is the orthogonal projection of \( W \) on \( R(H) \), and
2. \( \langle \lambda, \hat{\theta} \rangle \) is the least squares estimate of \( \langle \lambda, \theta \rangle \).

In the next section we shall further prove that, in certain random models, the estimators obtained from the above equation are identical to the ones obtained by the symmetric sum method proposed by Koch (1967).

In estimation of variance components Seely proposed a quadratic unbiased estimator without invariance for the general mixed model. He first introduced a linear operator \( T_s \) on \( A \) defined by

\[ T_s A = (I - P)A + A(I - P) \quad \text{for every } A \text{ in } A \] (4.1)
and proved that the range space of this linear operator coincides with
the subspace \( \{ A \in A : X'AX = 0 \} \) of \( A \). He further showed that \( T_s \) is sym-
metric and non-negative definite, i.e., if \( (A, T_s A) = 0 \), then \( T_s A = 0 \).
Also he showed that

\[
R(T_s) = R(T_s H) \oplus N(H^*)
\]

and hence the existence of \( \rho_2 \) such that \( H_s^T H_s G_2 = \ell_2 \) is a necessary
and sufficient condition for \( \langle \ell_2, \theta_2 \rangle \) to be estimable.

Further, Seely stated that if \( \hat{\theta}_2 \) is a solution of the equation

\[
H_s^T H_s \hat{\theta}_2 = H_s^T W
\]

then \( \langle \ell_2, \hat{\theta}_2 \rangle \) is an unbiased estimate of \( \langle \ell_2, \theta_2 \rangle \) if \( \langle \ell_2, \theta_2 \rangle \) is estimable.
However, the consistency of the equation (4.3) was not verified.

We found that the equality in (4.2) and the consistency of equation
(4.3) are equivalent. Therefore, the existence of a solution of (4.3)
is assured. In proving this, we first present the following.

**Theorem 4.1.** If the linear operator \( L \) on \( A \) is such that \( R(L) \supseteq N(H^*) \),
then the following statements are equivalent.

\[
(1) \quad R(LH) + N(H^*) = R(L)
\]
\[
(2) \quad R(H^*LH) = R(H^*L)
\]

**Proof.** Statement (1) implies that for every \( A \) in \( A \) there exist \( \rho \)
in \( R^M \) and \( A_0 \) in \( N(H^*) \) such that

\[
LA = LH\rho + A_0
\]

therefore \( H^*LA = H^*LH\rho \)
and this implies

\[ R(H^*LH) \supseteq R(H^*L) . \]

The reversed containment is obvious. Therefore

\[ R(H^*LH) = R(H^*L) . \]

Conversely, statement (2) implies that for every \( A \) in \( A \) there exists \( \rho \) in \( R^M \) such that

\[ H^*LA = H^*LHp . \]

Therefore \( (LA - LH\rho) \) is in \( N(H^*) \). Write

\[ LA = LH\rho + (LA - LH\rho) \]

for every \( A \) in \( A \). This shows

\[ R(L) \subseteq R(LH) + N(H^*) . \]

But both \( R(LH) \) and \( N(H^*) \) are subspaces of \( R(L) \), so is \( R(LH) + N(H^*) \). Hence

\[ R(L) = R(LH) + N(H^*) . \]

Since \( R(T_1) = N(H^*) \supseteq N(H^*) \), the equivalence in the above theorem holds if we substitute \( T_1 \) for \( L \). Further, since \( R(H^*1) = \{ 0 \} \) and \( R(T_11) = R(T_1H_2) \), it is easily proved that

\[ R(H^*T_1H_2) = R(H^*T_1) \]

(4.8)
is equivalent to

\[ R(H^*T_sH) = R(H^*T_s). \]

This shows that the equality in (4.2) and the consistency of equation (4.3) are equivalent. Therefore the existence of solution of equation (4.3) is assured.

Next we shall investigate if the estimators of \(<\lambda_2, \theta_2>\) derived from equation (4.3) are identical to those from (3.3). Since (3.3) and (3.15) are consistent, we shall examine if (3.15) and (4.3) have common solution. First, it can be easily proved that

\[ T_s = T + T_2. \]  \hspace{1cm} (4.9)

If Equations (3.15) and (3.22) have common solutions, i.e., if there exists \(\hat{\theta}_2\) in \(\mathbb{R}^m\) such that both

\[ H^*T_2\hat{\theta}_2 = H^*T_2W \]

and

\[ H^*T_2H_2\hat{\theta}_2 = H^*T_2W \]

hold, then adding these two equations, we have

\[ H^*T_2^*H_2\hat{\theta}_2 = H^*T_2^*W. \]

This shows that (3.15) and (4.3) have common solutions if and only if (3.15) and (3.22) do. And this is when the estimators of variance components derived from (3.3) and (4.3) are identical.
4.2 Koch's Symmetric Sum Approach to the Estimation of Variance Components

The symmetric sum approach proposed by Koch (1967) provides quadratic unbiased estimates of variance components in random models. To begin with, we summarize Koch's method in the general random model as follows:

(1) The general random model is

\[ Y = \mu 1 + \sum_{i=1}^{m} U_i \epsilon_i, \]

where \( Y \) is the nx1 vector of observations, \( \mu \) is the overall mean and \( \epsilon_i \) \((i=1, 2, \ldots, m)\) are vectors of random effects. The entries in each of these vectors \( \epsilon_i \) are assumed to be independently and identically distributed with zero mean and variance \( \sigma_i^2 \) \((i=1, 2, \ldots, m)\). From these assumptions we have

\[ E(W) = \mu^2 J + \sum_{i=1}^{m} V_i \sigma_i^2, \]

where \( J \) is the nxm matrix of unity and \( V_i = U_i U_i' \) \((i=1, 2, \ldots, m)\).

(2) Divide the entries \( W_{ij} \) of \( W \) into groups such that all the \( W_{ij} \) with the same expectation are in the same group. Transform the \( W \) matrix into a column vector of \( N = n^2 \) entries

\[
Z = \begin{bmatrix}
Z_1 \\
Z_2 \\
\vdots \\
\vdots \\
Z_r
\end{bmatrix},
\]

(4.10)

where \( Z_t \) \((t=1, 2, \ldots, r)\) is the vector of \( N_i \) entries of the \( t \)-th group of \( W_{ij} \left( \sum_{i=1}^{r} N_i = N = n^2 \right) \).
(3) Calculate the average \( \overline{z}_t(t=1, 2, \ldots, r) \) of each vector \( z_t \) and let

\[
\overline{Z} = (\overline{z}_1 \overline{z}_2 \ldots \overline{z}_r)'
\]  

(4.11)

be the \( r \)-vector of the averages with expectation

\[
E(\overline{Z}) = L\theta .
\]  

(4.12)

(4) Solve the equation

\[
L\hat{\theta} = \overline{Z} .
\]  

(4.13)

The random models considered by Koch (1967) produced non-singular coefficient matrices \( L \). Therefore \( \hat{\theta} \) is the unbiased estimator of \( \theta \) by the symmetric sum approach. We shall prove that, when \( \operatorname{rank}(L) = r \leq m + 1 \) (i.e., the rows of \( L \) are linearly independent), \( L\hat{\theta} \) is the least squares estimator of \( L\theta \). The proof is as follows. Let \( D \) be an \( N \times r \) matrix defined by

\[
D = \begin{bmatrix}
1_1 & 0 & \ldots & 0 \\
0 & 1_2 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1_r \\
\end{bmatrix}
\]

where \( 1_t \) is the \( N_t \)-vector of unity \((t=1, \ldots, r)\), and let

\[
F = DL.
\]
Then we have

\[ E(Z) = DL\theta = F\theta . \quad (4.14) \]

Set the equation

\[ \frac{3}{\theta} \| Z - F\theta \|^2 = 0 , \quad (4.15) \]

which is equivalent to

\[ F'F\hat{\theta} = F'Z . \quad (4.16) \]

Observe that \( F'F = L'D'DL \), \( D'D = \text{Diag}(N_1 \ldots N_r) \) and \( F'Z = L'D'Z = L'DD\bar{Z} \).

By the assumption that \( L \) is \( r \times (m+1) \) of rank \( r \), we know the inverse \((LL')^{-1}\) of \( LL' \) exists. Pre-multiplying both sides of the normal equation

\[ F'F\hat{\theta} = F'Z \]

by \((LL')^{-1}L\), we get

\[ D'DL\hat{\theta} = D'D\bar{Z} . \]

Since \( D'D = \text{Diag}(N_1 \ldots N_r) \) is non-singular, this is equivalent to

\[ L\hat{\theta} = \bar{Z} . \]

This shows that \( L\hat{\theta} \) is the least squares estimate of \( L\theta \) if \( L \) is a set of linearly independent functions of \( \theta \).

Further, the transformation from \( W \) to \( Z \) is a one-to-one correspondence between \( A \) and \( R^N \) that preserves the inner product, i.e.,

\[ \| Z - F\theta \|^2 = \| W - H\theta \|^2 \quad (4.17) \]
for every \( \theta \in \mathbb{R}^{m+1} \). Therefore the least squares estimate \( \langle \hat{\tau}, \hat{\theta} \rangle \) of \( \langle \tau, \theta \rangle \) obtained from equations

\[
\frac{\partial}{\partial \hat{\theta}} \left\| \mathbf{W} - \mathbf{H} \hat{\theta} \right\|^2 = 0 \tag{4.18}
\]

and

\[
\frac{\partial}{\partial \hat{\theta}} \left\| \mathbf{Z} - \mathbf{F} \hat{\theta} \right\|^2 = 0 \tag{4.19}
\]

are identical if and only if \( \langle \tau, \theta \rangle \) is estimable.

Forhoffer and Koch (1974) modified the method and extended it to mixed models. The general approach of the modified method can be illustrated by an example of two-way classification mixed model. First, the differences among the observations within each level of the fixed effect factor are calculated. If the fixed effect factor has \( f \) levels and there are \( n_i \) observations within the \( i \)th level, then there are \( \sum_{i=1}^{f} \frac{n_i}{2} \) differences altogether. The differences can be expressed in a vector form \( \mathbf{D} \mathbf{Y} \) with a suitably defined matrix \( \mathbf{D} \). Then the vector \( \mathbf{G} \) is formed by collecting the upper triangle entries of \( \mathbf{D} \mathbf{Y}' \mathbf{D}' \) excluding those whose expectation is identically zero. Denote the expectation of \( \mathbf{G} \) by \( \mathbf{A} \sigma \), where \( \mathbf{A} \) is the known coefficient matrix and \( \sigma \) the vector of variance components. The estimators are derived by least squares procedures based on the model \( \mathbf{E}(\mathbf{G}) = \mathbf{A} \sigma \). The estimators are invariant with respect to the fixed effect parameters \( \beta \). However, the invariant estimators given by Forhoffer and Koch and those obtained in Section 3.4 are not identical in general, since the transformation from \( \mathbf{T}_2 \mathbf{W} \) to \( \mathbf{G} \) is not one-to-one and the equations
\[ \frac{\partial}{\partial \theta_2} \| T_2 \hat{W} - T_2 H_2 \hat{\theta}_2 \|^2 = 0 \]

and \[ \frac{\partial}{\partial \sigma} \| G - A \sigma \|^2 = 0 \]

are not equivalent.

4.3 A Least Squares Approach to the MINQUE

In this section we shall discuss a general approach of estimation of variance components known as the MINQUE developed by Rao (1971). We shall demonstrate, with a suitable linear transformation, that the MINQUE can be obtained by the least squares approach developed in Chapter 3. In the process, we shall also discuss the estimability by the MINQUE and its equivalence to quadratic estimability defined in Sections 3.3 and 3.4.

We first consider the MINQUE with invariance. Let \( X_i, V_i, i = 1, 2, \ldots, m \) be the same as given in (2.2). Define \( V = \sum_{i=1}^{m} V_i \). Since \( V \) is positive definite, there exists an \( n \times n \) non-singular matrix \( C \) such that \( C'VC = I \) or \( V = C^{-1}C' \). (In Rao's notation, \( C \) and \( C' \) are both denoted by \( V^{-1/2} \), but \( C \) is not symmetric unless \( V \) is diagonal.) Let \( Z = C'X \) and \( P_Z \) the orthogonal projection matrix on the range space of \( Z \), i.e.,

\[ P_Z = Z(Z'Z)^{-1}Z' = C'X(X'V^{-1}X)^{-1}X'C. \quad (4.20) \]

The procedure of the MINQUE is to find \( A_* \) in the subset

\[ A_p = \{ A \in A: AX = 0 \quad \text{and} \quad H_A = p, p \in \mathbb{R}^m \} \]

of \( A \) such that the minimum of \( \| C^{-1}A'C^{-1} \| \) is attained at \( A_* \). It should be noted that the subset \( A_p \) of \( A \) is non-empty if and only if \( <p, \theta_2> \) is
invariant estimable. The MINQUE with invariance $Y'A_\ast Y$ given by Rao (1971) is

$$A_\ast = \sum_{i=1}^{m} \lambda_i C(I - P_z)C'V_i C(I - P_z)C',$$

where $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is so chosen such that $H_2^* A_\ast = p$. One problem that needs clarification is the existence of such a vector $\lambda$ in $\mathbb{R}^m$. In other words, assume that $A_p$ is non-empty, is there a member of $A_p$ expressible in the form (4.21)?

We shall answer the question by showing that the MINQUE with invariance can be obtained by the least squares approach developed in Section 3.4.

First we define a linear operator $G$ on $A$ by

$$GA = C'AC$$

(4.22)

for every $A$ in $A$. Let $W_M = GW$. Then

$$E(W_M) = H_M \theta$$

(4.23)

where $H_M = GH$, we make the partition

$$H_M = H_{M1} + H_{M2}$$

(4.24)

just as we did in partitioning $H = H_1 + H_2$.

Since $C$ is non-singular, $G$ is invertible and for every $A$ in $A$, there exists $B$ in $A$ such that $(A, W)$ and $(B, W'_M)$ have the same expectation for every value of $\theta$. (Actually $(A, W) = (B, W'_M)$ if, and only if, $B = C^{-1*}A$, where $C^{-1*}$ is the adjoint of the inverse of $G$.) This means that estimability based on the model $E(W) = H \theta$ is equivalent to that based on the model (4.23).
Let \( T_M \) be a linear operator on \( A \) defined by

\[
T_M A = (I - P_z)A(I - P_z)
\]  \hspace{1cm} (4.25)

for every \( A \) in \( A \). Then by analogy in Lemma 3.7 we have

\[
T_M = T_M^2 = T_M^*
\]

and

\[
R(T_M) = \{ A \in A : AZ = 0 \}.
\]  \hspace{1cm} (4.26)

The least squares estimators based on the model

\[
E(T_M w_M) = T_M M^2 \theta^2
\]

are obtained from the equation\(^5\)

\[
\frac{\partial}{\partial \theta^2} \| T_M w_M - T_M M^2 \theta^2 \|^2 = 0
\]

or

\[
H^* T_M M^2 \theta^2 = H^* T_M w_M.
\]  \hspace{1cm} (4.27)

By analogy of Lemma 3.5, a necessary and sufficient condition that

\(<p, \theta^2>\) is invariant estimable is the existence of \( \lambda \in \mathbb{R}^m \) such that

\[
H^* T_M M^2 \lambda = p.
\]  \hspace{1cm} (4.28)

Combining the above results and Lemma 3.5, we have the following.

\[^5\text{Pukelsheim proved that the MINQUE with invariance can be obtained from the normal equation}
D'(M \Theta M \cdot T \cdot T \cdot M \Theta M)'D = D'(M \Theta M \cdot T \cdot T \cdot M \Theta M)'Y \theta Y.\]
Lemma 4.1. \( R(H^2 M^2 T H^2) = R(H^2 M^2 T H M^2) \).

Let \( \theta_2 \) and \( p \) be given in (4.27) and (4.28), then \( <p, \theta_2> \) is the least squares estimate of \( <p, \theta_2> \) based on the model (4.26). If \( \lambda \) is a solution of (4.28), then

\[
<p, \theta_2> = <\lambda, H^* T M^2 N M^* M>.
\]

(4.29)

By some elementary but tedious algebraic operation, the right side of (4.29) can be written as \( (A_\lambda, W) \) or \( Y'A_\lambda Y \) where \( A_\lambda \) is of the form (4.21).

The above development shows that the MINQUE with invariance is identical to the least squares estimate based on the transformed model (4.26). Since the existence of \( A_\lambda \) in \( A_p \) depends on the consistency of Equation (4.28), this, together with Lemma 4.1, shows that the MINQUE exists for every invariant estimable function \( <p, \theta_2> \).

Under the assumption that the random variables \( \varepsilon_i, i=1, ..., m \) are normally distributed, Rao (1971b) showed that, when \( V \) is redefined as

\[
V = \sum_{i=1}^{m} r_i V_i
\]

(4.30)

where \( r_i, i=1, ..., m \) are positive numbers, the MINQUE is the locally minimum variance quadratic unbiased estimator (MIVQUE) at the points where \( \theta_2' = (\sigma_1^2 \ldots \sigma_m^2) \) is a multiple of \( r' = (r_1 \ldots r_m) \). This suggests that the MIVQUE can be obtained by weighted least squares procedures by introducing the weights \( r_i \) into the model (4.26).

In the case of the MINQUE without invariance, the problem is to minimize \( \text{tr}(A(V + 2XX')AV) \) under the restriction that \( A \) is in
\[ B_p = \{ A \in A : X'AX = 0, H_p A = p \in \mathbb{R}^m \} \]  \hspace{1cm} (4.31)

where \( B_p \) is non-empty if and only if \( \langle p, \theta_2 \rangle \) is estimable.

The procedures of applying the least squares method developed in Section 3.2 to obtain the MINQUE without invariance are given as follows.

It can be easily proved that if \( X'AX = 0 \), then

\[ \text{tr}(A(V + 2XX')AV) = \text{tr}(A(V + XX')A(V + XX')) \] .

Since \( (V + XX') \) is positive definite, there exists a non-singular matrix \( N \) such that

\[ N'(V + XX')N = I \]

or

\[ V + XX' = N^{-1}N^{-1} \] .

Therefore the problem is to minimize \( \| N^{-1}_A N^{-1} \|^2 \), \( A \) in \( B_p \).

Let

\[ W_N = N'WN \]

\[ U = N'X \]

\[ P_U = U(U'U)^{-1}U' \]

and define linear transformations \( H_{N2} : \mathbb{R}^m \rightarrow A \) and \( T_N : A \rightarrow A \)

by

\[ H_{N2} \lambda = N' (H_2 \lambda) N = \sum_{i=1}^{m} \lambda_i N'V_i N, \quad \lambda \in \mathbb{R}^m \]

and

\[ T_N A = A - P_U A P_U', \quad A \in A \] .
respectively. Then we have

\[ E(T^TW) = T_N^N_N^Z \theta \cdot \tag{4.32} \]

Set the equation

\[ \frac{\partial}{\partial \theta} \left\| T_N^N_N^Z - T_N^N_N^Z \theta \right\|^2 = 0 \]

or

\[ H_N^Z T_N^N_N^Z \theta = H_N^Z T_N^W \tag{4.33} \]

and solve for \( \theta \).

Let \( \lambda \) be the solution of the equation

\[ H_N^Z T_N^N_N^Z \lambda = p \cdot \tag{4.34} \]

Then \( \langle p, \theta \rangle = \langle \lambda, H_N^Z T_N^W \rangle \) is the least squares estimate of \( \langle p, \theta \rangle \) based on the model (4.32).

Let \( A_\lambda = \sum_{i=1}^{m} \lambda_i (V + XX')^{-1} (V - Q(V + XX')^{-1} V) (V + XX')^{-1} \)

where \( Q = X(X'(V + XX')^{-1} X) X' \), then,

\[ \text{Min}_{A \in \mathcal{B}_p} \text{tr}(A(V + 2XX')AV) = \text{tr}(A_\lambda (V + 2XX')A_\lambda V) \]

and \( \langle \lambda, H_N^Z T_N^W \rangle = \langle A_\lambda, W \rangle \)

is also the MINQUE of \( \langle p, \theta \rangle \) without invariance.

The matrix \( A_\lambda \) given in (4.35) can also be obtained from applying Lemma 3.10, Rao (1971) with \( V \) replaced by \( V + XX' \).
However, in deriving the MINQUE with invariance, Rao (1971) did not use this lemma and the matrix he obtained in minimizing \( \text{tr}(A(V + XX')AV) \) is different in form from the matrix \( A_\star \) in (4.35). Pringle (1974) also obtained a different form of the matrix \( A \) that minimizes \( \text{tr}(A(V + 2XX')AV) \).

The algebraic identity of the three different forms of the matrix has not been established but in the next section we will prove that \( A_\star \) in (4.35) does minimize \( \text{tr}(A(V + 2XX')AV) \) and that the MINQUE without invariance is unique for every estimable function \( \langle p, \theta_2 \rangle \).

4.4 A Geometric Approach to Quadratic Least Squares Estimation and the MINQUE

In the last section we have demonstrated that the MINQUE can be obtained by the least squares method developed in Chapter 3. In this section we shall discuss the relationship between the two methods in a geometric approach. Even though the following discussion is mostly on unbiased estimation without invariance, its implication to invariant estimation should be easily seen.

Let \( p \) be a fixed vector in \( \mathbb{R}^m \) such that \( B_p \) is not empty. By the definition of \( B_p \) in (4.31) we know that \( A \) is in \( B_p \) if and only if

\[
A = TA = A - PAP
\]

and

\[
H\ast TA = p.
\]

By Lemma 3.5, there exists \( \lambda \) in \( \mathbb{R}^m \) such that

\[
H\ast TH_2 \lambda = p.
\]
Therefore $\mathbf{TH}_2^\lambda$ is in $\mathcal{B}_p$. Observe that

$$
\mathbf{TH}_2^\lambda = \mathbf{TH}_2(\mathbf{H}_2^\lambda \mathbf{TH}_2)\mathbf{p} \\
= \mathbf{TH}_2(\mathbf{H}_2^\lambda \mathbf{TH}_2)\mathbf{H}_2^\lambda \mathbf{T} \mathbf{A}
$$

for every $\mathbf{A}$ in $\mathcal{B}_p$. By Theorem 3.3 $\mathbf{TH}_2^\lambda$ is the orthogonal projection on $\mathbf{R(TH}_2)$ of every $\mathbf{A}$ in $\mathcal{B}_p$.

If $\mathbf{TH}_2^\rho$ is in $\mathcal{B}_p$ then

$$
\mathbf{H}_2^\rho \mathbf{TH}_2^\rho = \mathbf{p} = \mathbf{H}_2^\rho \mathbf{TH}_2^\lambda
$$

and $(\mathbf{TH}_2^\rho - \mathbf{TH}_2^\lambda)$ is in $\mathbf{N(TH}_2)$. But $(\mathbf{TH}_2^\rho - \mathbf{TH}_2^\lambda)$ is also in $\mathbf{R(TH}_2)$. Therefore $\mathbf{TH}_2^\rho = \mathbf{TH}_2^\lambda$ and it is the unique point in $\mathbf{R(TH}_2)\mathbf{\cap B}_p$.

Let $\mathbf{A}$ be in $\mathcal{B}_p$. Then

$$
\mathbf{H}_2^\lambda \mathbf{T} \mathbf{A} = \mathbf{p} = \mathbf{H}_2^\lambda \mathbf{TH}_2^\lambda
$$

and $(\mathbf{A} - \mathbf{TH}_2^\lambda)$ is in $\mathbf{N(H}_2^\lambda \mathbf{T})$. Therefore

$$
||\mathbf{A}||^2 = ||\mathbf{TH}_2^\lambda||^2 + ||\mathbf{A} - \mathbf{TH}_2^\lambda||^2
$$

$$
> ||\mathbf{TH}_2^\lambda||^2
$$

if $\mathbf{A} \neq \mathbf{TH}_2^\lambda$.

Summarizing the above discussion, we have the following.

**Theorem 4.2.** If $\mathcal{B}_p$ is not empty, then

1. There exists $\lambda$ in $\mathbb{R}^m$ such that $\mathbf{TH}_2^\lambda$ is in $\mathcal{B}_p$;
2. $\mathcal{B}_p = \{\mathbf{TH}_2^\lambda\} \oplus \mathbf{N(H}_2^\lambda)$ where $\lambda$ satisfies

$$
\mathbf{H}_2^\lambda \mathbf{TH}_2^\lambda = \mathbf{p} ;
$$
(3) \( \| T H_2 \lambda \| ^2 \leq \| A \| ^2 \) for every \( A \) in \( B_p \). The equality holds if, and only if, \( A = TH_2 \lambda \).

From Theorem 4.2 we have

\[
\min_{A \in B_p} \| A \| ^2 = \| TH_2 \lambda \| ^2
\]

where \( \lambda \) satisfies \( H^* TH_2 \lambda = p \). Since

\[
(TH_2 \lambda, W) = \langle \lambda, H^*TW \rangle
\]

\[
= \langle \lambda, H^* TH_2 \hat{\theta}_2 \rangle
\]

\[
= \langle p, \hat{\theta}_2 \rangle
\]

where \( \hat{\theta}_2 \) satisfies \( H^* TH_2 \hat{\theta}_2 = H^*TW \). Therefore \( (A, W) \) is the least squares estimator of \( \langle p, \theta_2 \rangle \) if \( A \) is in \( B_p \) and has the minimum norm.

Conversely if \( (A, W) \) is the least squares estimator of \( \langle p, \theta_2 \rangle \), i.e.,

\[
(A, W) = (p, \hat{\theta}_2)
\]

where \( \hat{\theta}_2 \) satisfies \( H^* TH_2 \hat{\theta}_2 = H^*TW \), then

\[
(A, W) = \langle H^* TH_2 \lambda, \hat{\theta}_2 \rangle
\]

\[
= \langle \lambda, H^*TW \rangle
\]

\[
= (TH_2 \lambda, W)
\]

for every \( W \). Therefore \( A = TH_2 \lambda \) and this shows that \( \| A \| ^2 \) is minimized if \( (A, W) \) is the least squares estimator of \( \langle p, \theta_2 \rangle \) .
We now conclude the above discussion in the following.

**Theorem 4.3.** \((A,W)\) is the least squares estimator of \(<p, \theta_2>\) if, and only if, \(A\) is in \(B_p\) and

\[
||A||^2 = \min_{B \in B_p} ||B||^2.
\]

In the unbiased estimation with invariance, the statements in Theorems 4.2 and 4.3 are still true if \(T\) and \(N(H^*)\) are replaced by \(T_2\) and \(R(T_2) \cap N(H^*)\), respectively.

In order to obtain the MINQUE through least squares procedures, we have to make a linear transformation on the data. In the case of the MINQUE without invariance, the linear transformation \(G\) was defined as \(GA = N'AN\) so that

\[
\text{tr}(A(V + 2XX')AV) = ||N^{-1}AN^{-1}||^2. \tag{4.36}
\]

Corresponding transformations to be made are

\[
U = N'X \quad \text{or} \quad T_NA = A - P_UA\overline{P_U}
\]

\[
V_{N1} = N'V_iN \quad \text{or} \quad H_{N2} = G\overline{H_2}.
\]

With this set up we define

\[
B_p = \{A \in A: U'N^{-1}AN^{-1}U = 0, H_{N2}^2 = p\}
\]

and

\[
C_p = \{B \in A: U'BU = 0, H_{N2}^2 = p\}.
\]

Then there is a one-to-one correspondence between \(B_p\) and \(C_p\) and

\[
\min_{A \in B_p} (\text{tr}(A(V + 2XX')AV)) = \min_{B \in C_p} ||B||^2.
\]
This, together with Theorems 4.2 and 4.3, shows that the MINQUE from the model $E(W) = H\theta$ is the least squares estimator from the model

$E(T_N W_N) = T_N H_N \theta_2$. 
5. CONCLUSIONS

The application of the theory of least squares in quadratic estimation in the general mixed model has been developed. In reviewing the results presented in this dissertation, we find it interesting to note the analogy and the dissimilarity between linear and quadratic estimation in application of least squares principles. In both cases, we consider the observation as a member or a point in a finite dimensional inner product space. Its expected value lies in a subspace which we refer to as estimation space. The estimation space is spanned by a finite subset specified by the model. In the model $E(Y) = X\beta$, the spanning subset consists of the columns of $X$. In the model $E(W) = H\theta$, it is the collection of $B_{ij}$ and $V_i$. A linear function of parameters is estimable if and only if it can be expressed as a linear function of the expected value of the observation. The least squares estimator of such a parametric function is a linear function of the orthogonal projection of the observation on the estimation space. In linear estimation, if $C'Y$ is unbiased for $l'\beta$, then $C'\hat{Y}$ is the least squares estimate, where $\hat{Y}$ is the orthogonal projection of $Y$ on $R(X)$. Similarly, if $(A,W)$ is unbiased for $<l,\theta>$, then $(A,W_1)$ is the least squares estimate if $W_1$ is the orthogonal projection of $W$ on $R(H)$.

In estimation of variance components, we make the partition $H\theta = H_1\theta_1 + H_2\theta_2$ like the partition $X\beta = X_1\beta_1 + X_2\beta_2$ in linear models. One dissimilarity between the two cases is observed in unbiased estimation with invariance. In linear estimation, if $a'Y$ is unbiased for $a'\beta_2$, then $a$ is in the null space of $X_1'$ and $a'Y$ is invariant with respect to the translation of $\beta_1$. However, in quadratic estimation
that \((A, W)\) is unbiased for \(<\ell_2, \theta_2>\) does not necessarily mean that it is invariant with respect to \(\theta_1\). Invariance is a desirable property in quadratic estimation of variance components. However, invariant estimates do not always exist for quadratic estimable functions of the form \(<\ell_2, \theta_2>\). Least squares estimators of variance components derived from the model \(E(TW) = TH_2\theta_2\) are proved to be identical with those from the model \(W = H\theta\). These estimators are unbiased without invariance. It is proved that Koch's (1967) symmetric sum approach to estimation of variance components is a special case of least squares estimation base on the model \(E(W) = H\theta\). Invariant estimators are derived from the model \(E(T_2W) = T_2H_2\theta_2\). Since the range spaces of the linear transformations \(TH_2\) and \(T_2H_2\) do not necessarily coincide, the projections of \(W\) on these two subspaces are not identical in general.

It is proved that Rao's (1971a) MINQUE can be derived by the weighted least squares approach. If \(<p, \theta_2>\) is quadratic estimable with invariance, then the MINQUE of \(<p, \theta_2>\) with invariance is the quadratic estimator \((A_*, W)\) such that

\[
\|C'A_\ast C\|^2 = \min_{A \in A_p} \|C'AC\|^2.
\]

Where \(A_p = \{A \in A: AX = 0 \text{ and } HX = p\}\)

and \(C\) is the non-singular matrix such that \(\sum_{i=1}^{m} C'V_i C = I\).

If \((\sigma_1^2 \ldots \sigma_m^2)\) is a multiple of \((r_1 \ldots r_m)\) and \(C\) is such that

\[
\sum_{i=1}^{m} C'V_i C r_i = I,
\]

then the MINQUE is the locally minimum variance unbiased estimator under the assumption of normality. This is analogous to the case in linear estimation where the least squares estimate \(a_\ast Y\) of \(\lambda'\beta\) is
such that $a^*_\mathbf{x}$ minimizes $\mathbf{a}'\mathbf{a}$ among all linear unbiased estimates $\mathbf{a}'\mathbf{Y}$ and $a^*_\mathbf{Y}$ is the best linear unbiased estimate if $a^*_\mathbf{x}$ minimizes $\mathbf{a}'\mathbf{Va}$ and $\mathbf{V} = \text{Var}(\mathbf{Y})$.

The theory of least squares leads to minimum variance unbiased estimator in linear estimation. In the present work, the least squares approach to quadratic estimation of variance components is developed and the analogy and dissimilarity between linear and quadratic estimation are outlined in the hope that minimum variance unbiased quadratic estimators will be developed in the future research. Other topics of suggested future research are the admissibility of invariant quadratic estimators, specification of the class of experimental designs where all variance components are quadratic estimable with invariance and the experimental designs for quadratic least squares estimation of variance components.
LIST OF REFERENCES


LIST OF REFERENCES (continued)


