UNIVERSITY OF SOUTHAMPTON

Faculty of Mathematical Studies

Mathematics

Estimation of Variance Components With Applications in Sample Surveys

by

Sujuan Gao

Thesis submitted for the degree of Doctor of Philosophy

April, 1991

University of Southampton

Abstract

Faculty of Mathematical Studies

Mathematics

Doctor of Philosophy

Estimation of Variance Components
With Applications in Sample Surveys

by Sujuan Gao

This thesis is concerned with the problem of variance components estimation and its applications in sample surveys.

The MINQUE (minimum norm invariant quadratic unbiased estimator) was proposed for a general variance components model, but its optimality requires normality assumption and correct prior values. A sufficient condition for optimality is given in the thesis as an alternative condition to the normality assumption. A necessary and sufficient condition is proved for MINQUE to be independent of prior values and a simplified condition is given for the balanced analysis of variance models.

There are several modified versions of MINQUE that yield nonnegative estimates for the variance components. In this thesis the nonexistence of a nonnegative minimum biased quadratic estimator across the parameter space is proved. A nonnegative estimator, which has minimum variance among all the estimators minimizing an upper bound of the bias function, is proposed. Numerical and empirical studies are carried out and suggestions are made on the use of these nonnegative estimators.

MINQUE is applied to estimate the interviewer's variance in a complex sample survey and its efficiency is compared with some existing estimators. An optimal design with a specified cost constraint is given and an unbiased estimator for the variance of the estimator of the mean is derived.

Acknowledgements

I would like to thank Prof. T. M. F. Smith sincerely for his insight, encouragement and help in supervising this research.

I would also like to thank Dr. S. M. Lewis deeply for her encouragement and help in various stages of my research.

Many members of staff and students in the Faculty have given me their generous help which enables me to do a research in statistics and makes my stay in Southampton a pleasant one. I would like to thank them all. My special thanks go to: Prof. A. G. Howson, Dr. R. P. Mercier and Prof. H. B. Griffiths.

I am grateful to my parents for their encouragement. I have to express my deep gratitude to my husband Yongshe for his understanding and encouragement all the time.

This research is supported by Sino-British Friendship Scholarship administered by the State Education Commission of China and the British Council.

Contents

Chapter 1

INTRODUCTION

1.1 The problem of estimation of variance components

Since Fisher (1925) introduced the terms 'variance' and 'analysis of variance' into the literature and implicitly employed variance components models several variance components models and methods of estimating variance components have been developed and used in many fields of statistics.

In a sample survey different operations are associated with different stages of the survey process and it is impossible to avoid the occurrence of error in at least some of the operations. For example, in the process of data collection there are many potential sources of error: the interviewers, respondents, nonrespondents, and so forth. These errors affect the total variance of the survey estimators. In large-scale surveys it is not uncommon that the contribution of interviewer's error to total variance is larger than the contribution of sampling error (U. S. Bureau of Census, 1979). In a survey using interviews the conventional method of estimating the variance of the survey estimators which only takes sampling error into account overestimates the accuracy of the survey. A variance components model should be applied in this situation and estimation of important components such as the interviewer's variance is needed.

In experimental design because of the complexity of experimental conditions estimation of variance components is also needed. Many factors in the experiment are random factors. They do not affect the biases of the treatment effects but they affect the estimates of variance of the treatment effects and hence the efficiency of the treatment effects. Since many optimality criteria in experimental design depend upon the variance of the treatment effects, estimation of variance components may change the design. Searle (1971) gives examples of random factors in the design of experiments.

The purpose of estimation of variance components is twofold. First, the estimates of variance components give individual variance values for different error sources which will help to identify the poorly performing stages in the operation. Second, using the estimates of variance components we can obtain more accurate estimates of the variances of desired estimators such as population means and treatment effects.

The problems relating to the estimation of variance components are:

1. The construction of a proper variance components model.

This problem is closely related to the experience of previous operations and the understanding of data. It is very important to use a proper model to start the estimation. The emphasis of this thesis, however, is not on this problem. In the next section we shall introduce some commonly used variance components models. Throughout the thesis we assume that the model is predetermined before we start the estimation.

2. The choice of an optimal estimator.

The choice of an optimal estimator is usually determined by the objectives of the estimation, the optimality criteria and the structure of the data. We shall introduce several estimators in the following sections and will investigate the properties of some of these estimators.

1.2 Variance components models

In the early stage of the development of methods of estimating variance components specific models such as the one-way random model and the two-way nested random model were often used. For the convenience of theoretical work there is a need for a general variance components model which will include these specific models as special cases. Rao and Kleffe (1980) give a list of papers on the early use of variance components models. We shall follow the notation of Rao and Kleffe and call the following model the general variance components model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\boldsymbol{\xi}_1 + \cdots + \mathbf{U}_k\boldsymbol{\xi}_k, \tag{1.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector containing the observed values, $\boldsymbol{\beta}$ is a $p \times 1$ fixed effect parameter vector, $\mathbf{X}$ is an $n \times p$ design matrix for the fixed effect, $\mathbf{U}_i$ is an $n \times p_i$

design matrix for the variance component, and $\boldsymbol{\xi}_i$ is a $p_i \times 1$ random effect vector, $i = 1, \ldots, k$.

The following assumptions are imposed on model (1.1):

$$\mathrm{E}(\boldsymbol{\xi}_i) = 0, \quad i = 1, \ldots, k, \tag{1.2}$$

$$\mathrm{V}(\boldsymbol{\xi}_i) = \sigma_i^2 \mathbf{I}, \quad i = 1, \ldots, k, \tag{1.3}$$

$$\mathrm{cov}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) = 0, \quad i, j = 1, \ldots, k, i \neq j. \tag{1.4}$$

We also assume that finite third and fourth moments exist for all random variables and that the third and fourth moments are equal for all variables in a given vector $\boldsymbol{\xi}_i$, $i = 1, \ldots, k$.

Assumption (1.4) assumes no correlation between different variance component vectors. Throughout the thesis whenever model (1.1) is used the above assumptions are imposed. The associated statistical problems with model (1.1) are:

1. Estimation of $\boldsymbol{\beta}$.
2. Estimation of σ_i^2, $i = 1, \ldots, k$.
3. Prediction of $\boldsymbol{\xi}_i$, $i = 1, \ldots, k$.

This thesis concentrates on dealing with the second problem: estimation of σ_i^2, $i = 1, \ldots, k$.

In model (1.1) the vector $\mathbf{y}$ is determined by the observations from a survey or an experiment, and $\mathbf{X}, \mathbf{U}_1, \ldots, \mathbf{U}_k$ are the design matrices for the fixed and random effects. Different forms of design matrices will yield different models.

The earliest proposed and the most commonly used models are the analysis of variance models (ANOVA models) which assume an overall mean for all the observations and the numbering of the cells are in lexicographical order. Suppose there are n random factors each at s_j levels. We use $p = (p_1 p_2 \ldots p_n)$ to label a cell, t_p is the number of observations in each cell. There are two special classes of ANOVA models, nested ANOVA models and crossed ANOVA models, which are of particular interest.

An n-way nested ANOVA model can be written as:

$$y_{p_1 p_2 \ldots p_n h} = \mu + a_{p_1} + a_{p_1 p_2} + a_{p_1 p_2 p_3} + \cdots + a_{p_1 p_2 \ldots p_n} + e_{p_1 p_2 \ldots p_n h}, \tag{1.5}$$

$p_j = 1, \ldots, s_j$, $j = 1, \ldots, n$, $h = 1, \ldots, t_p$, where μ is the overall mean and $a_{p_1 \ldots p_k}$ is a random term with mean zero and variance σ_k^2, $k = 1, \ldots, n$, $e_{p_1 \ldots p_n h}$ is the random error with mean zero and variance σ_e^2.

A crossed n-way ANOVA model can be written as:

$$\begin{aligned} y_{p_1p_2\cdots p_nh} &= \mu + a^{(1)}_{p_1} + a^{(2)}_{p_2} + \cdots + a^{(n)}_{p_n} + a^{(12)}_{p_1p_2} + \cdots + a^{(n-1,n)}_{p_{n-1}p_n} \\ &\quad + \cdots + a^{(12\ldots n)}_{p_1p_2\cdots p_n} + e_{p_1p_2\cdots p_nh}, \end{aligned} \tag{1.6}$$

$$p_j = 1,\ldots,s_j,\ j = 1,\ldots,n,\ h = 1,\ldots,t_p.$$

where μ is the overall mean and the other terms are random factors.

Example 1.1:

A two-way nested ANOVA model:

$$y_{ijk} = \mu + a_i + b_{ij} + e_{ijk},$$

where μ is the overall mean, a_i, b_{ij} and e_{ijk} are random terms with variance components σ^2_a, σ^2_b and σ^2_e respectively.

Example 1.2:

A two-way crossed model without interaction

$$y_{ijk} = \mu + a_i + b_j + e_{ijk},$$

where a_i, b_j and e_{ijk} are random terms with variance components σ^2_a, σ^2_b and σ^2_e respectively.

Example 1.3:

Two-way crossed model with interaction:

$$y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk},$$

where a_i, b_j, c_{ij} and e_{ijk} are random terms with variance components σ^2_a, σ^2_b, σ^2_c and σ^2_e respectively. □

In practice we often need to know about more general fixed effects than an overall mean. For instance, in a survey we may be interested in knowing the subclass means. Hence the above ANOVA models are not adequate and another class of models is needed. For the convenience of presenting the work we shall call these models extended ANOVA models (E-ANOVA models) which are formed by changing the overall mean in the ANOVA models into a subclass mean and retaining all the corresponding random terms. The E-ANOVA models will include the ANOVA models as special cases.

Example 1.4:

An E-ANOVA model used in Chapter 6:

$$y_{\gamma tjs} = \eta_{\gamma t} + b_{\gamma j} + e_{\gamma tjs},$$

$\gamma = 1, \ldots, l$, $t, j = 1, \ldots, k$, $s = 1, \ldots, f$, where $\eta_{\gamma t}$ is a fixed effect, $b_{\gamma j}$ and $e_{\gamma tjs}$ are random terms with variance components σ_b^2 and σ_e^2 respectively. □

The E-ANOVA models are special cases of the general variance components model (1.1) because the design matrices $\mathbf{X}$ and $\mathbf{U}_i$ for the E-ANOVA models only have 0 and 1 as its elements and the element 1 appears only once in a row of the design matrices. Many of the results obtained in this thesis are obtained for the general variance components model (1.1), but the E-ANOVA models are used extensively either to simplify the conditions or as examples of how to apply the results.

A balanced data set is one in which there are equal numbers of observations in all the subclasses which are formed by the various combinations of the factor's levels. If there are unequal numbers of observations the data set is referred to as unbalanced. A linear model fitted to a balanced data set will be called for short a balanced model. A linear model fitted to an unbalanced data set will therefore be called an unbalanced model.

1.3 Estimators of variance components

The choice of good estimators in statistics is sometimes subjective. Generally speaking, there are two ways to obtain an estimator. One is by maximizing or minimizing a function of the parameters. Examples of estimators obtained this way are the maximum likelihood estimator (Section 1.3.2) and the restricted maximum likelihood estimator (Section 1.3.3). Another way of obtaining an estimator is by setting up some desired properties of the estimator as constraints and solving the equation system to obtain the estimator. Examples are the BLUE (Best Linear Unbiased Estimator) and the MIVQUE (Section 1.3.4).

Considering that we are estimating variance components which bear the properties of variance we may want our estimators to be:

1. invariant with respect to the fixed effect, *i.e.* the estimates of the variance components will not change when the mean vector changes;

2. nonnegative;

3. unbiased;

4. have minimum variance among all the unbiased estimators;
5. have minimum mean squared error among all estimators.

In the process of looking for good estimators we often need to limit the class of the estimators in order to secure a solution to the problem. For example, we often look for the best unbiased linear estimator for the mean factor. When estimating variance components we can restrict ourselves to quadratic estimators.

Sometimes we cannot find an estimator which will satisfy all the requirements we impose. Mathematically this is the situation when we have more equations than unknowns and some of the equations are contradictory of others, hence there will be no solutions to the equations. LaMotte (1973b) and Pukelsheim (1981) show that there does not always exist an unbiased nonnegative estimator with minimum variance, so the constraints 2 and 3 are not compatible. We have to decide what objective is the most important one in the estimation and then adopt suitable steps.

There are several types of estimators proposed in the literature. A brief introduction to the major estimators is given below and an investigation of the properties of some of the estimators is carried out in the following chapters.

1.3.1 The ANOVA estimator

The ANOVA estimator is the earliest estimator developed to estimate variance components for the ANOVA models. The idea of ANOVA is to put various sample sums of squares equal to their corresponding expected values and to solve the resulting equations for the variance components.

Following the notation we used in section 1.2 on ANOVA models, let $y_{...}$ denote the overall mean of the observations. We may partition the total sum of squares $\sum(y_{n_1 \dots n_k} - y_{...})^2$ into t nonnegative quadratic forms: $Q_1, Q_2, \dots, Q_t$ such that:

$$\sum(y_{n_1 \dots n_k} - y_{...})^2 = \sum_{j=1}^{t} Q_j.$$

Let f_j represent the degree of freedom associated with the sums of squares Q_j and let L_j represent the expected value of Q_j/f_j. Then L_j is a linear function of the variance components σ_i^2, *i.e.* $L_j = \sum_{w=1}^{m} p_w(j)\sigma_w^2$. ANOVA is often put into a table:

Table 1.1: the analysis of variance table

source of variation	sum of squares	degree of freedom	expected mean squares
total	$\sum(y_{n_1 \ldots n_k} - y_{\ldots})^2$	f_0	
due to $A^{(1)}$	Q_1	f_1	L_1
due to $A^{(2)}$	Q_2	f_2	L_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
due to error	Q_t	f_t	L_t

We then solve the following equations for the variance components provided they have a solution:

$$Q_j/f_j = L_j, \quad j = 1, \ldots, t.$$

Notice only when $t = k$ do the above equations yield unique solutions for the σ_i^2. We use the solutions of the equations as the estimates of the variance components.

Example 1.5:

Consider a one-way random model:

$$y_{ij} = \mu + a_i + e_{ij} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

where μ is the mean, a_i is a random term with variance σ_a^2, e_{ij} is the random error with variance σ_e^2. The ANOVA estimators for σ_a^2 and σ_e^2 are:

$$\hat{\sigma}_a^2 = \frac{1}{m-1} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{mn(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2,$$

$$\hat{\sigma}_e^2 = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2.$$

Now for $m = 2$, $n = 3$, the observed data values are:

$i \backslash j$	1	2	3	total	mean
1	19	17	15	51	17
2	25	5	15	45	15
				96	16

The standard one-way analysis of variance table gives:

Source	d.f.	S.S.	M.S.	Expected mean squares
mean	1	1536	1536	
classes	1	6	6	$3\sigma_a^2 + \sigma_e^2$
residual error	4	208	52	σ_e^2
total	6	1750		

The two estimating equations are:

$$3\sigma_a^2 + \sigma_e^2 = 6, \quad \sigma_e^2 = 52;$$

which give estimates:

$$\hat{\sigma}_a^2 = -15\frac{1}{3}, \quad \hat{\sigma}_e^2 = 52. \quad \square$$

Example 1.5 not only demonstrated how to use the ANOVA estimator but also shows that the ANOVA estimator can take negative values.

The ANOVA estimators are usually derived using the analysis of variance table. It is known that for balanced data the ANOVA estimator is unbiased. There are various methods to extend the ANOVA estimator to unbalanced data, see Searle (1971), but the unbiasedness property is not usually preserved. For balanced data there have been studies on the optimality of the ANOVA estimator. Graybill and Wortham (1956) used the concept of sufficient statistics to prove the following theorem:

Theorem 1.1 (Graybill and Wortham, 1956) *If the subscript $n_1, \ldots, n_k$ in the analysis of variance model are such that the quantity*

$$\sum (y_{n_1 \ldots n_k} - y_{\ldots})^2$$

can be partitioned into nonnegative quadratic forms $Q_1, \ldots, Q_t$ as indicated in Table 1.1 such that

(a). Q_j/L_j $(i = 1, \ldots, t)$ are distributed independently as Chi-squares with f_i degrees of freedom respectively,

(b). The L_i are linearly independent linear function of the σ_j^2, so that the equations $Q_i/f_i = L_i$ $(i = 1, \ldots, t)$ have unique solutions for the σ_j^2 $(j = 1, \ldots, t)$,

then the uniformly best (minimum variance) unbiased estimator of any linear function of the L_j is given by the same linear function of the Q_j/f_j.

When conditions (a) and (b) are met then the ANOVA estimator is uniformly the best unbiased estimator of the variance components.

For the balanced ANOVA models it is possible to partition the sum of squares into nonnegative quadratic forms. If in addition normality is assumed for the distribution of the data, then from Theorem 1.1 the ANOVA estimator is the best unbiased estimator. Therefore the following three conditions are sufficient for the optimality of the ANOVA estimator:

1. the model is an ANOVA model,
2. the data are normally distributed,
3. the data are balanced.

Graybill (1954) considered the optimality of ANOVA estimators without the normal distribution assumption. He calculated the variance of a quadratic unbiased estimator for a balanced two-way nested model and showed that the ANOVA estimator has the minimum variance among all the quadratic unbiased estimators. He extended his result to multi-way balanced nested ANOVA models.

Theorem 1.2 (Graybill, 1954) *For balanced nested multi-way ANOVA model (1.5) the ANOVA estimator is the best quadratic unbiased estimator of the variance components.*

A closely related problem is the design problem for the ANOVA estimator. Since there are various extensions for the ANOVA estimator to the unbalanced data situations, there should be various design concerns associated with each extension. Mukerjee and Huda (1988) considered the design problem for the unweighted analysis of variance estimator following Searle (1971).

Theorem 1.3 (Mukerjee and Huda, 1988) *If (1) a crossed ANOVA model is considered, (2) an unweighted analysis of variance estimator is used, (3) the total number of observations is fixed, then a balanced design (equal observation in each cell) is optimal.*

The advantage of the ANOVA estimator is its simplicity for balanced ANOVA models and its optimality under certain conditions. The disadvantage of the ANOVA estimator is that it is only available for the ANOVA models and even for ANOVA models the ANOVA estimator lacks optimality when dealing with unbalanced data.

1.3.2 The maximum likelihood estimator

The maximum likelihood estimation approach was proposed as early as the ANOVA estimator, but was neglected because of the difficulty in solving the likelihood equations. Maximum likelihood (ML) estimators have become more popular recently with the increasing power of computers.

In the literature when the ML approach is mentioned it means maximum likelihood estimation for data coming from a normal distribution. No other forms of distribution has been used by this approach. We shall therefore use the term ML estimator meaning ML estimators for normally distributed data. Harville (1977) gives a thorough review of the maximum likelihood (ML) and the restricted maximum likelihood (REML) approach. The REML approach will be discussed in the next subsection.

Consider the general variance components model (1.1). Since $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, let $\mathbf{V}_i = \mathbf{U}_i\mathbf{U}_i'$, then $\mathrm{V}(\mathbf{y}) = \mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$ is the variance components matrix of the data vector $\mathbf{y}$. Let $\Theta = (\sigma_1^2, \ldots, \sigma_k^2)'$. When the observed data is normally distributed the log-likelihood function of model (1.1) is proportional to:

$$L(\beta, \Theta; \mathrm{y}) = -\frac{1}{2}\log|\mathrm{V}| - \frac{1}{2}(\mathrm{y} - \mathrm{X}\beta)'\mathrm{V}^{-1}(\mathrm{y} - \mathrm{X}\beta). \tag{1.7}$$

The ML estimators of β and Θ are the values β_0 and Θ_0 such that:

$$L(\beta_0, \Theta_0; \mathrm{y}) = \sup_{\beta,\Theta} L(\beta, \Theta; \mathrm{y}).$$

Usually in the ML approach we obtain the likelihood equations:

$$\frac{\partial L}{\partial \beta} = 0, \quad \frac{\partial L}{\partial \Theta} = 0. \tag{1.8}$$

If β_0 and Θ_0 are the solutions to (1.8) and also

$$L(\beta_0, \Theta_0; \mathrm{y}) = \sup_{\beta,\Theta} L(\beta, \Theta; \mathrm{y})$$

then β_0 and Θ_0 are the maximum likelihood estimates of β and Θ.

Maximum likelihood estimators do not exist in some cases. The following example is given by Rao and Kleffe (1980). It is a bit artificial because it has less observations than the number of parameters.

Example 1.6:

There are two random variables: $y_1 = \mu + \varepsilon_1$, $y_2 = \mu + \varepsilon_2$, $\mathrm{E}(\varepsilon_1) = \mathrm{E}(\varepsilon_2) = 0$, $\mathrm{E}(\varepsilon_1^2) = \sigma_1^2$, $\mathrm{E}(\varepsilon_2^2) = \sigma_2^2$, $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0$. The log-likelihood function of y_1 and y_2 is:

$$L = -\frac{1}{2}\log\sigma_1^2 - \frac{1}{2}\log\sigma_2^2 - \frac{(y_1-\mu)^2}{2\sigma_1^2} - \frac{(y_2-\mu)^2}{2\sigma_2^2}.$$

The function L is unbounded because by choosing $\mu = y_1$ and $\sigma_1^2 \to 0$ we can make L arbitrarily large. But the likelihood equations give:

$$\hat{\mu} = \frac{y_1+y_2}{2}, \quad \hat{\sigma}_1^2 = \hat{\sigma}_2^2 = (\frac{y_1-y_2}{2})^2.$$

In fact these values do not maximize the likelihood function L because L has its maximum at a boundary point. Thus the ML equations fail to provide acceptable estimates. □

Example 1.6 shows that the solutions to the likelihood equations are not necessarily the ML estimates. After solving the likelihood equations careful study of the likelihood function is needed before accepting the solutions as the ML estimates.

Considering the log-likelihood function (1.7) and the corresponding likelihood equations (1.8), Hartley and Rao (1967) showed that the likelihood equations for model (1.1) when $\mathbf{y}$ has a normal distribution are:

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.9}$$

$$\left(\mathrm{Tr}\,\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j\right)\Theta = \left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right), \tag{1.10}$$

where $\mathrm{Tr}\,\mathbf{A}$ stands for the trace of matrix $\mathbf{A}$, $(\mathrm{Tr}\,\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j)$ is a matrix with the (i,j)th element equals to $\mathrm{Tr}\,\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j$, and $((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}))$ is a vector with the ith element equal to $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$.

Let $\mathbf{P}_V = \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}$. $\mathbf{P}_V$ is a projection matrix into $M(\mathbf{X})$ which is the subspace generated by the column vectors of $\mathbf{X}$ and the inner product of the space is defined as $(\mathbf{x},\mathbf{y}) = \mathbf{x}'\mathbf{V}\mathbf{y}$ where $\mathbf{x}$ and $\mathbf{y}$ are vectors in the space. We can then rewrite (1.9) as:

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{P}_V\mathbf{y}. \tag{1.11}$$

Substituting (1.11) into the right hand side of (1.10) we have:

$$\left(\mathrm{Tr}\,\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j\right)\Theta = \left(\mathbf{y}'(\mathbf{I}-\mathbf{P}_V)'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)\mathbf{y}\right). \tag{1.12}$$

Equations (1.11) and (1.12) do not solve β and Θ explicitly. Explicit solution to the likelihood equations (1.11) and (1.12) means that the solution β_0 and Θ_0 do not depend on the unknown parameters β and Θ. As shown by Herbach (1959) explicit solutions for Θ does not exist for the balanced two-way crossed model with interaction. Szatrowski and Miller (1980) give a procedure for determining whether or not the explicit ML estimator exists for balanced mixed models.

When there is no explicit ML estimator or when the data are unbalanced iterative computing must be used to solve the maximum likelihood equations. So far the properties of convergence and the speed of convergence of such iteration have not been studied analytically.

Hartley and J. N. K. Rao (1967) considered the asymptotic property of Θ assuming Θ is the global maximum of the likelihood. Miller (1977) considered the asymptotic property of the solutions to the likelihood equations. These authors have assumed some conditions for the level of factors and the number of observations so that situations as in Example 1.6 could not happen. Since the large-sample properties of the estimators are not the emphasis of this thesis, we shall not go into the details of asymptotic theory.

Example 1.7:

Consider a one-way balanced random model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

where μ is the mean, a_i and e_{ij} are random terms with variances σ_a^2 and σ_e^2 respectively.

The solutions to the likelihood equations are:

$$\hat{\sigma}_a^2 = \frac{1}{m}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{y}_{..})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2,$$

$$\hat{\sigma}_e^2 = \frac{1}{m(n-1)}\sum_{i-1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2.$$

With the data set given in Example 1.5 we have:

$$\hat{\sigma}_a^2 = -16\frac{1}{3}, \; \hat{\sigma}_e^2 = 52.$$

Since the variance component cannot be negative, $\hat{\sigma}_a^2 = -16\frac{1}{3}$ is not the ML estimate of σ_a^2. Herbach (1959) studied the likelihood function of the one-way balanced model and concluded that if a negative value appears the ML estimate should be zero and adjustment should also be made to $\hat{\sigma}_e^2$, giving:

$$\hat{\sigma}_a^2 = 0, \ \hat{\sigma}_e^2 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = 35\frac{2}{3}. \ \square$$

As we notice in Example 1.7, the ML estimate of σ_a^2 is different from the ANOVA estimate given in Example 1.5. The $\hat{\sigma_a^2}$ in Example 1.7 is biased. In the following section we shall introduce the restricted maximum likelihood estimators which is proposed to overcome the biasedness of the ML estimators.

1.3.3 The restricted maximum likelihood estimator

The restricted maximum likelihood estimator (REML) estimator was proposed by Patterson and Thompson (1971, 1974). It is observed from the likelihood equations (1.11) and (1.12) that the maximum likelihood estimate of Θ is a function of $(\mathbf{I} - \mathbf{P}_V)\mathbf{y}$ and the ML estimate of $\boldsymbol{\beta}$ is a function of $\mathbf{P}_V\mathbf{y}$. Patterson and Thompson suggested maximizing the likelihood function of $\mathbf{Sy}$ for the estimation of Θ and the likelihood function of $\mathbf{Qy}$ for the estimation of $\boldsymbol{\beta}$ where $\mathbf{Sy}$ and $\mathbf{Qy}$ are statistically independent and satisfy the conditions imposed by Patterson and Thompson (1971):

(1). The matrix $\mathbf{S}$ is of rank $n - t$ and $\mathbf{Q}$ is a matrix of rank t.

(2). The two parts are statistically independent, *i.e.* $\text{cov}(\mathbf{Sy}, \mathbf{Qy}) = 0$. This condition is met if $\mathbf{SVQ}' = 0$.

(3). The matrix $\mathbf{S}$ is chosen so that $\mathrm{E}(\mathbf{Sy}) = 0$, *i.e.* $\mathbf{SX} = 0$.

(4). The matrix $\mathbf{QX}$ is of rank t.

Since $\mathbf{Sy}$ and $\mathbf{Qy}$ are statistically independent and $\mathbf{Sy}$ has expectation zero, thus the likelihood function L' of $\mathbf{Sy}$ depends on Θ only, hence the likelihood function of $\mathbf{y}$ can be decomposed into the sum of the likelihood functions of $\mathbf{Sy}$ and $\mathbf{Qy}$, *i.e.*

$$L(\boldsymbol{\beta}, \Theta; \mathrm{y}) = L'(\Theta; \mathrm{Sy}) + L''(\boldsymbol{\beta}, \Theta; \mathrm{Qy}) \tag{1.13}$$

In Patterson and Thompson (1971) they specified $\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{Q} = \mathbf{X}'\mathbf{V}^{-1}$. The REML approach is thus specified to be the estimators derived using the specific choices of $\mathbf{S}$ and $\mathbf{Q}$. In deriving the likelihood functions of $\mathbf{Sy}$ and $\mathbf{Qy}$ since $\mathbf{SVS}'$ and $\mathbf{QVQ}'$ are singular, the conventional form of log-likelihood function is not applicable here. Patterson and Thompson (1971) have developed a way of writing the log-likelihood functions of $\mathbf{Sy}$ and $\mathbf{Qy}$.

Since $\mathbf{S}^2 = \mathbf{S}$, there exists an $n \times (n-t)$ matrix $\mathbf{A}$ such that $\mathbf{AA}' = \mathbf{S}$ and $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Now let $\mathbf{B}$ be an orthogonal matrix which diagonalizes $\mathbf{A}'\mathbf{VA}$, and let $\mathbf{P} = \mathbf{AB}$, then

$$\mathbf{P}'\mathbf{P} = \mathbf{I}, \quad \mathbf{PP}' = \mathbf{S}, \quad \mathbf{P}'\mathbf{VP} = diag(\tau_1, \ldots, \tau_{n-k}),$$

hence

$$L'(\Theta; \mathbf{Sy}) = \text{constant} - \frac{1}{2}\log|\mathbf{P}'\mathbf{VP}| - \frac{1}{2}\mathbf{y}'(\mathbf{SVS})^{-1}\mathbf{y}, \tag{1.14}$$

$$\begin{aligned} L''(\beta, \Theta; \mathbf{Qy}) &= \text{constant} - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\ &\quad -\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta). \end{aligned} \tag{1.15}$$

$$\frac{\partial L''}{\partial \beta} = 0 \text{ and } \frac{\partial L'}{\partial \Theta} = 0 \text{ give :}$$

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.16}$$

$$(\mathrm{Tr}\,\mathbf{RV}_i\mathbf{RV}_j)\,\Theta = (\mathbf{y}'\mathbf{RV}_i\mathbf{Ry}), \tag{1.17}$$

where $\mathbf{R} = \mathbf{V}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})$, $(\mathrm{Tr}\,\mathbf{RV}_i\mathbf{RV}_j)$ is the matrix with the (i,j)th element equals to $\mathrm{Tr}\,\mathbf{RV}_i\mathbf{RV}_j$ and $(\mathbf{y}'\mathbf{RV}_i\mathbf{Ry})$ is the vector with the ith element equal to $\mathbf{y}'\mathbf{RV}_i\mathbf{Ry}$.

It should be noticed that equation (1.16) for the REML estimate is the same as equation (1.9), *i.e.* REML gives the same estimate as ML estimate for β. But equation (1.17) is different from equation (1.10). When the ML estimate does not exist, *i.e.* the likelihood function $L(\beta, \Theta; \mathbf{y})$ is unbounded, then from (1.13), either L' or L'' or both are unbounded. Thus either the REML of Θ or the REML of β or the REML of both β and Θ do not exist. In other words, the REML approach does not improve the chances for the existence of the REML estimate over that of the ML estimate.

If equations (1.16) and (1.17) give explicit solutions for β and Θ then the REML estimates of β and Θ are unbiased. For nested balanced ANOVA models REML of Θ is identical to the ANOVA estimator. For unbalanced data REML also requires iterative computing as in the case of ML. REML also has unknown convergence properties.

It is not known if REML is robust to non-normal distributions. Since REML is identical to the ANOVA estimator for balanced nested ANOVA models and the ANOVA estimator has optimality in this case regardless of the distribution of data as shown in theorem 1.2, we suspect that REML is more robust than the ML estimator.

It is a very inspiring idea of Patterson and Thompson to split the original data into two independent data sets and maximize their corresponding likelihood functions. Geometrically, the REML approach projects the maximum likelihood function $L(\beta, \Theta; \mathrm{y})$ into two orthogonal subspaces to obtain $L'(\beta, \Theta; \mathrm{Sy})$ and $L''(\beta, \Theta; \mathrm{Qy})$ each of which has a lower dimension. The REML approach results in unbiasedness of the estimating equations. It would be interesting to see if we can extend the REML approach to estimate moments higher than the second and if the property of unbiasedness of the estimating equations is still preserved.

1.3.4 MIVQUE and MINQUE

MIVQUE stands for minimum variance invariant quadratic unbiased estimator and MINQUE stands for minimum norm invariant quadratic unbiased estimator. On the analogy of unbiased linear estimators for the means, Rao (1970, 1971a, 1971b, 1972) and LaMotte (1973) suggested using unbiased quadratic estimators for the variance components. To find a quadratic estimator is to determine a symmetric matrix $\mathbf{A}$ and use $\mathbf{y}'\mathbf{A}\mathbf{y}$ as the estimator.

Consider the general variance components model (1.1).

Let $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_k]$, $\boldsymbol{\xi} = [\xi_1', \xi_2', \ldots, \xi_k']'$, then the model can be written as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U}\boldsymbol{\xi},$$

where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$, $\mathrm{V}(\mathbf{y}) = \mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{U}_i \mathbf{U}_i' = \mathbf{U}\Delta_1\mathbf{U}'$, where

$$\Delta_1 = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{p_1} & & \\ & \ddots & \\ & & \sigma_k^2 \mathbf{I}_{p_k} \end{bmatrix}.$$

We now define what we mean by invariance.

Definition 1.1 *Assume* $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$ *and* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is a quadratic estimator of the variance component* $\sum_{i=1}^{k} q_i\sigma_i^2$. *If*

$$(\mathbf{y} - \mathbf{X}\beta)'\mathbf{A}(\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}'\mathbf{A}\mathbf{y},$$

then $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is invariant to the value of* β.

Throughout this thesis we shall say $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant for short if $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant to the mean vector of $\mathbf{y}$. We use the following lemmas and theorems to state the main properties of a quadratic estimator and the properties of MIVQUE and MINQUE.

Lemma 1.1 *Consider the general variance components model (1.1). If and only if* $\mathbf{AX} = 0$, *then* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is an invariant quadratic estimator.*

Proof: $\Rightarrow$ If $\mathbf{AX} = 0$, then

$$\begin{aligned}(\mathrm{y}-\mathrm{X}\beta)'\mathrm{A}(\mathrm{y}-\mathrm{X}\beta) &= \mathrm{y}'\mathrm{A}\mathrm{y}-\mathrm{y}'\mathrm{A}\mathrm{X}\beta-\beta'\mathrm{X}'\mathrm{A}\mathrm{y}+\beta'\mathrm{X}'\mathrm{A}\mathrm{X}\beta \\ &= \mathrm{y}'\mathrm{A}\mathrm{y},\end{aligned}$$

hence $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant.

$\Leftarrow$ If $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant, then from Definition 1.1,

$$(\mathrm{y}-\mathrm{X}\beta)'\mathrm{A}(\mathrm{y}-\mathrm{X}\beta) = \mathrm{y}'\mathrm{A}\mathrm{y},$$

for any random vector $\mathbf{y}$, *i.e.*

$$\mathrm{y}'\mathrm{A}\mathrm{X}\beta + \beta'\mathrm{X}'\mathrm{A}\mathrm{y} = \beta'\mathrm{X}\mathrm{A}\mathrm{X}\beta.$$

From the randomness of $\mathbf{y}$, the above equality holds only if $\mathbf{AX} = 0$. $\square$

Now we prove some lemmas for a more general model than model (1.1)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\eta}, \tag{1.18}$$

where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathrm{E}(\boldsymbol{\eta}) = 0$,

$$\mathrm{V}(\boldsymbol{\eta}) = \begin{bmatrix} \theta_1^2 & & \\ & \ddots & \\ & & \theta_N^2 \end{bmatrix} \stackrel{\text{def}}{=} \Delta_\theta,$$

where $N = \sum_{i=1}^k p_i$, then model (1.1) becomes a special case of model (1.18) when $\Delta_\theta = \Delta_1$.

Lemma 1.2 *If model (1.18) is considered and* $\mathbf{AX} = 0$, *then the expectation of the quadratic estimator* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is:*

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sum_{i=1}^{N} b_{ii}\theta_i^2, \tag{1.19}$$

where $\mathbf{B} = \mathbf{U}'\mathbf{A}\mathbf{U} = (b_{ij})_{N\times N}$.

Proof: Since $\mathbf{AX} = 0$, then

$$\begin{aligned}\mathbf{y}'\mathbf{A}\mathbf{y} &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\eta}'\mathbf{U}'\mathbf{A}\mathbf{U}\boldsymbol{\eta} \\ &= \boldsymbol{\eta}'\mathbf{B}\boldsymbol{\eta} = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\eta_i\eta_j.\end{aligned}$$

Thus

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\,\mathrm{E}(\eta_i\eta_j) = \sum_{i=1}^{N} b_{ii}\theta_i^2. \qquad \square$$

The diagonal elements of $\mathbf{B}$ play an important role in expressing the expectation and the variance of $\mathbf{y}'\mathbf{A}\mathbf{y}$ and will be used later, so we give it a notation. Let $\tilde{\mathbf{B}}$ denote the diagonal matrix with the diagonal elements equal to those of $\mathbf{B}$, *i.e.* if $\mathbf{B} = (b_{ij})_{N\times N}$, then $\tilde{\mathbf{B}} = diag(b_{11}, \ldots, b_{NN})$.

Corollary 1.1 *If the general variance components model (1.1) is considered and* $\mathbf{AX} = 0$, *then the expectation of the quadratic estimator* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is:*

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sum_{i=1}^{k} \sigma_i^2 \,\mathrm{Tr}\,\mathbf{A}\mathbf{V}_i.$$

Proof: Notice that model (1.1) is a special case of model (1.18) with $\Delta_\theta = \Delta_1$, from Lemma 1.2,

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sum_{i=1}^{N} b_{ii}\theta_i^2 = \sum_{i=1}^{k}\left(\sum_{j=p_{i-1}+1}^{p_i} b_{jj}\right)\sigma_i^2,$$

where $\sum_{j=p_{i-1}+1}^{p_i} b_{jj}$ is the partial summation of the diagonal elements of $\mathbf{B}$. But

$$\mathbf{B} = \mathbf{U}'\mathbf{A}\mathbf{U} = \begin{bmatrix}\mathbf{U}_1' \\ \vdots \\ \mathbf{U}_k'\end{bmatrix} \mathbf{A}\,[\,\mathbf{U}_1, \ldots, \mathbf{U}_k\,],$$

therefore

$$\sum_{j=p_{i-1}+1}^{p_i} b_{jj} = \mathrm{Tr}\,\mathbf{U}_i'\mathbf{A}\mathbf{U}_i = \mathrm{Tr}\,\mathbf{A}\mathbf{V}_i,$$

which proves the corollary. $\square$

Corollary 1.2 *If the general variance components model (1.1) is considered and* $\mathbf{AX} = 0$, *then* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is unbiased for* $\sum_{i=1}^{k} q_i\sigma_i^2$ *if and only if*

$$\mathrm{Tr}\,\mathbf{A}\mathbf{V}_i = q_i, \quad i = 1, \ldots, k.$$

Proof: $\Rightarrow$ If $\mathrm{Tr}\,\mathbf{AV}_i = q_i,\ \ i = 1, \ldots, k$, then from Corollary 1.1,

$$\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sum_{i=1}^{k} \sigma_i^2 \,\mathrm{Tr}\,\mathbf{AV}_i = \sum_{i=1}^{k} q_i \sigma_i^2,$$

thus $\mathbf{y}'\mathbf{A}\mathbf{y}$ is unbiased for $\sum_{i=1}^{k} q_i\sigma_i^2$.

$\Leftarrow$ If $\mathbf{y}'\mathbf{A}\mathbf{y}$ is unbiased for $\sum_{i=1}^{k} q_i\sigma_i^2$, then from Corollary 1.1 we should have:

$$\sum_{i=1}^{k} \sigma_i^2 \,\mathrm{Tr}\,\mathbf{AV}_i = \sum_{i=1}^{k} q_i \sigma_i^2.$$

The above equality holds for $\sigma_i^2 \geq 0,\ i = 1, \ldots, k$, hence

$$\mathrm{Tr}\,\mathbf{AV}_i = q_i,\ \ i = 1, \ldots, k. \quad \square$$

Now let

$$\Delta_\gamma = \begin{bmatrix} \theta_1^4 \gamma_1 & & \\ & \ddots & \\ & & \theta_N^4 \gamma_N \end{bmatrix},$$

where $\gamma_i = \mathrm{E}(\eta_i^4)/\theta_i^4 - 3$. The following lemma considers the covariance between two quadratic estimators.

Lemma 1.3 *Consider model (1.18). Let* $\mathbf{A}$ *and* $\mathbf{N}$ *be symmetric matrices with* $\mathbf{AX} = 0$ *and* $\mathbf{NX} = 0$, $\mathbf{B} = \mathbf{U}'\mathbf{A}\mathbf{U}$, $\mathbf{M} = \mathbf{U}'\mathbf{N}\mathbf{U}$, *then*

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) = 2\,\mathrm{Tr}\,\mathbf{B}\Delta_\theta\mathbf{M}\Delta_\theta + \mathrm{Tr}\,\tilde{\mathbf{B}}\Delta_\gamma\tilde{\mathbf{M}}. \tag{1.20}$$

Proof: Since $\mathbf{AX} = 0$ and $\mathbf{NX} = 0$,

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \boldsymbol{\eta}'\mathbf{U}'\mathbf{A}\mathbf{U}\boldsymbol{\eta} = \boldsymbol{\eta}'\mathbf{B}\boldsymbol{\eta},$$

$$\mathbf{y}'\mathbf{N}\mathbf{y} = \boldsymbol{\eta}'\mathbf{U}'\mathbf{N}\mathbf{U}\boldsymbol{\eta} = \boldsymbol{\eta}'\mathbf{M}\boldsymbol{\eta}.$$

Let $\mathbf{B} = (b_{ij})_{N\times N}$, $\mathbf{M} = (m_{ij})_{N\times N}$, $\boldsymbol{\eta}' = (\eta_1, \ldots, \eta_N)$. Then

$$\begin{aligned} \mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) &= \mathrm{cov}(\boldsymbol{\eta}'\mathbf{B}\boldsymbol{\eta}, \boldsymbol{\eta}'\mathbf{M}\boldsymbol{\eta}) \\ &= \mathrm{cov}\left(\sum_{i,j} b_{ij}\eta_i\eta_j, \sum_{k,l} m_{kl}\eta_k\eta_l\right) \\ &= \sum_{i,j}\sum_{k,l} b_{ij}m_{kl}\,\mathrm{cov}\,(\eta_i\eta_j, \eta_k\eta_l) \\ &= \sum_{(i,j)=}\sum_{(k,l)} b_{ij}m_{kl}\,\mathrm{V}(\eta_i\eta_j) + \sum_{(i,j)\neq}\sum_{(k,l)} b_{ij}m_{kl}\,\mathrm{cov}\,(\eta_i\eta_j, \eta_k\eta_l) \end{aligned}$$

$$
\begin{aligned}
&= \sum_i b_{ii}m_{ii}\left[\mathrm{E}(\eta_i^4) - \theta_i^4\right] + \sum_{i\neq j} b_{ij}m_{ij}\theta_i^2\theta_j^2 \\
&\quad + \sum_{\substack{(i,j)\neq \\ i=l,}} \sum_{\substack{(k,l) \\ j=k}} b_{ij}m_{kl}\operatorname{cov}(\eta_i\eta_j, \eta_j\eta_i) \\
&\quad + \sum_{\substack{(i,j)\neq \\ (i,j)\neq}} \sum_{\substack{(k,l) \\ (l,k)}} b_{ij}m_{kl}\operatorname{cov}(\eta_i\eta_j, \eta_k\eta_l) \\
&= \sum_i b_{ii}m_{ii}\left[\mathrm{E}(\eta_i^4) - \theta_i^4\right] + \sum_{i\neq j} b_{ij}m_{ij}\theta_i^2\theta_j^2 + \sum_{i\neq j} b_{ij}m_{ij}\theta_i^2\theta_j^2 \\
&= 2\sum_i\sum_j b_{ij}m_{ij}\theta_i^2\theta_j^2 + \sum_i b_{ii}m_{ii}\theta_i^4\gamma_i \\
&= 2\,\mathrm{Tr}\,\mathbf{B}\Delta_\theta\mathbf{M}\Delta_\theta + \mathrm{Tr}\,\tilde{\mathbf{B}}\Delta_\gamma\tilde{\mathbf{M}},
\end{aligned}
$$

because $\gamma_i = \mathrm{E}(\eta_i^4)/\theta_i^4 - 3$. □

Now let

$$
\Delta_2 = \begin{bmatrix} \sigma_1^4\gamma_1\mathbf{I}_{p_1} & & & \\ & \sigma_2^4\gamma_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^4\gamma_k\mathbf{I}_{p_k} \end{bmatrix},
$$

it is noticed that Δ_2 is a special case of Δ_γ. The following corollary gives the variance of the quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$.

Corollary 1.3 *Consider the general variance components model (1.1). If* $\mathbf{AX} = 0$, *then the variance of the quadratic estimator* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is:*

$$
\mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2\,\mathrm{Tr}\,\mathbf{B}\Delta_1\mathbf{B}\Delta_1 + \mathrm{Tr}\,\tilde{\mathbf{B}}\Delta_2\tilde{\mathbf{B}} \tag{1.21}
$$

Proof: Since model (1.1) is a special case of model (1.18) with $\Delta_\theta = \Delta_1$, $\Delta_\gamma = \Delta_2$, let $\mathbf{A} = \mathbf{N}$, the conclusion follows from (1.20). □

It can be seen from Corollary 1.3 that the variance of a quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ needs information up to the fourth moment of the distribution of $\mathbf{y}$. However, if $\mathbf{y}$ comes from a normal distribution, then the kurtosis $\gamma_i = 0$, $i = 1, \ldots, k$, hence $\Delta_2 = 0$, and then (1.21) reduces to:

$$
\mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2\,\mathrm{Tr}\,\mathbf{B}\Delta_1\mathbf{B}\Delta_1.
$$

Now we have established the equivalent mathematical conditions for the statistical constraints of a quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$:

1. invariant $\leftrightarrow$ $\mathbf{AX} = 0$;

2. unbiased $\leftrightarrow$ $\operatorname{Tr}\mathbf{AV}_i = q_i$, $i = 1, \ldots, k$;
3. minimum variance $\leftrightarrow$ $\mathbf{A}$ minimizes $2\operatorname{Tr}\mathbf{B}\Delta_1\mathbf{B}\Delta_1 + \operatorname{Tr}\tilde{\mathbf{B}}\Delta_2\tilde{\mathbf{B}}$.

We can then define MIVQUE and MINQUE in terms of mathematical equations.

Definition 1.2 *Consider the general variance components model (1.1). If* $\mathbf{A}$ *is a symmetric matrix satisfying the following conditions:*

$$\begin{cases} \mathbf{AX} = 0, \\ \operatorname{Tr}\mathbf{AV}_i = q_i, \quad i = 1, \ldots, k, \\ \mathbf{A} \text{ minimizes } 2\operatorname{Tr}\mathbf{B}\Delta_1\mathbf{B}\Delta_1 + \operatorname{Tr}\tilde{\mathbf{B}}\Delta_2\tilde{\mathbf{B}}, \end{cases} \tag{1.22}$$

then $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is the MIVQUE of* $\sum_{i=1}^k q_i\sigma_i^2$.

C. R. Rao (1973, *p*303) pointed out that a natural estimator of $\sum_{i=1}^k q_i\sigma_i^2$ is $\boldsymbol{\xi}'\Delta\boldsymbol{\xi}$, where

$$\Delta = \begin{bmatrix} \frac{q_1}{n_1}\mathbf{I}_{p_1} & & & \\ & \frac{q_2}{n_2}\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \frac{q_k}{n_k}\mathbf{I}_{p_k} \end{bmatrix}.$$

Now let $\mathbf{W}_i = \alpha_i\mathbf{U}_i$, $\mathbf{W} = [\mathbf{W}_1|\ldots|\mathbf{W}_k]$. It can be verified that if there exists a symmetric matrix $\mathbf{A}$ such that $\mathbf{W}'\mathbf{A}\mathbf{W} = \Delta$, then $\mathbf{A}$ is the solution to the last two constraints of (1.22). In practice, such an $\mathbf{A}$ does not always exist. Since the proposed quadratic estimator is

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \boldsymbol{\xi}'\mathbf{W}'\mathbf{A}\mathbf{W}\boldsymbol{\xi},$$

Rao (1973, *p*303) argued that by minimizing the norm $\| \mathbf{W}'\mathbf{A}\mathbf{W} - \Delta \|$, we make the difference between the proposed estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ and the natural estimator small according to a suitable chosen norm.

Rao (1973) proposed MINQUE for general forms of norm. In practice, the solution of $\mathbf{A}$ is only found for the MINQUE using a Euclidean norm. In this thesis we shall restrict ourselves to the MINQUE using the Euclidean norm, therefore the term MINQUE used in the following context means MINQUE using the Euclidean norm.

Definition 1.3 *If* $\mathbf{A} = (a_{ij})_{N\times N}$, *then the Euclidean norm of* $\mathbf{A}$ *is defined as:*

$$\| \mathbf{A} \|_E^2 = \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2.$$

We can therefore define MINQUE in terms of Euclidean norm.

When $\mathbf{A}$ is symmetric then $\| \mathbf{A} \|_E^2 = \operatorname{Tr} \mathbf{AA}$, hence

$$\| \mathbf{W'AW} - \Delta \|_E^2 = \operatorname{Tr} \mathbf{W'AWW'AW} - 2 \operatorname{Tr} \mathbf{W'AW}\Delta + \operatorname{Tr} \Delta\Delta.$$

Now since $\operatorname{Tr} \mathbf{AW}_i\mathbf{W}_i' = q_i$, $i = 1, \ldots, k$, from the unbiasedness requirement we have

$$\begin{aligned} \operatorname{Tr} \mathbf{W'AW}\Delta &= \operatorname{Tr} \mathbf{AW}\Delta\mathbf{W'} \\ &= \operatorname{Tr} \mathbf{A} \left(\sum_{i=1}^{k} \frac{q_i}{p_i} \mathbf{W}_i \mathbf{W}_i' \right) \\ &= \sum_{i=1}^{k} \frac{q_i^2}{p_i} = \operatorname{Tr} \Delta\Delta. \end{aligned}$$

Therefore

$$\begin{aligned} \| \mathbf{W'AW} - \Delta \|_E^2 &= \operatorname{Tr} \mathbf{W'AWW'AW} - \operatorname{Tr} \Delta\Delta \\ &= \operatorname{Tr} \mathbf{AWW'AWW'} - \operatorname{Tr} \Delta\Delta \\ &= \| \mathbf{AV} \|_E^2 - \| \Delta \|_E^2 . \end{aligned}$$

Since $\| \Delta \|_E^2$ is a constant, minimizing $\| \mathbf{W'AW} - \Delta \|_E^2$ is equivalent to minimizing $\| \mathbf{AV} \|_E^2$.

Definition 1.4 *Consider the general variance components model (1.1). If a symmetric matrix* $\mathbf{A}$ *satisfies:*

$$\begin{cases} \mathbf{AX} = 0, \\ \operatorname{Tr} \mathbf{AV}_i = q_i, \quad i = 1, \ldots, k, \\ \mathbf{A} \text{ minimizes } \| \mathbf{AV} \|_E^2, \end{cases} \tag{1.23}$$

then $\mathbf{y'Ay}$ *is the MINQUE of* $\sum_{i=1}^{k} q_i\sigma_i^2$.

Theorem 1.4 (Rao, 1971b) *Consider the general variance components model (1.1), then the MINQUE of* $\sum_{i=1}^{k} q_i\sigma_i^2$ *is* $\mathbf{y'Ay}$, *where*

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{RV}_i\mathbf{R}, \tag{1.24}$$

$$\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X'V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X'V}^{-1}, \tag{1.25}$$

and λ_i*'s satisfy the following equations:*

$$\sum_{i=1}^{k} \lambda_i \operatorname{Tr} \mathbf{RV}_i\mathbf{RV}_j = q_j, \quad j = 1, \ldots, k. \tag{1.26}$$

A proof can be found in Rao (1973, $p65$).

Equations (1.24)—(1.26) will be referred in later chapters as the MINQUE formulas or the MINQUE equations.

The existence of the MINQUE depends on that of the solutions to equation (1.26). If the matrix $(\operatorname{Tr} \mathbf{RV}_i\mathbf{RV}_j)$ is nonsingular, then there will be solutions for the λ_i's. Otherwise, there will be no MINQUE for the variance components.

MINQUE has its appeal in dealing with normally distributed data because it is also the MIVQUE. The fact is proved in the following corollary.

Corollary 1.4 *Consider model (1.1). If the data vector* $\mathbf{y}$ *is normally distributed with* $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathrm{V}(\mathbf{y}) = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$, *then the MINQUE of* $\sum_{i=1}^{k} q_i\sigma_i^2$ *is also the MIVQUE of* $\sum_{i=1}^{k} q_i\sigma_i^2$.

Proof: If $\mathbf{y}$ is normally distributed, then the kurtosis of the distribution $\gamma_i = 0$, and then $\Delta_2 = 0$. Hence

$$\mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2 \operatorname{Tr} \mathbf{B}\Delta_1\mathbf{B}\Delta_1 = 2 \parallel \mathbf{B}\Delta_1 \parallel_E^2 = 2 \parallel \mathbf{AV} \parallel_E^2,$$

hence the MINQUE $\mathbf{y}'\mathbf{A}\mathbf{y}$ of $\sum_{i=1}^{k} q_i\sigma_i^2$ is also the MIVQUE of $\sum_{i=1}^{k} q_i\sigma_i^2$.
□

The following lemma states the additivity of MINQUE.

Lemma 1.4 (Rao, 1970) *If* S_1 *is the MINQUE of* $\sum_{i=1}^{k} c_i\sigma_i^2$ *and* S_2 *of* $\sum_{i=1}^{k} d_i\sigma_i^2$, *then* $S_1 + S_2$ *is the MINQUE of* $\sum_{i=1}^{k} (c_i + d_i)\sigma_i^2$.

The additivity property of MINQUE eases the task of estimating linear combinations of the variance components. It is sufficient to estimate each single variance component and use these estimates to estimate the linear combinations.

Although MINQUE is not MIVQUE for distributions other than normal, MINQUE has been used as an algorithm for any distribution, in the same way as the ML estimator derived for the normal distribution has been used for other distributions.

Looking at the MINQUE formulas (1.24)-(1.26), we notice that the computation of the matrix $\mathbf{R}$ uses $\mathbf{V}^{-1}$, which is the inverse of the variance covariance of $\mathbf{y}$, and hence requires the σ_i^2's. Since the true values of the variance components are never known, the computation of MINQUE requires prior values of the variance components. As we shall show in Example 1.8 sometimes the σ_i^2's cancel out and therefore the MINQUE is a function of the data only. In Chapter 3 there will be further investigations on the conditions for the σ_i^2's to cancel out. In the situations where σ_i^2's do not cancel out we are left with an infinite number of choices for the prior values of the σ_i^2 provided $\sigma_i^2 \geq 0$. MINQUE is not only a function of the data but also a function of the predetermined prior values. Different prior values then lead to different estimators. This is a very big disadvantage of the MINQUE.

Two approaches have been used to 'overcome' this disadvantage of MINQUE. Hartley, LaMotte and J. N. K. Rao (1978) have suggested a synthesis-based MINQUE which assumes the prior value for the variance of the random error to be 1, *i.e.* $\sigma_e^2 = 1$, and all the other prior values are chosen to be zero. The MINQUE obtained using this specific set of prior values is called the synthesis-based MINQUE. By using this specific set of prior values the variance covariance matrix used in the MINQUE formulas (1.24)–(1.26) is the identity matrix, hence the computation involved in deriving the MINQUE is greatly reduced. The synthesis-based MINQUE has its appeal in computational simplicity. But by gaining this simplicity we must trade off the optimality of MINQUE. Swallow and Monahan (1984) carried out Monte Carlo comparisons of the existing estimators for the one-way random model. Their results show that when $\sigma_a^2/\sigma_e^2 \geq 1$, the synthesis-based MINQUE performs poorly even in mildly unbalanced cases. We suggest in Chapter 3 that only when we can decide that MINQUE is independent of the prior values, which can be examined using the necessary and sufficient conditions given in Chapter 3, should the synthesis-based MINQUE be used. In that case we are enjoying the computational simplicity without losing optimality of the estimator.

When the σ_i^2's do not cancel out in the computation of MINQUE the only way that may lead to an unique estimate is by iterative computing, provided that the process converges to a unique point from any starting point. The estimate obtained from iterative computing will no longer be a quadratic function of the data vector $\mathbf{y}$ although the estimating equation is a quadratic function of $\mathbf{y}$, and it also loses the properties of unbiasedness and minimum variance. As in the discussions on ML and REML estimators there have been no systematic results on the properties of the convergence. We virtually have no control over the speed of the convergence of the process and those who do iterative computing have to rely on their 'good fortune' for the process to converge quickly to a stationary point and also hope that this is the unique converging point.

The MINQUE formulas (1.24)—(1.26) result in the same estimating equations as the REML equation for the estimation of variance components (1.17). To show this, suppose we want to estimate $\Theta = (\sigma_1^2, \ldots, \sigma_k^2)'$ using MINQUE. We will have:

$$\hat{\Theta} = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2)' = (\mathbf{y}'\mathbf{A}_1\mathbf{y}, \ldots, \mathbf{y}'\mathbf{A}_k\mathbf{y})',$$

where $\mathbf{A}_i = \sum_{j=1}^k \lambda_j^{(i)}\mathbf{R}\mathbf{V}_j\mathbf{R}$, and $\lambda_j^{(i)}$ satisfy:

$$\sum_{j=1}^k \lambda_j^{(i)} \operatorname{Tr}\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Let $\mathbf{S} = (s_{ij})_{k\times k}$, where $s_{ij} = \operatorname{Tr}\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j$, and let $\mathbf{t} = (\mathbf{y}'\mathbf{R}\mathbf{V}_1\mathbf{R}\mathbf{y}, \ldots, \mathbf{y}'\mathbf{R}\mathbf{V}_k\mathbf{R}\mathbf{y})'$, then the MINQUE of Θ can be written as

$$\mathbf{S}\Theta = \mathbf{t}, \tag{1.27}$$

or equivalently,

$$(\operatorname{Tr}\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j)\Theta = (\mathbf{y}'\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{y}). \tag{1.28}$$

Equation (1.28) gives the same estimating equation as (1.17) of REML. The difference between MINQUE and REML is that MINQUE assumes that the equation (1.28) does not contain any unknown values of the variance components, apart from Θ, because prior values are used in place of the variance components. In other words MINQUE regards (1.28) as a conventional expression for a estimator:

$$\Theta = (\operatorname{Tr}\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j)^{-1}(\mathbf{y}'\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{y}).$$

and every term in the right hand side of the above equation does not contain unknowns. The REML approach gives simultaneous estimates for β and the variance components. Since the estimation of variance components in REML is independent of the estimation of β, equation (1.17) itself can be a estimating equation. REML regards (1.17) as an estimating equation with all the variance components involved in both sides of the equation being treated as unknowns. Therefore iterative computing is a way of solving the equations and obtaining estimates. Iterative computing gives both approaches the same result. But REML will take it as its estimate while MINQUE regard it as a way of solving the problem of prior values. Hence iterative computation is more justified for the REML approach than the MINQUE approach.

Example 1.8:

Swallow and Searle (1978) give explicit expressions for the MINQUE of σ_a^2 and σ_e^2 for the one-way unbalanced random model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i$$

where μ is the mean, a_i and e_{ij} are random terms with variance components σ_a^2 and σ_e^2. Let

$$k_i = \frac{n_i}{\sigma_e^2 + n_i\sigma_a^2}, \; k = \frac{1}{\sum_{i=1}^m k_i}, \; \bar{y}_{i.} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i},$$

$$S_{11} = \sum_{i=1}^m k_i^2 - 2k\sum_{i=1}^m k_i^3 + k^2\left(\sum_{i=1}^m k_i^2\right)^2,$$

$$S_{12} = \sum_{i=1}^m \frac{k_i^2}{n_i} - 2k\sum_{i=1}^m \frac{k_i^3}{n_i} + k^2\sum_{i=1}^m k_i^2 \sum_{i=1}^m \frac{k_i^2}{n_i},$$

$$S_{22} = \frac{N-m}{\sigma_e^2} + \sum_{i=1}^m \frac{k_i^2}{n_i^2} - 2k\sum_{i=1}^m \frac{k_i^3}{n_i^2} + k^2\left(\sum_{i=1}^m \frac{k_i^2}{n_i}\right)^2,$$

$$u_1 = \sum_{i=1}^m k_i^2\left(\bar{y}_{i.} - k\sum_{i=1}^m k_i\bar{y}_{i.}\right)^2,$$

$$u_2 = \frac{1}{\sigma_e^2}\left(\sum_{i=1}^m\sum_{j=1}^{n_1} y_{ij}^2 - \sum_{i=1}^m n_i\bar{y}_{i.}^2\right) + \sum_{i=1}^m \frac{k_i^2}{n_i}\left(\bar{y}_{i.} - k\sum_{i=1}^m k_i\bar{y}_{i.}\right)^2.$$

Then the MINQUE of σ_a^2 and σ_e^2 are:

$$\hat{\sigma}_a^2 = \frac{1}{|S|}(S_{22}u_1 - S_{12}u_2), \tag{1.29}$$

$$\hat{\sigma}_e^2 = \frac{1}{|S|}(-S_{12}u_1 + S_{11}u_2), \tag{1.30}$$

where $|S| = S_{11}S_{22} - S_{12}^2$.

When we use balanced data, *i.e.* $n_i = n$, the σ_a^2 and σ_e^2 will cancel out in (1.29) and (1.30) and $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$ are identical to the ANOVA estimators given in Example 1.5. When we use unbalanced data, $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$ are functions of σ_a^2 and σ_e^2, we need to assign prior values to σ_a^2 and σ_e^2 to obtain MINQUE estimates.

For the data set in Example 1.5 the MINQUE estimates are the same as the ANOVA estimates:

$$\hat{\sigma}_a^2 = -15\frac{1}{3}, \; \hat{\sigma}_e^2 = 52. \quad \square$$

Another disadvantage of MINQUE is that MINQUE can give negative estimates for variance components.

1.3.5 Goldstein's method

Goldstein (1986) observed that by writting the matrix $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'$ in vector form we can transform the problem of quadratic estimation into that of linear estimation.

First we state the well known result on least squares estimation. Suppose we have a linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.31}$$

where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathrm{V}(\mathbf{y}) = \mathbf{V}$, then if $\mathbf{V}$ does not contain unknown parameters and if $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ exists, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{1.32}$$

Variance is used as the optimality criterion. $\hat{\boldsymbol{\beta}}$ has the minimum variance among all linear unbiased estimators of $\boldsymbol{\beta}$.

To transform quadratic forms into linear forms we need the following definition to allow matrices to be transformed into vectors.

Definition 1.5 *Let* $\mathbf{A} = (a_{ij})_{n\times n}$ *be a symmetric matrix and let* $\mathrm{vec}(\mathbf{A})$ *be a vector formed by connecting each column of* $\mathbf{A}$, *similarly let* $\mathrm{vech}(\mathbf{A})$ *be a vector formed by connecting each column of the upper triangle of* $\mathbf{A}$, *i.e.*

$$\mathrm{vec}(\mathbf{A}) = (a_{11}, a_{21}, \ldots, a_{n1}, a_{12}, \ldots, a_{nn})',$$

$$\mathrm{vech}(\mathbf{A}) = (a_{11}, a_{12}, a_{22}, a_{13}, a_{23}, a_{33}, \ldots, a_{nn})'.$$

Now we assume that $\mathrm{V}(\mathbf{y}) = \mathbf{V} = f(\boldsymbol{\Theta})$, where $\mathrm{V}(\mathbf{y})$ is the variance covariance matrix of $\mathbf{y}$, $\boldsymbol{\Theta}$ is a vector containing the unknown parameters and f is a linear function in $\boldsymbol{\Theta}$.

For model (1.31) $\mathrm{E}\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\right) = \mathbf{V}$, let

$$\mathbf{y}^* = \mathrm{vec}\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\right),\ \mathbf{V}^* = \mathrm{vec}(\mathbf{V}) = \mathbf{X}^*\boldsymbol{\Theta},$$

where $\mathbf{X}^*$ is the design matrix which relates $\mathbf{y}^*$ to $\boldsymbol{\Theta}$. From model (1.31) we can generate another 'linear' model:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\Theta} + \boldsymbol{\varepsilon}^*, \tag{1.33}$$

where $\mathrm{E}(\mathbf{y}^*) = \mathbf{X}^*\boldsymbol{\Theta}$, and $\mathrm{V}(\mathbf{y}^*) = \mathbf{V}^{**}$. If $\boldsymbol{\beta}$ is known, then $\mathbf{y}^*$ is a known vector, then for model (1.33), the 'BLUE' of $\boldsymbol{\Theta}$ is:

$$\hat{\Theta} = \left(\mathbf{X}^{*\prime}\mathbf{V}^{**-1}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{**-1}\mathbf{y}^{*}. \tag{1.34}$$

$\hat{\Theta}$ is linear in the form of $\mathbf{y}^*$, but since $\mathbf{y}^*$ is a quadratic form of the original data $\mathbf{y}$, (1.34) in fact gives a quadratic estimator for the unknown parameter Θ.

Goldstein's method can not only be used for the estimation of variance components, which is the case when Θ contains the variance components, but can also be used to estimate a wider range of parameters involving higher moments of the distributions provided the variance covariance matrix can be written as linear function of such parameters. We shall discuss how Goldstein's method can be used to estimate variance components later in this section.

We have assumed $\mathbf{V}$ is known in (1.32) and $\boldsymbol{\beta}$ is known in (1.34). In practice both $\mathbf{V}$ and $\boldsymbol{\beta}$ will contain unknowns. Goldstein suggested combining (1.32) and (1.34) to form a system of equations. Hence iterative computing can be used to find simultaneous estimates for $\boldsymbol{\beta}$ and Θ.

Goldstein proved that in the situation where Θ contains variance components his method is equivalent to maximum likelihood estimation.

Goldstein (1989) proposed another restricted version of his method which is equivalent to the restricted maximum likelihood estimation approach in the normal distribution case. Since the restricted version is parallel to the REML method we discussed above, we shall restrict ourselves to discuss the method in Goldstein (1986).

To show how Goldstein's method can be used to estimate variance components from the general variance components model, we have to define the Kronecker product of matrices which we shall also use in later chapters.

Definition 1.6 *Let* $\mathbf{A}$ *be an* $m_1 \times n_1$ *matrix and* $\mathbf{B}$ *be an* $m_2 \times n_2$ *matrix, then the Kronecker product of* $\mathbf{A}$ *and* $\mathbf{B}$ *which we write as* $\mathbf{A} \otimes \mathbf{B}$ *is an* $m_1 m_2 \times n_1 n_2$ *matrix defined by:*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n_1}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n_1}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m_1 1}\mathbf{B} & a_{m_1 2}\mathbf{B} & \dots & a_{m_1 n_1}\mathbf{B} \end{bmatrix}$$

This definition is sometimes called the right Kronecker product of matrices.

Now we consider applying Goldstein's method to estimate variance components. Recall the notations we used for the general variance components model (1.1)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\boldsymbol{\xi}_1 + \cdots + \mathbf{U}_k\boldsymbol{\xi}_k$$

Let $\mathbf{V}_i = \mathbf{U}_i\mathbf{U}_i'$, $\mathbf{V} = \sum_{i=1}^k \sigma_i^2 \mathbf{V}_i$ and $\Theta = (\sigma_1^2, \ldots, \sigma_k^2)'$. Following the notations we introduced in this section, $\mathbf{y}^* = \mathrm{vec}((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})')$, and using the concept of Kronecker product we can see that $\mathbf{y}^* = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \otimes (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$. $\mathbf{X}^* = (\mathrm{vec}\,\mathbf{V}_1, \ldots, \mathrm{vec}\,\mathbf{V}_k)$, *i.e.* $\mathbf{X}^*$ is a matrix with the ith column equal to $\mathrm{vec}\,\mathbf{V}_i$. Then

$$\mathbf{V}^* = \mathrm{vec}(\mathbf{V}) = \sum_{i=1}^k \sigma_i^2\,\mathrm{vec}(\mathbf{V}_i) = \mathbf{X}^*\Theta.$$

In (1.32) and (1.34) the only structure we need now is the variance covariance matrix of $\mathbf{y}^*$, all the other terms are either design matrices or are known function of design matrices.

$$\begin{aligned} \mathbf{V}^{**} &= \mathrm{V}(\mathbf{y}^*) = \mathrm{V}\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \otimes (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right) \\ &= \mathrm{V}\left(\mathbf{U}\boldsymbol{\xi} \otimes \mathbf{U}\boldsymbol{\xi}\right) = (\mathbf{U} \otimes \mathbf{U})\,\mathrm{V}(\boldsymbol{\xi} \otimes \boldsymbol{\xi})\,(\mathbf{U}' \otimes \mathbf{U}') \end{aligned}$$

where $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_k)$, which depends on the design matrices $\mathbf{U}_i$ only, and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_k')'$. We need to give an expression for $\mathrm{V}(\boldsymbol{\xi} \otimes \boldsymbol{\xi})$.

To simplify the derivation of $\mathrm{V}(\boldsymbol{\xi} \otimes \boldsymbol{\xi})$ we start to work with $\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$ where $\boldsymbol{\eta}$ comes from model (1.18) with $\mathrm{E}(\boldsymbol{\eta}) = 0$, $\mathrm{V}(\boldsymbol{\eta}) = \begin{bmatrix} \theta_1^2, & & \\ & \ddots & \\ & & \theta_N^2 \end{bmatrix}$. The general variance components model will then be a special case of model (1.18) with

$$\mathrm{V}(\boldsymbol{\xi}) = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{p_1} & & \\ & \ddots & \\ & & \sigma_k^2 \mathbf{I}_{p_k} \end{bmatrix},$$

i.e. the variance components matrix of $\boldsymbol{\xi}$ is a patterned diagonal matrix.

On the diagonal of $\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$,

$$\mathrm{V}(\eta_i\eta_j) = \begin{cases} \theta_i^4\gamma_i + 2\theta_i^4, & i = j, \\ \theta_i^2\theta_j^2, & i \neq j, \end{cases}$$

where $\gamma_i = \mathrm{E}(\eta_i^4)/\sigma_i^4 - 3$.

Off the diagonal of $\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$

$$\mathrm{cov}(\eta_i\eta_j, \eta_k\eta_l) = \begin{cases} \theta_i^2\theta_j^2, & \text{if } i = l, j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$$

$$
= \begin{bmatrix}
\theta_1^4\gamma_1 + 2\theta_1^4 & & & & & & \\
& \theta_1^2\theta_2^2 & & \theta_1^2\theta_2^2 & & & \\
& & \ddots & & & & \\
& & \theta_1^2\theta_N^2 & & & & \\
& \theta_1^2\theta_2^2 & & \theta_2^2\theta_1^2 & & & \\
& & & & \theta_2^4\gamma_2 + 2\theta_2^4 & & \\
& & & & & \ddots & \\
& & & & & & \theta_N^4\gamma_N + 2\theta_N^4
\end{bmatrix} \tag{1.35}
$$

$\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$ is a matrix the elements of which are functions of the variance components and the kurtosis of the distribution of $\mathbf{y}$. Imposing the patterns of $\mathrm{V}(\boldsymbol{\xi})$ into $\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$ we obtain $\mathrm{V}(\boldsymbol{\xi} \otimes \boldsymbol{\xi})$, Hence $\mathbf{V}^{**}$.

From (1.32) and (1.34) substituting the matrices we obtained so far gives the estimating equations. Notice that if $\mathbf{y}$ has a normal distribution then $\gamma_i = 0$ in the expression for $\mathrm{V}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$, $i = 1, \ldots, k$. Then the estimating equations contain $\boldsymbol{\beta}$ and σ_i^2 only. Hence iterative computing can be used to solve the equation system. When $\mathbf{y}$ does not have a normal distribution, then $\gamma_i \neq 0$, and the estimating equations in this case contain not only $\boldsymbol{\beta}$ and σ_i^2, but also the kurtosis of the distribution. It may be helpful to study the distribution of the data carefully and specify the γ_i in $\mathbf{V}^{**}$ so that the estimating equations only contain $\boldsymbol{\beta}$ and σ_i^2, hence iterative computing can be used. If no information on γ_i is available, we can use $\gamma_i = 0$, $i = 1, \ldots, k$, *i.e.* assume a normal distribution for the data vector $\mathbf{y}$ and carry out the iterative computing on σ_i^2.

Comparing (1.32) and (1.34) where $\mathbf{V}^{**}$ is given by (1.35) with the ML estimating equations (1.9)–(1.10), the REML estimating equations (1.16)–(1.17), and the MINQUE formulas (1.24)–(1.26), we notice that Goldstein's method is the only estimating procedure which takes the kurtosis of the distribution into consideration. In other words, Goldstein's method has an adjustment which allows the distribution of $\mathbf{y}$ to vary while the other methods are all designed for the normal distribution only. If the γ_i's we choose are not far from the true values of the kurtosis, we expect the estimates given by Goldstein's method to be better than all the estimators we have discussed so far.

Bradley (1973) proved the equivalence of maximum likelihood and weighted least square estimates for a member of the exponential family with one parameter. It will be interesting to investigate whether for a certain class of distributions (exponential family or otherwise?) Goldstein's method is equivalent to maximum likelihood estimation. In that case, Goldstein's method will be an extension of the maximum likelihood estimation to the non-normal case.

1.3.6 Nonnegative estimators

All the estimators we have introduced so far have the problem of giving negative estimates.

Theoretically, ML and REML estimators should always be nonnegative because we can restrict the parameter space to be nonnegative. In practice what we obtain is the solution to the likelihood equations which can be negative. When a negative solution appears, we know that it is not the ML/REML estimate and we need to look for another nonnegative estimate. Herbach (1959) showed that for the balanced one-way random model if a negative value appears then zero is the ML/REML estimate and modification is needed for the other estimates. Such results need analytical investigation of the likelihood function. For other models we do not have this result. Though it is a common practice to use zero as the value of a variance component if the corresponding solution to the likelihood equation is negative, this approach has not been justified except in the case of the balanced one-way random model.

It is also a common practice to put the negative ANOVA or MINQUE estimates to zero. By doing so we change the original estimator into a new one which will lose some of the optimality of the original estimator. For example, the ANOVA and MINQUE estimators will be biased, so that minimum variance may not be appropriate as an optimality criterion.

Several estimators have considered the constraint of nonnegativity.

J.N.K. Rao and Chaubey (1978) considered the problem of MINQUE having negative estimates and modified the MINQUE formulas to construct a nonnegative estimator. Since this estimator is biased and has a form very similar to MINQUE, Rao and Chaubey called the estimator MINQE (MINQUE without unbiasedness). MINQE was constructed by minimizing variance as MINQUE did. It is nonnegative and has additivity. But since it is biased we can argue that variance alone is not an appropriate optimality criterion. Another deficiency of MINQE is that whenever a nonnegative unbiased estimator is possible, for example, in the balanced ANOVA model case the ANOVA estimator for the random error is nonnegative and unbiased, then MINQE may not coincide with it. In other words, MINQE does not always give the 'best' possible nonnegative estimates.

Chaubey (1983) used the spectral decomposition approach and derived a new nonnegative estimator, CMINQUE (the estimator closest to MINQUE). Instead of using $\mathbf{y}'\mathbf{A}\mathbf{y}$ as in MINQUE theory, Chaubey proposed to use the positive eigenvalues and the corresponding eigenvectors of $\mathbf{A}$ to form a new nonnegative matrix $\mathbf{B}$. He then uses $\mathbf{y}'\mathbf{B}\mathbf{y}$ as the CMINQUE estimator. This method is very natural but is intuitive. CMINQUE coincides with MINQUE whenever MINQUE is both nonnegative and unbiased. The deficiency of CMINQUE is that its existence depends on MINQUE. When MINQUE does not exist, CMINQUE does not exist.

Hartung (1981) proposed another nonnegative estimator which he called minimum biased MINQ. This estimator is constructed by first minimizing a function of the bias and then minimizing the variance of a quadratic estimator. It is proved in Chapter 4 that there does not exist a globally minimum biased estimator in the whole parameter space. In this thesis Hartung's 'minimum biased MINQ' will be referred to as 'Hartung's estimator'. Theoretically, Hartung's estimator always exists. It will coincide with the nonnegative unbiased MINQUE if such an estimator exists. Practically, we only managed to obtain explicit formulas for the balanced nested ANOVA models.

In Chapter 4 another nonnegative estimator is proposed. It is derived by minimizing an upper bound of the bias function of the estimator. It is therefore called minimum range MINQ. In some sense it makes the bias small. This estimator always exists. It is shown that this new estimator is Hartung's estimator if the parameter $\gamma = 1$. The explicit formulas to obtain minimum range MINQ estimator for the balanced nested ANOVA models are given.

These estimators are constructed using different optimality criteria. It would be desirable to assess their performance by statistical measures such as bias and mean squared error. A numerical comparison for these nonnegative estimators is carried out in Chapter 5. As expected none of the estimators has an overall better performance than the others over the whole parameter space. The efficiencies of these estimators vary greatly for different positions in the parameter space. Some suggestions on the use of the estimators are given.

1.3.7 Some other estimators

There are other estimators existing in the literature. There are many papers on Bayesian estimators, see Tiao and Tan (1965, 1966), Tiao and Box (1967), etc. Rather than considering any specific model this thesis is concerned with estimating problems for a general class of models. There will be no discussion on Bayesian estimators in this thesis.

Browne (1974) derived a generalized least squares estimator for the variance components by minimizing a specially defined distance between the sample variance covariance matrix and the true variance covariance matrix in terms of variance components. This estimator has not received much attention in the literature and will not be discussed in this thesis.

In this section we have introduced several estimators. This thesis is mainly concerned with quadratic estimation of variance components, hence ML, REML and Goldstein's method will not be further discussed. It is hoped that this section gives a general review of the major methods available in the subject of estimation of variance components.

1.4 Outline of thesis

This thesis can be divided into three parts.

1. Chapters 2 and 3 concern the quadratic unbiased estimator, MINQUE.

As mentioned in section 1.3.4 MINQUE has minimum variance only when the following conditions are met:

(1). the data have a normal distribution;

(2). the prior values are correct values of the variance components.

Chapter 2 aims to weaken condition 1. We assume condition 2 holds and give a sufficient condition for the model under which condition 1 is no more required for MINQUE to have minimum variance. We also examine some models and conclude that for some ANOVA and E-ANOVA models, MINQUE is MIVQUE without the normality assumption.

Chapter 3 deals with the problem of prior values in the computation of MINQUE. We give necessary and sufficient conditions for MINQUE to be independent of any prior values. Therefore any starting point will give the same results after one round of iteration.

2. Chapters 4 and 5 consider nonnegative estimators.

In Chapter 4 it is proved that there is no globally minimum biased nonnegative estimator for variance components. We investigate the properties of several existing estimators and propose another nonnegative estimator.

Since there is no one single estimator globally better than the others in the whole parameter space, we report on numerical and empirical comparisons in Chapter 5 and give recommendations on the use of these estimators.

3. Chapter 6 considers an application of variance components models in sample surveys. In particular, we consider the problem of estimating the interviewer's variance in a complex survey. After constructing a variance components model, we choose MINQUE as the estimator and argue that this is a better estimator than the previously used estimators. A design problem for the optimal number of interpenetrated interviewers is also considered. An unbiased estimator is given for the variance of the estimator of the population mean.

Chapter 2

OPTIMALITY CONDITIONS FOR MINQUE

In section 1.3.4, it is shown that the optimality of MINQUE needs the following two assumptions:

1. the data are normally distributed;
2. the prior values are the true variance components values.

In practice both assumptions are very restrictive. In this chapter we shall assume that assumption 2 holds and try to find other conditions to substitute the condition of normality.

The notation used in this chapter follows that in Chapter 1.

2.1 Optimality conditions for MINQUE

To find the optimality conditions for MINQUE we need a theorem by C.R. Rao.

Theorem 2.1 (Rao, 1973, *p*317) *A necessary and sufficient condition for an unbiased estimator* $\mathbf{T}$ *of* $\mathrm{g}(\boldsymbol{\theta})$ *to have minimum variance at the value* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ *is that* $\mathrm{cov}(\mathbf{T}, \mathbf{f}|\boldsymbol{\theta}_0) = 0$ *for every* $\mathbf{f}$ *such that* $\mathrm{E}(\mathbf{f}|\boldsymbol{\theta}) = 0$ *provided that* $\mathbf{V}(\mathbf{f}|\boldsymbol{\theta}_0) < \infty$ *and* $\mathbf{V}(\mathbf{T}|\boldsymbol{\theta}_0) < \infty$. *All the other forms of unbiased minimum variance estimator of* $g(\boldsymbol{\theta})$ *differ from* $\mathbf{T}$ *only on a set of samples with probability measure zero.*

To prove the major theorem in this chapter some results given by Rao in matrix theory are needed. These results are presented in the form of lemmas. Some of the proofs were outlined in Rao (1971a, 1971b). For completeness a proof is given after each lemma.

Recall model (1.1) as defined in Chapter 1,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\boldsymbol{\xi}_1 + \mathbf{U}_2\boldsymbol{\xi}_2 + \cdots + \mathbf{U}_k\boldsymbol{\xi}_k, \tag{2.1}$$

where $\mathbf{y}$ is the $n\times 1$ vector of observed values, $\boldsymbol{\beta}$ is the $p\times 1$ vector of fixed effect parameter, $\mathbf{X}$ is the $n\times p$ design matrix for the fixed effect, $\mathbf{U}_1$, $\mathbf{U}_2$, $\cdots$, $\mathbf{U}_k$ are the $n\times p_1$, $n\times p_2$, $\cdots$, $n\times p_k$ design matrices for the variance components, $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, $\cdots$, $\boldsymbol{\xi}_k$ are $p_1\times 1$, $p_2\times 1$, $\cdots$, $p_k\times 1$ vectors of variance components.

The following assumptions are imposed on model (2.1):

$$\mathbf{E}(\boldsymbol{\xi}_i) = 0 \quad i = 1, \ldots, k, \tag{2.2}$$

$$\mathbf{V}(\boldsymbol{\xi}_i) = \sigma_i^2\mathbf{I} \quad i = 1, \ldots, k, \tag{2.3}$$

$$\mathrm{cov}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) = 0 \quad i, j = 1, \ldots, k, \ i \neq j. \tag{2.4}$$

The class of estimators we are considering is that of quadratic estimators. Applying Theorem 2.1 to find the optimality conditions for MINQUE requires the knowledge of the covariance between two quadratic estimators. The covariance of two quadratic estimators for model (1.18) which includes model (2.1) as a special case is given in Lemma 1.3. The following lemma gives the covariance of two quadratic estimators for model (2.1).

Let $\mathbf{U} = [\mathbf{U}_1|\ldots|\mathbf{U}_k]$.

Lemma 2.1 (Rao, 1971a) *If model* (2.1) *is considered,* $\mathbf{X}$ *and* $\mathbf{U}$ *are the design matrices defined in model* (2.1) $\mathbf{A}$, N *are symmetric matrices with* $\mathbf{AX}=0$, $\mathbf{NX}=0$, *then:*

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathrm{N}\mathbf{y}) = 2\mathrm{Tr}\mathbf{B}\Delta_1\mathrm{M}\Delta_1 + \mathrm{Tr}\tilde{\mathbf{B}}\Delta_2\widetilde{\mathrm{M}} \tag{2.5}$$

where $\mathbf{B} = \mathbf{U}'\mathbf{A}\mathbf{U}$, $\mathrm{M} = \mathbf{U}'\mathrm{N}\mathbf{U}$. $\tilde{\mathbf{B}}$ *is the diagonal matrix with the diagonal elements equal to those on the diagonal of* $\mathbf{B}$. $\widetilde{\mathrm{M}}$ *is similarly defined as* $\tilde{\mathbf{B}}$.

$$\Delta_1 = \begin{bmatrix} \sigma_1^2\mathbf{I}_{p_1} & & & \\ & \sigma_2^2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^2\mathbf{I}_{p_k} \end{bmatrix} \quad \Delta_2 = \begin{bmatrix} \sigma_1^4\gamma_1\mathbf{I}_{p_1} & & & \\ & \sigma_2^4\gamma_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^4\gamma_k\mathbf{I}_{p_k} \end{bmatrix}$$

where $\gamma_i = E(\xi_i^4)/\sigma_i^4 - 3$ *is the kurtosis of the distribution.*

Lemma 2.1 is a special case of Lemma 1.3 with $\boldsymbol{\Delta}_\theta = \Delta_1$ and $\Delta_\gamma = \Delta_2$. The conclusion follows from Lemma 1.3.

Lemma 2.2 will be used in proving Lemma 2.3 and will not be directly used in the proof of the main result.

Lemma 2.2 (Rao, 1971a) *Let* $\mathbf{A}_i$ *be an* $n\times m_i$ *matrix of rank* r_i, $i = 1,\ldots,s$, Λ *be a symmetric positive definite matrix and* $(\mathbf{A}_i'\Lambda\mathbf{A}_i)^-$ *be a generalized inverse of* $\mathbf{A}_i'\Lambda\mathbf{A}_i$. *If* $\sum_{i=1}^s r_i = n$, *and* $\mathbf{A}_i'\Lambda\mathbf{A}_j = 0$ *for* $i\neq j$, *then*

$$\Lambda^{-1} = \sum_{i=1}^s \mathbf{A}_i(\mathbf{A}_i'\Lambda\mathbf{A}_i)^-\mathbf{A}_i'.$$

Proof: Let $\mathbf{B}_i = [\mathbf{A}_i|0]$ be an $n \times n$ matrix and let $\mathbf{C}_i = \Lambda^{1/2}\mathbf{B}_i$.

We have $\text{rank}(\mathbf{B}_i) = \text{rank}(\mathbf{A}_i) = r_i$ and $\text{rank}(\mathbf{C}_i) = r_i$ because Λ is positive definite, thus nonsingular.

Now $\mathbf{C}_i'\mathbf{C}_j = \mathbf{B}_i'\Lambda\mathbf{B}_j = [\mathbf{A}_i|0]'\Lambda[\mathbf{A}_j|0] = \begin{bmatrix} \mathbf{A}_i'\Lambda\mathbf{A}_j & 0 \\ 0 & 0 \end{bmatrix} = 0$, if $i \neq j$.

Let $\mathbf{C} = \sum_{i=1}^s \mathbf{C}_i$, since $\mathbf{C}_i'\mathbf{C}_j = 0$, then $\text{rank}(\mathbf{C}) = \sum_{i=1}^s \text{rank}(\mathbf{C}_i) = n$, therefore $\mathbf{C}$ is a square nonsingular matrix.

Let $\mathbf{D} = \sum_{i=1}^s \mathbf{A}_i(\mathbf{A}_i'\Lambda\mathbf{A}_i)^-\mathbf{A}_i'$.

$$\begin{aligned}
\mathbf{C}'\Lambda^{1/2}\mathbf{D}\Lambda^{1/2}\mathbf{C} &= \sum_{i=1}^s \left\{(\textstyle\sum_{j=1}^s \mathbf{B}_j')\Lambda\mathbf{A}_i(\mathbf{A}_i'\Lambda\mathbf{A}_i)^-\mathbf{A}_i\Lambda(\textstyle\sum_{j=1}^s \mathbf{B}_j)\right\} \\
&= \sum_{i=1}^s \left\{[\mathbf{A}_i'\Lambda\mathbf{A}_i|0]'(\mathbf{A}_i'\Lambda\mathbf{A}_i)^-[\mathbf{A}_i'\Lambda\mathbf{A}_i|0]\right\} \\
&= \sum_{i=1}^s \begin{bmatrix} \mathbf{A}_i'\Lambda\mathbf{A}_i & 0 \\ 0 & 0 \end{bmatrix} \\
&= \sum_{i=1}^s \mathbf{C}_i'\mathbf{C}_i = \mathbf{C}'\mathbf{C}. \qquad (2.6)
\end{aligned}$$

Note that both $\mathbf{C}$ and Λ are nonsingular, thus from both sides of (2.6) we have: $\mathbf{D} = \Lambda^{-1}$. □

The results in Lemma 2.3 and 2.4 will be used in the proof of Theorem 2.2.

Lemma 2.3 (Rao, 1971a) *Given an* $n\times m$ *matrix* $\mathbf{X}$ *of rank* r *and a symmetric positive definite matrix* $\mathbf{V}$ *of order* n, *then there exists an* $n\times(n-r)$ *matrix* $\mathbf{G}$ *of rank* $(n-r)$ *such that:* $\mathbf{G}'\mathbf{X} = 0$, $\mathbf{G}'\mathbf{V}\mathbf{G} = \mathbf{I}$, *and* $\mathbf{G}\mathbf{G}' = \mathbf{R}$, *where* $\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$.

Proof: From Rao (1973, *p*25) we know that $\text{rank}(\mathbf{X}\mathbf{X}') = \text{rank}(\mathbf{X}) = r$.

Since $\mathbf{V}$ is nonsingular, $\text{rank}(\mathbf{X}\mathbf{X}'\mathbf{V}^{-1/2}) = \text{rank}(\mathbf{X}\mathbf{X}') = r$, thus $\mathbf{X}\mathbf{X}'\mathbf{V}^{-1/2}$ has zero as its eigenvalues with order $n - r$.

Let β_1, ..., β_{n-r} be the normalized orthogonal eigenvectors of $\mathbf{XX'V}^{-1/2}$ corresponding to the eigenvalue zero, *i.e.* $\beta_i'\beta_i = 1$, $\beta_i'\beta_j = 0$, $i \neq j$.

Let $\boldsymbol{\alpha}_i = \mathbf{V}^{-1/2}\boldsymbol{\beta}_i$, then:

$$\boldsymbol{\alpha}_i'\mathbf{V}\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i'\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\boldsymbol{\beta}_i = \boldsymbol{\beta}_i'\boldsymbol{\beta}_i = 1, \tag{2.7}$$

$$\boldsymbol{\alpha}_i'\mathbf{V}\boldsymbol{\alpha}_j = \boldsymbol{\beta}_i'\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\boldsymbol{\beta}_j = \boldsymbol{\beta}_i'\boldsymbol{\beta}_j = 0, \quad \text{for } i \neq j. \tag{2.8}$$

Further, $\mathbf{XX'}\boldsymbol{\alpha}_i = \mathbf{XX'V}^{-1/2}\boldsymbol{\beta}_i = 0$, *i.e.* $\boldsymbol{\alpha}_i'\mathbf{XX'} = 0$, and so $\boldsymbol{\alpha}_i'\mathbf{XX'}\boldsymbol{\alpha}_i = 0$, thus

$$\mathbf{X'}\boldsymbol{\alpha}_i = 0. \tag{2.9}$$

Let $\mathbf{G} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{n-r})$, then from (2.7) and (2.8) we have $\mathbf{G'VG} = \mathbf{I}$.

Also from (2.9) we have $\mathbf{X'G} = 0$, *i.e.* $\mathbf{G'X} = 0$.

In Lemma 2.2 let $\Lambda = \mathbf{I}$, $\mathbf{A}_1 = \mathbf{V}^{1/2}\mathbf{G}$ and $\mathbf{A}_2 = \mathbf{V}^{-1/2}\mathbf{X}$, we have:

$$\begin{aligned}\mathbf{I} &= \mathbf{V}^{1/2}\mathbf{G}(\mathbf{G'VG})^{-}\mathbf{G'V}^{1/2} + \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1/2} \\ &= \mathbf{V}^{1/2}\mathbf{GG'V}^{1/2} + \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1/2},\end{aligned}$$

therefore,

$$\mathbf{GG'} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-}\mathbf{X'V}^{-1} = \mathbf{R}. \quad \square$$

Lemma 2.4 (Rao, 1971b) *Let* $\mathbf{A}$ *and* $\mathbf{N}$ *be* $n\times n$ *symmetric matrices with* $\mathbf{AX} = 0$, $\mathbf{NX} = 0$, *then there exist matrices* $\mathbf{C}$ *and* $\mathbf{D}$ *of order* $(n-r)\times(n-r)$ *such that:* $\mathbf{A} = \mathbf{GCG'}$, $\mathbf{N} = \mathbf{GDG'}$, *where* $\mathbf{G}$ *is the matrix in Lemma* 2.3.

Proof: Let $\mathbf{C} = \mathbf{G'VAVG}$, $\mathbf{D} = \mathbf{G'VNVG}$.

$$\begin{aligned}\mathbf{GCG'} &= \mathbf{GG'VAVGG'} \\ &= \mathbf{RVAVR} \\ &= [\mathbf{I} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'}]\mathbf{A}[\mathbf{I} - \mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}] \\ &= [\mathbf{I} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'}]\mathbf{A} \\ &= \mathbf{A}, \ (\textit{because } \mathbf{AX} = 0).\end{aligned}$$

Similarly, $\mathbf{GDG'} = \mathbf{N}$. $\square$

Lemma 2.5 *Consider model* (2.1). *Let* $\mathbf{u}_{i_j}$ *be the jth column in the design matrix* $\mathbf{U}_i$. *If* $\mathbf{A}$ *is a symmetric matrix, then:*

$$\mathbf{u}'_{i_j}\mathbf{A}\mathbf{u}_{i_j} = c_i, \quad for\ i = 1, \ldots, k,\ j = 1, \ldots, p_i \tag{2.10}$$

is equivalent to:

$$\widetilde{\mathbf{U}'_i\mathbf{A}\mathbf{U}_i} = c_i\mathbf{I}, \quad i = 1, \ldots, k. \tag{2.11}$$

where c_i is a constant which is fixed for all the columns in $\mathbf{U}_i$.

Proof: Note $\mathbf{u}'_{i_j}\mathbf{A}\mathbf{u}_{i_j}$ is the jth element on the diagonal of matrix $\mathbf{U}'_i\mathbf{A}\mathbf{U}_i$, so (2.10) is equivalent to (2.11). □

It is known from section 1.3.4 that MINQUE has optimality when the data arise from a normal distribution. It is not known how well MINQUE performs if the normality assumption is not valid. Theorem 2.2 is proved using theorem 2.1 of Rao and gives an alternative condition (2.14) for MINQUE's optimality.

Theorem 2.2 *Consider model* (2.1). *If a symmetric matrix* $\mathbf{A}$ *satisfies:*

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i\ \mathbf{R}\mathbf{V}_i\mathbf{R}, \tag{2.12}$$

where λ_i's satisfy:

$$\sum_{i=1}^{k} \lambda_i\ \mathrm{Tr}\ \mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j = q_j,\ j = 1, \ldots, k, \tag{2.13}$$

and

$$\widetilde{\mathbf{U}'_i\mathbf{A}\mathbf{U}_i} = c_i\mathbf{I}, \quad i = 1, \ldots, k. \tag{2.14}$$

where c_i is a constant related to the design matrix $\mathbf{U}_i$, then

1. $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is the MIVQUE of* $\sum_{i=1}^{k} q_i\sigma_i^2$.

2. *All the other forms of minimum variance estimator among the class of invariant quadratic unbiased estimators differ from* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *only on a set of samples with probability measure zero.*

Proof: Since $\mathbf{R}\mathbf{X} = \mathbf{G}\mathbf{G}'\mathbf{X} = 0$ from Lemma 2.3, then $\mathbf{A}\mathbf{X} = \sum_{i=1}^{k} \lambda_i\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{X} = 0$, so that $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant.

Also $\mathbf{A}\mathbf{V}_j = \sum_{i=1}^{k} \lambda_i\mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j = q_j$, so $\mathrm{Tr}\,\mathbf{A}\mathbf{V}_j = q_j$, for $j = 1, \ldots, k$, thus $\mathbf{y}'\mathbf{A}\mathbf{y}$ is unbiased.

What we need to prove is that for any symmetric matrix $\mathbf{N}$ such that $\mathbf{N}\mathbf{X} = 0$, and $\mathbf{E}(\mathbf{y}'\mathbf{N}\mathbf{y}) = 0$, we have: $\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y},\ \mathbf{y}'\mathbf{N}\mathbf{y}) = 0$, then from Theorem 2.1, we know that $\mathbf{y'Ay}$ is the minimum variance estimator among all quadratic invariant unbiased estimators.

From Lemma 2.4, we can write: $\mathbf{A} = \mathbf{GCG'}$ and $\mathbf{N} = \mathbf{GDG'}$.
Denote the ith column of $\mathbf{G'U}$ by $\boldsymbol{\alpha}_i$. Now

$$\begin{aligned}
\mathbf{E}(\mathbf{y'Ny}) &= \mathbf{E}(\mathbf{y'GDG'y}) \\
&= \mathrm{Tr}\,\{(\mathbf{GDG'})\mathbf{V}\} \\
&= \mathrm{Tr}\,\mathbf{DG'VG} \\
&= \mathrm{Tr}\,\mathbf{DG'U}\Delta_1\mathbf{U'G} \\
&= \mathrm{Tr}\,\mathbf{D}\,[\boldsymbol{\alpha}_1, \;|\ldots,|\;\boldsymbol{\alpha}_N] \begin{bmatrix} \sigma_1^2\mathbf{I}_{p_1} & & & \\ & \sigma_2^2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^2\mathbf{I}_{p_k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \\ \vdots \\ \boldsymbol{\alpha}_k' \end{bmatrix} \\
&= \mathrm{Tr}\,\mathbf{D}\left\{(\sum_{i=1}^{k}\sigma_i^2(\sum \boldsymbol{\alpha}_j\boldsymbol{\alpha}_j')\right\}.
\end{aligned}$$

But $\mathbf{E}(\mathbf{y'Ny}) = 0$ for all σ_i^2 implies

$$\mathrm{Tr}\,\mathbf{D}(\sum_{i=1}^{N}\nu_i\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i')\equiv 0, \tag{2.15}$$

where $N = \sum_{i=1}^{k} p_i$, ν_i are scalars with the following pattern:

$$\begin{bmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_N \end{bmatrix} = \begin{bmatrix} \varepsilon_1\mathbf{I}_{p_1} & & & \\ & \varepsilon_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \varepsilon_k\mathbf{I}_{p_k} \end{bmatrix}, \tag{2.16}$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ are arbitrary scalars.

From Lemma 2.1:

$$\mathrm{cov}(\mathbf{y'Ay}, \mathbf{y'Ny}) = 2\;\mathrm{Tr}\;\mathbf{U'AU}\Delta_1\mathbf{U'NU}\Delta_1 + \mathrm{Tr}\;\widetilde{\mathbf{U'AU}}\Delta_2\widetilde{\mathbf{U'NU}}. \tag{2.17}$$

Note $\mathbf{V} = \mathbf{U}\Delta_1\mathbf{U'}$, $\mathbf{A} = \mathbf{GCG'}$, $\mathbf{G'VG} = \mathbf{I}$. Thus

$$\begin{aligned}
\mathrm{Tr}\;\mathbf{AU}\Delta_1\mathbf{U'NU}\Delta_1\mathbf{U'} &= \mathrm{Tr}\;\mathbf{AVNV} \\
&= \mathrm{Tr}\;\mathbf{GCG'VGDG'V} \\
&= \mathrm{Tr}\;\mathbf{CD}.
\end{aligned}$$

Now, Tr $\mathbf{U'\widetilde{A}U\Delta_2U'\widetilde{N}U}$ = Tr $(\mathbf{U'G\widetilde{C}G'U})\Delta_2(\mathbf{U'G\widetilde{D}G'U})$, and

$$\mathbf{U'GCG'U} = \begin{bmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \\ \vdots \\ \boldsymbol{\alpha}_N' \end{bmatrix} \mathbf{C}\,[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_N].$$

The diagonal form of $\mathbf{U'GCG'U}$ is

$$\mathbf{U'G\widetilde{C}G'U} = \begin{bmatrix} \boldsymbol{\alpha}_1'\mathbf{C}\boldsymbol{\alpha}_1 & & & \\ & \boldsymbol{\alpha}_2'\mathbf{C}\boldsymbol{\alpha}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\alpha}_N'\mathbf{C}\boldsymbol{\alpha}_N \end{bmatrix}.$$

Similarly,

$$\mathbf{U'G\widetilde{D}G'U} = \begin{bmatrix} \boldsymbol{\alpha}_1'\mathbf{D}\boldsymbol{\alpha}_1 & & & \\ & \boldsymbol{\alpha}_2'\mathbf{D}\boldsymbol{\alpha}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\alpha}_N'\mathbf{D}\boldsymbol{\alpha}_N \end{bmatrix}.$$

Since $\Delta_2 = \begin{bmatrix} \sigma_1^4\gamma_1\mathbf{I}_{p_1} & & & \\ & \sigma_2^4\gamma_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^4\gamma_k\mathbf{I}_{p_k} \end{bmatrix}$, then

$$\begin{aligned}
&\mathrm{Tr}\mathbf{U'\widetilde{A}U\Delta_2U'\widetilde{N}U} \\
&= \mathrm{Tr}\begin{bmatrix} \sigma_1^4\gamma_1\boldsymbol{\alpha}_1'\mathbf{C}\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1'\mathbf{D}\boldsymbol{\alpha}_1 & & & \\ & \sigma_1^4\gamma_1\boldsymbol{\alpha}_2'\mathbf{C}\boldsymbol{\alpha}_2\boldsymbol{\alpha}_2'\mathbf{D}\boldsymbol{\alpha}_2 & & \\ & & \ddots & \\ & & & \sigma_k^4\gamma_k\boldsymbol{\alpha}_N'\mathbf{C}\boldsymbol{\alpha}_N\boldsymbol{\alpha}_N'\mathbf{D}\boldsymbol{\alpha}_N \end{bmatrix} \\
&= \sum_{i=1}^N \delta_i\boldsymbol{\alpha}_i'\mathbf{C}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\mathbf{D}\boldsymbol{\alpha}_i \\
&= \sum_{i=1}^N \delta_i\mathrm{Tr}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\mathbf{C}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\mathbf{D} \\
&= \sum_{i=1}^N \delta_i\mathrm{Tr}\ \mathbf{D}(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\mathbf{C}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i') \\
&= \mathrm{Tr}\ \mathbf{D}(\sum_{i=1}^N \delta_i\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\mathbf{C}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'),
\end{aligned}$$

where $\begin{bmatrix} \delta_1 & & & \\ & \delta_2 & & \\ & & \ddots & \\ & & & \delta_N \end{bmatrix} = \begin{bmatrix} \sigma_1^4\gamma_1\mathbf{I}_{p_1} & & & \\ & \sigma_2^4\gamma_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \sigma_k^4\gamma_k\mathbf{I}_{p_k} \end{bmatrix}.$

Therefore,

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) = \mathrm{Tr}\ \mathbf{D}(2\mathbf{C} + \sum_{i=1}^{N} \delta_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i' \mathbf{C} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i'). \tag{2.18}$$

Let $N_i = \sum_{w=1}^{i} p_w$. Since $\mathbf{A} = \sum_{i=1}^{k} \lambda_i\ \mathbf{R}\mathbf{V}_i\mathbf{R} = \sum_{i=1}^{k} \lambda_i\ \mathbf{G}\mathbf{G}'\mathbf{V}_i\mathbf{G}\mathbf{G}'$, therefore,

$$\mathbf{C} = \sum_{i=1}^{k} \lambda_i\ \mathbf{G}'\mathbf{V}_i\mathbf{G} = \sum_{i=1}^{k} \lambda_i\ (\sum_{j=N_{i-1}+1}^{N_i} \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j') = \sum_{i=1}^{N} \tau_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i', \tag{2.19}$$

where τ_i has the pattern:

$$\begin{bmatrix} \tau_1 & & & \\ & \tau_2 & & \\ & & \ddots & \\ & & & \tau_N \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{I}_{p_1} & & & \\ & \lambda_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \lambda_k\mathbf{I}_{p_k} \end{bmatrix}.$$

Suppose $i = \sum_{m=1}^{j} p_m + t$, then $\boldsymbol{\alpha}_i = \mathbf{G}'\mathbf{u}_{j_t}$.
Since condition (2.14) is equivalent to condition (2.10), thus:

$$\boldsymbol{\alpha}_i'\mathbf{C}\boldsymbol{\alpha}_i = \mathbf{u}_{j_t}'\mathbf{G}\mathbf{C}\mathbf{G}'\mathbf{u}_{j_t} = \mathbf{u}_{j_t}'\mathbf{A}\mathbf{u}_{j_t} = c_j.$$

Therefore (2.18) becomes:

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}\mathbf{N}\mathbf{y}) = \mathrm{Tr}\ \mathbf{D}(\sum_{i=1}^{N} \psi_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i'),$$

where $\psi_i = 2\tau_i + \delta_i c_i$.

ψ_i also has the pattern:

$$\begin{bmatrix} \psi_1 & & & \\ & \psi_2 & & \\ & & \ddots & \\ & & & \psi_N \end{bmatrix} = \begin{bmatrix} \omega_1\mathbf{I}_{p_1} & & & \\ & \omega_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \omega_k\mathbf{I}_{p_k} \end{bmatrix},$$

because τ_i, δ_i and c_i have the pattern.

Recalling that the assumption of $\mathbf{N}$ makes (2.15) hold for any choice of ν_i which satisfies (2.16), we have:

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) = 0.$$

Thus $\mathbf{y}'\mathbf{A}\mathbf{y}$ has the minimum variance among all the invariant quadratic unbiased estimators.

The second part of the theorem follows from theorem 2.1. □

From the proof of theorem 2.2, we should notice that condition (2.14) is a sufficient condition.

Note that without condition (2.14), the quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ given in theorem 2.2 is the MINQUE of $\sum_{i=1}^{k} q_i\sigma_i^2$. In the proof of theorem 2.2 we did not use the fact that $\mathbf{y}'\mathbf{A}\mathbf{y}$ minimizes the norm of a matrix. In Corollary 1.4 we have used the concept of Euclidean norm to prove that MINQUE achieves minimum variance when data comes from a normal distribution . Now we give an alternative proof using theorem 2.2.

Corollary 2.1 *Consider model* (2.1). *If the data vector* $\mathbf{y}$ *is normally distributed with* $\mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{V}(\mathbf{y}) = \sum_{i=1}^{k} \sigma_i^2\mathbf{V}_i$, *then the MINQUE* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *of* $\sum_{i=1}^{k} q_i\sigma_i^2$ *is also a MIVQUE.*

Proof: If $\mathbf{y}$ is normally distributed, then the kurtosis of the distribution $\gamma_i = 0$, thus $\Delta_2 = 0$.

From (2.17),

$$\begin{aligned} \mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) &= 2\mathrm{Tr}\ \mathbf{U}'\mathbf{A}\mathbf{U}\Delta_1\mathbf{U}'\mathbf{N}\mathbf{U}\Delta_1 \\ &= \mathrm{Tr}\ \mathbf{D}(2\mathbf{C}). \end{aligned}$$

From (2.19) and the assumption on matrix $\mathbf{N}$, we have:

$$\mathrm{cov}(\mathbf{y}'\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{N}\mathbf{y}) = 0.$$

From theorem 2.1 we know that $\mathbf{y}'\mathbf{A}\mathbf{y}$ is the MIVQUE of $\sum_{i=1}^{k} q_i\sigma_i^2$. □

It is well known that MINQUE achieves optimality when the underlying distribution is normal. In theorem 2.2 we have proved that without the normality assumption but with condition (2.14), MINQUE still has optimality.

The result of Theorem 2.2 gives us a new perspective of the MINQUE estimator. Instead of making assumptions on the distribution of the data, we can examine the design matrices to see if they satisfy condition (2.14). If condition (2.14) is satisfied, then MINQUE is MIVQUE.

In the following section we shall look at some classes of models and examine if condition (2.14) is satisfied.

2.2 Models satisfying condition (2.14)

Theorem 2.2 gives a sufficient condition (2.14) for the MINQUE estimator to be optimal. The question which follows is what kinds of model satisfy this condition, if any.

Condition (2.14) can be examined for any specified model. We use a simple model to demonstrate how to apply theorem 2.2 in practice.

Example 2.1

We consider if the MINQUE for a balanced one-way random model satisfies condition (2.14). The model is

$$y_{ij} = \mu + a_i + e_{ij}, \qquad i = 1,2,\ j = 1,2. \tag{2.20}$$

The model can be written in matrix form:

$$\mathbf{y} = \mathbf{X}\mu + \mathbf{U}_1\mathbf{a} + \mathbf{U}_e\mathbf{e},$$

where the design matrices are:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{U}_e = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now using the MINQUE formulas (1.24)–(1.26) to estimate σ_a^2, we can obtain the MINQUE matrix $\mathbf{A}$:

$$\mathbf{A} = -\frac{1}{4}\mathbf{I}_4 + \frac{3}{8}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Now

$$\begin{aligned} \mathbf{U}_1'\mathbf{A}\mathbf{U}_1 &= \mathbf{I}_2 - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \end{aligned}$$

Hence $\widetilde{\mathbf{U}_1'\mathbf{A}\mathbf{U}_1} = \frac{1}{2}\mathbf{I}_2$.

Similarly, we can show that $\widetilde{\mathbf{U}_e'\mathbf{A}\mathbf{U}_e} = 0$. Therefore condition (2.14) is satisfied for the MINQUE $\hat{\sigma}_a^2$, where

$$\hat{\sigma}_a^2 = \sum_{i=1}^{2}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y}_{i.})^2.$$

$\hat{\sigma}_a^2$ is the best quadratic unbiased estimator of σ_a^2. □

The calculation involved in Example 2.1 can be shortened if we use Kronecker products of matrices $\mathbf{I}_n$ and $\mathbf{1}_n$, where $\mathbf{I}_n$ is the identity matrix of order n and $\mathbf{1}_n$ is the $n \times 1$ vector with all elements equal to 1, to express the design matrices of the model.

Example 2.2

Using Kronecker products the design matrices in Example 2.1 can be written as:

$$\mathbf{X} = \mathbf{1}_4, \quad \mathbf{U}_1 = \mathbf{I}_2 \otimes \mathbf{1}_2, \quad \mathbf{U}_e = \mathbf{I}_2 \otimes \mathbf{I}_2. \tag{2.21}$$

Let $\mathbf{J}_n$ be the matrix of order n with all elements equal to 1, thus

$$\mathbf{V}_1 = \mathbf{U}_1\mathbf{U}_1' = \mathbf{I}_2 \otimes \mathbf{J}_2, \;\; \mathbf{V}_e = \mathbf{U}_e\mathbf{U}_e' = \mathbf{I}_2 \otimes \mathbf{I}_2.$$

The variance covariance matrix is

$$\mathbf{V} = \sigma_a^2\mathbf{I}_2 \otimes \mathbf{J}_2 + \sigma_e^2\mathbf{I}_4,$$

and the inverse of the variance covariance matrix can be written as:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_e^2}\mathbf{I}_4 - \frac{\sigma_a^2}{\sigma_e^2(\sigma_e^2 + 2\sigma_a^2)}\mathbf{I}_2 \otimes \mathbf{J}_2.$$

The MINQUE matrix $\mathbf{A}$ for estimating σ_a^2 can then be calculated using the MINQUE formulas (1.24) -(1.26) in terms of Kronecker products of matrices as:

$$\mathbf{A} = -\frac{1}{4}\mathbf{I}_4 + \frac{3}{8}\mathbf{I}_2 \otimes \mathbf{J}_2 - \frac{1}{8}\mathbf{J}_2 \otimes \mathbf{J}_2.$$

Hence $\mathbf{U}_1'\mathbf{A}\mathbf{U}_1 = \mathbf{I}_2 - \frac{1}{2}\mathbf{J}_2$, thus

$$\widetilde{\mathbf{U}_1'\mathbf{A}\mathbf{U}_1} = \frac{1}{2}\mathbf{I}_2.$$

Also

$$\mathbf{U}_e'\mathbf{A}\mathbf{U}_e = -\frac{1}{4}\mathbf{I}_4 + \frac{3}{8}\mathbf{I}_2 \otimes \mathbf{J}_2 - \frac{1}{8}\mathbf{J}_2 \otimes \mathbf{J}_2,$$

$$\mathbf{U}_e'\widetilde{\mathbf{A}}\mathbf{U}_e = 0.$$

Therefore condition (2.14) is satisfied for $\hat{\sigma}_a^2$ which is $\mathbf{y}'\mathbf{A}\mathbf{y}$. $\mathbf{y}'\mathbf{A}\mathbf{y}$ is the best quadratic unbiased estimator of σ_a^2. □

Graybill (1954) calculated the variance of the ANOVA estimator for a two-way balanced nested analysis of variance model and proved that the ANOVA estimator is the best quadratic unbiased estimator of the variance components and he concluded that the result can be generalized to all balanced nested ANOVA models. In Example 1.8 we have shown that for the one-way balanced model the MINQUE of σ_a^2 and σ_e^2 are identical to the ANOVA estimators. Hence our conclusions in Example 2.1 and 2.2 coincide with Graybill's result.

Rather than examining condition (2.14) for the specific models, as we did in Example 2.1 and 2.2, we want to draw conclusions on all balanced E-ANOVA models. Our conclusions in this section will extend Graybill's result to more models as well as confirm his conclusions for the models he considered.

Condition (2.14) in theorem 2.2 is a condition to be examined for the MINQUE estimator. The design matrices $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_k$ are known from the structure of model, but the matrix $\mathbf{A}$ which is calculated from the MINQUE equations (1.24)–(1.26) is a complicated function of the design matrices. To examine condition (2.14) for balanced ANOVA and E-ANOVA models it is desirable to obtain the matrix $\mathbf{A}$ or at least know the structure of $\mathbf{A}$.

Looking at Example 2.2 we notice that the design matrices are Kronecker products of $\mathbf{I}$ and $\mathbf{1}$'s, and $\mathbf{U}_i'\mathbf{A}\mathbf{U}_i$ is a linear combination of Kronecker products of $\mathbf{I}$ and $\mathbf{J}$'s, $i = 1, 2$. Since condition (2.14) does not require exact values of c_i's, condition (2.14) is satisfied as long as $\mathbf{U}_i'\mathbf{A}\mathbf{U}_i$ is a linear combination of Kronecker products of $\mathbf{I}$ and $\mathbf{J}$'s.

Kurkjian and Zelen (1962), Zelen and Federer (1964) have developed a calculus for the design matrices of balanced designs. We adopt their way of expressing design matrices.

In the following we shall use $\otimes$ to denote the Kronecker product of two matrices and $\bigotimes_{i=1}^m$ to denote the Kronecker products of m matrices, *i.e.*

$$\bigotimes_{i=1}^{m} \mathbf{V}_i = \mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_m.$$

Lemma 2.6 and 2.7 are established to investigate the structure of the MINQUE matrix $\mathbf{A}$.

Before proceeding to the lemmas we state the following properties which we shall use in later context. Property 1 can be found in Chapter 8 of Graybill (1983) and the other properties can be easily verified.

Property 1: If $\mathbf{B} = \otimes_{i=1}^{m} \mathbf{B}_i$, $\mathbf{C} = \otimes_{i=1}^{m} \mathbf{C}_i$, then $\mathbf{BC} = \otimes_{i=1}^{m}(\mathbf{B}_i\mathbf{C}_i)$.

Property 2: $\mathbf{I}_m\mathbf{I}_m = \mathbf{I}_m$, $\mathbf{J}_m\mathbf{J}_m = m\mathbf{J}_m$, $\mathbf{I}_m\mathbf{J}_m = \mathbf{J}_m\mathbf{I}_m = \mathbf{J}_m$.

Property 3: $\mathbf{1}'_m\mathbf{I}_m\mathbf{1}_m = m$, $\mathbf{1}'_m\mathbf{J}_m\mathbf{1}_m = m^2$.

Property 4: $\mathbf{1}_m\mathbf{1}'_m = \mathbf{J}_m$.

Property 5: If $\mathbf{A} = (a_{ij})_{n\times n}$, then $\mathbf{1}'\mathbf{A}\mathbf{1} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$.

Property 6: Let $\mathbf{V}_x = (\otimes_{i=1}^{m} \mathbf{V}_i^{x_i}) \otimes \mathbf{J}_f$, where

$$\mathbf{V}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 0, \end{cases}$$

then $\mathbf{V}_x\mathbf{J} = \mathbf{J}\mathbf{V}_x = f\prod_{i=1}^{m} s_i^{(1-x_i)}\mathbf{J}$.

In the simple example considered in Example 2.1 there are only two variance components σ_a^2 and σ_e^2. If we consider a model with m random factors, we need an index set to label the variance components and their corresponding design matrices and express the variance covariance matrix of the model.

Suppose a model is chosen within the class of E-ANOVA models, so the design matrices for the fixed effect and random effects are known. For the time being we consider the design matrices for the random effects only.

Suppose there are m random factors, each factor has s_i observations, $i = 1, \ldots, m$. We consider a balanced design and assume that the observation in each cell has been replicated f times.

Definition 2.1 *For a balanced ANOVA or E-ANOVA model with m factors, let T be an 0, 1 m-digit set. Each element of T is a m-digit number corresponding to a random term in the model where the digit relates to the corresponding factor. A digit equals* 1 *if that subscript is present in the model and 0 if the subscript is absent.*

Example 2.3

Consider model:

$$y_{ijk} = \mu_i + \delta_i + \gamma_{ij} + e_{ijk}, \tag{2.22}$$

$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K.$

where μ_i is the fixed effect parameter, δ_i, γ_{ij} and e_{ijk} are the random terms. This model has two random factor at level I and J, respectively. Each observation is replicated K times. According to Definition 2.1, $T = \{10, 11\}$. □

In Section 1.2 we have described the ANOVA models and the E-ANOVA models. Using the index set T we can write the balanced E-ANOVA models and the balanced ANOVA models in matrix forms.

A balanced E-ANOVA model can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{x \in T} \mathbf{U}_x \boldsymbol{\xi}_x, \tag{2.23}$$

where

$$\mathbf{X} = \left(\bigotimes_{i=1}^{m} \mathbf{X}_i^{x_i} \right) \otimes \mathbf{1}_f,$$

where

$$\mathbf{X}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0, \end{cases} \tag{2.24}$$

and

$$\mathbf{U}_x = \left(\bigotimes_{i=1}^{m} \mathbf{U}_i^{x_i} \right) \otimes \mathbf{1}_f,$$

where

$$\mathbf{U}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0. \end{cases} \tag{2.25}$$

In other words, the design matrices of a balanced E-ANOVA model can be expressed as the Kronecker products of $\mathbf{I}$ and $\mathbf{1}$.

Particularly, if in the balanced E-ANOVA model, $\mathbf{X} = \mathbf{1}$, then the model is a balanced ANOVA model.

After defining the index set T we can obtain the Kronecker products expression for the design matrices and the variance covariance matrix using T.

Since

$$\mathbf{V}_x = \mathbf{U}_x \mathbf{U}_x' = \left(\bigotimes_{i=1}^{m} \mathbf{V}_i^{x_i} \right) \otimes \mathbf{J}_f,$$

where

$$\mathbf{V}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 0, \end{cases} \tag{2.26}$$

hence the variance covariance matrix is:

$$\mathbf{V} = \sum_{x \in T} \sigma_x^2 \mathbf{V}_x + \sigma_e^2 \mathbf{I}. \tag{2.27}$$

Example 2.4:

For the model considered in Example 2.3, since $T = \{10, 11\}$, then the design matrices for the random terms are:

$$\mathbf{U}_{10} = \mathbf{I}_I \otimes \mathbf{1}_J \otimes \mathbf{1}_K,\ \mathbf{U}_{11} = \mathbf{I}_I \otimes \mathbf{I}_J \otimes \mathbf{1}_K,\ \mathbf{U}_e = \mathbf{I}_{IJK},$$

and the variance covariance matrix of $\mathbf{y}$ is

$$\mathbf{V} = \sigma_{10}^2 \mathbf{I}_I \otimes \mathbf{J}_J \otimes \mathbf{J}_K + \sigma_{11}^2 \mathbf{I}_I \otimes \mathbf{I}_J \otimes \mathbf{J}_K + \sigma_e^2 \mathbf{I}_{IJK}. \quad \square$$

To obtain the structure of the MINQUE matrix $\mathbf{A}$, it is necessary to obtain the inverse of the variance covariance matrix $\mathbf{V}$ first. Corresponding to matrix multiplication we define a multiplication for the elements of set T.

Definition 2.2 *Define a multiplication $*$ among the elements of set T. If $x = x_1x_2...x_m \in T$, $y = y_1y_2...y_m \in T$, then $x * y = z_1z_2...z_m$, where $z_i = x_iy_i$, $i = 1, \ldots, m$.*

Definition 2.2 defines a unit by unit multiplication in the set T.

From Example 2.2 we suspect that for $\mathbf{V} = \sum_{x\in T} \sigma_x^2 \mathbf{V}_x + \sigma_e^2 \mathbf{I}$, where $\mathbf{V}_x$ is a Kronecker products of $\mathbf{I}$'s and $\mathbf{J}$'s, $\mathbf{V}^{-1} = \sum \alpha_x \mathbf{V}_x + \frac{1}{\sigma_e^2}\mathbf{I}$. In other words $\mathbf{V}^{-1}$ has the same structure as $\mathbf{V}$. But since for

$$\mathbf{V} = \sigma_e^2 \mathbf{I} + \sigma_{10}^2 \mathbf{I} \otimes \mathbf{J} \otimes \mathbf{J} + \sigma_{01}^2 \mathbf{J} \otimes \mathbf{I} \otimes \mathbf{J},$$

$$\mathbf{V}^{-1} = \frac{1}{\sigma_e^2}\mathbf{I} + \alpha_{10} \mathbf{I} \otimes \mathbf{J} \otimes \mathbf{J} + \alpha_{01} \mathbf{J} \otimes \mathbf{I} \otimes \mathbf{J} + \alpha_{00} \mathbf{J} \otimes \mathbf{J} \otimes \mathbf{J},$$

where α_{10}, α_{01} and α_{00} are suitably determined, we realized that $\mathbf{V}^{-1}$ is not necessarily indexed by the same index set T . Hence we introduce another index set T^* which is derived from T..

Definition 2.3 *Let T^* be the set containing all the products of elements in T under the multiplication $*$.*

Example 2.5:

Consider a two-way crossed random model:

$$y_{ijk} = \mu + a_i + b_j + e_{ijk}, \tag{2.28}$$

$i = 1, \ldots, I,\ \ j = 1, \ldots, J,\ \ k = 1, \ldots, K$, where μ is also the overall mean, a_i and b_j are the random terms, e_{ijk} is the random error. Then $T = \{10, 01\}$, according to Definition 2.3, $T^* = \{00, 10, 01\}$. □

Note that the element 00 is not included in T and it is the product of 10 and 01, hence it is in T^*.

In Lemma 2.6 we need to solve equations by successive substitution, hence we need to define an order in the set T^*.

Definition 2.4 *Define an order in set T^*. If $x = x_1x_2 \ldots x_m \in T^*$, $y = y_1y_2 \ldots y_m \in T^*$, we say $x \leq y$, if $x_i \leq y_i$, for $i = 1, \ldots, m$. $x < y$, if $x \leq y$, but $x \neq y$.*

The order in T^* is defined by comparing each digit of an element. For example, we should have $00 < 10 < 11$. We can also have $00 < 01 < 11$, but there is no order relation between 10 and 01.

Since each digit of the element of T^* has only two choices: 0 and 1, we can see that if $x * y = z$, then $x \geq z$, $y \geq z$. For example, $10 * 01 = 00$, then $10 > 00$, $01 > 00$.

Lemma 2.6 establishes the structure of the inverse of the variance covariance matrix of $\mathbf{y}$ for a balanced E-ANOVA model.

Lemma 2.6 *If $\mathbf{V} = \sum_{x \in T} \sigma_x^2 \mathbf{V}_x + \sigma_e^2 \mathbf{I}$, where $\mathbf{V}_x$ is defined in* (2.26), *then there exist a set of scalars α_x, $x \in T^*$, such that*

$$\mathbf{V}^{-1} = \sum_{x \in T^*} \alpha_x \mathbf{V}_x + 1/\sigma_e^2 \mathbf{I}.$$

Proof: Let

$$\mathbf{W} = \sum_{x \in T^*} \alpha_x \mathbf{V}_x + 1/\sigma_e^2 \mathbf{I}.$$

From the commutative properties of matrices $\mathbf{I}$ and $\mathbf{J}$ stated in Property 2, we know that $\mathbf{VW} = \mathbf{WV}$. We need to find α_x such that $\mathbf{VW} = \mathbf{I}$, thus $\mathbf{W}$ is $\mathbf{V}^{-1}$.

To make the index expression of the matrices easier in the following development we make some definition to write matrix $\mathbf{V}$ using index set T^*.

For $x \in T^*$, but $x \notin T$, define $\sigma_x^2 = 0$, thus we can write $\mathbf{V} = \sum_{x \in T^*} \sigma_x^2 \mathbf{V}_x + \sigma_e^2 \mathbf{I}$. Now

$$\mathbf{VW} = \mathbf{I} + \sigma_e^2 \sum_{y\in T^*} \alpha_y \mathbf{V}_y + (1/\sigma_e^2) \sum_{x\in T^*} \sigma_x^2 \mathbf{V}_x + \sum_{x\in T^*} \sum_{y\in T^*} \sigma_x^2 \alpha_y \mathbf{V}_x \mathbf{V}_y. \quad (2.29)$$

From the definition of $\mathbf{V}_x$ and $\mathbf{V}_y$ in (2.26) and property 1, we know that:

$$\mathbf{V}_x \mathbf{V}_y = f \left\{ \bigotimes_{i=1}^{m} (\mathbf{V}_i^{x_i} \mathbf{V}_i^{y_i}) \right\} \otimes \mathbf{J}_f,$$

where

$$\begin{aligned} \mathbf{V}_i^{x_i} \mathbf{V}_i^{y_i} &= \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1,\ y_i = 1, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 1,\ y_i = 0, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 0,\ y_i = 1, \\ s_i \mathbf{J}_{s_i}, & \text{if } x_i = 0,\ y_i = 0, \end{cases} \\ &= s_i^{(1-x_i)(1-y_i)} \mathbf{V}_i^{(x_i y_i)}. \end{aligned}$$

Let

$$s = \prod_{i=1}^{m} s_i,\ \tau(x) = \prod_{i=1}^{m} s_i^{x_i},\ \tau(y) = \prod_{i=1}^{m} s_i^{y_i},\ \tau(xy) = \prod_{i=1}^{m} s_i^{x_i y_i},$$

then

$$\prod_{i=1}^{m} s_i^{(1-x_i)(1-y_i)} = \frac{s\tau(xy)}{\tau(x)\tau(y)},$$

therefore,

$$\sum_{x\in T^*} \sum_{y\in T^*} \sigma_x^2 \alpha_y \mathbf{V}_x \mathbf{V}_y = \sum_{z\in T^*} f \left(\sum_{\substack{x*y=z \\ x,y\in T^*}} \sigma_x^2 \alpha_y \frac{s\tau(xy)}{\tau(x)\tau(y)} \right) \mathbf{V}_z,$$

where

$$\mathbf{V}_z = \left(\bigotimes_{i=1}^{m} \mathbf{V}_i^{z_i} \right) \otimes \mathbf{J}_f.$$

Therefore (2.29) can be written as:

$$\mathbf{VW} = \mathbf{I} + \sum_{z\in T^*} \left(\sigma_e^2 \alpha_z + \frac{\sigma_z^2}{\sigma_e^2} + f \sum_{\substack{x*y=z \\ x,y\in T^*}} \sigma_x^2 \alpha_y \frac{s\tau(xy)}{\tau(x)\tau(y)} \right) \mathbf{V}_z. \quad (2.30)$$

Putting the left side of (2.30) equal to $\mathbf{I}$ makes each coefficient of $\mathbf{V}_z$ equal to zero. *i.e.*

$$\sigma_e^2 \alpha_z + \frac{\sigma_z^2}{\sigma_e^2} + f \sum_{\substack{x*y=z \\ x,y\in T^*}} \sigma_x^2 \alpha_y \frac{s\tau(xy)}{\tau(x)\tau(y)} = 0 \qquad z \in T^*. \quad (2.31)$$

(2.31) is the fundamental equation for solving α_z.

Note that T^* is an ordered set. We can solve (2.31) by solving α_w first where w is the largest element in T^*. For example, $w = 1\ldots1$, from (2.31) we have:

$$\sigma_e^2\alpha_{1\ldots1} + \frac{\sigma_{1\ldots1}^2}{\sigma_e^2} + f\sigma_{1\ldots1}^2\alpha_{1\ldots1} = 0,$$

thus

$$\alpha_{1\ldots1} = -\frac{\sigma_{1\ldots1}^2}{\sigma_e^2(\sigma_e^2 + f\sigma_{1\ldots1}^2)}.$$

Generally, let w be the largest element of T^*. Now $w = w * w$ and this is the only choice of the elements of T^* whose product equal w. Since if there exist x and y, $x \neq w$, or $y \neq w$, but $x * y = w$, then we should have $x \geq w$, $y \geq w$, this contradicts the assumption that w is the largest element of T^*.

From (2.31):

$$\sigma_e^2\alpha_w + \frac{\sigma_w^2}{\sigma_e^2} + f\sigma_w^2\alpha_w\frac{s\tau(ww)}{\tau(w)\tau(w)} = 0,$$

thus

$$\alpha_w = -\frac{\sigma_w^2}{\sigma_e^2\left(\sigma_e^2 + f\frac{s\tau(ww)}{\tau(w)\tau(w)}\sigma_w^2\right)}. \tag{2.32}$$

Suppose we have obtained α_y such that $y > z$. Notice that if $x * y = z$, then there are two situations for y: $y = z$ and $y > z$. Split these two cases in (2.31) we have:

$$\sigma_e^2\alpha_z + \frac{\sigma_z^2}{\sigma_e^2} + f\sum_{x \geq z}\frac{s\tau(xz)}{\tau(x)\tau(z)}\sigma_x^2\alpha_z + f\sum_{\substack{x*y=z\\ y>z}}\frac{s\tau(xy)}{\tau(x)\tau(y)}\sigma_x^2\alpha_y = 0. \tag{2.33}$$

Solve for α_z from (2.33) we have:

$$\alpha_z = -\left(\frac{\sigma_z^2}{\sigma_e^2} + f\sum_{\substack{x*y=z\\ y>z}}\frac{s\tau(xy)}{\tau(x)\tau(y)}\sigma_x^2\alpha_y\right)\frac{1}{\left(\sigma_e^2 + f\sum_{x\geq z}\frac{s\tau(xz)}{\tau(x)\tau(z)}\sigma_x^2\right)}$$

$$z \in T^*. \tag{2.34}$$

Note that all σ^2's are known, and by our assumption α_y for $y > z$ are known too, therefore α_z can be solved from (2.34).

Suppose that the variance components for the random error is always positive, *i.e.* $\sigma_e^2 > 0$, then even if by our definition some of the σ_x^2, $x \in T^*$ may be zero, the denominator of (2.34) will always be greater than zero.

With (2.32) and (2.34) we can solve for α_z by successive substitution. Therefore W is $\mathbf{V}^{-1}$, *i.e.*

$$\mathbf{V}^{-1} = \sum_{x \in T^*} \alpha_x \mathbf{V}_x + \frac{1}{\sigma_e^2}\mathbf{I}. \quad \square$$

Lemma 2.6 showed that the inverse matrix of the variance covariance matrix of balanced design has a similar structure to the variance covariance matrix in the sense that both $\mathbf{V}$ and $\mathbf{V}^{-1}$ can be expressed as linear combinations of Kronecker products of matrices $\mathbf{I}$ and $\mathbf{J}$.

We give an example on how to solve for $\mathbf{V}^{-1}$.

Example 2.6:

Consider a two-way crossed random model:

$$y_{ijk} = \mu + a_i + b_j + e_{ijk},$$

$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K.$

Suppose a_i is a random term with variance σ_a^2, b_j with σ_b^2 and e_{ijk} with σ_e^2. The variance covariance matrix of the model is:

$$\mathbf{V} = \sigma_a^2 \mathbf{I} \otimes \mathbf{J} \otimes \mathbf{J} + \sigma_b^2 \mathbf{J} \otimes \mathbf{I} \otimes \mathbf{J} + \sigma_e^2 \mathbf{I}.$$

Now $T = \{10, 01\}$ and $T^* = \{00, 10, 01\}$. In our notation: $\sigma_{10}^2 = \sigma_a^2$, $\sigma_{01}^2 = \sigma_b^2$, and $\sigma_{00}^2 = 0$.

Notice that in this example both 10 and 01 are the largest elements according to Definition 2.4.

First let $w = 10$, from (2.32) we have:

$$\alpha_{10} = -\frac{\sigma_a^2}{\sigma_e^2\left(\sigma_e^2 + JK\sigma_a^2\right)}. \tag{2.35}$$

Next let w=01, also from (2.32) we have:

$$\alpha_{01} = -\frac{\sigma_b^2}{\sigma_e^2(\sigma_e^2 + IK\sigma_b^2)}. \tag{2.36}$$

Now using (2.34):

$$\alpha_{00} = -\left\{\frac{\sigma_{00}^2}{\sigma_e^2} + K\sigma_{10}^2\alpha_{01} + K\sigma_{01}^2\alpha_{10}\right\} \Big/ \left\{\sigma_e^2 + IJK\sigma_{00}^2 + JK\sigma_{10}^2 + IK\sigma_{01}^2\right\},$$

but $\sigma_{00}^2 = 0$, α_{10} and α_{01} are known from (2.35) and (2.36), thus

$$\alpha_{00} = \frac{K\sigma_a^2\sigma_b^2}{\sigma_e^2(\sigma_e^2 + JK\sigma_a^2 + IK\sigma_b^2)} \left\{ \frac{1}{\sigma_e^2 + JK\sigma_a^2} + \frac{1}{\sigma_e^2 + IK\sigma_b^2} \right\} \tag{2.37}$$

Hence we obtained:

$$\mathbf{V}^{-1} = \alpha_{00}\mathbf{J} \otimes \mathbf{J} \otimes \mathbf{J} + \alpha_{10}\mathbf{I} \otimes \mathbf{J} \otimes \mathbf{J} + \alpha_{01}\mathbf{J} \otimes \mathbf{I} \otimes \mathbf{J} + (\frac{1}{\sigma_e^2})\mathbf{I}.$$

where α_{00}, α_{10} and α_{01} are given in the above formulas. It can be verified that $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}$. □

The MINQUE equations involve the multiplication of $\mathbf{X}$ and $\mathbf{V}^{-1}$. We have worked out the structure of $\mathbf{V}^{-1}$ in Lemma 2.6 and need to see if the MINQUE matrix $\mathbf{A}$ for the balanced E-ANOVA model satisfies condition (2.14) of Theorem 2.2. Since $\mathbf{X}$ is a Kronecker product of $\mathbf{I}$ and $\mathbf{1}$'s, and $\mathbf{V}^{-1} = \sum_{x \in T^*} \alpha_x \mathbf{V}_x + \frac{1}{\sigma_e^2}\mathbf{I}$, thus $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_{x \in T^*} \alpha_x \mathbf{X}'\mathbf{V}_x\mathbf{X} + \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{X}$, where $\mathbf{V}_x$ is given by (2.26). From Property 3 we know that wherever $\mathbf{X}$ has $\mathbf{1}$, $\mathbf{X}'\mathbf{V}_x\mathbf{X}$ is a constant number, and hence the corresponding index digit of the elements in T^* will disappear. Following the idea of using T to index $\mathbf{V}$, we define T_x to index $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$.

Definition 2.5 *Suppose $x = x_1 \ldots x_m$ is a 0,1 m-digit number. T is the set defined in Definition 2.1. If $y \in T$, let $(x * y)|_x$ be the product of x and y under $*$ with all those digits where x_i is zero crossed out. Then define $T_x = \{(x * y)|_x \ |y \in T\}$ to be the projection of set T on x.*

Example 2.7:

Consider a two-way balanced E-ANOVA model:

$$y_{ijk} = \mu_i + a_i + b_j + e_{ijk}, \tag{2.38}$$

$i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, where μ_i is a subclass mean, a_i and b_j are the random terms, e_{ijk} is the random error. Then $x = 10$, $T = \{10, 01\}$, $T_x = \{1\ \cancel{0}, 0\ \cancel{0}\} = \{1, 0\}$.

Now for the model considered here, we have: $\mathbf{X} = \mathbf{I}_I \otimes \mathbf{1}_J \otimes \mathbf{1}_K$, and from Example 2.6, we know that

$$\mathbf{V}^{-1} = \alpha_{00}\mathbf{J} \otimes \mathbf{J} \otimes \mathbf{J} + \alpha_{10}\mathbf{I} \otimes \mathbf{J} \otimes \mathbf{J} + \alpha_{01}\mathbf{J} \otimes \mathbf{I} \otimes \mathbf{J} + \frac{1}{\sigma_e^2}\mathbf{I},$$

where α_{00}, α_{10} and α_{01} were given in (2.35)–(2.37). Hence from Property 3,

$$\begin{aligned} \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} &= J^2K^2\alpha_{00}\mathbf{J}_I + J^2K^2\alpha_{10}\mathbf{I}_I + JK\alpha_{01}\mathbf{J}_I + \frac{JK}{\sigma_e^2}\mathbf{I}_I \\ &= \left(J^2K^2\alpha_{10} + \frac{JK}{\sigma_e^2}\right)\mathbf{I}_I + \left(J^2K^2\alpha_{00} + JK\alpha_{01}\right)\mathbf{J}_I. \end{aligned}$$

Now we can also determine the structure of $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ using T_x. For $y \in T_x$, if $y = 1$, then $\mathbf{V}_y = \mathbf{I}$, if $y = 0$, then $\mathbf{V} = \mathbf{J}$. Thus $T_x = \{1, 0\}$ indicates that

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \tau_1\mathbf{I} + \tau_2\mathbf{J},$$

where τ_1 and τ_2 are determined by s_i, f, σ_{10}^2, σ_{01}^2 and σ_e^2. The result is confirmed by the above calculation of $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. $\square$

Since T_x is a 0, 1 set, we can have $(T_x)^*$ according to Definition 2.3. We can also have $(T^*)_x$ according to Definition 2.5.

From Property 1 and 2 we know that the multiplication of two Kronecker products of $\mathbf{I}$ and $\mathbf{J}$'s will result in a Kronecker product of $\mathbf{I}$ and $\mathbf{J}$. This fact will be used in the proof of Lemma 2.7.

Looking back at Example 2.2 we noticed that the coefficients in the expression of the MINQUE matrix $\mathbf{A}$ and $\mathbf{U}_i'\mathbf{A}\mathbf{U}_i$, $i = 1, 2$, are not important in examining condition (2.14). It is the fact that $\mathbf{A}$ and $\mathbf{U}_i'\mathbf{A}\mathbf{U}_i$ are linear combinations of the Kronecker products of $\mathbf{I}$ and $\mathbf{J}$ that ensures that condition (2.14) is satisfied. In Lemma 2.7 we shall prove that the MINQUE matrix $\mathbf{A}$ for any balanced E-ANOVA model is a linear combination of the Kronecker products of $\mathbf{I}$ and $\mathbf{J}$.

Lemma 2.7 *Consider the balanced E-ANOVA model. Let T be the set corresponding to the random terms in the model, then if the MINQUE matrix $\mathbf{A}$ exists for the model, then it has the following structure:*

$$\mathbf{A} = \sum_{y \in T^* \cup \{x\}} \varphi_y \mathbf{V}_y,$$

where φ_y are functions of s_i, f, σ_y^2 and σ_e^2, and $\mathbf{V}_y$ is given in (2.26).

Proof: The variance covariance matrix of a balanced E-ANOVA model is:

$$\mathbf{V} = \sum_{y \in T} \sigma_y^2 \mathbf{V}_y + \sigma_e^2 \mathbf{I}.$$

From Lemma 2.6 there exists a set of scalars α_y, $y \in T^*$, such that

$$\mathbf{V}^{-1} = \sum_{y \in T^*} \alpha_y \mathbf{V}_y + \frac{1}{\sigma_e^2}\mathbf{I}.$$

Then from Definition 2.5, $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ can be indexed by $(T^*)_x$,

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_{y \in (T^*)_x} \tau_y \mathbf{V}_y + \frac{c}{\sigma_e^2}\mathbf{I},$$

where τ_y are functions of s_i, f, σ_y^2 and σ_e^2 produced by the multiplications of $\mathbf{X}'$, $\mathbf{V}^{-1}$ and $\mathbf{X}$, c is produced by $\mathbf{X}'\mathbf{X} = c\mathbf{I}$.

Again from Lemma 2.6 there exists a set of scalars τ_y, $y \in [(T^*)_x]^*$, such that

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \sum_{y\in[(T^*)_x]^*} \nu_y \mathbf{V}_y + \frac{\sigma_e^2}{c}\mathbf{I}.$$

Now from Property 4,

$$\begin{aligned}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' &= \sum_{y\in T^*} \delta_y \mathbf{V}_y + \frac{\sigma_e^2}{c}\mathbf{X}\mathbf{X}' \\ &= \sum_{y\in T^*\cup\{x\}} \delta_y \mathbf{V}_y,\ (\delta_x = \frac{\sigma_e^2}{c}).\end{aligned}$$

Using Properties 1 and 2,

$$\begin{aligned}\mathbf{R} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \\ &= \sum_{y\in T^*\cup\{x\}} \omega_y \mathbf{V}_y.\end{aligned}$$

From the MINQUE formulas (1.24)–(1.26), if the MINQUE of $\sum_{y\in T} q_y\sigma_y^2$ exists, then solutions of λ_y's exist for the equations

$$\sum_{y\in T} \lambda_y \mathrm{Tr}\mathbf{R}\mathbf{V}_y\mathbf{R}\mathbf{V}_z = q_z,\ z \in T.$$

Again from Property 1 and 2,

$$\mathbf{A} = \sum_{y\in T} \lambda_y \mathbf{R}\mathbf{V}_y\mathbf{R} = \sum_{y\in T^*\cup\{x\}} \varphi_y \mathbf{V}_y,$$

where ν_y, δ_y, ω_y and φ_y are functions of s_i, f, σ_y^2 and σ_e^2. Hence we proved the lemma. □

Theorem 2.3 *Consider a balanced E-ANOVA model. If the MINQUE matrix* $\mathbf{A}$ *of* $\sum_{y\in T} q_y\sigma_y^2$ *exists, then* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is the MIVQUE of* $\sum_{y\in T} q_y\sigma_y^2$.

Proof: We need to prove that condition (2.14) holds.

From Lemma 2.7, we know that

$$\mathbf{A} = \sum_{y\in T^*\cup\{x\}} \varphi_y \mathbf{V}_y.$$

For a balanced E-ANOVA model the design matrix $\mathbf{U}_y$ is given by (2.25) which is a Kronecker products of $\mathbf{I}$ and $\mathbf{1}$. From Properties 1 and 3 the multiplication of $\mathbf{I}$ does not change the MINQUE matrix $\mathbf{A}$ and the multiplication of $\mathbf{1}$ only results in the change of coefficients in $\mathbf{U}_y'\mathbf{A}\mathbf{U}_y$, so $\mathbf{U}_y'\mathbf{A}\mathbf{U}_y$ is still a Kronecker product of $\mathbf{I}$ and $\mathbf{J}$'s, hence

$$\mathbf{U}_y'\widetilde{\mathbf{A}}\mathbf{U}_y = \pi_y \mathbf{I},$$

where π_y is function of s_i, f, σ_y^2 and σ_e^2.

Therefore condition (2.14) is satisfied. From Theorem 2.2, $\mathbf{y}'\mathbf{A}\mathbf{y}$ is the MIVQUE of $\sum_{y \in T} q_y \sigma_y^2$. □

In this section we have examined the balanced E-ANOVA models. If the MINQUE matrix $\mathbf{A}$ exists, then condition (2.14) is satisfied for these models, and then $\mathbf{y}'\mathbf{A}\mathbf{y}$ is the best quadratic unbiased estimator of the variance component.

In Chapter 6 we derive an estimator for the interviewer's variance. If an interpenetrated interview scheme is adopted for the survey, then the design is balanced. Hence from Theorem 2.3 the estimator we derive is the minimum variance invariant quadratic unbiased estimator of the interviewer's variance.

There are many situations in practice where the normality assumption is not appropriate to make, *e.g.* data collected from a social survey and data collected from an experiment with a small number of replications. If we can manage to have a balanced design, then without the normality assumption Theorem 2.3 secures that the MINQUE $\mathbf{y}'\mathbf{A}\mathbf{y}$ will still be MIVQUE.

Discussion 2.1 *In Chapter 1 it is known that when data come from a normal distribution MINQUE is the best quadratic unbiased estimator of variance components. Theorem 2.2 ensures for any model with design matrices satisfying condition (2.14) MINQUE will be the best quadratic unbiased estimator. Particularly, for balanced E-ANOVA models it is proved in this section that balance is an alternative condition to normality for the optimality of MINQUE. Our result in this chapter extends the situations where MINQUE can be used as an optimal estimator.*

After the investigation on balanced E-ANOVA models, we now use a simple unbalanced model to examine if condition (2.14) is satisfied. We shall only examine condition (2.14) at parameter values $\sigma_a^2 = 1$ and $\sigma_e^2 = 1$ because of the algebraic complexity involved in deriving an algebraic solution.

Example 2.8:

Consider a one-way unbalanced model:

$$y_{ij} = \mu + a_i + e_{ij} \quad i = 1,2, j_1 = 1,2, j_2 = 1,2,3.$$

The design matrices are:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{U}_2 = \mathbf{I}_5.$$

Now at parameter values $\sigma_a^2 = 1$ and $\sigma_e^2 = 1$, then from the MINQUE formulas (1.24)–(1.26), we can obtain the MINQUE matrix $\mathbf{A}$ for the estimation of σ_a^2:

$$\begin{aligned} \mathbf{A} &= \lambda_1 \begin{bmatrix} 0.062 & 0.062 & -0.042 & -0.042 & -0.042 \\ 0.062 & 0.062 & -0.042 & -0.042 & -0.042 \\ -0.042 & -0.042 & 0.028 & 0.028 & 0.028 \\ -0.042 & -0.042 & 0.028 & 0.028 & 0.028 \\ -0.042 & -0.042 & 0.028 & 0.028 & 0.028 \end{bmatrix} \\ &+\lambda_2 \begin{bmatrix} 0.526 & -0.474 & -0.017 & -0.017 & -0.017 \\ -0.474 & 0.526 & -0.017 & -0.017 & -0.017 \\ -0.017 & -0.017 & 0.678 & -0.322 & -0.322 \\ -0.017 & -0.017 & -0.322 & 0.678 & -0.322 \\ -0.017 & -0.017 & -0.322 & -0.322 & 0.678 \end{bmatrix}, \end{aligned}$$

where $\lambda_1 = 2.065$, $\lambda_2 = -0.139$.

Since $\mathbf{U}_2 = \mathbf{I}$, hence $\mathbf{U}_2'\mathbf{A}\mathbf{U}_2 = \mathbf{A}$, it can be seen that:

$$\widetilde{\mathbf{U}_2'\mathbf{A}\mathbf{U}_2} = \tilde{\mathbf{A}} = \begin{bmatrix} 0.055\mathbf{I}_2 & 0 \\ 0 & -0.036\mathbf{I}_3 \end{bmatrix},$$

therefore condition (2.14) is not satisfied. But we can not conclude that $\mathbf{y}'\mathbf{A}\mathbf{y}$ is not the best quadratic unbiased estimator of σ_a^2, because condition (2.14) is a sufficient condition. □

Discussion 2.2 *The investigation in this section seems to suggest that condition (2.14) is a condition on the balance of the design. We shall not carry out further study on unbalanced designs because:*

(1) When dealing with unbalanced designs we would lose the powerful expression of the Kronecker products of matrices. It is therefore impossible to examine condition (2.14) generally;

(2) Condition (2.14) is a sufficient condition, hence not satisfying condition (2.14) does not necessarily mean that the MINQUE is not MIVQUE.

Nelder (1965a, 1965b), Speed (1987), Speed and Bailey (1987) and Bailey (1991) have worked on the analysis of randomized experiments. Although they also considered the estimation of random effect parameters, their work differed from the study in this chapter mainly in the following ways. First, they did not use variance components models. Their emphasis was on the extension of the ANOVA estimator (*i.e.* the splitting of the total sum of squares into sums of squares due to different variances) to designs of experiments with different blocking structures. C. R. Rao's MINQUE is derived for the general variance components model (1.1) so that MINQUE is dependent on the assumption that model (1.1) holds. It is not known if there exist some designs for which an appropriate model cannot be expressed in the form of variance components models. In such cases MINQUE is not available and the extension of ANOVA estimator is necessary.

Second, Nelder *et al*'s research focus is on the analysis of the randomized designs, *i.e.* on developing methods to estimate the variance and covariance parameter, not on the optimality of the estimator they used. Speed (1987) wrote at the end of his article:"We have not discussed any of the many questions, which are both mathematically and statistically interesting, which arise when the array of random variables has an anova." While in this chapter we considered the optimality of MINQUE which can be regarded as an extension to the ANOVA estimator through the framework of the general variance components model (1.1).

Speed and Bailey's work is an extension of Nelder (1965a, 1965b) to discuss more blocking structures using algebraic notation. Here we give a brief comparison between Nelder's work and the study in this chapter.

Nelder (1965a) gave three forms for expressing the variance covariance matrix $\mathbf{V}$ (see Nelder, 1965a, Appendix). In his analysis he only used the spectral decompositional form for the variance covariance matrix $\mathbf{V} = \sum_{i=1}^{k} \xi_i \mathbf{C}_i$, where ξ_i is the eigenvalues of $\mathbf{V}$ and $\mathbf{C}_i = \mathbf{p}_i \mathbf{p}_i'$, where $\mathbf{p}_i$ is the normalized orthogonal eigenvector of $\mathbf{V}$. Nelder (1965a) showed that when $\mathbf{C}_i$ is a complete binary set, *i.e.* $\sum_{i=1}^{k} \mathbf{C}_i = \mathbf{I}$, then we can have $\mathbf{y}'\mathbf{y} = \sum_{i=1}^{k}(\mathbf{y}'\mathbf{C}_i\mathbf{y})$. In other words the splitting of the total sum of squares to different sums of squares is possible. By writing the variance covariance matrix in the spectral decomposition form it is easy to see that $\mathbf{V}^{-1} = \sum_{i=1}^{k} \xi_i^{-1} \mathbf{C}_i$, though Nelder only used $\mathbf{V}^{-1}$ in the likelihood interpretation of the ANOVA estimator. But it is difficult to see the relationship between the variance components, σ_i^2, and the eigenvalues ξ_i of $\mathbf{V}$. Since MINQUE depends on the assumption of variance components models, I used $\mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$ in terms of the variance components σ_i^2 and the design matrices

$\mathbf{U}_i$ for the random effects ($\mathbf{V}_i = \mathbf{U}_i\mathbf{U}_i'$). Adopting different ways of writing the variance covariance matrix Nelder and I achieved different goals. Nelder (1965a) showed when the splitting of the total sum of squares is possible, while I showed that the optimality condition (2.14) is satisfied. The result proved in Lemma 2.6 does not directly follow from $\mathbf{V}^{-1} = \sum_{i=1}^{k} \xi_i^{-1}\mathbf{C}_i$, because deriving the eigenvalues and eigenvectors of $\mathbf{V}$ and converting them back into the form of $\mathbf{V}^{-1}$ in Lemma 2.6 requires far more work than the proof of Lemma 2.6.

Nelder (1965a) considered null analysis of variance, *i.e.* he assumed an overall mean for the data. The simple block structure he mentioned can be expressed as balanced ANOVA models, although it is not known if the inverse is true.

Nelder (1965b) considered estimating treatment effect in the presence of random effects. He assumed $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ without restriction on the design matrix $\mathbf{X}$. The class of designs he considered can be put into the form of model (1.1) and therefore is larger than the class of balanced E-ANOVA models considered in Section 2.2. But Nelder mainly considered the generalized least squares estimation of $\boldsymbol{\beta}$ and he did not consider how to estimate the random effect parameters in the presence of fixed effects. I restrict model (1.1) to balanced E-ANOVA models because I want to show that within this class of models MINQUE is the optimal estimator of variance components.

To summarize, Nelder, Speed and Bailey are extending the class of designs for which the ANOVA estimator is available. While in this chapter I use balanced E-ANOVA models to show that MINQUE is optimal to this class of models and design.

2.3 Conclusions

The key result in this chapter is the proof of Theorem 2.2 which gives condition (2.14) as an alternative sufficient condition for MINQUE to be MIVQUE without a normality assumption. We examined balanced ANOVA and E-ANOVA models and found that if the MINQUE $\mathbf{y}'\mathbf{A}\mathbf{y}$ exists it satisfies condition (2.14), hence the MINQUE $\mathbf{y}'\mathbf{A}\mathbf{y}$ is MIVQUE without a normality assumption.

In the next chapter we shall look at the problem of prior values and find out what kind of models can have estimators without using prior values.

Chapter 3

CONDITIONS FOR MINQUE TO BE INDEPENDENT OF PRIOR VALUES

3.1 The problem of prior values in the computation of MINQUE

In section 1.3.4 it is shown that the MINQUE formulas (1.24)–(1.26) require the inversion of the variance covariance matrix $\mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$ where the σ_i^2's are the unknown parameters. When using numeric computation to obtain the MINQUE it is necessary to put numerical values for the σ_i^2's which are called prior values. One of the requirements for the optimality of MINQUE is that the prior values are the true variance components values which are never known in practice. If we calculate the MINQUE equation algebraically, as we did in Example 1.8, where for the one way random model the MINQUE formulas for σ_a^2 and σ_e^2 are given, it can be seen that when the data are balanced the prior values of σ_a^2 and σ_e^2 will cancel out. In this case any prior values in the parameter space will lead us to the same estimate. Due to the complication of the MINQUE formulas (1.24)–(1.26) it is impossible and unrealistic to calculate the MINQUE of the variance components algebraically for all the models. Therefore before commencing the computation of MINQUE by the computer it is not known if the prior values will cancel out. If we put arbitrary prior values into the MINQUE formulas and if the prior values do not cancel out we will have different estimates by using different prior values while estimating the same parameter from the same formula.

In the literature there are two approaches proposed to solve the problem of prior values. Hartley, Rao and LaMotte (1978) suggested a synthesis-based MINQUE approach. They used a special set of prior values for all the models, which is to let $\sigma_e^2 = 1$ and all the other variance components $\sigma_i^2 = 0$. By using these prior values in the MINQUE formulas they obtain estimates which they called synthesis-based MINQUE. This approach uses the variance covariance matrix $\mathbf{V} = \mathbf{I}$ instead of $\mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$ in the computation, hence it reduced the computation needed to derive the inverse of the variance covariance matrix. The synthesis-based MINQUE is simple and easy to use, but usually the optimality of MINQUE will not be preserved. In Chapter 6 we shall show that by using the synthesis-based MINQUE approach the estimator derived only depends on one part of the data set, and hence ignores the information contained in the other part of the data set. The reason for using the specific set of prior values chosen for the synthesis-based MINQUE has not yet been justified.

Another approach is to use iterative computing. The idea is to use an arbitrary set of prior values to obtain the estimates from the MINQUE formulas and use the estimates as a new set of prior values to obtain further estimates, then repeat the procedure until the estimates obtained converge. If we regard the σ_i^2 on both sides of the MINQUE equations as unknowns, the iterative computing approach is actually seeking a numerical solution to the MINQUE equations. The problems are: first, we do not know what optimality the estimates obtained from iterative computing have; second, we do not know if MINQUE will always converge; Third, the estimator is no longer quadratic.

Neither approach is completely satisfactory.

In Example 1.8 the variance components σ_a^2 and σ_e^2 were involved in the calculation of MINQUE and cancelled out at the end of the calculation,so in this case MINQUE is independent of prior values and an analytic solution of MINQUE is secured, therefore, in the computation different prior values will lead to the same estimator.

In this chapter we investigate the conditions under which the prior values will cancel out for the MINQUE of σ_i^2, hence MINQUE will be independent of prior values. If MINQUE is independent of prior values then we can use synthesis-based MINQUE in the computation of MINQUE and ensure that the optimality of MINQUE remains. If MINQUE is dependent on prior values it seems that iterative computing is the only way of solving the problem.

3.2 Two useful theorems

There are two theorems in Szatrowski (1980) which will be used to derive the necessary and sufficient conditions for MINQUE to be independent of prior values.

Theorem 3.1 (Szatrowski, 1980) *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $p \times p$ *symmetric positive definite matrices and let* $\mathbf{X}$ *be a* $p \times r$ *matrix of full rank,* $r \leq p$*. Then* $(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{-1} = (\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^{-1}$ *if and only if the columns of* $\mathbf{X}$ *are linear combinations of* r *eigenvectors of* $\mathbf{AB}^{-1}$.

A proof can be found in Szatrowski (1980).

We need some definitions before stating the second theorem.

Definition 3.1 *Let* $\mathbf{A}$ *be a symmetric* $p \times p$ *matrix.* $<\mathbf{A}>$ *is defined to be a column vector consisting of the upper triangle of elements of* $\mathbf{A}$ *written in the following order:*

$$<\mathbf{A}>= (a_{11}, a_{22}, \ldots, a_{pp}, a_{12}, a_{13}, \ldots, a_{1p}, a_{23}, \ldots, a_{pp})'.$$

$<\mathbf{A}>$ is different from vec($\mathbf{A}$) and vech($\mathbf{A}$) defined in section 1.3.5 because $<\mathbf{A}>$ puts the diagonal elements of $\mathbf{A}$ into the vector first and then inserts the rows of the remainder of the upper triangle of $\mathbf{A}$.

A characteristic of $<\mathbf{A}>$ is that $<>$ transforms a diagonal matrix into a special vector form. If $\mathbf{A} = \text{diag}(a_{11}, a_{22}, \ldots, a_{pp})$, then $<\mathbf{A}>= (a_{11}, a_{22}, \ldots, a_{pp}, 0, \ldots, 0)'$. vec($\mathbf{A}$) and vech($\mathbf{A}$) do not possess this property. $<\mathbf{A}>$ is useful in transforming the variance of a quadratic estimator so that Theorem 3.1 can be used to find the necessary and sufficient conditions for MINQUE to be independent of prior values.

Definition 3.2 *Let* $\mathbf{A}$ *be a* $p \times p$ *symmetric matrix and let* $\Phi(\mathbf{A}) =<\mathbf{A}> \star <\mathbf{A}>'$ *be a* $\left\{\frac{p(p+1)}{2}\right\} \times \left\{\frac{p(p+1)}{2}\right\}$ *symmetric matrix, where the multiplication of the elements of* $<\mathbf{A}>$ *is defined as:*

$$a_{ij} \star a_{kl} = a_{ik}a_{jl} + a_{il}a_{jk}.$$

Both $<\mathbf{A}>$ and $\Phi(\mathbf{A})$ were first used by T. W. Anderson (1969) to study the properties of variance covariance matrices.

According to Definition 3.2, $a_{kl} \star a_{ij} = a_{ki}a_{lj} + a_{kj}a_{li} = a_{ik}a_{jl} + a_{il}a_{jk}$, because $\mathbf{A}$ is symmetric, so $\Phi(\mathbf{A})$ is symmetric.

We use an example to demonstrate how $\Phi(\mathbf{A})$ is constructed from $\mathbf{A}$.

Example 3.1:

Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$ be a 2×2 symmetric matrix, from Definition 3.1, $< \mathbf{A} >= (a_{11}, a_{22}, a_{12})'$. From Definition 3.2,

$$\begin{aligned} \Phi(\mathbf{A}) &= \begin{bmatrix} a_{11} \\ a_{22} \\ a_{12} \end{bmatrix} \star \begin{bmatrix} a_{11} & a_{22} & a_{12} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} \star a_{11} & a_{11} \star a_{22} & a_{11} \star a_{12} \\ a_{22} \star a_{11} & a_{22} \star a_{22} & a_{22} \star a_{12} \\ a_{12} \star a_{11} & a_{12} \star a_{22} & a_{12} \star a_{12} \end{bmatrix} \\ &= \begin{bmatrix} 2a_{11}^2 & 2a_{12}^2 & 2a_{11}a_{12} \\ 2a_{12}^2 & 2a_{22}^2 & 2a_{12}a_{22} \\ 2a_{11}a_{12} & 2a_{12}a_{22} & a_{11}a_{22} + a_{12}^2 \end{bmatrix}. \end{aligned}$$

We can see that $\Phi(\mathbf{A})$ is symmetric from this example. Particularly, if $a_{12} = 0$, *i.e.* $\mathbf{A}$ is a diagonal matrix, then

$$\Phi(\mathbf{A}) = \begin{bmatrix} 2a_{11}^2 & 0 & 0 \\ 0 & 2a_{22}^2 & 0 \\ 0 & 0 & a_{11}a_{22} \end{bmatrix}. \quad \square$$

Generally, if $\mathbf{A} = \text{diag}(a_{11}, a_{22}, \dots, a_{pp})$, then

$$\Phi(\mathbf{A}) = \begin{bmatrix} 2a_{11}^2 & & & & \\ & 2a_{22}^2 & & & 0 \\ & & \ddots & & \\ & & & 2a_{pp}^2 & \\ & 0 & & & \boldsymbol{\Sigma}_1 \end{bmatrix},$$

where $\boldsymbol{\Sigma}_1$ is a $\left\{\frac{p(p-1)}{2}\right\} \times \left\{\frac{p(p-1)}{2}\right\}$ matrix and its elements are determined by $\mathbf{A}$.

Particularly,

$$\Phi(\mathbf{I}_n) = \begin{bmatrix} 2\mathbf{I}_n & 0 \\ 0 & \mathbf{I}_m \end{bmatrix}, \quad \Phi^{-1}(\mathbf{I}_n) = \begin{bmatrix} \frac{1}{2}\mathbf{I}_n & 0 \\ 0 & \mathbf{I}_m \end{bmatrix},$$

where $m = \frac{n(n-1)}{2}$.

Theorem 3.2 uses the notation of $\Phi(\mathbf{A})$.

Theorem 3.2 (Szatrowski, 1980) *If $\boldsymbol{\Sigma}$ is a positive definite matrix, $\mathbf{E}$ and $\mathbf{F}$ are $p \times p$ symmetric matrices, then*

$$\mathrm{Tr}\,\boldsymbol{\Sigma}^{-1}\mathbf{E}\boldsymbol{\Sigma}^{-1}\mathbf{F} = 2 < \mathbf{E} >' \Phi^{-1}(\Sigma) < \mathbf{F} > . \tag{3.1}$$

A proof can be found in Anderson (1969).

3.3 Conditions for MINQUE to be independent of prior values

In this section we consider the general variance components model (1.1).

Since MINQUE has additivity, it is sufficient to derive MINQUE estimators for the σ_i^2 only and all the estimators for linear combinations of the variance components in the form of $\sum_{i=1}^k q_i\sigma_i^2$ can be derived correspondingly.

In section 1.3.4 we have derived equation (1.27) for the MINQUE of Θ:

$$\mathbf{S}\Theta = \mathbf{t}, \tag{3.2}$$

where $\Theta = (\sigma_1^2, \ldots, \sigma_k^2)'$, $\mathbf{S} = (s_{ij})_{k\times k}$, where $s_{ij} = \mathrm{Tr}\,\mathbf{RV}_i\mathbf{RV}_j$, $\mathbf{t} = (t_i)_{k\times 1}$, where $t_i = \mathbf{y}'\mathbf{RV}_i\mathbf{Ry} = \mathrm{Tr}\,\mathbf{RV}_i\mathbf{RC}$, $\mathbf{C} = \mathbf{yy}'$.

Now we want to use Theorem 3.2 to transform equation (3.2) into an equivalent form which will enable us to use Theorem 3.1.

Let $\mathbf{P}_V = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, then $\mathbf{R} = \mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)$. Then using Theorem 3.2

$$\begin{aligned} s_{ij} &= \mathrm{Tr}\,\mathbf{RV}_i\mathbf{RV}_j \\ &= \mathrm{Tr}\,\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)\mathbf{V}_j \\ &= 2 < (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >' \Phi^{-1}(\mathbf{V}) < (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_j > . \end{aligned} \tag{3.3}$$

Similarly,

$$\begin{aligned} t_i &= \mathrm{Tr}\,\mathbf{y}'\mathbf{RV}_i\mathbf{Ry} \\ &= \mathrm{Tr}\,\mathbf{RV}_i\mathbf{RC} \\ &= \mathrm{Tr}\,\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P}_V)\mathbf{C} \\ &= 2 < (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >' \Phi^{-1}(\mathbf{V}) < (\mathbf{I}-\mathbf{P}_V)\mathbf{C} > . \end{aligned} \tag{3.4}$$

Let $\mathbf{X}^* = (< (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_1 >, \ldots, < \mathbf{I}-\mathbf{P}_V)\mathbf{V}_k >)$, *i.e.* let $< (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >$ be the ith column of $\mathbf{X}^*$, then (3.2) can be written as:

$$\left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V})\mathbf{X}^*\right)\Theta = \mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V}) < (\mathbf{I}-\mathbf{P}_V)\mathbf{C} >, \tag{3.5}$$

or equivalently,

$$\Theta = \left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V}) < (\mathbf{I}-\mathbf{P}_V)\mathbf{C} > . \tag{3.6}$$

With the equivalent MINQUE equation (3.6), we can use Theorem 3.1 to derive the necessary and sufficient conditions for the MINQUE of Θ to be independent of prior values.

Theorem 3.3 *The necessary and sufficient conditions for the MINQUE of Θ to be independent of prior values of unknown parameters are:*

1. $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$;

2. *each* $< (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >$ *can be expressed as linear combinations of k eigenvectors of* $\Phi(\mathbf{V})\Phi^{-1}(\mathbf{I})$, $i = 1, \ldots, k$.

Proof: $\Rightarrow$ If the MINQUE of Θ is independent of prior values, any prior values of the variance covariance matrix will lead to the MINQUE estimate of Θ. Particularly, $\mathbf{V} = \mathbf{I}$ will yield the same estimate.

Thus $\mathbf{P}_V = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}' = \mathbf{P}_I$, *i.e.* the first condition is satisfied and from (3.6):

$$\Theta = \left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I}) < (\mathbf{I}-\mathbf{P}_I)\mathbf{C} > . \tag{3.7}$$

Since (3.6) and (3.7) give the same estimate and $\mathbf{C} = \mathbf{y}\mathbf{y}'$ where $\mathbf{y}$ is any data vector, we should have:

$$\left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{V}) = \left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I}). \tag{3.8}$$

Let $\mathbf{A} = \Phi(\mathbf{V})$ and $\mathbf{B} = \Phi(\mathbf{I})$ in Theorem 3.1, then if (3.8) holds, the columns of $\mathbf{X}^*$ are linear combinations of k eigenvectors of $\Phi(\mathbf{V})\Phi^{-1}(\mathbf{I})$. Note that $< (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >$ is the ith column of $\mathbf{X}^*$, so the second condition is satisfied, hence the two conditions are necessary.

$\Leftarrow$ Now assume the two conditions in the theorem are satisfied, then from the first condition, we know that $\mathbf{P}_V = \mathbf{P}_I = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'$ which is independent of prior values.

From the second condition we know that $< (\mathbf{I}-\mathbf{P}_V)\mathbf{V}_i >$ is the linear combinations of k eigenvectors of $\Phi(\mathbf{V})\Phi^{-1}(\mathbf{I})$, then using theorem 3.1 (3.6) becomes:

$$\Theta = \left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I}) < (\mathbf{I}-\mathbf{P}_I)\mathbf{C} > . \tag{3.9}$$

Since

$$\Phi^{-1}(\mathbf{I}) = \begin{bmatrix} \frac{1}{2}\mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix},$$

$\mathbf{C} = \mathbf{y}\mathbf{y}'$, and $\mathbf{P}_I = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, (3.9) does not contain any prior values of the unknown parameters, so MINQUE of Θ is independent of prior values. Hence we have proved that the conditions are sufficient.

Therefore the two conditions are necessary and sufficient. $\square$

Note if the MINQUE of Θ is independent of prior values, it has the following expression:

$$\Theta = \left(\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I})\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\Phi^{-1}(\mathbf{I}) < (\mathbf{I} - \mathbf{P}_I)\mathbf{C} >,$$

where $\mathbf{P}_I = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'$.

The necessary and sufficient conditions given in theorem 3.3 are very complicated because the set of models we are considering is large. For balanced E-ANOVA models the conditions can be simplified.

3.4 Simplified conditions for balanced E-ANOVA models

In addition to the six properties of Kronecker products of $\mathbf{I}_{s_i}$'s and $\mathbf{J}_{s_i}$'s in Chapter 2 there are four more properties needed in this section. Properties 8, 9 and 10 can be found in Chapter 8 of Graybill (1983). Property 7 can be derived using Property 1 in Chapter 2.

Property 7: If $\boldsymbol{\xi}_i$ is an eigenvector of $\mathbf{A}_i$ corresponding to the eigenvalue λ_i, $i = 1, \ldots, m$, then $\bigotimes_{i=1}^{m} \boldsymbol{\xi}_i$ is an eigenvector of $\bigotimes_{i=1}^{m} \mathbf{A}_i$ corresponding to the eigenvalue $\prod_{i=1}^{m} \lambda_i$.

Property 8: If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices, then $\mathbf{A} \otimes \mathbf{B}$ is nonsingular. Generally, if $\mathbf{A}_i$ is nonsingular matrix, $i = 1, \ldots, m$, then $\bigotimes_{i=1}^{m} \mathbf{A}_i$ is nonsingular.

Property 9: If $\mathbf{A}$ and $\mathbf{B}$ are matrices, a and b are scalars, then

$$ab(\mathbf{A} \otimes \mathbf{B}) = (a\mathbf{A}) \otimes (b\mathbf{B}).$$

Property 10: Let $\mathbf{A}$ be an $m_1 \times n_1$ matrix of rank r_1 and $\mathbf{B}$ be an $m_2 \times n_2$ matrix of rank r_2, then $\mathbf{A} \otimes \mathbf{B}$ has rank $r_1 r_2$.

In a balanced E-ANOVA model defined in Section 1.2 the design matrices for both fixed and random effects can be expressed as Kronecker products of $\mathbf{I}$'s and $\mathbf{1}$'s.

For the one-way balanced random model:

$$y_{ij} = \nu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

the design matrices are:

$$\mathbf{X} = \mathbf{1}_m \otimes \mathbf{1}_n, \quad \mathbf{U}_1 = \mathbf{I}_m \otimes \mathbf{1}_n, \quad \mathbf{U}_e = \mathbf{I}_m \otimes \mathbf{I}_n.$$

Thus $\mathbf{V}_1 = \mathbf{I}_m \otimes \mathbf{J}_n$, $\mathbf{V}_e = \mathbf{I}_m \otimes \mathbf{I}_n$, and the variance covariance matrix is:

$$\mathbf{V} = \sigma_a^2 \mathbf{I}_m \otimes \mathbf{J}_n + \sigma_e^2 \mathbf{I}_m \otimes \mathbf{I}_n.$$

Generally, assume a balanced E-ANOVA model is chosen and there are $m-1$ factors each at s_i levels and each observation is replicated s_n times, then the design matrix for the fixed effect can be expressed by:

$$\mathbf{X} = \bigotimes_{i=1}^{m} \mathbf{X}_i^{x_i},$$

where

$$\mathbf{X}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0. \end{cases} \tag{3.10}$$

The design matrix corresponding to σ_x^2 $(x = x_1 \ldots x_m)$ is:

$$\mathbf{U}_x = \bigotimes_{i=1}^{m} \mathbf{U}_i^{x_i},$$

where

$$\mathbf{U}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0. \end{cases} \tag{3.11}$$

Hence

$$\mathbf{V}_x = \bigotimes_{i=1}^{m} \mathbf{V}_i^{x_i},$$

where

$$\mathbf{V}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 0. \end{cases} \tag{3.12}$$

The variance covariance matrix is thus $\mathbf{V} = \sum \sigma_x^2 \mathbf{V}_x$.

Since the design matrices of balanced E-ANOVA models have this special structure, we can use an orthogonal matrix to diagonalise the matrices required in the conditions of Theorem 3.3 and derive simpler conditions than those of Theorem 3.3.

Now we choose the orthogonal matrix to be: $\boldsymbol{\Gamma}_n = (\gamma_{ij})_{n\times n}$, where

$$\gamma_{ij} = n^{-1/2}\left\{\cos\left[2\pi n^{-1}(i-1)(j-1)\right] + \sin\left[2\pi n^{-1}(i-1)(j-1)\right]\right\}.$$

The columns of Γ_n are the n roots of unity. Example of Γ_n for $n = 2$ is:

$$\boldsymbol{\Gamma}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

For $n = 3$,

$$\boldsymbol{\Gamma}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -\frac{1}{2}+\frac{\sqrt{3}}{2} & -\frac{1}{2}-\frac{\sqrt{3}}{2} \\ 1 & -\frac{1}{2}-\frac{\sqrt{3}}{2} & -\frac{1}{2}+\frac{\sqrt{3}}{2} \end{bmatrix}.$$

It can be verified that the columns of $\boldsymbol{\Gamma}_n$ are orthogonal eigenvectors of $\mathbf{J}_n$ which is the matrix of order n with all elements equal to one. $\boldsymbol{\Gamma}_n$ is symmetric and nonsingular.

Since $\mathbf{I}_n$ is the identity matrix of order n, it follows that the columns of $\boldsymbol{\Gamma}_n$ are also the eigenvectors of $\mathbf{I}_n$.

Note that the first column of $\boldsymbol{\Gamma}_n$ contains only 1's and this column corresponds to the only nonzero eigenvalue of $\mathbf{J}_n$, namely, n. The rest of the columns of $\mathbf{J}_n$ correspond to the zero eigenvalues of $\mathbf{J}_n$, hence

$$\boldsymbol{\Gamma}_n\mathbf{J}_n\boldsymbol{\Gamma}_n = \begin{bmatrix} n & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}, \quad \boldsymbol{\Gamma}_n\mathbf{I}_n\boldsymbol{\Gamma}_n = \mathbf{I}_n.$$

$\boldsymbol{\Gamma}_n$ diagonalizes $\mathbf{I}$ and $\mathbf{J}$ given their eigenvalues on the diagonal.

Now let $\mathbf{P} = \bigotimes_{i=1}^{m}\boldsymbol{\Gamma}_{s_i}$. Since the columns of $\boldsymbol{\Gamma}_{s_i}$ are eigenvectors of $\mathbf{I}_{s_i}$ and $\mathbf{J}_{s_i}$, from Property 7 we know that the columns of $\mathbf{P}$ are eigenvectors of $\mathbf{V}_i$ and $\mathbf{V}$. $\mathbf{P}$ is useful because it diagonalizes $\mathbf{X}$, $\mathbf{V}_i$ and $\mathbf{V}$.

Now we prove that for balanced E-ANOVA models the first condition of Theorem 3.3 is satisfied.

Theorem 3.4 *For balanced E-ANOVA models,*

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Proof: Now for balanced E-ANOVA models the design matrix $\mathbf{X}$ is given by (3.10): $\mathbf{X} = \bigotimes_{i=1}^{n} \mathbf{X}_i^{x_i}$, where

$$\mathbf{X}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0. \end{cases}$$

Let $\mathbf{B} = \bigotimes_{i=1}^{n} \mathbf{B}_i^{x_i}$, where

$$\mathbf{B}_i^{x_i} = \begin{cases} \boldsymbol{\Gamma}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0. \end{cases}$$

and let $\mathbf{C} = \bigotimes_{i=1}^{n} \mathbf{C}_i^{x_i}$, where

$$\mathbf{C}_i^{x_i} = \begin{cases} \boldsymbol{\Gamma}_{s_i}, & \text{if } x_i = 1, \\ 1, & \text{if } x_i = 0. \end{cases}$$

From Property 8 we know that $\mathbf{C}$ is nonsingular.

Since $\boldsymbol{\Gamma}_{s_i}$ is orthogonal and symmetric,

$$\mathbf{BC} = \left(\bigotimes_{i=1}^{m} \mathbf{B}_i^{x_i}\right)\left(\bigotimes_{i=1}^{m} \mathbf{C}_i^{x_i}\right) = \bigotimes_{i=1}^{m} (\mathbf{B}_i^{x_i}\mathbf{C}_i^{x_i}).$$

But

$$\mathbf{B}_i^{x_i}\mathbf{C}_i^{x_i} = \begin{cases} \boldsymbol{\Gamma}_{s_i}\boldsymbol{\Gamma}_{s_i} = \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{1}_{s_i}, & \text{if } x_i = 0, \end{cases}$$

thus $\mathbf{B}_i^{x_i}\mathbf{C}_i^{x_i} = \mathbf{X}_i^{x_i}$, hence

$$\mathbf{BC} = \mathbf{X}. \tag{3.13}$$

Since the columns of $\boldsymbol{\Gamma}_{s_i}$ are eigenvectors of $\mathbf{J}_{s_i}$, $\mathbf{1}_{s_i}$ is the first column of $\boldsymbol{\Gamma}_{s_i}$ and is therefore an eigenvector of $\mathbf{J}_{s_i}$, hence from Property 7 the columns of $\mathbf{B}$ are eigenvectors of $\mathbf{V}$. The number of columns of $\mathbf{B}$ is the same as the number of columns of $\mathbf{X}$, hence (3.13) means that the columns of $\mathbf{X}$ are linear combinations of r eigenvectors of $\mathbf{V}$, where $r = \text{rank}(\mathbf{B}) = \text{rank}(\mathbf{X})$.

From theorem 3.1, we have:

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

therefore

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \quad \square$$

Theorem 3.4 showed that all balanced E-ANOVA models satisfy the first condition for MINQUE to be independent of prior values. The remaining task is to simplify the second condition of Theorem 3.3 for the balanced E-ANOVA models.

Since the symmetric orthogonal matrix $\mathbf{P}$ can transform $\mathbf{V}_i$ and the variance covariance matrix $\mathbf{V}$ into diagonal forms containing the eigenvalues on the diagonal, we then use $\mathbf{A}^0 = \mathbf{PAP}$ to denote the diagonalised matrix of $\mathbf{A}$ by $\mathbf{P}$. So $\mathbf{V}^0 = \mathbf{PVP}$, $(\mathbf{I} - \mathbf{P}_V)^0 = \mathbf{P}(\mathbf{I} - \mathbf{P}_V)\mathbf{P}$, etc.

Using the diagonalised matrices and $\mathbf{PP} = \mathbf{I}$, we can transform the MINQUE equation (3.2) into another form. Now

$$\begin{aligned} s_{ij} &= \operatorname{Tr}\mathbf{RV}_i\mathbf{RV}_j \\ &= \operatorname{Tr}\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P}_V)\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P}_V)\mathbf{V}_j \\ &= \operatorname{Tr}\mathbf{PV}^{-1}\mathbf{PP}(\mathbf{I} - \mathbf{P}_V)\mathbf{PPV}_i\mathbf{PPV}^{-1}\mathbf{PP}(\mathbf{I} - \mathbf{P}_V)\mathbf{PPV}_j\mathbf{P} \\ &= \operatorname{Tr}(\mathbf{V}^0)^{-1}(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_i^0(\mathbf{V}^0)^{-1}(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_j^0. \end{aligned}$$

Similarly,

$$\begin{aligned} t_i &= \operatorname{Tr}\mathbf{RV}_i\mathbf{RC} \\ &= \operatorname{Tr}\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P}_V)\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P}_V)\mathbf{C} \\ &= \operatorname{Tr}\mathbf{PV}^{-1}\mathbf{PP}(\mathbf{I} - \mathbf{P}_V)\mathbf{PPV}_i\mathbf{PPV}^{-1}\mathbf{PP}(\mathbf{I} - \mathbf{P}_V)\mathbf{PPCP} \\ &= \operatorname{Tr}(\mathbf{V}^0)^{-1}(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_i^0(\mathbf{V}^0)^{-1}(\mathbf{I} - \mathbf{P}_V)^0\mathbf{PCP}. \end{aligned}$$

Let $\mathbf{X}^{*0} = (< (\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_1^0 >, \ldots, < (\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_k^0 >)$.

Using Theorem 3.2 the MINQUE equation (3.2) can be written as:

$$\Theta = \left(\mathbf{X}^{*0\prime}\Phi^{-1}(\mathbf{V}^0)\mathbf{X}^{*0}\right)^{-1}\mathbf{X}^{*0\prime}\Phi^{-1}(\mathbf{V}^0) < (\mathbf{I} - \mathbf{P}_V)^0\mathbf{PCP} > . \tag{3.14}$$

Theorem 3.5 *For the balanced E-ANOVA model, a necessary and sufficient condition for the MINQUE of $\Theta = (\sigma_1^2, \ldots, \sigma_k^2)'$ to be independent of prior values is that $< (\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_i^0 >$, $i = 1, \ldots, k$, can be expressed as linear combinations of k eigenvectors of $\Phi(\mathbf{V}^0)\Phi_I^{-1}$.*

Proof:

From Theorem 3.4 we know that for balanced E-ANOVA models,

$$\mathbf{P}_V = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}_I.$$

Hence condition 1 of Theorem 3.3 is satisfied.

Now substituting the MINQUE equation (3.6) used in the proof of Theorem 3.3 with the MINQUE equation (3.14) and following the same reasoning as in the proof of Theorem 3.3, we can prove that condition 2 of Theorem 3.3 is satisfied. Therefore this theorem is proved. □

Now we use an example to demonstrate how Theorem 3.5 works.

Example 3.2

Consider a one way balanced model:

$$y_{ij} = \mu + a_i + e_{ij} \quad i = 1,2, \;\; j = 1,2.$$

Now we have $\mathbf{X} = \mathbf{1}_4$, $\mathbf{V} = \sigma_e^2\mathbf{I}_4 + \sigma_a^2\mathbf{I}_2 \otimes \mathbf{J}_2$.

$$\mathbf{V}^{-1} = \frac{1}{\sigma_e^2}\mathbf{I}_4 - \frac{\sigma_a^2}{\sigma_e^2(\sigma_e^2 + 2\sigma_a^2)}\mathbf{I}_2 \otimes \mathbf{J}_2, \quad \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \frac{4}{\sigma_e^2 + 2\sigma_a^2}.$$

Hence

$$\begin{aligned}
\mathbf{P}_V &= \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1} \\
&= \frac{\sigma_e^2 + 2\sigma_a^2}{4}\mathbf{J}_4\left[\frac{1}{\sigma_e^2}\mathbf{I}_4 - \frac{\sigma_a^2}{\sigma_e^2(\sigma_e^2 + 2\sigma_a^2)}\mathbf{I}_2 \otimes \mathbf{J}_2\right] \\
&= \frac{1}{4}\mathbf{J}_4.
\end{aligned}$$

Now $\mathbf{P}_I = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \frac{1}{4}\mathbf{J}_4$, therefore $\mathbf{P}_V = \mathbf{P}_I$, as proved in Theorem 3.4 for all balanced E-ANOVA models.

Now

$$(\mathbf{I} - \mathbf{P}_V)\mathbf{V}_1 = \mathbf{I}_2 \otimes \mathbf{J}_2 - \frac{1}{2}\mathbf{J}_4,$$

$$(\mathbf{I} - \mathbf{P}_V)\mathbf{V}_2 = \mathbf{I}_4 - \frac{1}{4}\mathbf{J}_4.$$

Therefore the diagonalised forms are:

$$\begin{aligned}
(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_1^0 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 2 & \\ & & & 0 \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}_2^0 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}.
\end{aligned}$$

Hence

$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_1^0 >= (0,0,2,0,\ldots,0)',$$
$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_2^0 >= (0,1,1,1,0,\ldots,0)'.$$

Now the diagonalised form of the variance covariance matrix is:

$$\begin{aligned}\mathbf{V}^0 &= \sigma_e^2\begin{bmatrix}1&&&\\&1&&\\&&1&\\&&&1\end{bmatrix}+\sigma_a^2\begin{bmatrix}2&&&\\&0&&\\&&2&\\&&&0\end{bmatrix}\\ &= \begin{bmatrix}2\sigma_a^2+\sigma_e^2&&&\\&\sigma_e^2&&\\&&2\sigma_a^2+\sigma_e^2&\\&&&\sigma_e^2\end{bmatrix}.\end{aligned}$$

Therefore,

$$\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I}) = \begin{bmatrix}(2\sigma_a^2+\sigma_e^2)^2&&&&\\&\sigma_e^4&&&0\\&&(2\sigma_a^2+\sigma_e^2)^2&&\\&&&\sigma_e^4&\\&0&&&\Sigma_1\end{bmatrix}.$$

Let $\eta_1 = (0,0,1,0,\ldots,0)'$ and $\eta_2 = (0,1,0,1,0,\ldots,0)'$.
It can be verified that η_1 and η_2 are eigenvectors of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$. Now

$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_1^0 >= 2\eta_1,$$
$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_2^0 >= \eta_1+\eta_2.$$

From Theorem 3.5 for this model the MINQUEs of σ_a^2 and σ_e^2 are independent of prior values. □

From Example 3.2 we can see that examining the condition in Theorem 3.5 requires a lot of algebraic calculations. The following lemmas are aimed to simplify the condition in Theorem 3.5. Lemma 3.1 and Lemma 3.2 will be used in the proof of Lemma 3.3.

Lemma 3.1 *Let* $\mathbf{A} = \begin{bmatrix}\lambda_1&&\\&\ddots&\\&&\lambda_N\end{bmatrix}$, *and* $\Sigma = \begin{bmatrix}\lambda_1^2&&\\&\ddots&\\&&\lambda_N^2\end{bmatrix}$, *where* λ_i*'s are distinct nonzero real numbers,* $i=1,\ldots,N$, *if* η *is an eigenvector of* $\mathbf{A}$, *then*

1. η *is an eigenvector of* Σ;
2. $(\eta',0)'$ *is an eigenvector of* $\Phi(\mathbf{A})\Phi^{-1}(\mathbf{I})$.

Proof:

1. Since λ_i's are the distinct eigenvalues of $\mathbf{A}$, suppose $\boldsymbol{\eta} = (\eta_i)_{N\times 1}$ is the eigenvector corresponding to λ_i, then

$$\mathbf{A}\boldsymbol{\eta} = \begin{bmatrix} \lambda_1\eta_1 \\ \vdots \\ \lambda_i\eta_i \\ \vdots \\ \lambda_N\eta_N \end{bmatrix} = \lambda_i\boldsymbol{\eta} = \begin{bmatrix} \lambda_i\eta_1 \\ \vdots \\ \lambda_i\eta_i \\ \vdots \\ \lambda_i\eta_N \end{bmatrix}$$

gives $\eta_j = 0$, for $j \neq i$. Hence $\boldsymbol{\eta} = (0 \quad \ldots \quad 0 \quad 1 \quad 0 \quad \ldots \quad 0)'$. Therefore

$$\boldsymbol{\Sigma\eta} = (0, \ldots, 0, \lambda_i^2, 0, \ldots, 0)' = \lambda_i^2\boldsymbol{\eta}.$$

Hence $\boldsymbol{\eta}$ is an eigenvector of $\boldsymbol{\Sigma}$.

2. Since

$$\Phi(\mathbf{A}) = \begin{bmatrix} 2\lambda_1^2 & & & \\ & \ddots & & 0 \\ & & 2\lambda_N^2 & \\ & 0 & & \boldsymbol{\Sigma}_1 \end{bmatrix}, \quad \Phi^{-1}(\mathbf{I}) = \begin{bmatrix} \frac{1}{2}\mathbf{I}_N & 0 \\ 0 & \mathbf{I} \end{bmatrix},$$

thus

$$\Phi(\mathbf{A})\Phi^{-1}(\mathbf{I}) = \begin{bmatrix} \lambda_1^2 & & & \\ & \ddots & & 0 \\ & & \lambda_N^2 & \\ & 0 & & \boldsymbol{\Sigma}_1 \end{bmatrix}.$$

As proved in part 1 $\boldsymbol{\eta}$ is an eigenvector of $\boldsymbol{\Sigma}$, hence $(\boldsymbol{\eta}', 0)'$ is an eigenvector of $\Phi(\mathbf{A})\Phi^{-1}(\mathbf{I})$. □

Using Lemma 3.1 we can focus our attention on the eigenvectors of $\mathbf{A}$ in order to determine the eigenvectors of $\Phi(\mathbf{A})\Phi^{-1}(\mathbf{I})$ which is more complicated than $\mathbf{A}$, if $\mathbf{A}$ is a diagonal matrix.

Lemma 3.2 *Let*

$$\mathbf{A} = \begin{bmatrix} 0 & & & \\ & \tau_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \tau_k\mathbf{I}_{p_k} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \tau_1\mathbf{I}_{p_1} & & & \\ & \tau_2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \tau_k\mathbf{I}_{p_k} \end{bmatrix},$$

where $\tau_i \neq \tau_j$, *for* $i \neq j$. *then*

$$\boldsymbol{\xi}_i = (0, \ldots, 0, \mathbf{1}_{p_i}', 0, \ldots, 0)', \ i = 2, \ldots, k,$$

are eigenvectors of both $\mathbf{A}$ *and* $\mathbf{B}$.

The lemma can be verified by showing $\mathbf{A}\boldsymbol{\xi}_i = \tau_i\boldsymbol{\xi}_i$, and $\mathbf{B}\boldsymbol{\xi}_i = \tau_i\boldsymbol{\xi}_i$, where $i = 2, \ldots, k$.

Now we need to define another way of transforming a diagonal matrix into a vector. The definition is very helpful in the proof of Lemma 3.3.

Definition 3.3 *If* $\mathbf{A}$ *is a* $p \times p$ *diagonal matrix, let* $d(\mathbf{A})$ *be the vector containing all the diagonal elements of* $\mathbf{A}$, *i.e.*

$$d(\mathbf{A}) = (a_{11}, a_{22}, \ldots, a_{pp})'.$$

Remember that $< \mathbf{A} >$ also transforms a matrix into a vector, the relationship between $< \mathbf{A} >$ and $d(\mathbf{A})$ for a diagonal matrix $\mathbf{A}$ is: $< \mathbf{A} >= \begin{bmatrix} d(\mathbf{A}) \\ 0 \end{bmatrix}$.

To make the proof of Lemma 3.3 easier we now investigate the properties of $(\mathbf{I} - \mathbf{P}_V)^0$ and $\mathbf{V}_i^0$.

From Theorem 3.4 we know that for balanced E-ANOVA models $\mathbf{P}_V = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, where $\mathbf{X}$ is given by (3.10). For a chosen design matrix $\mathbf{X}$ the combination $x = x_1 \ldots x_m$ is determined. Recall the notation we used in Chapter 2: $\tau(x) = \prod_{i=1}^m s_i^{x_i}$, $s = \prod_{i=1}^m s_i$.

For example, if $\mathbf{X} = \mathbf{I}_{s_1} \otimes \mathbf{1}_{s_2} \otimes \mathbf{1}_{s_3}$, then $x = 100$, $\tau(x) = s_1^1 s_2^0 s_3^0 = s_1$.

The notation $\tau(x)$ enables us to write: $\mathbf{X}'\mathbf{X} = \frac{s}{\tau(x)}\mathbf{I}$. In the above example, $\mathbf{X}'\mathbf{X} = s_2 s_3 \mathbf{I}_{s_1} = \frac{s}{s_1}\mathbf{I} = \frac{s}{\tau(100)}\mathbf{I}$.

Hence $(\mathbf{X}'\mathbf{X})^{-1} = \frac{\tau(x)}{s}\mathbf{I}$. Therefore

$$\mathbf{P}_V = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \frac{\tau(x)}{s}\left(\bigotimes_{i=1}^{m} \bar{\mathbf{X}}_i^{x_i}\right),$$

where

$$\bar{\mathbf{X}}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \mathbf{J}_{s_i}, & \text{if } x_i = 0. \end{cases}$$

Since $\mathbf{P} = \bigotimes_{i=1}^m \boldsymbol{\Gamma}_{s_i}$,

$$\mathbf{P}_V^0 = \mathbf{P}\mathbf{P}_V\mathbf{P} = \frac{\tau(x)}{s}\left[\bigotimes_{i=1}^{m} \left(\boldsymbol{\Gamma}_{s_i}\bar{\mathbf{X}}_i^{x_i}\boldsymbol{\Gamma}_{s_i}\right)\right].$$

Since $\boldsymbol{\Gamma}_{s_i}$ is the orthogonal matrix diagonalising $\mathbf{J}_{s_i}$,

$$\boldsymbol{\Gamma}_{s_i}\bar{\mathbf{X}}_i^{x_i}\boldsymbol{\Gamma}_{s_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \begin{bmatrix} s_i & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix}, & \text{if } x_i = 0. \end{cases}$$

Taking the coefficient $\frac{\tau(x)}{s}$ into account and using Property 9, we have:

$$\mathbf{P}_V^0 = \bigotimes_{i=1}^{m} \mathbf{D}_i^{x_i},$$

where

$$\mathbf{D}_i^{x_i} = \begin{cases} \mathbf{I}_{s_i}, & \text{if } x_i = 1, \\ \begin{bmatrix} 1 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix}, & \text{if } x_i = 0. \end{cases}$$

In the example we should have $\mathbf{P}_V = \frac{1}{s_2 s_3}\mathbf{I}_{s_1} \otimes \mathbf{J}_{s_2} \otimes \mathbf{J}_{s_3}$, and

$$\begin{aligned} \mathbf{P}_V^0 &= \frac{1}{s_2 s_3}\mathbf{I}_{s_1} \otimes \begin{bmatrix} s_2 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \otimes \begin{bmatrix} s_3 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \\ &= \mathbf{I}_{s_1} \otimes \begin{bmatrix} 1 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix}. \end{aligned}$$

Since $(\mathbf{I} - \mathbf{P}_V)^0 = \mathbf{P}(\mathbf{I} - \mathbf{P}_V)\mathbf{P} = \mathbf{I} - \mathbf{P}\mathbf{P}_V\mathbf{P} = \mathbf{I} - \mathbf{P}_V^0$, and $\mathbf{P}_V^0$ only has 1's and 0's on its diagonal, $(\mathbf{I} - \mathbf{P}_V)^0$ only has 0's and 1's on its diagonal.

Similarly, for the $\mathbf{V}_i$ given by (3.12) we can see that $\mathbf{V}_i^0$ only has $\frac{s}{\tau(i)}$'s and 0's on its diagonal, where i is the index of $\mathbf{V}_i$.

Lemma 3.3 *For balanced E-ANOVA models, the following two conditions are equivalent:*

1. *each* $< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0 >$, $i = 1, \ldots, k$, *can be expressed as linear combinations of* k *eigenvectors of* $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$.

2. $(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}^0$ *contains exactly* k *distinct linear independent nonzero combinations of* σ_i^2*'s.*

Proof: $2 \Rightarrow 1$

Assume that $\tau_1, \ldots, \tau_k$ are the k distinct nonzero linear combination of σ_i^2's contained on the diagonal of $(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}^0$.

Let $\boldsymbol{\xi}_i$ be the vector having 1 wherever $(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}^0$ has τ_i and 0 otherwise, $i = 1, \ldots, k$.

From Lemma 3.2 we know that $\boldsymbol{\xi}_i$ is the eigenvector of $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$ corresponding to τ_i, $i=1,\ldots,k$. Since $(\mathbf{I}-\mathbf{P}_V)^0$ only have 1's and 0's on its diagonal, $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$ has the same nonzero elements as $\mathbf{V}^0$. From Lemma 3.2 $\boldsymbol{\xi}_i$'s are eigenvectors of $\mathbf{V}^0$, $i=1,\ldots,k$.

Since $\mathbf{V}=\sum_{i=1}^k \sigma_i^2\mathbf{V}_i$, then $\mathbf{V}^0=\sum_{i=1}^m \sigma_i^2\mathbf{V}_i^0$,

$$(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0=\sum_{i=1}^k \sigma_i^2(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0. \tag{3.15}$$

From the investigation on $(\mathbf{I}-\mathbf{P}_V)^0$ and $\mathbf{V}_i^0$ before the lemma we know that $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0$ can only have $\frac{s}{\tau(i)}$ and 0 on its diagonal. The equality in (3.15) indicates that the elements of $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0$ are the coefficients of σ_i^2's which form the combinations on the diagonal of $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$. In other words when a variance component, say σ_i^2, is present in a combination τ_j, $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0$ must have $\frac{s}{\tau(i)}$ in the same position as τ_j on the diagonal of $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$. Otherwise, $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0$ must have zero.

We then have:

$$d((\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0)=\sum_{j=1}^k b_j^{(i)}\boldsymbol{\xi}_j, \tag{3.16}$$

where

$$b_j^{(i)}=\begin{cases}\frac{s}{\tau(i)}, & \text{if } \sigma_i^2 \text{ is present in } \tau_j,\\ 0, & \text{otherwise.}\end{cases}$$

Let $\boldsymbol{\eta}_i=(\boldsymbol{\xi}_i',0)'$, then by Lemma 3.1 we know that $\boldsymbol{\eta}_i$ is the eigenvector of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$. Since

$$\begin{aligned}<(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0> &= \left(d((\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0)',0\right)'\\ &= \left(\sum_{j=1}^k b_j^{(i)}\boldsymbol{\xi}_j',0\right)'\\ &= \sum_{j=1}^k b_j^{(i)}(\boldsymbol{\xi}_i',0)'\\ &= \sum_{j=1}^k b_j^{(i)}\boldsymbol{\eta}_j,\ i=1,\ldots,k,\end{aligned} \tag{3.17}$$

then $<(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0>$ can be expressed as linear combinations of $\boldsymbol{\eta}_j$'s, $i,j=1,\ldots,k$.

Now $\mathbf{V}_k=\mathbf{I}$, so σ_e^2 is present in all the nonzero linear combinations of σ_i^2 in $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$, therefore,

$$b_j^{(k)}=1,\quad j=1,\ldots,k.$$

In other words, all the k $\boldsymbol{\eta}_i$'s are needed for expressing $< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_k^0 >$, therefore, $< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0 >$ can be expressed as linear combinations of k eigenvectors of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$.

$1 \Rightarrow 2$

Assume that $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$ contains m distinct combinations of σ_i^2's, $m \neq k$.

Then by the same reasoning as in part 1 of the proof, we can define $\boldsymbol{\xi}_i$ and $\boldsymbol{\eta}_i$ so that $\boldsymbol{\eta}_i$ is the eigenvectors of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$ and

$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0 > = \sum_{j=1}^{m} b_j^{(i)}\boldsymbol{\eta}_j.$$

Therefore $< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_i^0 >$ can be expressed as a linear combinations of m eigenvectors of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$, $m \neq k$. This contradicts condition 1, hence $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$ consists of exactly k linear independent nonzero combinations of the σ_i^2's. □

Example 3.3:

The construction of $\boldsymbol{\xi}_i$ and η_i in the proof of Lemma 3.3 can be demonstrated by using the model in Example 3.2.

From Example 3.2 we know that

$$(\mathbf{I}-\mathbf{P}_V)^0 = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix},$$

hence

$$(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0 = \begin{bmatrix} 0 & & & \\ & \sigma_e^2 & & \\ & & 2\sigma_a^2+\sigma_e^2 & \\ & & & \sigma_e^2 \end{bmatrix}.$$

So there are two distinct nonzero combinations of σ_a^2 and σ_e^2, namely, $\tau_1 = \sigma_e^2$, $\tau_2 = 2\sigma_a^2 + \sigma_e^2$, thus we define

$$\boldsymbol{\xi}_1 = (0,1,0,1)', \quad \boldsymbol{\xi}_2 = (0,0,1,0)'.$$

It can be seen that $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are eigenvectors of $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}^0$, they are also eigenvectors of $\mathbf{V}^0$.

With $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_1^0$ and $(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_2^0$ given in Example 3.2, we can write:

$$d\left[(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_1^0\right] = (0,0,2,0)' = 2\boldsymbol{\xi}_2, \quad (b_1^{(1)} = 0, b_2^{(1)} = 2),$$

$$d\left[(\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_2^0\right] = (0,1,1,1)' = \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2, \quad (b_1^{(2)} = 1, b_2^{(2)} = 1).$$

Now let $\boldsymbol{\eta}_1 = (\boldsymbol{\xi}_1', 0)'$, $\boldsymbol{\eta}_2 = (\boldsymbol{\xi}_2', 0)'$, $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are eigenvectors of $\Phi(\mathbf{V}^0)\Phi^{-1}(\mathbf{I})$. As shown in Example 3.2 we have:

$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_1^0 >= 2\boldsymbol{\eta}_2,$$

$$< (\mathbf{I}-\mathbf{P}_V)^0\mathbf{V}_2^0 >= \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2.$$

So the MINQUEs of σ_a^2 and σ_e^2 are independent of prior values. □

The next lemma establishes another equivalent condition for the second condition in lemma 3.3, but we need some definitions before the lemma.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be row vectors corresponding to the random terms in a balanced E-ANOVA model. The elements of $\mathbf{v}_i$ contain 0 and 1 only. $\mathbf{v}_i$ has 1 in the jth element if the jth factor is present in the ith random term in the model and 0 otherwise. $\mathbf{v}_1, \ldots, \mathbf{v}_k$ will determine the random effect part of the model. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the row vector corresponding to the fixed effect term in the E-ANOVA model and $\mathbf{x}$ is similarly defined as the $\mathbf{v}_i$'s.

Example 3.4:

Two way crossed model with interaction:

$$y_{ijl} = \mu + a_i + b_j + c_{ij} + e_{ijl},$$

$$i = 1, \ldots, I, \; j = 1, \ldots, J, \; l = 1, \ldots, L.$$

Thus $k = 4$, $\mathbf{v}_1 = (1,0,0)$, $\mathbf{v}_2 = (0,1,0)$, $\mathbf{v}_3 = (1,1,0)$ and $\mathbf{v}_4 = (1,1,1)$. $\mathbf{x} = (0,0,0)$. □

Denote the $k \times n$ matrix $\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{bmatrix}$ by $\mathbf{W}$. The rows of $\mathbf{W}$ correspond to the random effects in the model and the columns are indexed by the factors. We notice that only balanced E-ANOVA models allow us to use the notation of $\mathbf{W}$.

Define the vector multiplication as: $\mathbf{x} * \mathbf{v}_i = (x_1 v_{i_1}, \ldots, x_n v_{i_n})$, which is a Hadamard product of matrices, see Rao (1973, $p30$). Hence $\mathbf{x} * \mathbf{v}_i$'s are vectors.

Let

$$\mathbf{W}_X = \begin{bmatrix} \mathbf{x} * \mathbf{v}_1 \\ \mathbf{x} * \mathbf{v}_2 \\ \vdots \\ \mathbf{x} * \mathbf{v}_k \end{bmatrix},$$

$\mathbf{W}$ and $\mathbf{W}_X$ are $k \times n$ matrices. The column vectors of $\mathbf{W}$ and $\mathbf{W}_X$ can multiply each other using the rule of the Hadamard product. A column vector is said to be generated by the column vectors of $\mathbf{W}$ if it is a Hadamard product of two or more column vectors of $\mathbf{W}$. Let $\mathbf{Z}$ be the matrix containing all columns of $\mathbf{W}$ and the columns generated by the columns of $\mathbf{W}$. If $\mathbf{W} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n]$, then

$$\mathbf{Z} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n, \boldsymbol{\alpha}_1 * \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{n-1} * \boldsymbol{\alpha}_n, \boldsymbol{\alpha}_1 * \boldsymbol{\alpha}_2 * \boldsymbol{\alpha}_3, \ldots, \boldsymbol{\alpha}_1 * \cdots * \boldsymbol{\alpha}_n]$$

$\mathbf{Z}_X$ is similarly defined.

Definition 3.4 *The number of distinct nonzero columns generated by the columns of* $\mathbf{Z} - \mathbf{Z}_X$ *is defined as* $N(\mathbf{Z} - \mathbf{Z}_X)$.

Example 3.5:

Continued from Example 3.4,

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \ \mathbf{W}_X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{Z}_X = 0.$$

then $N(\mathbf{Z} - \mathbf{Z}_X) = 4$. □

Example 3.5 demonstrated that $N(\mathbf{Z} - \mathbf{Z}_X)$ can be larger than the number of columns of $\mathbf{W}$ and $\mathbf{W}_X$.

For balanced E-ANOVA models matrices $\mathbf{W}$ and $\mathbf{W}_X$ are determined when a model is chosen, hence $N(\mathbf{Z} - \mathbf{Z}_X)$ is known from a model without using design matrices. The next lemma establishes another equivalence which will eventually allow us to use $N(\mathbf{Z} - \mathbf{Z}_X)$ as an indicator of examining the necessary and sufficient condition of Theorem 3.5.

The result of Lemma 3.4 will be used in proving Lemma 3.5.

Lemma 3.4 *Let* $\Delta_i = \begin{bmatrix} \delta_{i1} & & \\ & \ddots & \\ & & \delta_{iN} \end{bmatrix}$ *be an* $N \times N$ *diagonal matrix,* $i = 1, \ldots, k$, *if* $\mathbf{A} = \sum_{i=1}^{k} \sigma_i^2 \Delta_i$, *then* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = diag\left\{(\sigma_1^2, \ldots, \sigma_k^2)\Delta\right\},$$

where

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1N} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2N} \\ \vdots & \vdots & & \vdots \\ \delta_{k1} & \delta_{k2} & \dots & \delta_{kN} \end{bmatrix},$$

i.e. the ith row of Δ *is* $d(\Delta_i)'$.

Proof: Since $\mathbf{A} = \sum_{i=1}^{k} \sigma_i^2 \Delta_i$,

$$\begin{aligned} \mathbf{A} &= \sigma_1^2 \begin{bmatrix} \delta_{11} & & \\ & \ddots & \\ & & \delta_{1N} \end{bmatrix} + \cdots + \sigma_k^2 \begin{bmatrix} \delta_{k1} & & \\ & \ddots & \\ & & \delta_{kN} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^{k} \sigma_i^2 \delta_{i1} & & \\ & \ddots & \\ & & \sum_{i=1}^{k} \sigma_i^2 \delta_{iN} \end{bmatrix}. \end{aligned}$$

Let $\Theta = (\sigma_1^2, \dots, \sigma_k^2)'$, $\boldsymbol{\delta}_i = (\delta_{1i}, \dots, \delta_{ki})'$, $i = 1, \dots, N$, it can be seen that $\boldsymbol{\delta}_i$ is the ith column of Δ. Then

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \Theta'\boldsymbol{\delta}_1 & & \\ & \ddots & \\ & & \Theta'\boldsymbol{\delta}_N \end{bmatrix} \\ &= diag\left\{\Theta'(\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_N)\right\} \\ &= diag\left\{\Theta'\Delta\right\} \\ &= diag\left\{(\sigma_1^2, \dots, \sigma_k^2)\Delta\right\}. \quad \square \end{aligned}$$

Lemma 3.5 *For balanced E-ANOVA models, the following two conditions are equivalent:*

1. $(\mathbf{I} - \mathbf{P}_V)^0\mathbf{V}^0$ *contains exactly* k *distinct linear combinations of* σ_i^2*'s;*
2. $N(\mathbf{Z} - \mathbf{Z}_X) = k$.

Proof: We want to show that the number of distinct linear combinations of σ_i^2 is the same as $N(\mathbf{Z} - \mathbf{Z}_X)$, hence the equivalence.

The diagonal form of $\mathbf{V}_i$ is:

$$\mathbf{V}_i^0 = \begin{bmatrix} 1 & \\ & \mathbf{d}_{i_1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & \\ & \mathbf{d}_{i_n} \end{bmatrix},$$

where $\mathbf{d}_{i_j}$ is a diagonal matrix of which the diagonal elements are either of all 1 or all 0. Since the order of the matrix is known, a scalar d_{i_j} which is 1 or 0 according to $\mathbf{d}_{i_j}$ can represent $\mathbf{d}_{i_j}$. For simplicity's reason we assume that d_{i_j} is a scalar of 1 or 0. Hence,

$$\mathbf{V}_i^0 = \begin{bmatrix} 1 & \\ & d_{i_1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & \\ & d_{i_n} \end{bmatrix}.$$

Therefore,

$$\mathbf{V}_i^0 = \operatorname{diag}(1, d_{i_n}, d_{i_{(n-1)}}, d_{i_n} d_{i_{(n-1)}}, d_{i_{(n-2)}}, \ldots, \prod_{j=1}^{n} d_{i_j}).$$

Also

$$\mathbf{P}_V^0 = \begin{bmatrix} 1 & \\ & \mathbf{x}_1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & \\ & \mathbf{x}_n \end{bmatrix},$$

where $\mathbf{x}_i$ is a diagonal matrix of which the diagonal elements are either all 1 or all 0. For the same reason as $\mathbf{V}_i^0$ we use

$$\mathbf{P}_V^0 = \begin{bmatrix} 1 & \\ & x_1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & \\ & x_n \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{P}_V^0 \mathbf{V}_i^0 &= \begin{bmatrix} 1 & \\ & x_1 d_{i_1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & \\ & x_n d_{i_n} \end{bmatrix} \\ &= \operatorname{diag}(1, x_n d_{i_n}, x_{n-1} d_{i_{(n-1)}}, x_n d_{i_n} x_{n-1} d_{i_{(n-1)}}, \\ &\quad x_{n-2} d_{i_{(n-2)}}, \ldots, \prod_{j=1}^{n} x_j d_{i_j}). \end{aligned}$$

Since $(\mathbf{I}-\mathbf{P}_V)^0 \mathbf{V}^0 = \sum_{i=1}^k \sigma_i^2 (\mathbf{V}_i^0 - \mathbf{P}_V^0 \mathbf{V}_i^0)$, in Lemma 3.4, let $\Delta_i = \mathbf{V}_i^0 - \mathbf{P}_V^0 \mathbf{V}_i^0$, then

$$(\mathbf{I} - \mathbf{P}_V)^0 \mathbf{V}^0 = \operatorname{diag}\left\{(\sigma_1^2, \ldots, \sigma_k^2)(\mathbf{Z} - \mathbf{Z}_X)\right\}, \tag{3.18}$$

where

$$\mathbf{Z} = \begin{bmatrix} 1 & d_{1_n} & d_{1_{(n-1)}} & d_{1_n} d_{1_{(n-1)}} & d_{1_{(n-2)}} & \cdots & \prod_{j=1}^n d_{1_j} \\ 1 & d_{2_n} & d_{2_{(n-1)}} & d_{2_n} d_{2_{(n-1)}} & d_{2_{(n-2)}} & \cdots & \prod_{j=1}^n d_{2_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & d_{k_n} & d_{k_{(n-1)}} & d_{k_n} d_{k_{(n-1)}} & d_{k_{(n-2)}} & \cdots & \prod_{j=1}^n d_{k_j} \end{bmatrix},$$

$$\mathbf{Z}_X = \begin{bmatrix} 1 & x_n d_{1_n} & x_{n-1} d_{1_{(n-1)}} & x_n d_{1_n} x_{n-1} d_{1_{(n-1)}} & \cdots & \prod_{j=1}^n x_j d_{1_j} \\ 1 & x_n d_{2_n} & x_{n-1} d_{2_{(n-1)}} & x_n d_{2_n} x_{n-1} d_{2_{(n-1)}} & \cdots & \prod_{j=1}^n x_j d_{2_j} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n d_{k_n} & x_{n-1} d_{k_{(n-1)}} & x_n d_{k_n} x_{n-1} d_{k_{(n-1)}} & \cdots & \prod_{j=1}^n x_j d_{k_j} \end{bmatrix}.$$

From (3.18) the number of distinct combinations of σ_i^2 is the same as the number of the distinct columns of $\mathbf{Z} - \mathbf{Z}_X$.

Hence we proved the theorem. $\Box$

Using the equivalent conditions in Lemma 3.3 and 3.5 we are able to derive the following theorem using Theorem 3.5.

Theorem 3.6 *For balanced E-ANOVA models, the necessary and sufficient condition for MINQUE of $\theta = (\sigma_1^2, \ldots, \sigma_k^2)'$ to be independent of prior values is that $N(\mathbf{Z} - \mathbf{Z}_X) = k$.*

Now we demonstrate how to apply Theorem 3.6 in practice.

Example 3.6

Consider a one way random model:

$$y_{ij} = \mu + a_i + e_{ij} \quad i = 1, \ldots, I, \ j = 1, \ldots, J.$$

Now $k = 2$, $\mathbf{x} = (0, 0)$. $\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, $\mathbf{W}_X = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, and $\mathbf{Z}_X = 0$, then $N(\mathbf{Z} - \mathbf{Z}_X) = 2 = k$, thus the MINQUE of σ_a^2 and σ_e^2 are independent of prior values. $\Box$

This result confirms the result in Example 1.8 where Swallow and Searle calculated the MINQUEs of σ_a^2 and σ_e^2 explicitly and showed that they are independent of prior values.

Corollary 3.1 *For balanced nested ANOVA models, the MINQUE of σ_i^2's are independent of prior values.*

Proof: In this case

$$\mathbf{W} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix}_{k \times k}$$

$$\mathbf{W}_X = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

It is obvious that the generating of the columns of $\mathbf{W}$ will not creat new columns, thus $N(\mathbf{W} - \mathbf{W}_X) = k$. From theorem 3.6 we prove the corollary. $\Box$

Example 3.7

Consider two way crossed random model without interaction:

$$y_{ijl} = \mu + a_i + b_j + e_{ijl}$$

$$i = 1, \ldots, I,\ j = 1, \ldots, J,\ l = 1, \ldots, L.$$

Now $k = 3$

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix},\ \mathbf{W}_X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$N(\mathbf{Z} - \mathbf{Z}_X) = 3 = k.$

From theorem 3.6 the MINQUE of σ_a^2, σ_b^2 and σ_e^2 are independent of prior values. □

Corollary 3.2 *For balanced crossed ANOVA models without interaction, the MINQUE of σ_i^2's are independent of prior values.*

Proof:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix},\ \mathbf{W}_X = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

Therefore, $N(\mathbf{Z} - \mathbf{Z}_X) = k$, from theorem 3.6, we proved the corollary. □

Example 3.8

Two way crossed random model with interaction:

$$y_{ijl} = \mu + a_i + b_j + c_{ij} + e_{ijl}$$

$$i = 1, \ldots, I,\ j = 1, \ldots, J,\ l = 1, \ldots, L.$$

From Example 3.5 we know that $N(\mathbf{Z} - \mathbf{Z}_X) = 4 = k$, then from theorem 3.6 the MINQUE of σ_a^2, σ_b^2, σ_c^2 and σ_e^2 are independent of prior values. □

The result in this section shows that MINQUE for balanced ANOVA models are independent of prior values. From Theorem 2.4 we know MINQUE for balanced ANOVA models is MIVQUE. Therefore for general balanced ANOVA models we know that the MINQUE is identical to the ANOVA estimators. In this sense MINQUE is an extension of the ANOVA estimator to the unbalanced data cases.

Example 3.9

Interviewer's variance model (Biemer and Stokes, 1985). This model is used in Chapter 6 for interpenetrated interview data:

$$y_{\gamma tjs} = \eta_{\gamma t} + b_{\gamma j} + e_{\gamma tjs},$$

$$\gamma = 1, \ldots, l, \quad t, j = 1, \ldots, J, \quad s = 1, \ldots, f.$$

Now $k = 2$, $\mathbf{x} = (1, 1, 0, 0)$.

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \ \mathbf{W}_X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

It can be verified that $N(\mathbf{Z} - \mathbf{Z}_X) = 2 = k$, so the MINQUE of σ_b^2 and σ_e^2 are independent of prior values. □

In Chapter 6 the MINQUE formulas for σ_b^2 and σ_e^2 will be given and it can then be confirmed that the formulas are independent of the prior values of σ_b^2 and σ_e^2.

Example 3.10

Consider a balanced E-ANOVA model:

$$y_{ijl} = \mu_i + b_j + c_{ij} + e_{ijl},$$

$$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad l = 1, \ldots, L,$$

where μ_i is the fixed effect parameter, b_j, c_{ij} and e_{ijl} are random terms with variances σ_b^2, σ_c^2 and σ_e^2 respectively.

Now $k = 3$, $\mathbf{x} = (1, 0, 0)$

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{W}_X = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \ \mathbf{Z}_X = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus $N(\mathbf{Z} - \mathbf{Z}_X) = 3 = k$.

Therefore from Theorem 3.6 we know that the MINQUE of $(\sigma_b^2, \sigma_c^2, \sigma_e^2)$ are independent of the prior values. □

Discussion 3.1 *In Chapter 2 we have proved that all balanced E-ANOVA models satisfy condition (2.14). In this section we derived that for balanced E-ANOVA models* $N(\mathbf{Z} - \mathbf{Z}_X) = k$ *is the necessary and sufficient condition for the MINQUE to be independent of prior values. Therefore combining the conclusions in Chapter 2 and Chapter 3 we have the following conclusion: for any balanced E-ANOVA model if* $N(\mathbf{Z} - \mathbf{Z}_X) = k$*, then the MINQUE is the globally best quadratic unbiased estimator of the variance components; If* $N(\mathbf{Z} - \mathbf{Z}_X) \neq k$*, then the MINQUE is the locally best quadratic unbiased estimator.*

3.5 Conclusions

This chapter considered the problem of prior values in the computation of MINQUE. For the general variance components models Theorem 3.3 gives necessary and sufficient conditions for MINQUE to be independent of prior values .

For balanced E-ANOVA models the conditions can be greatly simplified. Using the simplified condition in theorem 3.6 it is proved that the MINQUE for all balanced ANOVA models are independent of prior values.

When MINQUE is independent of prior values synthesis-based MINQUE can be used to reduce the amount of computation and we are sure that the optimality of MINQUE is preserved by the estimate.

When MINQUE does need prior values the only sensible thing to do seems to be iterative computing. Further research is therefore needed to investigate the convergence property of MINQUE. As we pointed out in Section 1.3.4 the iterative MINQUE yields the same estimate as REML. Hence REML is an alternative approach if iterative MINQUE is to be used.

Chapter 4

NONNEGATIVE BIASED QUADRATIC ESTIMATORS

4.1 The problem of negative estimates

In section 1.3.1 we have given Example 1.1 to show that the ANOVA estimator sometimes gives negative estimates of variance components. By definition an estimate of a variance component should be always nonnegative. The maximum likelihood and restricted maximum likelihood approach do not produce negative estimates because the parameter space over which the maximum is sought is restricted to the positive part of each axis. In practice it is common that the solutions of the likelihood equations over the whole parameter space are taken to be the ML or REML estimates without restricting the parameter space. Herbach (1959) and Thompson (1962) investigated the properties of the maximum likelihood function and the restricted maximum likelihood function for the balanced one-way model and concluded that when a negative solution to the likelihood equations appears, zero should be used as the estimate for that variance component. There has been no such results for the general variance components model in the literature.

Chapter 2 and Chapter 3 have been devoted to MINQUE which is derived without the nonnegativity constraint. Ideally, we would like to use a quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ which has the following properties:

1. $\mathbf{y}'\mathbf{A}\mathbf{y}$ is invariant,
2. $\mathbf{y}'\mathbf{A}\mathbf{y}$ is unbiased,
3. $\mathbf{y}'\mathbf{A}\mathbf{y}$ is nonnegative,

4. The estimator has minimum variance,

4'. **AV** has minimum norm.

The MINQUE estimator satisfies 1, 2 and 4'. When the data come from a normal distribution, 4 and 4' are equivalent.

A question arises: Can we impose constraint 2 on the MINQUE matrix **A** so that we can obtain a nonnegative MINQUE which will satisfy 1-4? The answer is: we can not except in some special cases.

LaMotte (1973) used a balanced ANOVA model to demonstrate that the only variance component which can have an unbiased and nonnegative estimator is σ_e^2, the variance of the random error term. Generally, Pukelsheim (1981) showed that with a commutative quadratic subspace condition the concept of unbiasedness and nonnegativity are incompatible. In general we should not expect that a quadratic estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ will satisfy constraints 1-4. In other words, if we want nonnegative estimators of variance components, we have to drop the unbiasedness constraint.

When a biased estimator is considered, the optimality criterion often used in statistics is the mean squared error of the estimator. Suppose $\hat{\theta}$ is a biased estimator of θ, then the mean squared error of $\hat{\theta}$ is:

$$\mathrm{MSE}(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = \left[E(\hat{\theta})-\theta\right]^2 + E\left[\hat{\theta}-E(\hat{\theta})\right]^2 \tag{4.1}$$

where the term $\left[E(\hat{\theta})-\theta\right]^2$ is the bias of $\hat{\theta}$ and $E\left[\hat{\theta}-E(\hat{\theta})\right]^2$ is the variance of $\hat{\theta}$.

Sometimes it is impossible to obtain an estimator satisfying certain constraints and also minimizing the mean squared error. It is not known if there are any such estimators for variance components. In the literature there are different nonnegative quadratic estimators of variance components using different optimality criteria. I shall describe several existing approaches in the following sections.

4.2 Rao-Chaubey's MINQE

Rao and Chaubey (1978) proposed a modified MINQUE which is nonnegative and biased, hence the estimator is called MINQE —MINQUE without unbiasedness.

Recall all the notation and assumptions for model (1.1). In section 1.3.4 we have shown that when $\mathbf{y}'\mathbf{A}\mathbf{y}$ is an unbiased estimator of $\sum_{i=1}^k q_i\sigma_i^2$, then to minimize $\| \mathbf{W}'\mathbf{A}\mathbf{W} - \Lambda \|^2$ is equivalent to minimize $\mathrm{Tr}\,\mathbf{AVAV}$, where $\mathbf{W} =$

$[\sigma_1 \mathbf{U}_1, | \ldots, |\sigma_k \mathbf{U}_k]$, $\mathbf{V}_i = \mathbf{U}_i\mathbf{U}_i'$, $\mathbf{V} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i$, and

$$\Lambda = \begin{bmatrix} \frac{q_1}{p_1}\sigma_1^2\mathbf{I}_{p_1} & & & \\ & \frac{q_2}{p_2}\sigma_2^2\mathbf{I}_{p_2} & & \\ & & \ddots & \\ & & & \frac{q_k}{p_k}\sigma_k^2\mathbf{I}_{p_k} \end{bmatrix}.$$

When a normal distribution can be assumed for the data vector $\mathbf{y}$, $\mathrm{Tr}\,\mathbf{AVAV}$ is the variance of $\mathbf{y}'\mathbf{Ay}$, so with the unbiasedness constraint to minimize $\| \mathbf{W}'\mathbf{AW} - \Lambda \|^2$ is equivalent to minimize $V(\mathbf{y}'\mathbf{Ay})$. Rao and Chaubey's modification is that without the unbiasedness constraint the matrix $\mathbf{A}$ minimizing $\| \mathbf{W}'\mathbf{AW} - \Lambda \|^2$ is still used as the matrix to form the quadratic estimator. We now follow Rao and Chaubey to derive such a matrix and then comment on their approach.

Lemma 4.1 *Let* $\mathbf{X}$ *and* $\mathbf{C}$ *be* $n \times n$ *symmetric matrices, then* $\mathrm{Tr}\,\mathbf{XX} - 2\,\mathrm{Tr}\,\mathbf{XC}$ *is minimized if* $\mathbf{X}=\mathbf{C}$.

Proof: Let $\mathbf{X} = (x_{ij})_{n\times n}$, $\mathbf{C} = (c_{ij})_{n\times n}$, then

$$\mathrm{Tr}\,\mathbf{XX} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}^2 = \sum_{i=1}^{n} x_{ii}^2 + 2\sum_{i<j}^{n} x_{ij}^2,$$

$$\mathrm{Tr}\,\mathbf{XC} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}c_{ji} = \sum_{i=1}^{n} x_{ii}c_{ii} + 2\sum_{i<j}^{n} x_{ij}c_{ij}.$$

Let $L = \mathrm{Tr}\,\mathbf{XX} - 2\,\mathrm{Tr}\,\mathbf{XC}$, then

$$\frac{\partial L}{\partial x_{ii}} = 2x_{ii} - 2c_{ii} = 0,$$

gives $x_{ii} = c_{ii}$, $i = 1, \ldots, n$.

$$\frac{\partial L}{\partial x_{ij}} = 4x_{ij} - 4c_{ij} = 0,$$

gives $x_{ij} = c_{ij}$, $i < j$, $i = 1, \ldots, n$.

$$\frac{\partial^2 L}{\partial x_{ii}^2} = 2, \quad i = 1, \ldots, n,$$

$$\frac{\partial^2 L}{\partial x_{ij}^2} = 4, \quad i < j.$$

So L is minimized when $\mathbf{X} = \mathbf{C}$. $\square$

To derive the matrix which minimizes $\| \mathbf{W'AW} - \Lambda \|^2$, we need to express $\| \mathbf{W'AW} - \Lambda \|^2$ in another form so that Lemma 4.1 can be used. Now

$$\begin{aligned} \| \mathbf{W'AW} - \Lambda \|^2 &= \mathrm{Tr}(\mathbf{W'AW} - \Lambda)(\mathbf{W'AW} - \Lambda) \\ &= \mathrm{Tr}\,\mathbf{W'AWW'AW} - 2\,\mathrm{Tr}\,\mathbf{W'AW}\Lambda + \mathrm{Tr}\,\Lambda\Lambda \qquad (4.2) \end{aligned}$$

Note

$$\mathbf{WW'} = \sum_{i=1}^{k} \sigma_i^2 \mathbf{U}_i \mathbf{U}_i' = \sum_{i=1}^{k} \sigma_i^2 \mathbf{V}_i = \mathbf{V}.$$

Let $\mathbf{B} = \mathbf{V}^{1/2}\mathbf{A}\mathbf{V}^{1/2}$, thus $\mathbf{A} = \mathbf{V}^{-1/2}\mathbf{B}\mathbf{V}^{-1/2}$.

Let $\mathbf{Q} = \mathbf{I} - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1/2}$.

Since $\mathbf{AX} = 0$,

$$\begin{aligned} \mathbf{QBQ} &= \mathbf{QV}^{1/2}\mathbf{AV}^{1/2}\mathbf{Q} \\ &= \left\{\mathbf{V}^{1/2} - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'}\right\} \mathbf{A} \left\{\mathbf{V}^{1/2} - \mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1/2}\right\} \\ &= \mathbf{V}^{1/2}\mathbf{AV}^{1/2} = \mathbf{B}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathrm{Tr}\,\mathbf{W'AWW'AW} - 2\,\mathrm{Tr}\,\mathbf{W'AW}\Lambda \\ &= \mathrm{Tr}\,\mathbf{BB} - 2\,\mathrm{Tr}\,\mathbf{BQV}^{-1/2}\mathbf{W}\Lambda\mathbf{W'V}^{-1/2}\mathbf{Q}. \qquad (4.3) \end{aligned}$$

Also

$$\begin{aligned} \mathbf{QV}^{-1/2} &= \mathbf{V}^{-1/2} - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1} \\ &= \mathbf{V}^{1/2}\mathbf{R}, \end{aligned}$$

where $\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}$.

Similarly, $\mathbf{V}^{-1/2}\mathbf{Q} = \mathbf{RV}^{1/2}$, and

$$\begin{aligned} \mathbf{W}\Lambda\mathbf{W'} &= [\sigma_1\mathbf{U}_1, |\ldots, |\sigma_k\mathbf{U}_k] \begin{bmatrix} \frac{q_1}{p_1}\sigma_1^2\mathbf{I}_{p_1} & & \\ & \ddots & \\ & & \frac{q_k}{p_k}\sigma_k^2\mathbf{I}_{p_k} \end{bmatrix} \begin{bmatrix} \sigma_1\mathbf{U}_1' \\ \vdots \\ \sigma_k\mathbf{U}_k' \end{bmatrix} \\ &= \sum_{i=1}^{k} \frac{q_i}{p_i}\sigma_i^4\mathbf{U}_i\mathbf{U}_i' \\ &= \sum_{i=1}^{k} \frac{q_i}{p_i}\sigma_i^4\mathbf{V}_i. \end{aligned}$$

From (4.2) and (4.3)

$$\begin{aligned}\| \mathbf{W'AW} - \Lambda \|^2 &= \operatorname{Tr}\mathbf{BB} - 2\operatorname{Tr}\mathbf{BV}^{1/2}\mathbf{R}\left(\sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{V}_i\right)\mathbf{RV}^{1/2} + \operatorname{Tr}\Lambda\Lambda \\ &= \operatorname{Tr}\mathbf{BB} - 2\operatorname{Tr}\mathbf{BV}^{1/2}\left(\sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{RV}_i\mathbf{R}\right)\mathbf{V}^{1/2} + \operatorname{Tr}\Lambda\Lambda.\end{aligned} \tag{4.4}$$

Theorem 4.1 (Rao and Chaubey, 1978) *The invariant nonnegative quadratic estimator* $\mathbf{y'Ay}$ *of* $\sum_{i=1}^{k} q_i\sigma_i^2$ $q_i \geq 0,\ i = 1,\ldots,k,$ *which minimizes* $\| \mathbf{W'AW} - \Lambda \|^2$ *is given by*

$$\hat{\mathbf{A}} = \sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{RV}_i\mathbf{R}. \tag{4.5}$$

Note: Rao and Chaubey did not give a detailed proof for the above theorem.

Proof: Since $\mathbf{RX} = 0$, from the form of (4.5) we know $\hat{\mathbf{A}}\mathbf{X} = 0$, hence $\mathbf{y'}\hat{\mathbf{A}}\mathbf{y}$ is invariant.

Since $\frac{q_i}{p_i}\sigma_i^4 > 0,\ i = 1,\ldots,k,$

$$\begin{aligned}\mathbf{y'}\hat{\mathbf{A}}\mathbf{y} &= \sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{y'RV}_i\mathbf{Ry} \\ &= \sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4(\mathbf{U}_i'\mathbf{Ry})'(\mathbf{U}_i'\mathbf{Ry}) \\ &\geq 0.\end{aligned}$$

Hence $\mathbf{y'}\hat{\mathbf{A}}\mathbf{y}$ is nonnegative.

We then need to prove that $\hat{\mathbf{A}}$ minimizes $\| \mathbf{W'AW} - \Lambda \|^2$.

From (4.4) and using Lemma 4.1 we know that $\| \mathbf{W'AW} - \Lambda \|^2$ is minimized when

$$\hat{\mathbf{B}} = \mathbf{V}^{1/2}\left(\sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{RV}_i\mathbf{R}\right)\mathbf{V}^{1/2},$$

i.e.

when

$$\hat{\mathbf{A}} = \sum_{i=1}^{k}\frac{q_i}{p_i}\sigma_i^4\mathbf{RV}_i\mathbf{R}.$$

Hence we prove the theorem. □

In practice we do not know σ_i^2, so we have to choose prior values for the σ_i^2 in (4.5).

It is easy to see from (4.5) that Rao and Chaubey's MINQE has additivity. If $\mathbf{y}'\mathbf{A}_1\mathbf{y}$ is the MINQE of $\sum_{i=1}^{k} q_i\sigma_i^2$, and $\mathbf{y}'\mathbf{A}_2\mathbf{y}$ is the MINQE of $\sum_{i=1}^{k} t_i\sigma_i^2$, then $\mathbf{y}'(\mathbf{A}_1+\mathbf{A}_2)\mathbf{y}$ is the MINQE of $\sum_{i=1}^{k}(q_i+t_i)\sigma_i^2$.

Example 4.1:

Consider a balanced one way model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

where μ is the overall mean, a_i and e_{ij} are random effects with variance σ_a^2 and σ_e^2, respectively.

Let α_a and α_e be the prior values of σ_a^2 and σ_e^2, and $\hat{\gamma} = \alpha_a/\alpha_e$, then the MINQE of σ_a^2 and σ_e^2 are:

$$\hat{\sigma}_a^2(M) = \frac{n^2\hat{\gamma}^2}{m(1+n\hat{\gamma})^2}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2, \tag{4.6}$$

$$\hat{\sigma}_e^2(M) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 + \frac{1}{m(n\hat{\gamma}+1)^2}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2. \tag{4.7}$$

If we choose $\hat{\gamma} = 1$, then for the data set in Example 1.1, $m = 2$, $n = 3$,

$$\hat{\sigma}_a^2(M) = 0.5625, \quad \hat{\sigma}_e^2(M) = 34.75.$$

Notice that both $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$ are different from the ANOVA estimates given in Example 1.1. □

MINQE is built with $\| \mathbf{W}'\mathbf{A}\mathbf{W} - \Lambda \|^2$ as its optimality criterion. Without the unbiasedness constraint it is not known how $\| \mathbf{W}'\mathbf{A}\mathbf{W} - \Lambda \|^2$ relates to the usual statistical measures of optimality, namely, bias, variance and mean squared error. The following example demonstrates that MINQE does not always possess statistical optimality.

Example 4.2:

Consider the one way balanced model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots .n.$$

It is proved in section 2.2 that the ANOVA estimator $\hat{\sigma}_e^2$ is the best quadratic unbiased estimator for the above model:

$$\hat{\sigma}_e^2 = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2, \tag{4.8}$$

$\hat{\sigma}_e^2$ is nonnegative and unbiased. For normally distributed data

$$V(\hat{\sigma_e^2}) = \frac{2}{m(n-1)}\sigma_e^4. \tag{4.9}$$

From (4.7) the MINQE of σ_e^2 is:

$$\tilde{\sigma}_e^2(M) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + \frac{\alpha_e^2}{m(\alpha_e+n\alpha_a)^2}\sum_{i=1}^{m}(\bar{y}_{i.}-\bar{\bar{y}}_{..})^2, \tag{4.10}$$

$$\text{bias}\left(\tilde{\sigma}_e^2(M)\right) = \left\{\frac{(m-1)\alpha_e^2(n\sigma_a^2+\sigma_e^2)}{mn(n\alpha_a+\alpha_e)^2} - \frac{\sigma_e^2}{n}\right\}^2. \tag{4.11}$$

With normality assumption,

$$\mathrm{V}\left(\tilde{\sigma}_e^2(M)\right) = \frac{2}{mn}\left\{\frac{(n-1)\sigma_e^4}{n} + \frac{(m-1)\alpha_e^4(n\sigma_a^2+\sigma_e^2)^2}{mn(n\alpha_a+\alpha_e)^4}\right\}. \tag{4.12}$$

Now assume $m=10$, $n=8$, at $\sigma_a^2=\sigma_e^2=1$, then

$$\mathrm{MSE}(\hat{\sigma}_e^2) = 0.02857.$$

For MINQE, it is usual to choose the prior values to be: $\alpha_a=\alpha_e=1$. then:

$$\text{bias}\left(\tilde{\sigma}_e^2(M)\right) = 0.01266, \ \mathrm{V}\left(\tilde{\sigma}_e^2(M)\right) = 0.02191.$$

Therefore, $\mathrm{MSE}\left(\tilde{\sigma}_e^2(M)\right) = 0.03457 > \mathrm{MSE}(\hat{\sigma}_e^2)$.

We conclude that when mean squared error is used as the measure of optimality for the estimators, MINQE does not give the best possible nonnegative quadratic estimator. □

There are two unsatisfactory facts about MINQE. First, its optimality criterion is not a well recognized statistical measure of optimality. There is a need to compare MINQE with the other nonnegative estimators on a common statistical measure, *e.g.* mean squared error. Second, if a nonnegative unbiased estimator with minimum variance exists MINQE may not coincide with it.

4.3 Chaubey's CMINQUE

Chaubey (1983) proposed a nonnegative quadratic estimator of variance components which is called CMINQUE (the estimator closest to MINQUE).

Theorem 4.2 (Chaubey, 1983) *Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of the MINQUE matrix* $\mathbf{A}$ *and* $\mathbf{p}_1, \ldots, \mathbf{p}_N$ *be the corresponding normalised orthogonal eigenvectors, then* $\hat{\mathbf{D}} = \sum_{\lambda_i > 0} \lambda_i \mathbf{p}_i \mathbf{p}_i'$ *is the nonnegative matrix minimizing* $\| \mathbf{D} - \mathbf{A} \|_E^2$, *where* $\| \ \|_E^2$ *is the Euclidean norm of a matrix.*

Proof of this theorem can be found in Rao (1973), *p*63.

Chaubey called the nonnegative estimator $\mathbf{y}'\hat{\mathbf{D}}\mathbf{y}$ CMINQUE which can be regarded as the truncated MINQUE using the spectral decomposition. It is easy to see from theorem 4.2 that CMINQUE has additivity.

One explanation of CMINQUE's optimality criterion is the following:

$$\begin{aligned}
\text{bias}(\mathbf{y}'\mathbf{D}\mathbf{y}) &= [\mathrm{Tr}(\mathbf{DV} - \mathbf{AV})]^2 = [\mathrm{Tr}(\mathbf{D} - \mathbf{A})\mathbf{V}]^2 \\
&= \left[\sum_{i,j} (d_{ij} - a_{ij}) v_{ij} \right]^2 \\
&\leq \left(\sum_{i,j} (d_{ij} - a_{ij})^2 \right) \left(\sum_{i,j} v_{ij}^2 \right). \qquad (4.13)
\end{aligned}$$

Therefore the approach of minimizing $\| \mathbf{D} - \mathbf{A} \|^2$ has the effect of minimizing the bias range in (4.13), hence in some sense makes the bias of the estimator small.

Now we discuss a situation where the MINQUE matrix $\mathbf{A}$ and the variance covariance matrix $\mathbf{V}$ have a common set of normalised orthogonal eigenvectors, say, $\mathbf{p}_1, \ldots, \mathbf{p}_N$. From lemma 2.7, we know that balanced ANOVA models satisfy this assumption.

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of MINQUE matrix $\mathbf{A}$, $\tau_1, \ldots, \tau_N$ be those of $\mathbf{V}$.

Since $\mathbf{A} = \sum_{i=1}^{N} \lambda_i \mathbf{p}_i \mathbf{p}_i'$,

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= \mathrm{Tr}\,\mathbf{AV} = \sum_{i=1}^{N} \lambda_i \,\mathrm{Tr}\, \mathbf{p}_i \mathbf{p}_i' \mathbf{V} \\
&= \sum_{i=1}^{N} \lambda_i \mathbf{p}_i' \mathbf{V} \mathbf{p}_i = \sum_{i=1}^{N} \lambda_i \tau_i. \qquad (4.14)
\end{aligned}$$

Now $\mathbf{D} = \sum_{\lambda_i > 0} \lambda_i \mathbf{p}_i \mathbf{p}_i'$,

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\mathbf{D}\mathbf{y}) &= \mathrm{Tr}\,\mathbf{DV} = \sum_{\lambda_i > 0} \lambda_i \,\mathrm{Tr}\, \mathbf{p}_i \mathbf{p}_i' \mathbf{V} \\
&= \sum_{\lambda_i > 0} \lambda_i \mathbf{p}_i' \mathbf{V} \mathbf{p}_i = \sum_{\lambda_i > 0} \lambda_i \tau_i. \qquad (4.15)
\end{aligned}$$

Since $\mathbf{V}$ is a nonnegative matrix, thus $\tau_i \geq 0$, $i = 1, \ldots, N$.

From (4.14) and (4.15), we can see that:

$$\mathrm{E}(\mathbf{y}'\mathbf{D}\mathbf{y}) \geq \mathrm{E}(\mathbf{y}'\mathbf{A}\mathbf{y}).$$

This means that the CMINQUE gives estimates not less than the true values of the variance components. In other words, CMINQUE always has nonnegative bias. If a normal distribution is assumed for the data vector $\mathbf{y}$, then:

$$\begin{aligned} \mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}) &= 2\,\mathrm{Tr}\,\mathbf{AVAV} \\ &= 2\,\mathrm{Tr}\left(\sum_{i=1}^{N} \lambda_i \mathbf{p}_i \mathbf{p}_i' \mathbf{V}\right)\left(\sum_{j=1}^{N} \lambda_j \mathbf{p}_j \mathbf{p}_j' \mathbf{V}\right) \\ &= 2\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j (\mathbf{p}_i' \mathbf{V} \mathbf{p}_j)(\mathbf{p}_j' \mathbf{V} \mathbf{p}_i) \\ &= 2\sum_{i=1}^{N} \lambda_i^2 \tau_i^2. \end{aligned} \tag{4.16}$$

Similarly,

$$\mathrm{V}(\mathbf{y}'\mathbf{D}\mathbf{y}) = 2\,\mathrm{Tr}\,\mathbf{DVDV} = 2\sum_{\lambda_i > 0} \lambda_i^2 \tau_i^2. \tag{4.17}$$

From (4.16) and (4.17) it is obvious that:

$$\mathrm{V}(\mathbf{y}'\mathbf{D}\mathbf{y}) \leq \mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}).$$

This means that CMINQUE has less variance than MINQUE. But since CMINQUE is biased, its mean squared error may be greater than that of MINQUE.

CMINQUE uses $\| \mathbf{D} - \mathbf{A} \|^2$ as its optimality criterion which is not a commonly used statistical measure. When there exists a nonnegative unbiased MINQUE estimator CMINQUE will coincide with it. Apart from the unclear role of the optimality criterion it used, CMINQUE does not always exist. The existence of CMINQUE depends on the existence of MINQUE.

Example 4.3:

Consider the one way balanced model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n.$$

CMINQUE of σ_e^2 is identical to the ANOVA estimator, and the CMINQUE of σ_a^2 is:

$$\hat{\sigma}_a^2(C) = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2. \tag{4.18}$$

For the data set given in Example 1.1,
$\hat{\sigma}_a^2(C) = 2$, $\hat{\sigma}_e^2(C) = 52$. $\square$.

4.4 The nonexistence of a globally minimum biased nonnegative estimator

In section 4.1 it is shown that nonnegativity and unbiasedness are incompatible. In section 4.2 and 4.3 two nonnegative estimators are introduced. They are modified forms of MINQUE. When comparing estimators, we ought to use a common optimality measure to assess the performance of the estimators. For biased estimators the most commonly used optimality criterion is the mean squared error. An estimator with minimum mean squared error will be the best choice. However, in the problem of quadratic estimation of variance components, so far the attempt to obtain an estimator with minimum mean squared error has failed and we are left with two choices: (1) use some other measures for optimality, like the CMINQUE in section 4.3, and hope by doing so the mean squared error is small; (2) use bias and variance as separate optimality criteria.

The first choice leaves us with arbitrary numbers of possible optimality measures and no control on the mean squared error. It seems more sensible to use the second choice and by controlling bias and variance of the estimator we gain control over the mean squared error.

In the search for nonnegative quadratic estimators of variance components, neither bias nor variance should be used alone as the optimality criterion, because bias or variance alone cannot form enough constraints for an unique estimator and there is also the risk of leaving one term in the mean squared error uncontrolled while minimizing the other term.

To obtain a quadratic estimator of $\sum_{i=1}^{k} q_i\sigma_i^2$, some possible constraints are:

1. invariance, $\mathbf{AX} = 0$,
2. nonnegativity, $\mathbf{x}'\mathbf{Ax} \geq 0$, $\mathbf{x} \neq 0$,
3. unbiasedness, $\sum_{i=1}^{k} \sigma_i^2 \operatorname{Tr}\mathbf{AV} = \sum_{i=1}^{k} q_i\sigma_i^2$,
4. minimum bias,

$$\left\{\sum_{i=1}^{k} \sigma_i^2(\operatorname{Tr}\mathbf{AV} - q_i)\right\}^2 \quad \text{is minimized,} \tag{4.19}$$

5. minimum variance, $V(\mathbf{y}'\mathbf{A}\mathbf{y})$ is minimized.

Different estimators are derived by using different combination of constraints in different orders. For example, MINQUE is derived using (1,3,5) with normality assumption. It is known that (1,2,3,5) are too many contraints to obtain a matrix solution in general, but (1,2,4) and (1,2,5) are too few constraints to obtain an unique matrix solution. By giving up unbiasedness we can try to use (1,2,4,5) as the set of constraints to derive a nonnegative estimator with some optimality.

Before we start the search for an estimator with constraints (1,2,4,5), we prove a theorem which says that if an estimator satisfying constraints 2 and 4 exists, it can only be a locally minimum biased estimator.

Theorem 4.3 *Consider the general variance components model (1.1). If there does not exist a nonnegative unbiased quadratic estimator for $\sum_{i=1}^{k} q_i\sigma_i^2$, where $q_i \geq 0$, $i = 1, \ldots, k$, then there does not exist a nonnegative biased quadratic estimator which achieves global minimum bias across the parameter space.*

Proof: Let $\mathbf{A}$ be a symmetric matrix, and let PSD denote the set of all nonnegative symmetric matrices.

We know that:

$$\text{bias}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \left[E(\mathbf{y}'\mathbf{A}\mathbf{y}) - \sum_{i=1}^{k} q_i\sigma_i^2\right]^2 = \left[\sum_{i=1}^{k} \sigma_i^2(\operatorname{Tr}\mathbf{A}\mathbf{V}_i - q_i)\right]^2.$$

Assume that there exists a nonnegative matrix $\hat{\mathbf{A}}$ such that

$$\text{bias}(\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}) = \min_{\mathbf{A}\in PSD} \text{bias}(\mathbf{y}'\mathbf{A}\mathbf{y}), \quad \text{at all parameter values.} \tag{4.20}$$

(4.20) is equivalent to:

$$\left[\sum_{i=1}^{k} \sigma_i^2(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)\right]^2 = \min_{\mathbf{A}\in PSD} \left[\sum_{i=1}^{k} \sigma_i^2(\operatorname{Tr}\mathbf{A}\mathbf{V}_i - q_i)\right]^2, \tag{4.21}$$

$$\sigma_i^2 \geq 0, \; i = 1, \ldots, k.$$

We shall find a particular set of values for the σ_i^2, $i = 1, \ldots, k$, such that (4.21) does not hold, thus proving the theorem.

Since there does not exist a nonnegative unbiased quadratic estimator of $\sum_{i=1}^{k} q_i\sigma_i^2$, the equations

$$\operatorname{Tr}\mathbf{A}\mathbf{V}_i = q_i, \quad i = 1, \ldots, k,$$

do not have a solution in PSD. In other words there exists at least one q_j, such that $\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j \neq q_j$.

Since $\mathbf{V}_i = \mathbf{U}_i\mathbf{U}_i'$, then $\mathrm{Tr}\,\mathbf{V}_i > 0$, $i = 1, \ldots, k$.

Let $\mathbf{A}^0 = \frac{1}{2}\hat{\mathbf{A}} + \frac{q_j}{2\,\mathrm{Tr}\,\mathbf{V}_j}\mathbf{I}$, hence $\mathbf{A}^0 \neq \hat{\mathbf{A}}$.

For any $\mathbf{x} \neq 0$,

$$\mathbf{x}'\mathbf{A}^0\mathbf{x} = \frac{1}{2}\mathbf{x}'\hat{\mathbf{A}}\mathbf{x} + \frac{q_j}{2\,\mathrm{Tr}\,\mathbf{V}_j}\mathbf{x}'\mathbf{x} \geq 0,$$

So $\mathbf{A}^0 \in PSD$.

$$\begin{aligned} \mathrm{Tr}\,\mathbf{A}^0\mathbf{V}_j &= \frac{1}{2}\,\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j + \frac{q_j}{2\,\mathrm{Tr}\,\mathbf{V}_j}\,\mathrm{Tr}\,\mathbf{V}_j \\ &= \frac{1}{2}\,\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j + \frac{1}{2}q_j, \end{aligned}$$

then

$$\mathrm{Tr}\,\mathbf{A}^0\mathbf{V}_j - q_j = \frac{1}{2}(\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j - q_j).$$

Now at $\sigma_j^{2^0} = 1$ and $\sigma_i^{2^0} = 0$, for $i \neq j$,

$$\begin{aligned} \mathrm{bias}(\mathbf{y}'\mathbf{A}^0\mathbf{y}) &= \left[\sum_{i=1}^{k}\sigma_i^{2^0}(\mathrm{Tr}\,\mathbf{A}^0\mathbf{V}_i - q_i)\right]^2 \\ &= \frac{1}{4}(\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j - q_j)^2 \\ &< (\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_j - q_j)^2 \\ &= \left[\sum_{i=1}^{k}\sigma_i^{2^0}(\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_i - q_i)\right]^2 \\ &= \mathrm{bias}(\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}). \end{aligned}$$

This is contradictory to the assumption in (4.20). Therefore we proved the theorem. □

To summarize, when deriving a quadratic estimator of the variance components, we either have a nonnegative unbiased estimator or a locally nonnegative minimum biased estimator.

4.5 Hartung's estimator

4.5.1 The estimator

Hartung (1981) attempted to look for minimum biased and minimum variance estimators of variance components and he called the estimator "nonnegative minimum biased invariant estimator". From theorem 4.3, we know that no nonnegative estimator can achieve minimum bias globally across the parameter space, therefore, in this thesis the estimator proposed by Hartung is referred as "Hartung's estimator".

Hartung's approach is first to look for the class of symmetric matrices which satisfy constraints 1,2 and also minimize $\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$, and among this class of matrices choose the unique matrix $\mathbf{A}$ which minimizes $\| \mathbf{A} \|_E^2$.

To impose invariance on the estimator, Hartung used

$$\mathbf{W}_i = (\mathbf{I} - \mathbf{XX}^+)\mathbf{V}_i(\mathbf{I} - \mathbf{XX}^+),$$

where $\mathbf{X}^+$ is the Moore-Penrose generalized inverse of $\mathbf{X}$. So instead of working with $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \sum_{i=1}^{k}\sigma_i^2\mathbf{V}_i)$, we are working with $\mathbf{y} \sim (0, \sum_{i=1}^{k}\sigma_i^2\mathbf{W}_i)$.

After the transformation of matrix $\mathbf{V}_i$ into $\mathbf{W}_i$, Hartung's constraints on the estimator are:

1. $\mathbf{A}$ is nonnegative,
2. $\mathbf{A}$ minimizes $\left[\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{AW}_i - q_i)\right]^2$,
3. $\mathbf{A}$ minimizes $\| \mathbf{A} \|^2$.

Any matrix $\mathbf{A}$ satisfying the above constraints with the order 1,2 and 3 is the matrix defined as Hartung's estimator. In other words, Hartung's estimator minimizes the measure in 2 with regards to all matrices satisfying 1 and then minimizes the measure in 3 for all matrices satisfying 1 and 2. The order itself is an important constraint because other orders of the same constraints may lead to different estimators. For instance, if we use constraint 3 to follow constraint 1, we shall have a null matrix which is not of interest. We use $\hat{\mathbf{A}}$ to denote a matrix satisfying the constraints in the order 1, 2, 3 and $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$ is then Hartung's estimator.

Theorem 4.4 (Hartung, 1981) $\hat{\mathbf{A}}$ *always exists and is uniquely determined.*

The proof can be found in Hartung (1981). The proof used the properties of a closed convex cone in functional analysis, and Hartung went on to give his estimator an analytical expression.

We need some notations to follow Hartung's development. Let $\mathbf{A} = (a_{ij})_{n\times n}$, then $\mathbf{A}_+ = (a_{ij}^+)_{n\times n}$, where

$$a_{ij}^+ = \begin{cases} a_{ij} & \text{if } a_{ij} \geq 0 \\ 0 & \text{if } a_{ij} < 0. \end{cases}$$

Theorem 4.5 (Hartung, 1981) *Let $\{r_n\}$ be a positive nullsequence, i.e. $r_n \to \infty$, when $n \to 0$. If $\mathbf{A}_n$ is the solution to the following equation:*

$$\mathbf{A}_n - (\mathbf{A}_n)_+ + r_n^2 \sum_{i=1}^{k} (\mathrm{Tr}\,\mathbf{A}_n\mathbf{W}_i)\mathbf{W}_i + r_n^3\mathbf{A}_n = r_n^2 \sum_{i=1}^{k} q_i\mathbf{W}_i, \tag{4.22}$$

then Hartung's estimation matrix is $\hat{\mathbf{A}} = \lim_{n\to\infty} \mathbf{A}_n$, and Hartung's estimator is $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$.

The proof is in Hartung (1981).

There are two difficulties in deriving Hartung's estimator for the general variance components model (1.1). The first is that the estimator depends on the solution to (4.22) and (4.22) is not such a 'regular' function of $\mathbf{A}_n$ that explicit analytical solution of $\mathbf{A}_n$ can be found. The second difficulty is that the matrix used in Hartung's estimator is a limit of a sequence of matrices and may be unobtainable in practice.

For a special class of models these two difficulties do not exist, because Hartung gives an explicit expression for his estimator in this case.

Some definitions and lemmas are needed before presenting Hartung's theorem.

Definition 4.1 *Let $\mathbf{W}_1, \ldots, \mathbf{W}_k$ be symmetric matrices, and let Σ be the subspace containing all the linear combinations of $\mathbf{W}_1, \ldots, \mathbf{W}_k$, denoted as:*

$$\Sigma = span\,[\mathbf{W}_1, \ldots, \mathbf{W}_k].$$

If for $\mathbf{W}_i, \mathbf{W}_j \in \Sigma$, implies $\mathbf{W}_i^2 \in \Sigma$ and $\mathbf{W}_i\mathbf{W}_j = \mathbf{W}_j\mathbf{W}_i$, $i \neq j$, $i,j = 1, \ldots, k$, then Σ is a commutative quadratic subspace.

Lemma 4.2 (Seely, 1971) *A necessary and sufficient condition for Σ to be a k-dimensional commutative quadratic subspace is that there exist k pairwise orthogonal projection matrices $\mathbf{P}_i, \ldots, \mathbf{P}_k$ to form a basis of Σ.*

If the basis of the subspace is given, then we can express the matrices $\mathbf{W}_1$, …,$\mathbf{W}_k$ in terms of $\mathbf{P}_1$, …, $\mathbf{P}_k$. Let $\mathbf{W}_i = \sum_{j=1}^{k} \phi_{ij}\mathbf{P}_j$, $i = 1, \ldots, k$, then there exists a nonsingular matrix $\Phi = (\phi_{ij})$, $i, j = 1, \ldots, k$, such that $\mathbf{W}_1, \ldots, \mathbf{W}_k$ can be determined by Φ and $\mathbf{P}_1, \ldots, \mathbf{P}_k$.

When Σ is a commutative quadratic subspace, Hartung derived an explicit estimator and the result is given in the following theorem.

Theorem 4.6 (Hartung, 1981) *If* $\mathbf{W}_1, \ldots, \mathbf{W}_k$ *forms a commutative quadratic subspace and* $\mathbf{P}_1, \ldots, \mathbf{P}_k$ *is the basis for the subspace, then Hartung's estimator of* $\sum_{i=1}^{k} q_i \sigma_i^2$ *is given by:*

$$\mathbf{y}'\hat{\mathbf{A}}\mathbf{y} = \sum_{i=1}^{k} \hat{d}_i \parallel \mathbf{P}_i \mathbf{y} \parallel^2 / \operatorname{Tr} \mathbf{P}_i,$$

where $\parallel \; \parallel^2$ *denotes Euclidean norm of a matrix, and*

$$\hat{\mathbf{A}} = \sum_{i=1}^{k} \hat{d}_i (\operatorname{Tr} \mathbf{P}_i)^{-1} \mathbf{P}_i, \tag{4.23}$$

where $\hat{\mathbf{d}} = (\hat{d}_1, \ldots, \hat{d}_k)'$ *is the unique solution to the following system:*

$$\textit{minimize } (\Phi \mathbf{d} - \mathbf{q})'(\Phi \mathbf{d} - \mathbf{q}), \quad \textit{subject to}: \tag{4.24}$$

$$\mathbf{d} \in R_+^k, \tag{4.25}$$

$$\Phi'(\Phi \mathbf{d} - \mathbf{q}) \in R_+^k, \tag{4.26}$$

$$\mathbf{d}'\Phi'(\Phi \mathbf{d} - \mathbf{q}) = 0, \tag{4.27}$$

where $\mathbf{q} = (q_1, \ldots, q_k)'$.

The proof can be found in Hartung (1981).

Although Theorem 4.6 gives explicit expressions for Hartung's estimator for some models, the estimator depends on the solution $\mathbf{d}$ from (4.24)—(4.27). In the following section we shall derive Hartung's estimator for balanced nested ANOVA models.

4.5.2 Explicit formulae for Hartung's estimator for balanced nested ANOVA models

In this section we shall prove that $\mathbf{W}_1, \ldots, \mathbf{W}_k$ from balanced nested ANOVA models form a commutative quadratic subspace and we shall derive the solution to the system (4.24)—(4.27). Hence using Theorem 4.6 we can obtain explicit formulae for Hartung's estimator.

A balanced nested ANOVA model can be written as:

$$y_{i_1 i_2 \ldots i_k} = \mu + \xi_{i_1} + \xi_{i_1 i_2} + \cdots + \xi_{i_1 i_2 \ldots i_k}, \tag{4.28}$$

where μ is the overall mean and ξ_{i_1}, $\xi_{i_1 i_2}$, ..., $\xi_{i_1 i_2 \ldots, i_k}$ are the random terms with variances σ_1^2, σ_2^2, ..., σ_k^2, respectively. We assume that each random factor is at s_i levels, $s_i > 1$, $i = 1, \ldots, k$.

We introduce some notations needed in the following derivation of estimators. Let $s = \prod_{i=1}^{k} s_i$, $\tau(i) = s_1 \ldots s_i$, thus $\frac{s}{\tau(i)} = s_{i+1} \ldots s_k$. s is the product of all numbers of level and $\tau(i)$ is the partial product of the numbers of level up to level i.

The corresponding design matrices for model (4.28) are $\mathbf{X} = \mathbf{1}_s$, $\mathbf{U}_i = \mathbf{I}_{s_1} \otimes \cdots \otimes \mathbf{I}_{s_i} \otimes \mathbf{1}_{s_{i+1}} \otimes \cdots \otimes \mathbf{1}_{s_k}$. For simplicity, we use subscript i instead of s_i to denote a matrix of order s_i, *e.g.* $\mathbf{I}_i$ is the identity matrix of order s_i, hence $\mathbf{V}_i = \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_i \otimes \mathbf{J}_{i+1} \otimes \cdots \otimes \mathbf{J}_k$.

It can be seen that $\mathbf{X}^+ = \frac{1}{s}\mathbf{1}_s$, $\mathbf{I} - \mathbf{X}\mathbf{X}^+ = \mathbf{I}_s - \frac{1}{s}\mathbf{J}_s$, therefore,

$$
\begin{aligned}
\mathbf{W}_i &= (\mathbf{I} - \mathbf{X}\mathbf{X}^+)\mathbf{V}_i(\mathbf{I} - \mathbf{X}\mathbf{X}^+) \\
&= \mathbf{V}_i - \frac{1}{s}\mathbf{J}_s\mathbf{V}_i \\
&= \mathbf{V}_i - \frac{1}{\tau(i)}\mathbf{J}_s, \quad i = 1, \ldots, k. \qquad (4.29)
\end{aligned}
$$

Now we look for a basis for the subspace spanned by $\mathbf{W}_1, \ldots, \mathbf{W}_k$ in (4.29). Let $\bar{\mathbf{J}}_i = \frac{1}{s_i}\mathbf{J}_i$, $\mathbf{K}_i = \mathbf{I}_i - \bar{\mathbf{J}}_i$, and let

$$
\mathbf{P}_i = \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_{i-1} \otimes \mathbf{K}_i \otimes \bar{\mathbf{J}}_{i+1} \otimes \cdots \otimes \bar{\mathbf{J}}_k, \quad i = 1, \ldots, k. \qquad (4.30)
$$

Since

$$
\mathbf{K}_i\bar{\mathbf{J}}_i = (\mathbf{I}_i - \frac{1}{s_i}\mathbf{J}_i)\frac{1}{s_i}\mathbf{J}_i = 0,
$$

it follows that $\mathbf{P}_1, \ldots, \mathbf{P}_k$ are pairwise orthogonal matrices. Also note that $\mathbf{P}_i^2 = \mathbf{P}_i$, $i = 1, \ldots, k$, so that $\mathbf{P}_1, \ldots, \mathbf{P}_k$ are projection matrices.

Now we prove that $\mathbf{P}_1, \ldots, \mathbf{P}_k$ form the basis of the subspace spanned from $\mathbf{W}_1, \ldots, \mathbf{W}_k$.

Lemma 4.3 *Let* $\mathbf{P}_1, \ldots, \mathbf{P}_k$ *be defined as in* (4.30), *the following equality holds for* $k \geq 2$:

$$
\begin{aligned}
\mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_i \otimes \mathbf{J}_{i+1} \otimes \cdots \otimes \mathbf{J}_k - \frac{1}{\tau(i)}\mathbf{J}_s &= \frac{s}{\tau(i)}\{\mathbf{P}_1 + \cdots + \mathbf{P}_i\} \\
& \quad i = 1, \ldots, k. \qquad (4.31)
\end{aligned}
$$

Proof: We use $\mathbf{P}_i^{(k)}$ for the ith matrix defined in (4.30) with value k and $s^{(k)} = s_1 \dots s_k$.

We use mathematical induction to prove the lemma.

First, assume $k = 2$.

$$\mathbf{P}_1^{(2)} = \mathbf{K}_1 \otimes \bar{\mathbf{J}}_2, \quad \mathbf{P}_2^{(2)} = \mathbf{I}_1 \otimes \mathbf{K}_2.$$

For $i = 1$,

$$\begin{aligned}
\text{the left hand side of (4.31)} &= \mathbf{I}_1 \otimes \mathbf{J}_2 - \frac{1}{s_1}\mathbf{J}_1 \otimes \mathbf{J}_2 \\
&= (\mathbf{I}_1 - \frac{1}{s_1}\mathbf{J}_1) \otimes \mathbf{J}_2 \\
&= s_2 \mathbf{K}_1 \otimes \bar{\mathbf{J}}_2 \\
&= \frac{s}{\tau(1)}\mathbf{P}_1.
\end{aligned}$$

For $i = 2$,

$$\begin{aligned}
\text{the left hand side of (4.31)} &= \mathbf{I}_1 \otimes \mathbf{I}_2 - \frac{1}{s_1 s_2}\mathbf{J}_1 \otimes \mathbf{J}_2 \\
&= \mathbf{I}_1 \otimes \mathbf{K}_2 + \mathbf{I}_1 \otimes \bar{\mathbf{J}}_2 - \frac{1}{s_1 s_2}\mathbf{J}_1 \otimes \mathbf{J}_2 \\
&= \mathbf{I}_1 \otimes \mathbf{K}_2 + \mathbf{K}_1 \otimes \bar{\mathbf{J}}_2 \\
&= \mathbf{P}_1 + \mathbf{P}_2 \\
&= \frac{s}{\tau(2)}\left\{\mathbf{P}_1 + \mathbf{P}_2\right\}.
\end{aligned}$$

We have shown that (4.31) holds for $k = 2$.

Assume that when $k = n$, (4.31) holds.

Now let $k = n + 1$.

From (4.30) we have

$$\mathbf{P}_i^{(n+1)} = \mathbf{P}_i^{(n)} \otimes \bar{\mathbf{J}}_{n+1}, \quad i = 1, \dots, n,$$

$$\mathbf{P}_{n+1}^{(n+1)} = \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_n \otimes \mathbf{K}_{n+1}.$$

For $i = 1, \dots, n$,

the left hand side of (4.31)

$$= \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_i \otimes \mathbf{J}_{i+1} \otimes \cdots \otimes \mathbf{J}_{n+1} - \frac{1}{\tau(i)}\mathbf{J}_s$$

$$= s_{n+1}\left\{\mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_i \otimes \mathbf{J}_{i+1} \otimes \cdots \otimes \mathbf{J}_n - \frac{1}{\tau(i)}\mathbf{J}_1 \otimes \cdots \otimes \mathbf{J}_n\right\} \otimes \bar{\mathbf{J}}_{n+1}$$

$$= s_{n+1}\frac{s^{(n)}}{\tau(i)}\left\{\mathbf{P}_1^{(n)}+\cdots+\mathbf{P}_i^{(n)}\right\}\otimes\bar{\mathbf{J}}_{n+1} \text{ (using assumption on k = n)}$$

$$= \frac{s^{(n+1)}}{\tau(i)}\left\{\mathbf{P}_1^{(n+1)}+\cdots+\mathbf{P}_i^{(n+1)}\right\}.$$

So (4.31) holds for $i = 1, \ldots, n$.

For $i = n+1$,

Since $\mathbf{I}_{n+1} = \mathbf{K}_{n+1} + \bar{\mathbf{J}}_{n+1}$, $\tau(n+1) = \tau(n)s_{n+1}$,

the left hand side of (4.31)

$$= \mathbf{I}_1\otimes\cdots\otimes\mathbf{I}_n\otimes\mathbf{I}_{n+1} - \frac{1}{\tau(n+1)}\mathbf{J}_1\otimes\cdots\otimes\mathbf{J}_n\otimes\mathbf{J}_{n+1}$$

$$= \mathbf{I}_1\otimes\cdots\otimes\mathbf{I}_n\otimes\mathbf{K}_{n+1} + \mathbf{I}_1\otimes\cdots\otimes\mathbf{I}_n\otimes\bar{\mathbf{J}}_{n+1} - \frac{1}{\tau(n+1)}\mathbf{J}_1\otimes\cdots\otimes\mathbf{J}_n\otimes\mathbf{J}_{n+1}$$

$$= \mathbf{P}_{n+1}^{(n+1)} + \left\{\mathbf{I}_1\otimes\cdots\otimes\mathbf{I}_n - \frac{1}{\tau(n)}\mathbf{J}_1\otimes\cdots\otimes\mathbf{J}_n\right\}\otimes\bar{\mathbf{J}}_{n+1}$$

$$= \mathbf{P}_{n+1}^{(n+1)} + \left\{\mathbf{P}_1^{(n)}+\cdots+\mathbf{P}_n^{(n)}\right\}\otimes\bar{\mathbf{J}}_{n+1}$$

$$= \mathbf{P}_1^{(n+1)}+\cdots+\mathbf{P}_n^{(n+1)}+\mathbf{P}_{n+1}^{(n+1)}.$$

Therefore, for $k = n+1$, (4.31) holds.

Hence we proved the theorem. □

Lemma 4.4 *For balanced nested ANOVA models, the subspace spanned by* $\mathbf{W}_1, \ldots, \mathbf{W}_k$ *is a commutative quadratic subspace.*

Proof: From lemma 4.2 we need to show that $\mathbf{P}_1, \ldots, \mathbf{P}_k$ forms a basis for the subspace spanned by $\mathbf{W}_i$. In other words, there exists a nonsingular matrix $\Phi = (\phi_{ij})$ such that $\mathbf{W}_i = \sum_{j=1}^{k}\phi_{ij}\mathbf{P}_j$.

From (4.29) we know that $\mathbf{W}_i = \mathbf{V}_i - \frac{1}{\tau(i)}\mathbf{J}_s$ and from lemma 4.3 we know that:

$$\mathbf{W}_i = \frac{s}{\tau(i)}\left[\mathbf{P}_1+\cdots+\mathbf{P}_i\right],$$

therefore,

$$\Phi = \begin{bmatrix} \frac{s}{\tau(1)} & 0 & \ldots & 0 \\ \frac{s}{\tau(2)} & \frac{s}{\tau(2)} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}. \tag{4.32}$$

It can be seen from (4.32) that Φ is nonsingular. □

Lemma 4.5 *Let Φ be defined in (4.32), and let $\mathbf{q}_i$ be the ith column of the identity matrix, then $\hat{\mathbf{d}} = (0, \ldots, 0, \hat{d}_i, 0, \ldots, 0)'$ is the solution to system (4.24)—(4.27), where*

$$\hat{d}_i = \frac{\frac{s}{\tau(i)}}{\sum_{r=i}^{k} \left[\frac{s}{\tau(r)}\right]^2}.$$

Proof: It is obvious that $\hat{\mathbf{d}} \in R_+^k$, so $\hat{\mathbf{d}}$ satisfies (4.25).

Let $\Lambda = \Phi'\Phi = (\lambda_{hj})$. Thus,

$$\Lambda = \begin{bmatrix} \sum_{r=1}^{k} \left[\frac{s}{\tau(r)}\right]^2 & \sum_{r=2}^{k} \left[\frac{s}{\tau(r)}\right]^2 & \cdots & \sum_{r=k-1}^{k} \left[\frac{s}{\tau(r)}\right]^2 & 1 \\ \sum_{r=2}^{k} \left[\frac{s}{\tau(r)}\right]^2 & \sum_{r=2}^{k} \left[\frac{s}{\tau(r)}\right]^2 & \cdots & \sum_{r=k-1}^{k} \left[\frac{s}{\tau(r)}\right]^2 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}. \tag{4.33}$$

Generally,

$$\lambda_{hj} = \sum_{r=max(h,j)}^{k} \left[\frac{s}{\tau(r)}\right]^2,$$

thus for $h \leq j$, $\lambda_{hj} = \lambda_{jj} = \sum\limits_{r=j}^{k} \left[\frac{s}{\tau(r)}\right]^2$.

We want to show that $\hat{\mathbf{d}}$ given in the lemma is the solution for (4.24)-(4.27).

Let $\mathbf{d} = (d_1, \ldots, d_k)'$ be any vector satisfying (4.24)-(4.27),

$\Phi'(\Phi\mathbf{d} - \mathbf{q}_i) = \Lambda\mathbf{d} - \Phi'\mathbf{q}_i$

$$= \begin{bmatrix} \sum_{j=1}^{k} \lambda_{1j} d_j - \frac{s}{\tau(i)} \\ \vdots \\ \sum_{i=1}^{k} \lambda_{ij} d_j - \frac{s}{\tau(i)} \\ \sum_{j=1}^{k} \lambda_{(i+1)j} d_j \\ \vdots \\ (d_1 + \cdots + d_k) \end{bmatrix}. \tag{4.34}$$

Since (4.25) and (4.26) require $\mathbf{d} \in R_+^k$, $\Phi'(\Phi\mathbf{d} - \mathbf{q}_i) \in R_+^k$, so (4.27) makes:

$$d_h \left[\sum_{j=1}^{k} \lambda_{hj} d_j - \frac{s}{\tau(i)}\right] = 0, \quad h \leq i, \tag{4.35}$$

$$d_h \left[\sum_{j=1}^{k} \lambda_{hj} d_j\right] = 0, \quad h > i. \tag{4.36}$$

From (4.34) we know that if $\mathbf{d} = 0$, then

$$\Phi'(\Phi\mathbf{d} - \mathbf{q}_i) = \left[-\frac{s}{\tau(i)}, \ldots, -\frac{s}{\tau(i)}, 0, \ldots, 0\right]' \notin R_+^k,$$

hence $\mathbf{d} \neq 0$.

Since $\lambda_{hj} > 0$, from (4.36) $d_h = 0$, $h > i$.

Assume d_w is the first nonzero value among $d_1, \ldots, d_i$. Since $d_1 = \cdots = d_{w-1} = 0$, and $d_{i+1} = \cdots = d_k = 0$, from (4.35) we should have:

$$\sum_{j=1}^{k} \lambda_{wj} d_j - \frac{s}{\tau(i)} = \sum_{j=w}^{i} \lambda_{wj} d_j - \frac{s}{\tau(i)} = 0. \tag{4.37}$$

From (4.33) we have:

$$\lambda_{ww} > \lambda_{w(w+1)} > \cdots > \lambda_{wi},$$

therefore,

$$\begin{aligned}
\sum_{j=w}^{i} \lambda_{wj} d_j &= \lambda_{ww} d_w + \lambda_{w(w+1)} d_{w+1} + \cdots + \lambda_{w(i-1)} d_{i-1} + \lambda_{wi} d_i \\
&= (\lambda_{ww} - \lambda_{wi}) d_w + (\lambda_{w(w+1)} - \lambda_{wi}) d_{w+1} + \cdots \\
&\quad + (\lambda_{w(i-1)} - \lambda_{wi}) d_{i-1} + \lambda_{wi}(d_w + \cdots + d_i).
\end{aligned}$$

Substitute the above equality into (4.37):

$$\begin{aligned}
& d_w + \cdots + d_i \\
&= \frac{s}{\tau(i)\lambda_{wi}} - \left\{ \left(\frac{\lambda_{ww}}{\lambda_{wi}} - 1\right) d_w + \cdots + \left(\frac{\lambda_{w(i-1)}}{\lambda_{wi}} - 1\right) d_{i-1} \right\}.
\end{aligned} \tag{4.38}$$

Now since (4.27) holds for $\mathbf{d}$,

$$\begin{aligned}
& (\Phi\mathbf{d} - \mathbf{q}_i)'(\Phi\mathbf{d} - \mathbf{q}_i) = -\mathbf{q}_i'(\Phi\mathbf{d} - \mathbf{q}_i) \\
&= 1 - \frac{s}{\tau(i)}(d_1 + \cdots + d_i) \\
&= 1 - \frac{s}{\tau(i)}(d_w + \cdots + d_i) \\
&= 1 - \frac{\left[\frac{s}{\tau(i)}\right]^2}{\lambda_{wi}} + \frac{s}{\tau(i)} \left\{ \left(\frac{\lambda_{ww}}{\lambda_{wi}} - 1\right) d_w + \cdots + \left(\frac{\lambda_{w(i-1)}}{\lambda_{wi}} - 1\right) d_{i-1} \right\} \\
&\geq 1 - \frac{\left[\frac{s}{\tau(i)}\right]^2}{\lambda_{wi}}.
\end{aligned} \tag{4.39}$$

and $d_w = \cdots = d_{i-1} = 0$ will make the equality hold in (4.39), so $d_j = 0$, $j \neq i$ will minimize (4.24).

From (4.35) we have:

$$d_i \left[\lambda_{ii} d_i - \frac{s}{\tau(i)} \right] = 0.$$

But $d_i \neq 0$, hence

$$d_i = \frac{\frac{s}{\tau(i)}}{\lambda_{ii}} = \frac{\frac{s}{\tau(i)}}{\sum_{r=i}^{k} \left[\frac{s}{\tau(r)}\right]^2}.$$

We have proved that $\hat{\mathbf{d}}$ is the solution to system (4.24)–(4.27). □

Now that we have obtained the solution $\hat{\mathbf{d}}$, we need to work out the other terms in Theorem 4.6 to derive the explicit formulae for Hartung's estimator.

From the definition of $\mathbf{P}_i$ in (4.30) we know that

$$\operatorname{Tr} \mathbf{P}_i = \tau(i)(1 - \frac{1}{s_i}),$$

$$\| \mathbf{P}_i \mathbf{y} \|^2 = \frac{s}{\tau(i)} \sum_{j_1=1}^{s_1} \cdots \sum_{j_{i-1}=1}^{s_{i-1}} \sum_{j_i=1}^{s_i} (\bar{y}_{j_1 \ldots j_{i-1} j_i \ldots} - \bar{\bar{y}}_{j_1 \ldots j_{i-1} \ldots})^2,$$

where $\bar{y}_{j_1 \ldots j_i \ldots}$ denotes the mean of $y_{j_1 \ldots j_k}$ over the dotted factors.

Theorem 4.7 *For balanced nested ANOVA models, Hartung's estimator for the ith variance component σ_i^2 is:*

$$\hat{\sigma}_i^2 = \frac{A_i}{B_i C_i} \sum_{j_1}^{s_1} \cdots \sum_{j_i=1}^{s_i} (\bar{y}_{j_1 \ldots j_i \ldots} - \bar{\bar{y}}_{j_1 \ldots j_{i-1} \ldots})^2, \tag{4.40}$$

where $A_i = \left[\frac{s}{\tau(i)}\right]^2$, $B_i = \tau(i-1)(s_i - 1)$, *and* $C_i = \sum_{r=i}^{k} \left[\frac{s}{\tau(r)}\right]^2$.

Proof: From Theorem 4.6 we know that

$$\hat{\sigma}_i^2 = \sum_{i=1}^{k} \hat{d}_i \| \mathbf{P}_i \mathbf{y} \|^2 / \operatorname{Tr} \mathbf{P}_i.$$

We have derived the solution of $\mathbf{d}$ in Lemma 4.5 which gives

$$\hat{\mathbf{d}} = (0, \ldots, 0, \hat{d}_i, 0, \ldots, 0)',$$

where

$$\hat{d}_i = \frac{\frac{s}{\tau(i)}}{\sum_{r=i}^{k} \left[\frac{s}{\tau(r)}\right]^2}.$$

Now $\mathrm{Tr}\,\mathbf{P}_i = \tau(i)(1-\frac{1}{s_i}) = \tau(i-1)(s_i-1)$, and

$$\| \mathbf{P}_i \mathbf{y} \|^2 = \frac{s}{\tau(i)} \sum_{j_1=1}^{s_1} \cdots \sum_{j_{i-1}=1}^{s_{i-1}} \sum_{j_i=1}^{s_i} (\bar{y}_{j_1 \ldots j_{i-1} j_i \ldots} - \bar{\bar{y}}_{j_1 \ldots j_{i-1} \ldots})^2.$$

Hence

$$\hat{\sigma}_i^2 = \frac{\frac{s}{\tau(i)}}{\sum_{r=i}^{k} \left[\frac{s}{\tau(r)}\right]^2} \frac{s}{\tau(i)} \frac{1}{\tau(i-1)(s_i-1)} \sum_{j_1=1}^{s_1} \cdots \sum_{j_{i-1}=1}^{s_{i-1}} \sum_{j_i=1}^{s_i} (\bar{y}_{j_1 \ldots j_{i-1} j_i \ldots} - \bar{\bar{y}}_{j_1 \ldots j_{i-1} \ldots})^2,$$

which is (4.40). □

Example 4.4:

Consider the one way balanced model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n.$$

Here $s_1 = m$ and $s_2 = n$.

From (4.40), $A_1 = s_2^2 = n^2$, $B_1 = (m-1)$, $C_1 = 1 + n^2$, therefore,

$$\hat{\sigma}_a^2(H) = \frac{n^2}{(m-1)(1+n^2)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2. \tag{4.41}$$

$A_2 = 1$, $B_2 = m(n-1)$ and $C_2 = 1$, thus,

$$\hat{\sigma}_e^2(H) = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2. \tag{4.42}$$

Notice that $\hat{\sigma}_e^2(H)$ is identical to the ANOVA estimator of σ_e^2.

For the data set in Example 1.1,

$$\hat{\sigma}_a^2(H) = 1.8, \quad \hat{\sigma}_e^2(H) = 52. \; \square$$

After deriving Hartung's estimator for the balanced nested ANOVA models, we use the following example to demonstrate that for the balanced crossed ANOVA models the set of matrices $\mathbf{P}_i$ defined in (4.30) is not the basis of the subspace spanned by $\mathbf{W}_1, \ldots, \mathbf{W}_k$.

Example 4.5

Consider the two way balanced crossed model without interaction:

$$y_{ijk} = \mu + a_i + b_j + e_{ijk}$$

$$i = 1, \ldots, m, \; j = 1, \ldots, n, \; k = 1, \ldots, f.$$

Now for this model,

$$\mathbf{W}_1 = \mathbf{I}_1 \otimes \mathbf{J}_2 \otimes \mathbf{J}_3 - \frac{1}{m}\mathbf{J}_s,$$

$$\mathbf{W}_2 = \mathbf{J}_1 \otimes \mathbf{I}_2 \otimes \mathbf{J}_3 - \frac{1}{n}\mathbf{J}_s,$$

$$\mathbf{W}_3 = \mathbf{I}_s - \frac{1}{s}\mathbf{J}_s.$$

It can be seen that there does not exist c_1, c_2 and c_3 such that: $\mathbf{W}_i = c_1\mathbf{P}_1 + c_2\mathbf{P}_2 + c_3\mathbf{P}_3$. But it can be verified that $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ satisfy Definition 4.1, hence they form a commutative quadratic subspace. Therefore, we have to look for a new basis of the subspace for this model so that Theorem 4.6 can be used to derive Hartung's estimator. □

4.5.3 Discussion of Hartung's estimator

By using the constraint on the bias to minimize $\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{A}\mathbf{V}_i - q_i)^2$, Hartung's estimator can be uniquely determined. Theoretically Hartung's estimator is nonnegative and always exists which is an advantage over MINQUE. Whenever an unbiased nonnegative estimator is possible Hartung's estimator is identical to the estimator.

Practically we have only managed to give explicit formulae for Hartung's estimator for balanced nested ANOVA models. It remains a problem to make Hartung's estimator computationally available for the general variance components model (1.1) and this problem is the biggest obstacle to the use of Hartung's estimator.

There are two other deficiencies for Hartung's estimator. First, Hartung's estimator does not possess additivity. Second, Hartung's estimator can be severely biased when the true variance components values are considerably different.

Now I use an example to demonstrate the first point: Hartung's estimator does not have additivity.

Consider a two way balanced nested ANOVA model:

$$y_{ijk} = \mu + a_i + b_{ij} + e_{ijk}$$

$$i = 1, \ldots, r, \; i = 1, \ldots, s, \; k = 1, \ldots, t.$$

For this model the matrices Λ and Φ are:

$$\Lambda = \begin{bmatrix} s^2t^2 + t^2 + 1 & t^2 + 1 & 1 \\ t^2 + 1 & t^2 + 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$\Phi = \begin{bmatrix} st & 0 & 0 \\ t & t & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

The vector $\mathbf{d}$ in (4.23) for σ_a^2 and σ_b^2 are:

$$\hat{\mathbf{d}}(a) = \left(\frac{st}{s^2t^2+t^2+1}, 0, 0\right)',$$

$$\hat{\mathbf{d}}(b) = \left(0, \frac{t}{t^2+1}, 0\right)'.$$

If Hartung's estimator has additivity, then we expect Hartung's estimator for $\sigma_a^2 + \sigma_b^2$ to be:

$$\hat{\sigma}_a^2 + \hat{\sigma}_b^2 = \sum_{i=1}^{3} \left\{\hat{d}_i(a) + \hat{d}_i(b)\right\} \frac{\| \mathbf{P}_i \mathbf{y} \|^2}{\operatorname{Tr} \mathbf{P}_i},$$

and $\hat{d}_i(a) + \hat{d}_i(b)$ should satisfy (4.24)–(4.27), for $\mathbf{q} = (1,1,0)'$.

Now we know that $\hat{\mathbf{d}}(a) + \hat{\mathbf{d}}(b) = \left(\frac{st}{s^2t^2+t^2+1}, \frac{t}{t^2+1}, 0\right)'$, therefore,

$$\Phi'\Phi\left\{\hat{\mathbf{d}}(a) + \hat{\mathbf{d}}(b)\right\} - \Phi'\mathbf{q} = \left(0, \frac{st(t^2+1)}{s^2t^2+t^2+1}, \frac{st}{s^2t^2+t^2+1} + \frac{t}{t^2+1}\right)',$$

and

$$\left\{\hat{\mathbf{d}}(a) + \hat{\mathbf{d}}(b)\right\}' \Phi' \left\{\Phi\left[\hat{\mathbf{d}}(a) + \hat{\mathbf{d}}(b)\right] - \mathbf{q}\right\} = \frac{st^2}{s^2t^2+t^2+1} \neq 0.$$

So obviously $\hat{\mathbf{d}}(a)+\hat{\mathbf{d}}(b)$ does not satisfy (4.27), and hence it is not the solution to (4.24)–(4.27).

Since Hartung's estimator is uniquely determined, therefore the estimator has to be the one solved from system (4.24)–(4.27), hence $\hat{\sigma}_a^2 + \hat{\sigma}_b^2$ is not Hartung's estimator for $\sigma_a^2 + \sigma_e^2$. This fact leads to the conclusion that Hartung's estimator does not have additivity.

The second deficiency we have pointed out is that Hartung's estimator gives severe bias when the true variance components are considerably different, *e.g.* $\sigma_a^2 \gg \sigma_e^2$, or $\sigma_a^2 \ll \sigma_e^2$. We demonstrate this point with a numerical example.

Consider the one way balanced ANOVA model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n.$$

From (4.41) we know that Hartung's estimator of σ_a^2 is:

$$\hat{\sigma}_a^2(H) = \frac{n^2}{(m-1)(n^2+1)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2.$$

The relative bias of Hartung's estimator is:

$$\text{Rbias}\left(\hat{\sigma}_a^2(H)\right) = \frac{n-\gamma}{(n^2+1)\gamma},$$

where $\gamma = \sigma_a^2/\sigma_e^2$.

Rao and Chaubey's MINQE of σ_a^2 is:

$$\hat{\sigma}_a^2(M) = \frac{n^2\alpha_a}{m(n\alpha_a+\alpha_e)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 .$$

The relative bias of MINQE $\hat{\sigma}_a^2(M)$ is:

$$\text{Rbias}\left(\hat{\sigma}_a^2(M)\right) = \frac{(m-1)n(n\gamma+1)}{\gamma m(n\alpha_a+\alpha_e)^2} - 1,$$

where $\gamma = \sigma_a^2/\sigma_e^2$, α_a and α_e are the prior values of σ_a^2 and σ_e^2, respectively.

Now assume that $\gamma = 0.2$, $\sigma_e^2 = 1$, $m = 10$, and choose $\alpha_a = 1$ and $\alpha_e = 1$ to be the prior values in $\hat{\sigma}_a^2(M)$. For $n = 2, 3, 5, 10$, we calculate the relative biases of the two estimators:

n	2	3	5	10
$\hat{\sigma}_a^2(H)$	1.80	1.40	0.92	0.49
$\hat{\sigma}_a^2(M)$	0.40	0.35	0.25	0.12

Comparing with MINQE at the above parameter values Hartung's estimator has a relatively large bias which will contribute to the mean squared error of the estimator. This numerical result is an indication that Hartung's estimator has severe bias when the true variance components are considerably different. The numerical and Empirical studies in Chapter 5 support this view. The aim of the next section is to find an estimator which overcomes this deficiency of Hartung's estimator.

4.6 The minimum bias range MINQ estimator

4.6.1 The estimator

In the last section we have discussed Hartung's estimator and discovered that Hartung's estimator does not have satisfactory performance when bias and mean squared error are used as the optimality criteria. The reason for the poor bias performance is that one of the optimality criteria Hartung has used is that of minimizing $\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$, while the bias of the estimator $\mathbf{y}'\mathbf{A}\mathbf{y}$ is:

$$\left\{\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)\right\}^2,$$

thus Hartung's optimality criterion does not allow for the extremely unequal variance components.

The estimator proposed in this section has only one difference from Hartung's estimator: instead of using $\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$ as the second constraint we use $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$ as the second constraint in solving for a nonnegative minimum norm estimator.

We can interpret the optimality criterion we have used as follows:

$$\mathrm{bias}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \left\{\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)\right\}^2 \le \left\{\sum_{i=1}^{k}\sigma_i^2\left(\mathrm{Tr}\,\mathbf{AV}_i - q_i\right)^2\right\}\left(\sum_{i=1}^{k}\sigma_i^2\right),$$

thus $\mathrm{bias}(\mathbf{y}'\mathbf{A}\mathbf{y})$ lies in the range

$$\left[0,\quad \left\{\sum_{i=1}^{k}\sigma_i^2\left(\mathrm{Tr}\,\mathbf{AV}_i - q_i\right)^2\right\}\left(\sum_{i=1}^{k}\sigma_i^2\right)\right].$$

The approach of minimizing $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$ is in fact minimizing the range of the bias, hence putting a control on the bias and in some sense making the bias small.

Theorem 4.8 *There exists an unique symmetric and nonnegative matrix* $\hat{\mathbf{A}}$ *such that* $\hat{\mathbf{A}}\mathbf{X} = 0$, $\hat{\mathbf{A}}$ *minimizes* $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$ *and* $\|\mathbf{A}\|_E^2$.

By changing the Euclidean norm used in Hartung's proof of Theorem 4.4 into a weighted Euclidean norm we can obtain the proof of the above theorem.

Now let $\mathbf{W}_i = (\mathbf{I} - \mathbf{XX}^+)\mathbf{V}_i(\mathbf{I} - \mathbf{XX}^+)$, similar to Theorem 4.6, we have the following theorem:

Theorem 4.9 *If* $\mathbf{W}_1, \ldots, \mathbf{W}_k$ *form a commutative quadratic subspace, then the minimum bias range MINQ estimator* $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$ *for* $\sum_{i=1}^{k} q_i\sigma_i^2$ *is given by:*

$$\hat{\mathbf{A}} = \sum_{i=1}^{k}\hat{d}_i(\mathrm{Tr}\,\mathbf{P}_i)^{-1}\mathbf{P}_i, \tag{4.43}$$

where $\hat{\mathbf{d}} = (\hat{d}_1, \ldots, \hat{d}_k)'$ *is the solution to the following system:*

$$\textit{minimize} \quad (\Phi\mathbf{d} - \mathbf{q})'\Delta(\Phi\mathbf{d} - \mathbf{q}) \quad \textit{subject to}: \tag{4.44}$$

$$\mathbf{d} \in R^k_+, \tag{4.45}$$

$$\Phi'\Delta(\Phi\mathbf{d} - \mathbf{q}) \in R^k_+, \tag{4.46}$$

$$\mathbf{d}'\Phi'\Delta(\Phi\mathbf{d} - \mathbf{q}) = 0, \tag{4.47}$$

where $\Delta = diag(\sigma_1^2, \ldots, \sigma_k^2)'$.

The estimator is:

$$\mathbf{y}'\hat{\mathbf{A}}\mathbf{y} = \sum_{i=1}^{k} \hat{d}_i \parallel \mathbf{P}_i\mathbf{y} \parallel^2 (\mathrm{Tr}\,\mathbf{P}_i)^{-1}. \tag{4.48}$$

Proof: Let $\Psi = \Delta^{1/2}\Phi$, $\mathbf{g} = \Delta^{1/2}\mathbf{q}$.

Let $\mathbf{M}_i = \sigma_i\mathbf{W}_i$, thus

$$\mathbf{M}_i = \sigma_i \sum_{j=1}^{k} \phi_{ij}\mathbf{P}_j = \sum_{j=1}^{k} \psi_{ij}\mathbf{P}_j.$$

So $\mathbf{P}_1, \ldots, \mathbf{P}_k$ are the basis of Σ^* spanned by $\mathbf{M}_1, \ldots, \mathbf{M}_k$.

If $\hat{\mathbf{d}}$ is the solution to (4.44)–(4.47), then it is the solution to the following system:

$$\textit{minimize} \quad (\Psi\mathbf{d} - \mathbf{g})'(\Psi\mathbf{d} - \mathbf{g}) \quad \textit{subject to}: \tag{4.49}$$

$$\mathbf{d} \in R^k_+, \tag{4.50}$$

$$\Psi'(\Psi\mathbf{d} - \mathbf{g}) \in R^k_+, \tag{4.51}$$

$$\mathbf{d}'\Psi'(\Psi\mathbf{d} - \mathbf{g}) = 0. \tag{4.52}$$

From Theorem 4.6 we know that $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$ given in (4.48) is Hartung's estimator of $\sum_{i=1}^{k} g_i\sigma_i^2$, where Σ^* is spanned by $\mathbf{M}_1, \ldots, \mathbf{M}_k$.

Recall the constraints for Hartung's estimator are:

1. $\hat{\mathbf{A}}$ is nonnegative;
2. $\hat{\mathbf{A}}$ minimizes $\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{A}\mathbf{M}_i - g_i)^2$;
3. $\hat{\mathbf{A}}$ minimizes $\parallel \mathbf{A} \parallel^2$.

Since

$$\sum_{i=1}^{k}(\mathrm{Tr}\,\mathbf{AM}_i - g_i)^2 = \sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AW}_i - q_i)^2,$$

So we have:

2'. $\hat{\mathbf{A}}$ minimizes $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AW}_i - q_i)^2$.

Notice that 1, 2' and 3 are the constraints for the minimum bias range MINQ estimator, so $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$ is the minimum bias range MINQ estimator of $\sum_{i=1}^{k} q_i\sigma_i^2$ with Σ being spanned by $\mathbf{W}_1$, ..., $\mathbf{W}_k$, hence we proved the theorem. □

Since we introduced the unknown variance components in the constraints of the estimator, before we make further investigation to the properties of the minimum bias range MINQ estimator for balanced nested ANOVA models, we would like to answer a question: Are we able to find an estimator with the global minimum bias range in the parameter space? The answer is negative.

4.6.2 The nonexistence of a globally minimum bias range MINQ estimator

Theorem 4.10 *For a balanced ANOVA model, if there is no nonnegative unbiased estimator for* $\sum_{i=1}^{k} q_i\sigma_i^2$, $q_i > 0$, $i = 1, \ldots, k$, *then there does not exist a nonnegative matrix which minimizes* $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$ *for all* $\sigma_i^2 > 0$, $i = 1, \ldots, k$.

Proof: Since there is no nonnegative unbiased estimator of $\sum_{i=1}^{k} q_i\sigma_i^2$, *i.e.* there is no nonnegative matrix $\mathbf{A}$ which is the solution of the equation:

$$\mathrm{Tr}\,\mathbf{AV}_i = q_i, \quad i = 1, \ldots, k.$$

Then there exists at least one j such that $\mathrm{Tr}\,\mathbf{AV}_j \neq q_j$.

When balanced ANOVA models are considered, $\mathrm{Tr}\,\mathbf{V}_i = \mathrm{Tr}\,\mathbf{V}_j$, $i, j = 1, \ldots, k$. Assume that $\hat{\mathbf{A}}$ is the nonnegative matrix minimizing $\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{AV}_i - q_i)^2$.

Let $\mathbf{A}^0 = \frac{1}{2}\hat{\mathbf{A}} + \frac{q_j}{2\,\mathrm{Tr}\,\mathbf{V}_j}\mathbf{I}$, $\mathbf{A}^0$ is nonnegative.

$$\sum_{i=1}^{k}\sigma_i^2(\mathrm{Tr}\,\mathbf{A}^0\mathbf{V}_i - q_i)^2 = \sum_{i=1}^{k}\sigma_i^2(\frac{1}{2}\,\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_i + \frac{q_j}{2} - q_i)^2$$

$$= \sum_{i=1}^{k}\sigma_i^2\left\{\frac{1}{2}(\mathrm{Tr}\,\hat{\mathbf{A}}\mathbf{V}_i - q_i) + \frac{1}{2}(q_j - q_i)\right\}^2$$

$$= \frac{1}{4}\sum_{i=1}^{k}\sigma_i^2(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2 + \frac{1}{2}\sum_{i=1}^{k}\sigma_i^2(q_j - q_i)(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)$$
$$+\frac{1}{4}\sum_{i=1}^{k}\sigma_i^2(q_j - q_i)^2. \tag{4.53}$$

Now we want to find specific values for $\sigma_1^2, \ldots, \sigma_k^2$ such that:

$$\sum_{i=1}^{k}\sigma_i^2(\operatorname{Tr}\mathbf{A}^0\mathbf{V}_i - q_i)^2 < \sum_{i=1}^{k}\sigma_i^2(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2. \tag{4.54}$$

From (4.53) we want:

$$\frac{1}{2}\sum_{i=1}^{k}\sigma_i^2(q_j - q_i)(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i) + \frac{1}{4}\sum_{i=1}^{k}\sigma_i^2(q_j - q_i)^2 < \frac{3}{4}\sum_{i=1}^{k}\sigma_i^2(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2.$$

Equivalently we want:

$$\sum_{i=1}^{k}\sigma_i^2\left\{3(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2 - 2(q_j - q_i)(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i) - (q_j - q_i)^2\right\} > 0. \tag{4.55}$$

Let $\sigma_i^{2^0} = 1$, $i \neq j$, then (4.55) becomes:

$$\sum_{i\neq j}\left\{3(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2 - 2(q_j - q_i)(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i) - (q_j - q_i)^2\right\} +$$
$$3(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_j - q_j)^2\sigma_j^2 > 0. \tag{4.56}$$

Thus we should have:

$$\sigma_j^2 > \frac{\sum_{i\neq j}\left\{(q_j - q_i)^2 + 2(q_j - q_i)(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i) - 3(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_i - q_i)^2\right\}}{3(\operatorname{Tr}\hat{\mathbf{A}}\mathbf{V}_j - q_j)^2} \stackrel{\text{def}}{=} c$$

Now let $\sigma_j^{2^0} = \max(c+1, 1)$, then $\sigma_j^{2^0} > c$ and $\sigma_j^{2^0} > 0$, such a choice of σ_i^2 will make (4.54) hold. That means the nonnegative estimator $\mathbf{y}'\mathbf{A}^0\mathbf{y}$ has smaller bias range than $\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}$ and this is contradictory to the assumption of $\hat{\mathbf{A}}$. Hence we proved the theorem. □

Theorem 4.10 shows that by adopting the bias range approach we can only have local minimization not a global minimization in the parameter space.

4.6.3 The formulae of the estimator for balanced nested ANOVA models

Before the derivation of the formulae we need a lemma.

Lemma 4.6 *Let Φ be defined in (4.32), $\mathbf{q}_i$ is the ith column of the identity matrix, $\hat{\mathbf{d}} = (0, \ldots, 0, \hat{d}_i, 0, \ldots, 0)'$ is the unique solution to (4.44)–(4.47) where*

$$\hat{d}_i = \frac{\frac{s}{\tau(i)}\sigma_i^2}{\sum_{r=i}^{k}\left[\frac{s}{\tau(r)}\sigma_r\right]^2}.$$

The proof follows from that of Lemma 4.5.

Theorem 4.11 *For balanced nested ANOVA models, let $A(i) = \left[\frac{s}{\tau(i)}\right]^2$, $B(i) = \tau(i-1)(s_i - 1)$, α_i be the prior values of σ_i^2, $i = 1, \ldots, k$, then the minimum bias range MINQ estimator of σ_i^2 is given by:*

$$\hat{\sigma}_i^2 = \frac{A(i)\alpha_i}{B(i)\left[\sum_{r=i}^{k} A(r)\alpha_r\right]} \sum_{j_1}^{s_1} \cdots \sum_{j_i}^{s_i} \left(\bar{y}_{j_1 \ldots j_i \ldots} - \bar{\bar{y}}_{j_1 \ldots j_{i-1} \ldots}\right)^2. \tag{4.57}$$

The derivation of the estimator is similar to that of Theorem 4.7.

Since the minimum bias range MINQ estimator needs prior values of σ_i^2, different prior values produce different estimators. Particularly, $\alpha_i = 1$, $i = 1, \ldots, k$ gives Hartung's estimator. Iterative computing can be used to obtain an unique solution from (4.57), hence eliminating the uncertainty of the prior values. But the properties of the iterative estimates are unknown.

Example 4.6:

For the one way balanced model,

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$

The minimum bias range MINQ estimator of σ_a^2 and σ_e^2 are:

$$\hat{\sigma}_a^2(R) = \frac{n^2\alpha_a}{(m-1)(n^2\alpha_a + \alpha_e)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{y}_{..})^2, \tag{4.58}$$

$$\hat{\sigma}_e^2(R) = \frac{1}{m(n-1)} \sum_{i=1}^{m}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2. \tag{4.59}$$

$\hat{\sigma}_e^2(R)$ is independent of prior values. By assuming $\alpha_a = \hat{\sigma}_a^2$ and $\alpha_e = \hat{\sigma}_e^2$ in (4.58) and (4.59) we can solve $\hat{\sigma}_a^2(R)$ from (4.58) which is the iterative estimate of σ_a^2:

$$\hat{\sigma}_a^2(R) = \begin{cases} A = \frac{1}{m-1}\sum_{i=1}^m(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{mn^2(n-1)}\sum_{i=1}^m\sum_{j=1}^n(y_{ij} - \bar{y}_{i.})^2, & \text{if } A > 0, \\ 0, & \text{if } A \leq 0. \end{cases}$$

Now we use the data set in Example 1.1 to derive estimates by choosing different prior values for α_a and α_e.

Choose $\alpha_a = \alpha_e = 1$, we have Hartung's estimator:

$$\hat{\sigma}_a^2 = 1.8, \ \hat{\sigma}_e^2 = 52.$$

Choose $\alpha_a = 0.5$, $\alpha_e = 1$, we have:

$$\hat{\sigma}_a^2 = 1.64, \ \hat{\sigma}_e^2 = 52.$$

Choose $\alpha_a = 1/2n = 1/6$, $\alpha_e = 1$, we have:

$$\hat{\sigma}_a^2 = 1.20, \ \hat{\sigma}_e^2 = 52.$$

Iterative computing starting with $\alpha_a = 1$ and $\alpha_e = 1$ gives:

$$\hat{\sigma}_a^2 = 0, \quad \hat{\sigma}_e^2 = 52. \ \square$$

The iterative computing starting from nonzero values of α_a and α_e can yield nonzero solutions. For example, for the balanced one-way model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, 2, \ j = 1, 2, 3,$$

with data given:

$i \setminus j$	1	2	3	mean
1	19	17	18	18
2	24	4	14	14
				16

The ANOVA estimate of σ_a^2 is: $\hat{\sigma}_a^2(A) = -8.83$. But the iterative minimum bias range MINQ estimate is $\hat{\sigma}_a^2(R) = 1.83$.

The fact that the iterative minimum bias range MINQ estimator yields nonzero solutions depends on the ratio:

$$\frac{\sum_{i=1}^m(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2}{\sum_{i=1}^m\sum_{j=1}^n(y_{ij} - \bar{y}_{i.})^2}.$$

If $\frac{\sum_{i=1}^{m}(\bar{y}_{i.}-\bar{y}_{..})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2} > \frac{m-1}{mn^2(n-1)}$, then the estimate is positive. Otherwise, the estimate is zero.

The following theorem gives the bias formula for the minimum bias range MINQ estimator.

Theorem 4.12 *For balanced nested ANOVA models, let α_r be the prior value of σ_r^2, $r = 1, \ldots, k$, the bias of the nonnegative minimum bias range MINQ estimator $\hat{\sigma}_i^2$ is:*

$$\text{bias}(\mathbf{y}'\hat{\mathbf{A}}_i\mathbf{y}) = \left\{ \frac{\left[\frac{s}{\tau(i)}\alpha_i\right]\left[\sum_{r=i}^{k}\frac{s}{\tau(r)}\sigma_r^2\right]}{\sum_{r=i}^{k}\alpha_r\left[\frac{s}{\tau(r)}\right]^2} - \sigma_i^2 \right\}^2. \tag{4.60}$$

Proof: From theorem 4.9, we know that:

$$\hat{\mathbf{A}}_i = \hat{d}_i(\operatorname{Tr}\mathbf{P}_i)^{-1}\mathbf{P}_i.$$

The variance covariance matrix for the balanced nested ANOVA model is:

$$\begin{aligned}\mathbf{V} &= \sigma_1^2\mathbf{I}_1 \otimes \mathbf{J}_2 \otimes \cdots \otimes \mathbf{J}_k + \sigma_2^2\mathbf{I}_1 \otimes \mathbf{I}_2 \otimes \mathbf{J}_3 \otimes \cdots \otimes \mathbf{J}_k \\ &\quad + \cdots + \sigma_k^2\mathbf{I}_1 \otimes \mathbf{I}_2 \otimes \cdots \otimes \mathbf{I}_{k-1} \otimes \mathbf{I}_k.\end{aligned}$$

Using the equality in Lemma 4.3:

$$\begin{aligned}\mathbf{V} &= \sigma_1^2\left\{\frac{s}{\tau(1)}\mathbf{P}_1 + \frac{1}{\tau(1)}\mathbf{J}_s\right\} + \sigma_2^2\left\{\frac{s}{\tau(2)}[\mathbf{P}_1 + \mathbf{P}_2] + \frac{1}{\tau(2)}\mathbf{J}_s\right\} \\ &\quad + \cdots + \sigma_k^2\left\{\frac{s}{\tau(k)}[\mathbf{P}_1 + \cdots + \mathbf{P}_k] + \frac{1}{\tau(k)}\mathbf{J}_s\right\} \\ &= \left\{\sum_{r=1}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_1 + \left\{\sum_{r=2}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_2 \\ &\quad + \cdots + \sigma_k^2\mathbf{P}_k + \left\{\sum_{r=1}^{k}\frac{1}{\tau(r)}\sigma_r^2\right\}\mathbf{J}_s.\end{aligned}$$

Since $\mathbf{P}_1, \ldots, \mathbf{P}_k$ are pairwise orthogonal projection matrices and $\mathbf{P}_i\mathbf{J}_s = 0$, $i = 1, \ldots, k$, we have:

$$\begin{aligned}\mathbf{P}_i\mathbf{V} &= \left\{\sum_{r=i}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_i + \left\{\sum_{r=1}^{k}\frac{1}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_i\mathbf{J}_s \\ &= \left\{\sum_{r=i}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_i,\end{aligned}$$

and

$$\operatorname{Tr}\mathbf{P}_i\mathbf{V} = \left\{\sum_{r=i}^{k} \frac{s}{\tau(r)}\sigma_r^2\right\} \operatorname{Tr}\mathbf{P}_i.$$

Now,

$$\begin{aligned}
\mathrm{E}(\mathbf{y}'\hat{\mathbf{A}}\mathbf{y}) &= \operatorname{Tr}\hat{\mathbf{A}}_i\mathbf{V} \\
&= \hat{d}_i(\operatorname{Tr}\mathbf{P}_i)^{-1}\operatorname{Tr}\mathbf{P}_i\mathbf{V} \\
&= \hat{d}_i\left\{\sum_{r=i}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}.
\end{aligned}$$

Recall that

$$\hat{d}_i = \frac{\frac{s}{\tau(i)}\alpha_i}{\sum_{r=i}^{k}\alpha_r\left[\frac{s}{\tau(r)}\right]^2},$$

therefore,

$$E(\mathbf{y}'\hat{\mathbf{A}}_i\mathbf{y}) = \frac{\alpha_i\frac{s}{\tau(i)}\left\{\sum_{r=i}^{k}\frac{s}{\tau(r)}\sigma_r^2\right\}}{\sum_{r=i}^{k}\alpha_r\left[\frac{s}{\tau(r)}\right]^2}.$$

So the bias of $\mathbf{y}'\mathbf{A}\mathbf{y}$ is given in (4.60). □

We use the following numerical result to show that the minimum bias range MINQ estimator can have less bias than Hartung's estimator.

Consider the one way balanced model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$

From Theorem 4.11, we can derive the relative bias of the estimator:

$$\operatorname{Rbias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \frac{n\hat{\gamma} - \gamma}{\gamma(n^2\hat{\gamma} + 1)}$$

Now if we choose $\hat{\gamma} = \frac{1}{2n}$, the bias of the two estimators are:

n	2	3	5	10
$\hat{\sigma}_a^2(H)$	1.80	1.40	0.92	0.49
$\hat{\sigma}_a^2(\frac{1}{2n})$	0.75	0.60	0.43	0.25

It can be seen that the bias has been reduced.

The next theorem gives a formula for the variance of the minimum bias range MINQ estimator.

Theorem 4.13 *For balanced nested ANOVA models, if a normal distribution is assumed for the data vector* $\mathbf{y}$, α_r *is the prior value of* σ_r^2, $i = 1, \ldots, k$, *then,*

$$\mathrm{V}(\mathbf{y}'\mathbf{A}\mathbf{y}) = \frac{2\alpha_i^2\left[\frac{s}{\tau(i)}\right]^2\left\{\sum_{r=i}^k \frac{s}{\tau(r)}\sigma_r^2\right\}^2}{\tau(i-1)(s_i-1)\left\{\sum_{r=i}^k \alpha_r\left[\frac{s}{\tau(r)}\right]^2\right\}^2}. \tag{4.61}$$

Proof: From Theorem 4.9,

$$\hat{\mathbf{A}}_i = \hat{d}_i(\operatorname{Tr}\mathbf{P}_i)^{-1}\mathbf{P}_i.$$

From the proof of Theorem 4.12,

$$\hat{\mathbf{A}}_i\mathbf{V} = \hat{d}_i(\operatorname{Tr}\mathbf{P}_i)^{-1}\left\{\sum_{r=i}^k \frac{s}{\tau(r)}\sigma_r^2\right\}\mathbf{P}_i.$$

Thus

$$\hat{\mathbf{A}}_i\mathbf{V}\hat{\mathbf{A}}_i\mathbf{V} = \hat{d}_i^2(\operatorname{Tr}\mathbf{P}_i)^{-2}\left\{\sum_{r=i}^k \frac{s}{\tau(r)}\sigma_r^2\right\}^2\mathbf{P}_i,$$

$$\begin{aligned}\operatorname{Tr}\hat{\mathbf{A}}_i\mathbf{V}\hat{\mathbf{A}}_i\mathbf{V} &= \hat{d}_i^2\left\{\sum_{r=i}^k \frac{s}{\tau(r)}\sigma_r^2\right\}^2(\operatorname{Tr}\mathbf{P}_i)^{-1}\\ &= \hat{d}_i^2\left\{\sum_{r=i}^k \frac{s}{\tau(r)}\sigma_r^2\right\}^2\frac{1}{\tau(i-1)(s_i-1)}.\end{aligned}$$

Since with normality assumption,

$$V(\mathbf{y}'\hat{\mathbf{A}}_i\mathbf{y}) = 2\operatorname{Tr}\hat{\mathbf{A}}_i\mathbf{V}\hat{\mathbf{A}}_i\mathbf{V},$$

substituting the values of $\hat{d}_i$ from Lemma 4.6 we obtain (4.61). □

After showing that the bias of the minimum bias range MINQ estimator is smaller than that of Hartung's estimator for the one-way balanced model at $\gamma = \sigma_a^2/\sigma_e^2 = 0.2$, we now compare the mean squared errors of these two estimators.

From Theorem 4.12,

$$\text{bias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \sigma_e^4\left\{\frac{n\hat{\gamma}-\gamma}{n^2\hat{\gamma}+1}\right\}.$$

From Theorem 4.13 with normality assumption,

$$\mathrm{V}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \sigma_e^4\left\{\frac{2n^2\hat{\gamma}^2(n\gamma+1)^2}{(m-1)(n^2\hat{\gamma}+1)^2}\right\},$$

$$\text{MSE}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \text{bias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) + \mathrm{V}\left(\hat{\sigma}_a^2(\hat{\gamma})\right).$$

We use $\hat{\sigma}_a^2(H)$ to denote Hartung's estimator which uses $\hat{\gamma} = 1$ and $\hat{\sigma}_a^2(\frac{1}{2n})$ to denote the range MINQ estimator using $\hat{\gamma} = \frac{1}{2n}$, then the mean squared errors of the two estimators are:

n	2	3	5	10
$\hat{\sigma}_a^2(H)$	1.25	1.74	1.78	1.44
$\hat{\sigma}_a^2(\frac{1}{2n})$	0.31	0.50	0.68	0.81

We can see that the mean squared error of $\hat{\sigma}_a^2(\frac{1}{2n})$ is smaller than that of $\hat{\sigma}_a^2(H)$.

4.6.4 An investigation of the bias function

In section 4.6.1 we derived the minimum bias range MINQ estimator by minimizing the range of the bias of the estimator. We showed that such an estimator depends on prior values of the variance components. Now we use the one way balanced random model to investigate the behaviour of the bias function of the estimator.

The model we used is:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n.$$

The minimum bias range MINQ estimator of σ_a^2 is:

$$\hat{\sigma}_a^2(\hat{\gamma}) = \frac{n^2\hat{\gamma}}{(m-1)(n^2\hat{\gamma}+1)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 , \tag{4.62}$$

where $\hat{\gamma}$ is the prior value of $\gamma = \sigma_a^2/\sigma_e^2$.

We know that when $\hat{\gamma} = \gamma$ the upper bound of the bias range in section 4.6.1 is minimized.

Now the relative bias of $\hat{\sigma}_a^2(\hat{\gamma})$ is:

$$\text{Rbias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \frac{n\hat{\gamma} - \gamma}{\gamma(n^2\hat{\gamma}+1)}. \tag{4.63}$$

It can be seen that when $\hat{\gamma} = \gamma/n$, $\text{Rbias}(\hat{\sigma}_a^2(\hat{\gamma})) = 0$. In other words, when $\hat{\gamma} = \gamma/n$, the bias function is minimized to zero.

We used the one way model to demonstrate that when the true values of variance components are used, the minimum bias range MINQ estimator does not have minimum bias. In this specific example, $\hat{\gamma} = \gamma/n$ leads to minimum bias.

4.6.5 Discussion of the estimator

The minimum bias range MINQ estimator preserved all the good properties of Hartung's estimator. It always exists and is uniquely determined. One of the optimality measures used to derive this estimator can be interpreted as an upper bound of the bias function of the estimator, hence minimizing this upper bound will prevent the bias from becoming unacceptably large. Unfortunately, it is proved in section 4.6.2 that such an estimator can only have local optimality, not global optimality. By choosing realistic prior values this estimator has smaller bias than Hartung's estimator.

4.7 Conclusions

In this chapter nonnegative quadratic estimation of variance components is considered.

It is proved that in most cases unbiasedness and nonnegativity are incompatible, so the estimators discussed in this chapter are usually biased. The optimality criterion commonly used for biased estimators is the mean squared error. Due to the difficulty of obtaining the estimator with minimum mean squared error, there are attempts to use other optimality measures.

Rao and Chaubey's MINQE and Chaubey's CMINQUE built their optimality criteria on the concept of 'closeness' to a matrix and it is not clear how these criteria relate to the commonly used statistical measures, *i.e.* bias, variance and mean squared error.

The constraints of invariance, nonnegativity and minimum variance are not enough to determine a quadratic estimator and using these constraints also has the risk of producing unacceptably large bias. It seems necessary to bring a control on the bias term of the estimator. The first attempt is minimizing the bias of the estimator while keeping the other constraints. This failed because of the difficulty of obtaining a minimum biased estimator satisfying all the constraints. It is also proved in this chapter that a nonnegative minimum biased estimator, if obtained, can be only locally obtained, not globally.

The minimum bias range MINQ estimator is proposed in this chapter and it includes Hartung's estimator as a special case. This estimator also attempts to control the bias of the estimator and it actually minimizes an upper bound of the bias function. We are able to obtain such an estimator. Formulae are given for the balanced nested ANOVA models. Unfortunately, it is proved that this estimator also has local optimality and it does not achieve global optimality.

So far we have been unable to compare the estimators because they are built with different optimality criteria and some of them only have local optimality. In Chapter 5 numerical and empirical studies are carried out to compare these estimators.

Chapter 5

COMPARISONS OF THE NONNEGATIVE ESTIMATORS

In Chapter 4 several nonnegative estimators of variance components have been derived. None of the estimators has been derived with the mean squared error as the optimality criterion and some of them only have local optimality. In this Chapter we intend to compare the estimators in terms of their bias and mean squared error. We first compare the relative bias and efficiency of the estimators numerically and then carry out an empirical study to compare the nonnegative estimators with some commonly used approaches such as putting the negative ANOVA estimates to zero and using the maximum likelihood estimator.

5.1 Numerical comparison

5.1.1 Model and estimators used in the comparison

We use the one way balanced model in the comparison:

$$y_{ij} = \mu + a_i + e_{ij} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{5.1}$$

where y_{ij} is the observed data, μ is the overall mean, a_i is the random term with variance σ_a^2, e_{ij} is the random term with variance σ_e^2.

We assume that $\mathbf{y}$ is normally distributed with mean μ and variance covariance matrix $\mathbf{V} = \sigma_a^2 \mathbf{I}_m \otimes \mathbf{J}_n + \sigma_e^2 \mathbf{I}_{mn}$.

We shall only compare the estimators for σ_a^2 because in most situations of variance components estimation the variance component in which we are interested is σ_a^2. Besides, the ANOVA estimator for σ_e^2 is the best unbiased nonnegative estimator and almost all the estimators in Chapter 4 are identical to the ANOVA estimator of σ_e^2 except MINQE.

In Chapter 4 we have introduced MINQE and CMINQUE and derived the minimum bias range MINQ estimator. MINQE and the minimum bias range MINQ estimator need prior values. In practice we cannot know the true values of the variance components so the first four estimators included in the tables are either using fixed prior values ($\hat{\sigma}_a^2(H)$, $\hat{\sigma}_a^2(\frac{1}{2n})$, $\hat{\sigma}_a^2(M)$) or without prior values ($\hat{\sigma}_a^2(C)$). The first estimator is Hartung's estimator $\hat{\sigma}_a^2(H)$ which is the minimum bias range MINQ estimator with prior value $\hat{\gamma} = 1$, where $\gamma = \sigma_a^2/\sigma_e^2$. The second estimator is the minimum bias range MINQ estimator $\hat{\sigma}_a^2(\frac{1}{2n})$ with prior value $\hat{\gamma} = \frac{1}{2n}$. The third estimator is MINQE with prior values $\hat{\gamma} = 1$. The fourth estimator is CMINQUE. These four estimators are usable in practice and are therefore compared numerically. Since the nonnegative ML and REML estimators depend on the data, they will be included in the empirical studies in the next section.

We include another three estimators which need the true variance components values for the purpose of comparing the effect of different prior values on the performance of the estimators. The fifth and the sixth estimators are the minimum bias range MINQ estimator with $\hat{\gamma} = \gamma$ and $\hat{\gamma} = \frac{\gamma}{n}$, respectively. From Chapter 4 we know that $\hat{\gamma} = \frac{\gamma}{n}$ will make the estimator have zero bias, so we exclude $\hat{\sigma}_a^2(\frac{\gamma}{n})$ from Table 5.1 where the relative bias of the estimators are listed. The last estimator included is $\hat{\sigma}_a^2(M\gamma)$ which is MINQE with prior value $\hat{\gamma} = \gamma$.

Taking the formulae we derived in Chapter 4 we have:

$$\hat{\sigma}_a^2(\hat{\gamma}) = \frac{n^2\hat{\gamma}}{(m-1)(n^2\hat{\gamma}+1)} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2, \tag{5.2}$$

$$\text{bias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \sigma_e^4 \left\{ \frac{n\hat{\gamma} - \gamma}{n^2\hat{\gamma}+1} \right\}^2 . \tag{5.3}$$

The relative bias of $\hat{\sigma}_a^2(\hat{\gamma})$ is:

$$\text{Rbias}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \frac{|n\hat{\gamma} - \gamma|}{\gamma(n^2\hat{\gamma}+1)}. \tag{5.4}$$

With normality assumption,

$$\mathrm{V}\left(\hat{\sigma}_a^2(\hat{\gamma})\right) = \sigma_e^4 \left\{ \frac{2n^2\hat{\gamma}^2(n\gamma+1)^2}{(m-1)(n^2\hat{\gamma}+1)^2} \right\}, \tag{5.5}$$

where $\gamma = \sigma_a^2/\sigma_e^2$ and $\hat{\gamma}$ is the prior value of γ. $\hat{\sigma}_a^2(H)$, $\hat{\sigma}_a^2(\frac{1}{2n})$, $\hat{\sigma}_a^2(\gamma)$ and $\hat{\sigma}_a^2(\frac{\gamma}{n})$ are derived from (5.2) by using prior values $\hat{\gamma} = 1$, $\frac{1}{2n}$, γ and $\frac{\gamma}{n}$, respectively.

From section 4.2 we derive the MINQE of σ_a^2:

$$\hat{\sigma}_a^2(M) = \frac{n^2\alpha_a^2}{m(n\alpha_a+\alpha_e)^2}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2, \tag{5.6}$$

$$\mathrm{bias}\left(\hat{\sigma}_a^2(M)\right) = \left\{ \frac{(m-1)n\alpha_a^2(n\sigma_a^2+\sigma_e^2)}{m(n\alpha_a+\alpha_e)^2} - \sigma_a^2 \right\}^2, \tag{5.7}$$

$$\mathrm{Rbias}\left(\hat{\sigma}_a^2(M)\right) = \left| \frac{(m-1)n\alpha_a^2(n\gamma+1)}{\gamma m(n\alpha_a+\alpha_e)^2} - 1 \right|, \tag{5.8}$$

with normality assumption,

$$\mathrm{V}\left(\hat{\sigma}_a^2(M)\right) = \frac{2(m-1)n^2\alpha_a^4(n\sigma_a^2+\sigma_e^2)^2}{m^2(n\alpha_a+\alpha_e)^4}, \tag{5.9}$$

where α_a and α_e are the prior values for σ_a^2 and σ_e^2 respectively.

In the comparison we use $\hat{\sigma}_a^2(M)$ to denote the MINQE with $\alpha_a = \alpha_e = 1$, $\hat{\sigma}_a^2(M\gamma)$ with $\alpha_a = \gamma$ and $\alpha_e = 1$.

From section 4.3 the CMINQUE of σ_a^2 is:

$$\hat{\sigma}_a^2(C) = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2, \tag{5.10}$$

$$\mathrm{bias}\left(\hat{\sigma}_a^2(C)\right) = \frac{\sigma_e^4}{n^2}, \tag{5.11}$$

$$\mathrm{Rbias}\left(\hat{\sigma}_a^2(C)\right) = \frac{1}{n\gamma}, \tag{5.12}$$

with normality assumption,

$$\mathrm{V}\left(\hat{\sigma}_a^2(C)\right) = \sigma_e^4\frac{2(n\gamma+1)^2}{(m-1)n^2}, \tag{5.13}$$

where $\gamma = \sigma_a^2/\sigma_e^2$.

We use $\sigma_e^2 = 1$ and compare the estimators for different γ, the ratio of the two variance components.

Table 5.1 lists the relative biases of the estimators ($\hat{\sigma}_a^2(\frac{\gamma}{n})$ is absent because it has zero bias). We use $m = 10$ for $\gamma = 0.2$ and $m = 100$ for other γ. The left two columns are the parameter values for γ and n.

Table 5.2 shows the relative efficiencies of the estimators. We use the mean squared error of each estimator over the variance of the ANOVA estimator of σ_a^2 as the measure of relative efficiency. The ANOVA estimator of σ_a^2 is:

$$\hat{\sigma}_a^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2, \tag{5.14}$$

with normality assumption,

$$\mathrm{V}(\hat{\sigma}_a^2) = \sigma_e^4\left\{\frac{2}{mn^2(n-1)} + \frac{2(n\gamma+1)^2}{n^2(m-1)}\right\}. \tag{5.15}$$

The parameter values of γ, m and n are chosen in the same way as in Table 5.1.

From Table 5.1 we can see that none of the first four estimators which are independent of prior values have a reasonably small bias across the parameter space, which confirms the conclusion of Theorem 4.3. When γ is small Hartung's estimator and CMINQUE have very large bias; when γ is large $\hat{\sigma}_a^2(\frac{1}{2n})$ and $\hat{\sigma}_a^2(M)$ have large bias.

$\hat{\sigma}_a^2(\gamma)$, which minimizes the upper bound of the bias function in section 4.6.1, has a relatively larger bias than $\hat{\sigma}_a^2(\frac{1}{2n})$ when $\gamma < 1$, and $\hat{\sigma}_a^2(H)$ when $\gamma \geq 1$.

From Table 5.2 we can see that Hartung's estimator and CMINQUE are not efficient when $\gamma < 1$. $\hat{\sigma}_a^2(\frac{1}{2n})$ and $\hat{\sigma}_a^2(M)$ are not efficient when $\gamma > 1$. $\hat{\sigma}_a^2(\frac{1}{2n})$ and $\hat{\sigma}_a^2(M)$ are much more efficient than the ANOVA estimator when $\gamma \ll 1$. The CMINQUE only does well when γ is large.

$\hat{\sigma}_a^2(\frac{\gamma}{n})$, which has zero bias, is more efficient than the ANOVA estimator. MINQE with the true variance components values as its prior values has a very poor performance when $\gamma \leq 1$ comparing with the MINQE using prior value 1.

Since no single estimator in the comparison dominates the others, it is therefore necessary to study the prior information on the variance components before choosing an estimator.

In practice it is common to use the ANOVA estimator and put the negative estimate into zero if such estimate appears. How does this approach compare to the first four estimators in Table 5.2? To answer this question we need a Monte Carlo study.

5.2 Empirical comparison

5.2.1 The estimators included in the comparison

We include the first four estimators in Table 5.2 because they are independent of prior values. In addition we include five more estimators: $\hat{\sigma}_a^2(A)$, $\hat{\sigma}_a^2(0)$, $\hat{\sigma}_a^2(A\frac{1}{2n})$, $\hat{\sigma}_a^2(ML)$ and $\hat{\sigma}_a^2(R)$.

$\hat{\sigma}_a^2(A)$ is the ANOVA estimator. $\hat{\sigma}_a^2(0)$ is a modified ANOVA estimator which puts the negative values to zero, *i.e.*

$$\hat{\sigma}_a^2(0) = \begin{cases} \hat{\sigma}_a^2(A), & \text{if } \hat{\sigma}_a^2(A) \geq 0, \\ 0, & \text{if } \hat{\sigma}_a^2(A) < 0. \end{cases}$$

$\hat{\sigma}_a^2(0)$ is identical to the REML estimator for model (5.1).

$\hat{\sigma}_a^2(A\frac{1}{2n})$ is another modified ANOVA estimator which puts the negative ANOVA estimates into the minimum bias range MINQ estimator with prior value $\hat{\gamma} = \frac{1}{2n}$, *i.e.*

$$\hat{\sigma}_a^2(A\frac{1}{2n}) = \begin{cases} \hat{\sigma}_a^2(A), & \text{if } \hat{\sigma}_a^2(A) \geq 0, \\ \hat{\sigma}_a^2(\frac{1}{2n}), & \text{if } \hat{\sigma}_a^2(A) < 0. \end{cases}$$

$\hat{\sigma}_a^2(ML)$ is the maximum likelihood estimator:

$$\hat{\sigma}_a^2(ML) = \frac{1}{m}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \tag{5.16}$$

If (5.16) gives negative value then 0 is the taken as the maximum likelihood estimator of σ_a^2.

$\hat{\sigma}_e^2(R)$ is the iterative solution for the minimum bias range MINQ estimator which is:

$$\hat{\sigma}_a^2(R) = \begin{cases} A = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{\bar{y}}_{..})^2 - \frac{1}{mn^2(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2, & \text{if } A > 0, \\ 0, & \text{if } A \leq 0. \end{cases}$$

$\hat{\sigma}_a^2(R)$ is independent of prior values.

For each set of the parameter values listed on the left hand side of Table 5.3 and 5.4 we generate 1000 samples which are normally distributed with mean 1 and variance covariance matrix $\mathbf{V} = \sigma_a^2\mathbf{I}_m \otimes \mathbf{J}_n + \sigma_e^2\mathbf{I}_{mn}$. We calculate the estimates using different estimators and list the relative biases and the relative efficiencies calculated from the sample in Table 5.3 and Table 5.4. Again we use the mean squared error of each estimate over the variance of the ANOVA estimator as the measure of efficiency. We also include the numbers of negative ANOVA estimates among the 1000 samples in the far right column of Table 5.4.

5.2.2 Conclusions

From Table 5.3 we can see that Hartung's estimator and CMINQUE gave very large bias when γ is small. It can be noticed that in the table the minimum bias range MINQ type estimators, $\hat{\sigma}_a^2(\frac{1}{2n})$ and $\hat{\sigma}_a^2(H)$, have their minimum bias at $\gamma = 0.5$ when m and n are fixed. From (5.4) we know that $\hat{\sigma}_a^2(\frac{1}{2n})$ achieves the minimum bias at $\gamma = 0.5$. $\hat{\sigma}_a^2(H)$ achieves the minimum bias range at $\gamma = 1$. Therefore the column of $\hat{\sigma}_a^2(H)$ in Table 5.3 demonstrated that $\hat{\sigma}_a^2(H)$ does not necessarily minimize the bias while minimizing the upper bound of the bias function. We calculated the standard errors for Table 5.3 and found that they generally decrease when γ increases, although they vary for different m and n. For example at $\gamma = 0.5$ the standard errors of the estimators are about 0.03. At $\gamma = 1.0$ the standard errors are about 0.02. We shall not include a complete list of standard errors for Table 5.3.

From Table 5.4 we concluded that Hartung's estimator and CMINQUE should not be used when γ is small. It is interesting to note that $\hat{\sigma}_a^2(\frac{1}{2n})$ and $\hat{\sigma}_a^2(M)$ almost dominate the ANOVA estimator across the parameter space and they should be used when $\gamma < 1$. The approach putting the negative ANOVA estimates to zero, which is also the REML approach, produces smaller mean squared error than the ANOVA estimator. Even smaller is the approach of putting the negative ANOVA estimates into the minimum bias range MINQ estimate with $\hat{\gamma} = \frac{1}{2n}$. The maximum likelihood is quite good in this case. However, for the multi-way models there is no theoretical result on the negative solutions to the likelihood equation, it is not yet known what approach should be adopted when having negative solutions from the maximum likelihood equation. The iterative minimum bias range MINQ estimator is quite good and can be used when there is no prior information available on σ_a^2 and σ_e^2.

To conclude when there is prior information available on both σ_a^2 and σ_e^2, we can use Hartung's estimator, $\hat{\sigma}_a^2(\frac{1}{2n})$ or $\hat{\sigma}_a^2(M)$ in different situations. If $\sigma_a^2 > \sigma_e^2$, then Hartung's estimator should be used; If $\sigma_a^2 < \sigma_e^2$, then $\hat{\sigma}_a^2(\frac{1}{2n})$ or $\hat{\sigma}_a^2(M)$ should be used. When there is no information available on σ_a^2 and σ_e^2, we can use the iterative minimum bias range MINQ estimator which does not need any prior values. Alternatively, we can start with the ANOVA estimator and if a negative value appears we can then change to use the minimum bias range MINQ estimator with $\hat{\gamma} = \frac{1}{2n}$ which is nonnegative and better than putting the negative ANOVA estimates into zero.

Table 5.1: Numerical Comparison

The relative bias of the nonnegative estimators

Relative bias of $\hat{\sigma}_a^2 = \frac{|E(\hat{\sigma}_a^2)-\sigma_a^2|}{\sigma_a^2}$. $\gamma = \sigma_a^2/\sigma_e^2$

γ	n	$\hat{\sigma}_a^2(H)$	$\hat{\sigma}_a^2(\frac{1}{2n})$	$\hat{\sigma}_a^2(M)$	$\hat{\sigma}_a^2(C)$	$\hat{\sigma}_a^2(\gamma)$	$\hat{\sigma}_a^2(M\gamma)$
	2	1.80	0.75	0.40	2.50	0.56	0.74
0.2	3	1.40	0.60	0.35	1.67	0.71	0.66
	5	0.92	0.43	0.25	1.00	0.67	0.55
	10	0.49	0.25	0.12	0.50	0.43	0.40
	2	0.80	0.13	0.01	1.25	0.38	0.56
0.4	3	0.65	0.10	0.02	0.83	0.43	0.46
	5	0.44	0.07	0.03	0.50	0.36	0.34
	10	0.24	0.04	0.02	0.25	0.22	0.21
	2	0.47	0.08	0.19	0.83	0.29	0.46
0.6	3	0.40	0.07	0.13	0.56	0.31	0.36
	5	0.28	0.05	0.08	0.33	0.25	0.26
	10	0.16	0.03	0.05	0.17	0.15	0.15
	2	0.30	0.19	0.29	0.63	0.24	0.39
0.8	3	0.27	0.15	0.21	0.42	0.24	0.30
	5	0.20	0.11	0.14	0.25	0.19	0.21
	10	0.11	0.06	0.08	0.13	0.11	0.12
	2	0.20	0.25	0.34	0.50	0.20	0.34
1.0	3	0.20	0.20	0.26	0.33	0.20	0.26
	5	0.15	0.14	0.18	0.20	0.15	0.18
	10	0.09	0.08	0.10	0.10	0.09	0.10
	2	0.00	0.38	0.45	0.25	0.11	0.21
2.0	3	0.05	0.30	0.35	0.17	0.11	0.15
	5	0.06	0.21	0.24	0.10	0.08	0.10
	10	0.04	0.13	0.14	0.05	0.04	0.06
	2	0.12	0.45	0.52	0.10	0.05	0.10
5.0	3	0.04	0.36	0.41	0.07	0.04	0.07
	5	0.00	0.26	0.29	0.04	0.03	0.05
	10	0.01	0.15	0.17	0.02	0.02	0.03

Table 5.2: Numerical Comparison

The relative efficiency of the estimators

Relative efficiency $e(\hat{\sigma}_a^2) = \frac{\text{MSE}(\hat{\sigma}_a^2)}{\text{V}(\hat{\sigma}_a^2(\text{ANOVA}))}$. $\gamma = \sigma_a^2/\sigma_e^2$

γ	n	$\hat{\sigma}_a^2(H)$	$\hat{\sigma}_a^2(\frac{1}{2n})$	$\hat{\sigma}_a^2(M)$	$\hat{\sigma}_a^2(C)$	$\hat{\sigma}_a^2(\gamma)$	$\hat{\sigma}_a^2(\frac{\gamma}{n})$	$\hat{\sigma}_a^2(M\gamma)$
	2	1.25	0.31	0.15	2.26	0.21	0.06	0.14
0.2	3	1.74	0.50	0.28	2.35	0.63	0.12	0.25
	5	1.78	0.68	0.44	2.01	1.13	0.24	0.37
	10	1.44	0.81	0.57	1.48	1.26	0.44	0.47
	2	5.28	0.31	0.15	2.47	1.40	0.15	2.38
0.4	3	6.38	0.46	0.29	10.19	3.08	0.27	2.91
	5	5.09	0.61	0.48	6.33	3.64	0.43	2.66
	10	2.76	0.75	0.68	2.97	2.47	0.64	1.76
	2	3.19	0.29	0.62	9.32	1.47	0.25	2.66
0.6	3	3.84	0.42	0.64	6.88	2.55	0.39	2.70
	5	3.09	0.56	0.66	4.03	2.58	0.55	2.12
	10	1.85	0.72	0.74	2.01	1.76	0.73	1.36
	2	2.03	0.79	1.50	7.26	1.43	0.33	2.62
0.8	3	2.57	0.88	1.35	5.07	2.15	0.48	2.38
	5	2.19	0.87	1.09	2.95	2.04	0.63	1.75
	10	1.49	0.85	0.92	1.61	1.46	0.79	1.17
	2	1.37	1.46	2.47	5.86	1.37	0.40	2.47
1.0	3	1.87	1.43	2.09	3.97	1.87	0.55	2.09
	5	1.73	1.20	1.51	2.36	1.73	0.69	1.51
	10	1.30	0.98	1.08	1.41	1.30	0.83	1.08
	2	0.62	4.53	6.36	2.87	1.14	0.62	1.70
2.0	3	0.89	3.60	4.73	1.99	1.29	0.73	1.35
	5	1.06	2.38	2.90	1.41	1.21	0.82	1.08
	10	1.05	1.40	1.56	1.11	1.08	0.91	0.95
	2	1.22	8.46	11.00	1.40	0.99	0.82	1.07
5.0	3	0.88	5.99	7.47	1.19	1.04	0.88	0.98
	5	0.92	3.54	4.19	1.07	1.03	0.92	0.94
	10	0.98	1.76	1.97	1.02	1.01	0.96	0.95

Table 5.3: Empirical Comparison

The relative bias of the estimators

Relative bias of $\hat{\sigma}_a^2 = \frac{|E(\hat{\sigma}_a^2)-\sigma_a^2|}{\sigma_a^2}$. $\gamma = \sigma_a^2/\sigma_e^2$

γ	m	n	$\hat{\sigma}_a^2(A)$	$\hat{\sigma}_a^2(0)$	$\hat{\sigma}_a^2(A\frac{1}{2n})$	$\hat{\sigma}_a^2(ML)$	$\hat{\sigma}_a^2(\frac{1}{2n})$	$\hat{\sigma}_a^2(H)$	$\hat{\sigma}_a^2(R)$	$\hat{\sigma}_a^2(M)$	$\hat{\sigma}_a^2(C)$
		2	0.51	0.55	1.38	0.19	1.70	3.32	2.07	1.16	4.39
0.1	10	3	0.37	0.17	0.83	0.10	1.38	2.56	1.86	1.00	2.96
		5	0.18	0.02	0.36	0.21	1.02	1.72	1.43	0.77	1.83
		10	0.09	0.04	0.06	0.22	0.60	0.91	0.82	0.43	0.92
		2	0.62	0.41	0.01	0.45	0.46	1.34	0.65	0.27	1.92
0.2	50	3	0.41	0.36	0.15	0.40	0.37	1.05	0.71	0.26	1.28
		5	0.25	0.24	0.19	0.27	0.26	0.69	0.56	0.20	0.76
	25	10	0.22	0.22	0.21	0.27	0.08	0.29	0.25	0.03	0.30
		2	0.62	0.56	0.43	0.58	0.30	0.12	0.11	0.39	0.40
0.5	50	3	0.44	0.43	0.38	0.45	0.25	0.12	0.02	0.31	0.24
		5	0.26	0.26	0.26	0.28	0.18	0.10	0.06	0.22	0.14
	25	10	0.22	0.22	0.22	0.26	0.18	0.02	0.03	0.22	0.01
		2	0.64	0.62	0.57	0.63	0.56	0.30	0.38	0.62	0.13
1.0	50	3	0.47	0.46	0.45	0.48	0.47	0.21	0.23	0.51	0.12
		5	0.28	0.28	0.28	0.30	0.34	0.11	0.12	0.37	0.08
	25	10	0.23	0.23	0.23	0.27	0.27	0.13	0.13	0.30	0.12
		2	0.66	0.65	0.64	0.67	0.70	0.52	0.53	0.74	0.40
2.0	50	3	0.49	0.49	0.49	0.50	0.59	0.38	0.37	0.62	0.31
		5	0.29	0.29	0.29	0.31	0.42	0.22	0.21	0.45	0.19
	25	10	0.24	0.24	0.24	0.27	0.31	0.19	0.18	0.35	0.18
		2	0.69	0.69	0.68	0.69	0.79	0.67	0.64	0.82	0.58
5.0	50	3	0.51	0.51	0.51	0.53	0.66	0.49	0.46	0.69	0.44
		5	0.31	0.31	0.31	0.32	0.48	0.29	0.27	0.50	0.27
	25	10	0.24	0.24	0.24	0.27	0.34	0.22	0.21	0.37	0.21

Table 5.4: Empirical Comparison

The relative efficiency of the estimators

Relative efficiency $e(\hat\sigma_a^2) = \frac{\mathrm{MSE}(\hat\sigma_a^2)}{\mathrm{V}(\hat\sigma_a^2(\mathrm{ANOVA}))}$. $\gamma = \sigma_a^2/\sigma_e^2$

γ	m	n	$\hat\sigma_a^2(0)$	$\hat\sigma_a^2(A\frac{1}{2n})$	$\hat\sigma_a^2(ML)$	$\hat\sigma_a^2(\frac{1}{2n})$	$\hat\sigma_a^2(H)$	$\hat\sigma_a^2(R)$	$\hat\sigma_a^2(M)$	$\hat\sigma_a^2(C)$	-no.
		2	0.45	0.46	0.31	0.41	1.37	0.99	0.22	2.33	448
0.1	10	3	0.53	0.50	0.37	0.67	1.99	1.49	0.41	2.59	426
		5	0.70	0.58	0.54	0.95	2.22	1.86	0.63	2.47	306
		10	0.87	0.72	0.73	1.11	1.93	1.78	0.76	1.99	152
		2	0.58	0.35	0.57	0.28	1.59	0.88	0.15	3.12	347
0.2	50	3	0.82	0.53	0.82	0.44	2.08	1.41	0.31	2.94	215
		5	0.96	0.82	0.96	0.57	1.80	1.51	0.47	2.07	70
	25	10	1.00	0.97	1.02	0.60	1.11	1.05	0.52	1.15	9
		2	0.83	0.60	0.84	0.23	0.34	0.53	0.27	0.69	277
0.5	50	3	0.96	0.83	0.97	0.35	0.56	0.66	0.38	0.77	117
		5	0.99	0.97	0.99	0.50	0.78	0.83	0.51	0.87	20
	25	10	1.00	1.00	1.03	0.69	0.76	0.78	0.71	0.78	0
		2	0.93	0.81	0.94	0.54	0.40	0.63	0.60	0.46	207
1.0	50	3	0.99	0.94	0.99	0.64	0.57	0.70	0.69	0.63	68
		5	1.00	0.99	1,00	0.71	0.76	0.83	0.73	0.80	6
	25	10	1.00	1.00	1.03	0.87	0.82	0.83	0.92	0.82	0
		2	0.97	0.92	0.98	0.80	0.63	0.77	0.86	0.60	164
2.0	50	3	1.00	0.98	1.00	0.87	0.72	0.81	0.92	0.74	47
		5	1.00	1.00	1.00	0.88	0.84	0.89	0.91	0.86	1
	25	10	1.00	1.00	1.03	1.01	0.89	0.90	1.07	0.89	0
		2	0.99	0.98	1.00	1.00	0.84	0.90	1.04	0.80	119
5.0	50	3	1.00	1.00	1.00	1.03	0.88	0.91	1.08	0.87	25
		5	1.00	1.00	1.00	0.99	0.92	0.95	1.02	0.93	0
	25	10	1.00	1.00	1.03	1.10	0.95	0.95	1.17	0.94	0

Chapter 6

THE USE OF MINQUE IN ESTIMATION OF INTERVIEWER'S VARIANCE IN COMPLEX SURVEY

In previous chapters quadratic estimators of variance components, especially MINQUE, have been considered. Properties of MINQUE have been discovered and discussed. In this chapter we consider an example of the use of MINQUE in practice. Particularly we consider using MINQUE to estimate the component of variance due to interviewers in a complex survey.

A model-based approach is adopted in this chapter.

6.1 The need for variance components estimation

In this section we use a simple model to demonstrate the effect of failing to estimate variance components which may be present when estimating the mean parameter. Assume that there are m observers and that each records n_i observations on a random variable, y. y_{ij} is the jth record by the ith observer. We may use the following model:

$$y_{ij} = \mu + \varepsilon_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i, \tag{6.1}$$

where $\mathrm{E}(y_{ij}) = \mu$, $\mathrm{V}(y_{ij}) = \sigma_\varepsilon^2$, and $\mathrm{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$, $i, j \neq i', j'$. We wish to estimate μ and $\mathrm{V}(\hat{\mu})$.

If in the process of observing y_{ij} there are random effects due to the observers, then instead of having y_{ij}, the true value, we have y^*_{ij} recorded. We assume that the measurement error structure is:

$$y^*_{ij} = y_{ij} + a_i + \varepsilon^*_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i, \tag{6.2}$$

where a_i is the independent random effect of the ith observer with $\mathrm{E}(a_i) = 0$ and $\mathrm{V}(a_i) = \sigma^2_a$. We call σ^2_a the component of variance due to observers. ε^*_{ij} is the independent random error at each recording with $\mathrm{E}(\varepsilon^*_{ij}) = 0$ and $\mathrm{V}(\varepsilon^*_{ij}) = \sigma^{*2}_\varepsilon$.

The relationship between y^*_{ij} and μ can be set up by combining (6.1) and (6.2):

$$y^*_{ij} = \mu + a_i + \varepsilon_{ij} + \varepsilon^*_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i. \tag{6.3}$$

It is usually impossible to estimate σ^2_ε and σ^{*2}_ε separately from (6.3). Letting $e_{ij} = \varepsilon_{ij} + \varepsilon^*_{ij}$, then $\mathrm{E}(e_{ij}) = 0$, and $\mathrm{V}(e_{ij}) = \mathrm{V}(\varepsilon_{ij} + \varepsilon^*_{ij}) \stackrel{\mathrm{def}}{=} \sigma^2_e$, (6.3) can then be rewritten as:

$$y^*_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i. \tag{6.4}$$

We shall employ model (6.4) for taking measurement errors into account. It can be seen that it is possible to estimate σ^2_e from model (6.4), but not σ^2_ε and σ^{*2}_ε separately.

Failing to include the measurement errors when they are present implies that we are using the improper model,

$$y^*_{ij} = \mu + \varepsilon_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n_i. \tag{6.5}$$

where $\mathrm{E}(\varepsilon_{ij}) = 0$, $\mathrm{V}(\varepsilon_{ij}) = \sigma^2_\varepsilon$ and $\mathrm{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$, $i, j \neq i', j'$.

If there are measurement errors in observing the random samples, model (6.4) is the correct model to use. Now we investigate the effect of using the improper model (6.5) in this situation. Suppose that a generalized least squares estimator is used to estimate μ, then under model (6.4):

$$\hat{\mu} = \frac{\sum_{i=1}^m w_i n_i \bar{y}_i}{\sum_{i=1}^m w_i n_i}, \tag{6.6}$$

where $w_i = \frac{\sigma^2_e}{\sigma^2_e + n_i \sigma^2_a}$, $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y^*_{ij}$ and the variance of the estimator $\hat{\mu}$ is:

$$\mathrm{V}(\hat{\mu}) = \frac{\sigma^2_e}{\sum_{i=1}^m w_i n_i}. \tag{6.7}$$

If we ignore the measurement errors and use the generalized least squares estimator to estimate μ from model (6.5), we have:

$$\tilde{\mu} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}^*, \tag{6.8}$$

and the variance of $\tilde{\mu}$ under model (6.5) is:

$$\mathrm{V}(\tilde{\mu}) = \frac{\sigma_\varepsilon^2}{N}, \tag{6.9}$$

where $N = \sum_{i=1}^{m} n_i$.

Comparing (6.6) and (6.8) it can be seen that because $\sigma_a^2 > 0$, $0 < w_i < 1$,

$$\begin{aligned}\hat{\mu} - \tilde{\mu} &= \frac{1}{\sum_{i=1}^{m} w_i n_i}\sum_{i=1}^{m} w_i n_i \bar{y}_i - \frac{1}{N}\sum_{i=1}^{m} n_i \bar{y}_i \\ &\begin{cases} = 0, & \text{when } n_i = n, \\ \neq 0, & \text{otherwise,}\end{cases}\end{aligned}$$

i.e. when there are equal observations the estimators of μ under the two models are identical. Otherwise, they yield different estimates. Comparing (6.7) and (6.9) it can be seen that

$$\begin{aligned}\mathrm{V}(\hat{\mu}) - \mathrm{V}(\tilde{\mu}) &= \frac{\sum_{i=1}^{m}(1-w_i)n_i}{N\sum_{i=1}^{m} w_i n_i}\sigma_\varepsilon^2 + \frac{1}{\sum_{i=1}^{m} w_i n_i}(\sigma_e^2 - \sigma_\varepsilon^2) \\ &> 0,\end{aligned}$$

because $\sigma_e^2 = \mathrm{V}(\varepsilon_{ij} + \varepsilon_{ij}^*) > \sigma_\varepsilon^2$.

In the balanced data case where $n_i = n$, for $i = 1, \ldots, m$,

$$\mathrm{V}(\hat{\mu}) - \mathrm{V}(\tilde{\mu}) = \frac{\sigma_e^2 - \sigma_\varepsilon^2}{mn} + \frac{\sigma_a^2}{m}.$$

It can be noted that when σ_a^2 is large $\mathrm{V}(\hat{\mu})$ can be very different from $\mathrm{V}(\tilde{\mu})$.

If we ignore the presence of measurement errors and adopt the standard textbook formulas, then $\frac{S^2}{N}$ is used to estimate $\mathrm{V}(\hat{\mu})$, where $S^2 = \frac{1}{N-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}^* - \bar{y}_{..}^*)^2$. If model (6.5) is the correct model, then under model (6.5) $\mathrm{E}(\frac{S^2}{N}) = \frac{\sigma_\varepsilon^2}{N} = \mathrm{V}(\tilde{\mu})$, *i.e.* $\frac{S^2}{N}$ is an unbiased estimator of $\mathrm{V}(\tilde{\mu})$. The problem is whether $\frac{S^2}{N}$ is an unbiased estimator of $\mathrm{V}(\hat{\mu})$ under model (6.4).

With the assumption that model (6.4) is the correct model we see that:

$$\mathrm{E}(S^2) = \sigma_e^2 + \frac{N^2 - \sum_{i=1}^{n_i} n_i^2}{N(N-1)}\sigma_a^2,$$

hence

$$E(\frac{S^2}{N}) = \frac{\sigma_e^2}{N} + \frac{N^2 - \sum_{i=1}^m n_i^2}{N^2(N-1)}\sigma_a^2,$$

which is not equal to $V(\hat{\mu})$ of (6.7). The difference is

$$V(\hat{\mu}) - E(\frac{S^2}{N}) = \frac{1}{\sum_{i=1}^m w_i n_i}\sigma_e^2 - \frac{\sigma_e^2}{N} - \frac{N^2 - \sum_{i=1}^m n_i^2}{N^2(N-1)}\sigma_a^2$$

$$= \frac{\sum_{i=1}^m (1-w_i)n_i}{N\sum_{i=1}^m (1-w_i)}\sigma_a^2 - \frac{N^2 - \sum_{i=1}^m n_i^2}{N^2(N-1)}\sigma_a^2$$

$$= \frac{N^2\sum_{i=1}^m (1-w_i)(n_i-2) + N\sum_{i=1}^m (1-w_i)(N-n_i) + (\sum_{i=1}^m n_i^2)\sum_{i=1}^m (1-w_i)}{N^2(N-1)\sum_{i=1}^m (1-w_i)}\sigma_a^2$$

$$> 0,$$

because for the estimation of σ_a^2 and σ_e^2, we assume $n_i \geq 2$.

In the balanced data case of $n_i = n$, for $i = 1, \ldots, m$,

$$V(\hat{\mu}) - E(\frac{S^2}{N}) = \frac{n-1}{mn-1}\sigma_a^2 > 0.$$

$\frac{S^2}{N}$ will underestimate $V(\hat{\mu})$ and the underestimation will be serious if σ_a^2 is large. It can be seen that the bias of $\frac{S^2}{N}$ is due to the presence of σ_a^2.

We have demonstrated that by ignoring the measurement errors we will underestimate the variance of the estimator of the mean, and hence overstate the accuracy of estimation. Thus it is necessary to take measurement errors into account and to estimate the variance components due to the measurement errors. By doing so it is hoped that we can then derive an unbiased estimator for the variance of the estimator of the mean and we can also assess the performance of the observers involved. In the next section we shall consider the estimation of interviewer's variance in surveys as an example of this type of measurement error.

6.2 Estimation of interviewer's variance in surveys

In a complex survey there are usually many operators involved at different stages of the survey, *e.g.* interviewers at the interview stage, coders at the data processing stage, etc. These operators can bring measurement errors into the survey and the contribution of these nonsampling errors to the total variance sometimes turns out to be larger than the contribution of sampling error. Hansen, Hurwitz and Bershad (1959) developed a general survey error model to identify different error sources and to measure their impact. Ignoring the measurement errors and applying "standard textbook formulas" for the estimation of the variances of survey estimates may lead to serious underestimation of the real variability, as we showed in Section 6.1. Therefore it is necessary to estimate the variance components due to different error sources. In this section we consider the estimation of the effect of one error source only: namely the interviewer's component of variance.

Consider a stratified multistage survey. It is only at the last stage that we use interviewers and need to design an interview scheme. If there are I interviewers available, select I groups of units from the last stage of the original survey such that the ith group contains M_i units.

Let Π={ units contained in the I groups }.

In this chapter we shall treat Π as our population and design an interview scheme for Π and estimate parameters relating to Π. All conclusions will therefore be conditional on the selection of Π. When we condition on Π the groups can be considered as strata. But since Π is the last stage in the original survey design, we can draw inference about the total population by estimation over the other stages of the survey.

The survey design at the interview stage is: use simple random sampling to choose m_i units from the M_i units of each stratum. By conditioning on Π and treating the I groups as strata the above design can be considered as a design using stratified simple random sampling. A natural way to assign the interviewers is to send the ith interviewer to the ith stratum to do the m_i interviews. We wish to draw inference on each stratum and Π. Let y_{is} be the result of the sth interview carried out by the ith interviewer in the ith stratum. Assume that the interviewers have a systematic random effect on the interview and also a random error is present at each interview. If we denote the true result of the interview by η_{is}, then with an additive error structure we have:

$$y_{is} = \eta_{is} + b_i + \varepsilon^*_{is}, \quad i = 1, \ldots, I, \; s = 1, \ldots, m_i, \tag{6.10}$$

where b_i is the random effect of the ith interviewer, with $\mathrm{E}(b_i) = 0$, $\mathrm{V}(b_i) = \sigma_b^2$, and ε^*_{is} is the random error with $\mathrm{E}(\varepsilon^*_{is}) = 0$ and $\mathrm{V}(\varepsilon^*_{is}) = \sigma_\varepsilon^{*2}$.

Now assume that the true results of the interviews are random samples from a distribution with $\mathrm{E}(\eta_{is}) = \eta_i$, $\mathrm{V}(\eta_{is}) = \sigma_\varepsilon^2(i)$, *i.e.* we assume that the true results of each stratum have a common mean and variance, then

$$\eta_{is} = \eta_i + \varepsilon_{is}, \quad i = 1, \ldots, I, \ j = 1, \ldots, m_i. \tag{6.11}$$

The model relating y_{is} to η_i is:

$$y_{is} = \eta_i + b_i + e_{is}, \quad i = 1, \ldots, I, \ s = 1, \ldots, m_i, \tag{6.12}$$

where $e_{is} = \varepsilon_{is} + \varepsilon^*_{is}$ with $\mathrm{E}(e_{is}) = \mathrm{E}(\varepsilon_{is}) + \mathrm{E}(\varepsilon^*_{is}) = 0$, and $\mathrm{V}(e_{is}) = \mathrm{V}(\varepsilon_{is} + \varepsilon^*_{is}) = \mathrm{V}(\eta_{is} - \eta_i) + \sigma^{*2}_\varepsilon \stackrel{\text{def}}{=} \sigma^2_e(i)$. Hence $\sigma^2_e(i)$ is the sum of sampling errors from the simple random sampling selection of the m_i units out of M_i units in each stratum and the random errors in the recording of each interview.

We wish to estimate η_i and σ^2_b. However, in model (6.12) the stratum effect η_i and the random effect b_i are confounded. It is impossible to estimate η_i and σ^2_b from model (6.12). To break the confounding an interpenetration interview scheme must be used.

6.3 Interpenetration interview scheme

Instead of assigning one interviewer to one stratum as in the interview scheme considered in Section 6.2, we could assign k interviewers to one stratum, $k \geq 2$. The most commonly used scheme is the pair interpenetration scheme where $k = 2$.

Let Π be selected as in Section 6.2, group the strata into l nonoverlapping blocks each containing k strata (assume that $l = I/k$ is an integer). Assign k interviewers to each block to share the interview workload in each of the strata of the block. In this chapter we shall only consider equal workload among the interviewers, *i.e.* each interviewer does $\frac{1}{k}$ of the interviews in each stratum. Unequal workload can be modeled, but the estimator of σ^2_b will be more difficult to derive.

The combination (γ, t) is used to locate the tth stratum in the γth block and (γ, j) locates the $[(\gamma - 1)k + j]$th interviewer among the I interviewers available. We assume each interviewer brings a systematic random effect $b_{\gamma j}$ to the interview result. $y_{\gamma tjs}$ is the sth recorded interview result by the jth interviewer in the (γ, t)th stratum.

The model for the interviewer's random effect is:

$$y_{\gamma tjs} = \eta_{\gamma tjs} + b_{\gamma j} + \varepsilon^*_{\gamma tjs}, \tag{6.13}$$

$$\gamma = 1, \ldots, l,\ t, j = 1, \ldots, k,\ s = 1, \ldots, \frac{m_{\gamma t}}{k},$$

where $\eta_{\gamma tjs}$ is the true value which should be recorded, $b_{\gamma j}$ is the random effect of the $[(\gamma - 1)k + j]$th interviewer and we assume that $b_{\gamma j}$ is a random sample from a distribution with $\mathrm{E}(b_{\gamma j}) = 0$, $\mathrm{V}(b_{\gamma j}) = \sigma_b^2$, $\varepsilon^*_{\gamma tjs}$ is the random error in the recording of the interview.

The model for the true mean of the stratum is:

$$\eta_{\gamma tjs} = \eta_{\gamma t} + \varepsilon_{\gamma tjs}. \tag{6.14}$$

Hence the model relating $y_{\gamma tjs}$ to $\eta_{\gamma t}$ is:

$$y_{\gamma tjs} = \eta_{\gamma t} + b_{\gamma j} + e_{\gamma tjs}, \tag{6.15}$$

$$\gamma = 1, \ldots, l,\ t, j = 1, \ldots, k,\ s = 1, \ldots, \frac{m_{\gamma t}}{k},$$

and where $e_{\gamma tjs} = \varepsilon_{\gamma tjs} + \varepsilon^*_{\gamma tjs}$.

It can be seen that by employing the interpenetration interview scheme the confounding of η_i and b_i in model (6.12) is broken in model (6.15). It is therefore possible to estimate $\eta_{\gamma t}$ and σ_b^2 separately from model (6.15).

The disadvantage of adopting the interpenetration interview scheme in practice is that it is usually expensive. For instance, if the strata are selected with geographical convenience, the interpenetration interview scheme sends each interviewer into k different areas while the non-interpenetration interview scheme sends one interviewer to one area only. Interpenetration interview scheme will usually increase the cost of the survey.

In the next section we shall design our survey to use a partial interpenetration interview scheme which can break the confounding of the stratum and the interviewer's effects and is cheaper than a total interpenetration interview scheme.

6.4 Design of the survey and the model used

Let Π be selected as in Section 6.2 and partitioned arbitrarily into L nonoverlapping blocks, each containing k strata (assume $L = I/k$ is an integer and $k \geq 2$). Suppose the (γ, t)th stratum contains $M_{\gamma t}$ units. We have two important assumptions to make in the design of survey:

1. Use simple random sampling to choose m units from the (γ, t)th stratum;
2. Select a random sample of l blocks from the L blocks.

Apply two different interview allocation schemes to the selected l blocks and the remaining $L-l$ blocks, respectively.

Pattern 1 (applied to the selected l blocks):

Adopt an interpenetration interview scheme in these blocks. The k interviewers split the workload equally in each stratum of a block.

Assume that each interviewer carries out f interviews, $f=\frac{m}{k}$.

Pattern 2 (applied to the remaining $L-l$ blocks):

Adopt non-interpenetration interview scheme. Allocate each interviewer to each stratum in the block.

From Section 6.3 for the strata applied pattern 1 scheme we have:

$$y^{(1)}_{\gamma tjs}=\eta_{\gamma t}+b_{\gamma j}+e_{\gamma tjs}, \tag{6.16}$$

$$\gamma=1,\ldots,l,\ t,j=1,\ldots,k,\ s=1,\ldots,f.$$

From Section (6.2) for the strata applied pattern 2 scheme we have

$$y^{(2)}_{\gamma ts}=\eta_{\gamma t}+b_{\gamma t}+e_{\gamma ts}, \tag{6.17}$$

$$\gamma=l+1,\ldots,L,\ t=1,\ldots,k,\ s=1,\ldots,m.$$

We shall assume that

$$\mathrm{V}(e_{\gamma tjs})=\mathrm{V}(e_{\gamma ts})\stackrel{\mathrm{def}}{=}\sigma_e^2(\gamma,t).$$

Therefore $\sigma_e^2(\gamma,t)$ is the sum of sampling errors from the selection of the m units in the survey and random errors caused by the recording of each interview. We shall show that if we estimate $\eta_{\gamma t}$ with the generalized least squares estimator and estimate σ_b^2 with MINQUE it is sufficient to estimate $\sigma_e^2(\gamma,t)$ and derive an unbiased estimator for $\mathrm{V}(\hat{\eta}_{\gamma t})$. Therefore there is no need to estimate the sampling errors and the random errors consisted in $\sigma_e^2(\gamma,t)$ separately.

Models (6.16) and (6.17) can be written in matrix forms as:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\eta}+\mathbf{U}_b\mathbf{b}+\sum_{\gamma=1}^{L}\sum_{t=1}^{k}\mathbf{U}_{\gamma t}\mathbf{e}_{\gamma t}, \tag{6.18}$$

where $\mathbf{X}$ is the design matrix for the fixed effect, $\mathbf{U}_b$ for the interviewer's effect, $\mathbf{U}_{\gamma t}$ for the errors, also:

$$\mathrm{E}(\mathbf{b})=0,\quad \mathrm{V}(\mathbf{b})=\sigma_b^2\mathbf{I}_{Lk},$$

$$\mathrm{E}(\mathbf{e}_{\gamma t})=0,\quad \mathrm{V}(\mathbf{e}_{\gamma t})=\sigma_e^2(\gamma,t)\mathbf{I}_{Lkm}.$$

Now we arrange the data vector $\mathbf{y}$ in such an order that the interpenetrated data come first and the non-interpenetration data follow. Within each data set the data are written in lexicographical order, then (6.18) can be rewritten as:

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}^{(1)} \\ \boldsymbol{\eta}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{U}_1^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{b}^{(1)} \\ \mathbf{b}^{(2)} \end{bmatrix} + \sum_{\gamma=1}^{L} \sum_{t=1}^{k} \mathbf{U}_{\gamma t} \mathbf{e}_{\gamma t}, \tag{6.19}$$

where $y^{(1)}$, $\boldsymbol{\eta}^{(1)}$, $\mathbf{b}^{(1)}$ and $\mathbf{e}^{(1)}$ are the parameters for the data set of design pattern 1, $y^{(2)}$, $\boldsymbol{\eta}^{(2)}$, $\mathbf{b}^{(2)}$ and $\mathbf{e}^{(2)}$ are those of design pattern 2. $\mathbf{X}^{(1)}$ and $\mathbf{U}_1^{(1)}$ are design matrices for pattern 1, $\mathbf{X}^{(2)}$ and $\mathbf{U}_1^{(2)}$ are design matrices for pattern 2. $\mathbf{U}_{\gamma t}$ is design matrix for the error term, $\gamma = 1, \ldots, L$, $t = 1, \ldots, k$. Throughout this chapter we shall use superscripts 1 and 2 to denote parameters for interpenetrated data set and non-interpenetrated data set, respectively, and use subscripts to denote the order of the design matrices in the model.

The design matrices are:

$$\mathbf{X}^{(1)} = \mathbf{I}_l \otimes \mathbf{I}_k \otimes \mathbf{1}_m, \quad \mathbf{X}^{(2)} = \mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{1}_m,$$

$$\mathbf{U}_1^{(1)} = \mathbf{I}_l \otimes \mathbf{1}_k \otimes \mathbf{I}_k \otimes \mathbf{1}_f, \; \mathbf{U}_1^{(2)} = \mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{1}_m,$$

$$\mathbf{U}_{\gamma t} = \begin{bmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \mathbf{I}_m & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix},$$

where $\mathbf{I}_m$ is the $[(\gamma-1)k+t]$th submatrix of order m.

The variance covariance matrix of $\mathbf{y}$ is:

$$\mathbf{V} = \sigma_b^2 \mathbf{V}_1 + \sum_{\gamma=1}^{L} \sum_{t=1}^{k} \sigma_e^2(\gamma, t) \mathbf{V}_{\gamma,t} \tag{6.20}$$

where

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{V}_1^{(1)} & \\ & \mathbf{V}_1^{(2)} \end{bmatrix},$$

and

$$\mathbf{V}_1^{(1)} = \mathbf{I}_l \otimes \mathbf{J}_k \otimes \mathbf{I}_k \otimes \mathbf{J}_f, \; \mathbf{V}_1^{(2)} = \mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{J}_m.$$

$$\mathbf{V}_{\gamma t} = \mathbf{U}_{\gamma t} \mathbf{U}'_{\gamma t}, \quad \gamma = 1, \ldots, L, \; t = 1, \ldots, k.$$

The task here is to estimate σ_b^2 efficiently and find the optimal design for the estimator used. In the following section we shall introduce the existing estimators in the literature and then use MINQUE in section 6.6.

6.5 Fellegi's estimator and Biemer and Stokes' estimators

Fellegi (1974) considered an interpenetration interview scheme for the case of $k = 2$ in the design of survey in section 6.4 and proposed an unbiased estimator of interviewer's variance. The estimator is a linear combination of two estimators where the first estimator used the interpenetrated data set only and the second estimator is derived using the whole data set. Fellegi speculated that the two estimators are independent of each other hence the composite estimator will be more efficient than either of the two estimators used alone.

Biemer and Stokes (1985) extended Fellegi's estimators from a pairwise interpenetration scheme into l groups of k interviewer assignments for multistage survey designs. The survey design in section 6.4 follows from Biemer and Stokes' set up. Biemer and Stokes proved that with a normality assumption the two estimators of Fellegi are independent and they proposed to use the variances of the estimators as the weights in the linear combination of the two estimators to form a composite estimator so that the composite estimator will be more efficient than either of the two estimators used alone.

Biemer and Stokes considered model (6.18) which is equivalent to model (6.19). Since the variance covariance matrix of $\mathbf{y}$ shown in (6.20) is complicated, Biemer and Stokes used the synthesis-based MINQUE, *i.e.* in the calculation of $\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ they used $\mathbf{V} = \mathbf{I}$ instead of the true variance covariance matrix. By using $\mathbf{V} = \mathbf{I}$ the derivation of a quadratic estimator from MINQUE equations is greatly simplified. Biemer and Stokes found that the estimator depends only on the interpenetrated data set. The estimator is:

$$\begin{aligned}\hat{\sigma}_b^2(BS1) &= \frac{1}{l\,[k(m-2)+1]}\left[\frac{k(m-1)}{k-1}\sum_{\gamma=1}^{l}\sum_{j=1}^{k}(\bar{y}_{\gamma\cdot j\cdot}^{(1)} - \bar{\bar{y}}_{\gamma\ldots}^{(1)})^2\right. \\ &\left.\quad -\frac{1}{m}\sum_{\gamma=1}^{l}\sum_{t=1}^{k}\sum_{j=1}^{k}\sum_{s=1}^{f}(y_{\gamma tjs}^{(1)} - \bar{y}_{\gamma t\ldots}^{(1)})^2\right], \qquad (6.21)\end{aligned}$$

where a dot in place of a subscript denotes summing over that subscript and where $\bar{\bar{y}}_{\gamma\ldots}^{(1)} = y_{\gamma\ldots}^{(1)}/km$, $\bar{y}_{\gamma\cdot j\cdot}^{(1)} = y_{\gamma\cdot j\cdot}^{(1)}/m$, $\bar{y}_{\gamma t\ldots}^{(1)} = y_{\gamma t\ldots}^{(1)}/m$.

The variance of $\sigma_b^2(BS1)$ is:

$$\begin{aligned} \mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right) &= \frac{2}{l(k-1)}\left\{\sigma_b^4 + \frac{2}{kml}\sigma_b^2\frac{l}{L}\sum_{\gamma=1}^{l}\sum_{t=1}^{k}\sigma_e^2(\gamma,t)\right\} \\ &+\frac{2(m-1)}{[(km-2k+1)ml]^2}\frac{l}{L}\sum_{\gamma=1}^{L}\left\{\frac{m-1}{k-1}\left[\sum_{t=1}^{k}\sigma_e^2(\gamma,t)\right]^2-\sum_{t=1}^{k}\sigma_e^4(\gamma,t)\right\} \end{aligned} \tag{6.22}$$

If $\sigma_e^2(\gamma,t)=\sigma_e^2$, $\gamma=1,\ldots,l$, $t=1,\ldots,k$, then

$$\mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right)=\frac{2}{l(k-1)}\left[\sigma_b^4+\frac{2}{m}\sigma_b^2\sigma_e^2+\frac{k(m-1)}{m^2(km-2k+1)}\sigma_e^4\right]. \tag{6.23}$$

Since this estimator only depends on the interpenetrated data set, Biemer and Stokes give a second estimator which depends on the whole data set:

$$\hat{\sigma}_b^2(BS2) = \frac{1}{k-1}\left[\frac{1}{L-l}\sum_{\gamma=l+1}^{L}\sum_{t=1}^{k}(\bar{y}_{\gamma t.}^{(2)}-\bar{\bar{y}}_{\gamma..}^{(2)})^2-\frac{1}{l}\sum_{\gamma=1}^{l}\sum_{t=1}^{k}(\bar{y}_{\gamma t..}^{(1)}-\bar{\bar{y}}_{\gamma...}^{(1)})^2\right]. \tag{6.24}$$

The composite estimator proposed by Biemer and Stokes is:

$$\hat{\sigma}_b^2(BS)=a\hat{\sigma}_b^2(BS1)+(1-a)\hat{\sigma}_b^2(BS2), \tag{6.25}$$

where

$$a=\frac{\mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right)}{\mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right)+\mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right)}.$$

With a normality assumption Biemer and Stokes proved that $\hat{\sigma}_b^2(BS1)$ and $\hat{\sigma}_b^2(BS2)$ are independent, and hence $\hat{\sigma}_b^2(BS)$ has smaller variance than either of $\hat{\sigma}_b^2(BS1)$ or $\hat{\sigma}_b^2(BS2)$.

$\hat{\sigma}_b^2(BS1)$, $\hat{\sigma}_b^2(BS2)$ and $\hat{\sigma}_b^2(BS)$ are unbiased estimators of σ_b^2 and that is the only optimality claimed for these estimators. A problem remains for the improvement of efficiency of $\hat{\sigma}_b^2(BS)$ over $\hat{\sigma}_b^2(BS1)$ and $\hat{\sigma}_b^2(BS2)$, because the weight a in (6.25) uses the variances of $\hat{\sigma}_b^2(BS1)$ and $\hat{\sigma}_b^2(BS2)$ which involve the unknown variance components σ_b^2 and $\sigma_e^2(\gamma,t)$. So in practice $\hat{\sigma}_b^2(BS)$ is unreachable. Putting an arbitrary weight in (6.25) would not necessarily lead to an improvement of efficiency.

Suppose we use an arbitrary value w, $w\neq 0,1$, in (6.25) to form the composite estimator of Biemer and Stokes,

$$\hat{\sigma}_b^2(BS)=w\hat{\sigma}_b^2(BS1)+(1-w)\hat{\sigma}_b^2(BS2).$$

Also assuming normality for $\mathbf{y}$, then since $\hat{\sigma}_b^2(BS1)$ and $\hat{\sigma}_b^2(BS2)$ are independent of each other,

$$\mathrm{V}\left(\hat{\sigma}_b^2(BS)\right) = w^2\,\mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right) + (1-w)^2\,\mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right).$$

When

$$\mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right) < \frac{(1-w)^2}{1-w^2}\,\mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right),$$

$\mathrm{V}\left(\hat{\sigma}_b^2(BS)\right) > \mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right)$, *i.e.* the composite estimator is less efficient than the first estimator used alone.

When

$$\mathrm{V}\left(\hat{\sigma}_b^2(BS1)\right) > \frac{1-(1-w)^2}{w^2}\,\mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right),$$

$\mathrm{V}\left(\hat{\sigma}_b^2(BS)\right) > \mathrm{V}\left(\hat{\sigma}_b^2(BS2)\right)$, *i.e.* the composite estimator is less efficient than the second estimator used alone.

6.6 The use of MINQUE

Ideally if we obtain the MINQUE for model (6.18) it will be the best quadratic unbiased estimator for σ_b^2 under the normality assumption, hence it will be at least as efficient as the composite estimator of Biemer and Stokes. With the complexity of the variance covariance matrix $\mathbf{V}$ of (6.20) it is very difficult to derive the MINQUE of σ_b^2 algebraically. It is also very difficult to assign prior values for the σ_b^2 and $\sigma_e^2(\gamma,t)$ apart from $\sigma_b^2 = 0$, $\sigma_e^2(\gamma,t) = 1$ chosen by Biemer and Stokes, because the number of variance components (which is $Lk+1$) is very large.

Instead of using the simplified estimator of σ_b^2 as Biemer and Stokes did we use a simplified model of (6.18) by assuming $\sigma_e^2(\gamma,t) = \sigma_e^2$, $\gamma = 1,\ldots,L$, $t = 1,\ldots,k$. This assumption is true when the strata are homogeneous. We use the following model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\eta} + \mathbf{U}_b\mathbf{b} + \mathbf{e} \tag{6.26}$$

where $\mathrm{E}(\mathbf{b}) = 0$, $\mathrm{V}(\mathbf{b}) = \sigma_b^2\mathbf{I}_{Lk}$, $\mathrm{E}(\mathbf{e}) = 0$, $\mathrm{V}(\mathbf{e}) = \sigma_e^2\mathbf{I}_{Lkm}$.

The design matrices for the variance components are:

$$\mathbf{U}_b = \begin{bmatrix} \mathbf{U}_b^{(1)} & \\ & \mathbf{U}_b^{(2)} \end{bmatrix},$$

where $\mathbf{U}_b^{(1)} = \mathbf{I}_l \otimes \mathbf{1}_k \otimes \mathbf{I}_k \otimes \mathbf{1}_f$, $\mathbf{U}_b^{(2)} = \mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{1}_m$,

$$\mathbf{U}_e = \begin{bmatrix} \mathbf{U}_e^{(1)} & \\ & \mathbf{U}_e^{(2)} \end{bmatrix},$$

where $\mathbf{U}_e^{(1)} = \mathbf{I}_{lkm}$, $\mathbf{U}_e^{(2)} = \mathbf{I}_{(L-l)km}$.

We use MINQUE to estimate σ_b^2 and σ_e^2.

Theorem 6.1 *If model* (6.26) *is used, the MINQUE of* σ_b^2 *is:*

$$\begin{aligned} \hat{\sigma}_b^2 &= \frac{1}{Lk(m-1)-l(k-1)} \left\{ \frac{Lk(m-1)}{l(k-1)} \sum_{\gamma=1}^{l} \sum_{j=1}^{k} (\bar{y}^{(1)}_{\gamma\cdot j\cdot} - \bar{\bar{y}}^{(1)}_{\gamma\cdots})^2 \right. \\ & \left. -\frac{1}{m} \sum_{\gamma=1}^{l} \sum_{t=1}^{k} \sum_{j=1}^{k} \sum_{s=1}^{f} (y^{(1)}_{\gamma tjs} - \bar{y}^{(1)}_{\gamma t\cdot\cdot})^2 - \frac{1}{m} \sum_{\gamma=l+1}^{L} \sum_{t=1}^{k} \sum_{s=1}^{m} (y^{(2)}_{\gamma ts} - \bar{y}^{(2)}_{\gamma t\cdot})^2 \right\} \end{aligned} \tag{6.27}$$

Proof: The variance covariance matrix of $\mathbf{y}$ in model (6.26) is

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{(1)} & \\ & \mathbf{V}^{(2)} \end{bmatrix},$$

where

$$\mathbf{V}^{(1)} = \sigma_b^2 \mathbf{I}_l \otimes \mathbf{J}_k \otimes \mathbf{I}_k \otimes \mathbf{J}_f + \sigma_e^2 \mathbf{I}_{lkm},$$

$$\mathbf{V}^{(2)} = \sigma_b^2 \mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{J}_m + \sigma_e^2 \mathbf{I}_{(L-l)km}.$$

Let $\Delta = m\sigma_b^2 + \sigma_e^2$, then from MINQUE formulas:

$$\begin{aligned} \mathbf{R} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \\ &= \begin{bmatrix} \mathbf{R}^{(1)} & \\ & \mathbf{R}^{(2)} \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{R}^{(1)} &= \mathbf{V}^{(1)^{-1}} - \mathbf{V}^{(1)^{-1}}\mathbf{X}^{(1)}(\mathbf{X}^{(1)}\mathbf{V}^{(1)^{-1}}\mathbf{X}^{(1)})^{-1}\mathbf{X}'^{(1)}\mathbf{V}^{(1)^{-1}} \\ &= \frac{1}{\sigma_e^2}\mathbf{I}_{lkm} - \frac{\sigma_b^2}{\sigma_e^2\Delta}\mathbf{I}_l \otimes \mathbf{J}_k \otimes \mathbf{I}_k \otimes \mathbf{J}_k \\ & \quad -\frac{1}{m\sigma_e^2}\mathbf{I}_l \otimes \mathbf{I}_k \otimes \mathbf{J}_m + \frac{\sigma_b^2}{k\sigma_e^2\Delta}\mathbf{I}_l \otimes \mathbf{J}_{km}, \end{aligned}$$

$$\begin{aligned} \mathbf{R}^{(2)} &= \mathbf{V}^{(2)^{-1}} - \mathbf{V}^{(2)^{-1}}\mathbf{X}^{(2)}(\mathbf{X}'^{(2)}\mathbf{V}^{(2)^{-1}}\mathbf{X}^{(2)})^{-1}\mathbf{X}'^{(2)}\mathbf{V}^{(2)^{-1}} \\ &= \frac{1}{\sigma_e^2}\mathbf{I}_{(L-l)km} - \frac{1}{m\sigma_e^2}\mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{J}_m. \end{aligned}$$

Now

$$\mathbf{RV}_b\mathbf{R} = \begin{bmatrix} \mathbf{R}^{(1)}\mathbf{V}_b^{(1)}\mathbf{R}^{(1)} & 0 \\ 0 & \mathbf{R}^{(2)}\mathbf{V}_b^{(2)}\mathbf{R}^{(2)} \end{bmatrix},$$

$$\begin{aligned} \mathbf{R}^{(1)}\mathbf{V}_b^{(1)}\mathbf{R}^{(1)} &= \frac{1}{\Delta^2}\mathbf{I}_l \otimes \mathbf{J}_k \otimes \mathbf{I}_k \otimes \mathbf{J}_f - \frac{1}{k\Delta^2}\mathbf{I}_l \otimes \mathbf{J}_{km}, \\ \mathbf{R}^{(2)}\mathbf{V}_b^{(2)}\mathbf{R}^{(2)} &= 0. \end{aligned}$$

Since

$$\mathbf{RV}_e\mathbf{R} = \begin{bmatrix} \mathbf{R}^{(1)}\mathbf{V}_e^{(1)}\mathbf{R}^{(1)} & 0 \\ 0 & \mathbf{R}^{(2)}\mathbf{V}_e^{(2)}\mathbf{R}^{(2)} \end{bmatrix},$$

$$\begin{aligned} \mathbf{R}^{(1)}\mathbf{V}_e^{(1)}\mathbf{R}^{(1)} &= \frac{1}{\sigma_e^4}\mathbf{I}_{lkm} - \frac{1}{m\sigma_e^4}\mathbf{I}_l \otimes \mathbf{I}_k \otimes \mathbf{J}_m \\ &\quad -\frac{m\sigma_b^4 + 2\sigma_b^2\sigma_e^2}{\sigma_e^4\Delta^2}\mathbf{I}_l \otimes \mathbf{J}_k \otimes \mathbf{I}_k \otimes \mathbf{J}_f + \frac{m\sigma_b^4 + 2\sigma_b^2\sigma_e^2}{k\sigma_e^4\Delta^2}\mathbf{I}_l \otimes \mathbf{J}_{km} \\ \mathbf{R}^{(2)}\mathbf{V}_e^{(2)}\mathbf{R}^{(2)} &= \frac{1}{\sigma_e^4}\mathbf{I}_{(L-l)km} - \frac{1}{m\sigma_e^4}\mathbf{I}_{L-l} \otimes \mathbf{I}_k \otimes \mathbf{J}_m. \end{aligned}$$

So we should have:

$$\begin{aligned} q_1 &= \mathbf{y}'\mathbf{RV}_b\mathbf{Ry} = \frac{m^2}{\Delta^2}\sum_{\gamma=1}^{l}\sum_{j=1}^{k}(\bar{y}_{\gamma.j.}^{(1)} - \bar{\bar{y}}_{\gamma...}^{(1)})^2 \\ q_2 &= \mathbf{y}'\mathbf{RV}_e\mathbf{Ry} = \frac{1}{\sigma_e^4}\sum_{\gamma=1}^{l}\sum_{t=1}^{k}\sum_{j=1}^{k}\sum_{s=1}^{f}(y_{\gamma tjs}^{(1)} - \bar{\bar{y}}_{\gamma t..}^{(1)})^2 \\ &\quad -\frac{m^2(m\sigma_b^4 + 2\sigma_b^2\sigma_e^2)}{\sigma_e^4\Delta^2}\sum_{\gamma=1}^{l}\sum_{j=1}^{k}(\bar{y}_{\gamma.j.}^{(1)} - \bar{\bar{y}}_{\gamma...}^{(1)})^2 \\ &\quad +\frac{1}{\sigma_e^4}\sum_{\gamma=l+1}^{L}\sum_{t=1}^{k}\sum_{s=1}^{m}(y_{\gamma ts}^{(2)} - \bar{y}_{\gamma t.}^{(2)})^2. \end{aligned}$$

Let $\tau_{11} = \operatorname{Tr}\mathbf{RV}_b\mathbf{RV}_b$, $\tau_{12} = \operatorname{Tr}\mathbf{RV}_b\mathbf{RV}_e$, $\tau_{22} = \operatorname{Tr}\mathbf{RV}_e\mathbf{RV}_e$. Then we can have:

$$\begin{aligned} \tau_{11} &= \frac{lk^2f^2(k-1)}{\Delta^2}, \\ \tau_{12} &= \frac{lkf(k-1)}{\Delta^2}, \\ \tau_{22} &= \frac{Lk(m-1)}{\sigma_e^4} - \frac{lkf(k-1)(m\sigma_b^4 + 2\sigma_b^2\sigma_e^2)}{\sigma_e^4\Delta^2}. \end{aligned}$$

From the MINQUE equations we must solve:

$$\begin{cases} \lambda_1\tau_{11} + \lambda_2\tau_{12} = 1 \\ \lambda_1\tau_{12} + \lambda_2\tau_{22} = 0 \end{cases}$$

for λ_1 and λ_2.

Solving the above equations:

$$\begin{aligned}
\lambda_1 &= \frac{\tau_{22}}{\tau_{11}\tau_{22}-\tau_{12}^2} = \frac{\Phi\Delta^2 + l(k-1)\sigma_e^4}{lm^2(k-1)\Phi}, \\
\lambda_2 &= -\frac{\tau_{12}}{\tau_{11}\tau_{22}-\tau_{12}^2} = -\frac{1}{m\Phi}\sigma_e^4,
\end{aligned}$$

where $\Phi = Lk(m-1) - l(k-1)$. From the MINQUE formula,

$$\begin{aligned}
\hat{\sigma}_b^2 &= \lambda_1 \mathbf{y}'\mathbf{R}\mathbf{V}_b\mathbf{R}\mathbf{y} + \lambda_2 \mathbf{y}'\mathbf{R}\mathbf{V}_e\mathbf{R}\mathbf{y} \\
&= \lambda_1 q_1 + \lambda_2 q_2,
\end{aligned}$$

which is (6.27). $\square$

From (6.27) it is obvious that $\hat{\sigma}_b^2$ depends on both interpenetrated and non-interpenetrated data. From section 1.3.4 we know that when normality is assumed then $\hat{\sigma}_b^2$ is the best quadratic unbiased estimator of σ_b^2 for model (6.26).

Theorem 6.2 *If model* (6.26) *is used, then with the normality assumption the variance of* $\hat{\sigma}_b^2$ *given in* (6.27) *is:*

$$\mathrm{V}(\hat{\sigma}_b^2) = \frac{2}{l(k-1)}\left\{\sigma_b^4 + \frac{2}{m}\sigma_b^2\sigma_e^2 + \frac{Lk(m-1)}{m^2\left[Lk(m-1)-l(k-1)\right]}\sigma_e^4\right\}. \quad (6.28)$$

Proof: From the proof of Theorem 6.1 we know that

$$\begin{aligned}
\hat{\sigma}_b^2 &= \mathbf{y}'\left[\lambda_1\mathbf{R}\mathbf{V}_b\mathbf{R} + \lambda_2\mathbf{R}\mathbf{V}_e\mathbf{R}\right]\mathbf{y} \\
&\stackrel{\text{def}}{=} \mathbf{y}'\mathbf{A}\mathbf{y}.
\end{aligned}$$

With the normality assumption we know from section 1.3.4 $\mathrm{V}(\hat{\sigma}_b^2) = 2\,\mathrm{Tr}\,\mathbf{AVAV}$, where $\mathbf{V}$ is the variance covariance matrix of $\mathbf{y}$ for model (6.26).

$$\begin{aligned}
\mathbf{A} &= \lambda_1\mathbf{R}\mathbf{V}_b\mathbf{R} + \lambda_2\mathbf{R}\mathbf{V}_e\mathbf{R} \\
&= \lambda_1\begin{bmatrix}\mathbf{R}^{(1)}\mathbf{V}_b^{(1)}\mathbf{R}^{(1)} & 0 \\ 0 & \mathbf{R}^{(2)}\mathbf{V}_b^{(2)}\mathbf{R}^{(2)}\end{bmatrix} + \lambda_2\begin{bmatrix}\mathbf{R}^{(1)}\mathbf{V}_e^{(1)}\mathbf{R}^{(1)} & 0 \\ 0 & \mathbf{R}^{(2)}\mathbf{V}_e^{(2)}\mathbf{R}^{(2)}\end{bmatrix} \\
&\stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{A}^{(1)} & 0 \\ 0 & \mathbf{A}^{(2)}\end{bmatrix}
\end{aligned}$$

Now

$$\begin{aligned}
\mathbf{A}^{(1)} &= -\frac{1}{m\Phi}\mathbf{I}_{lkm} + \frac{1}{m^2\Phi}\mathbf{I}_l\otimes\mathbf{I}_k\otimes\mathbf{J}_m \\
&\quad + \frac{Lk(m-1)}{lm^2(k-1)\Phi}\mathbf{I}_l\otimes\mathbf{J}_k\otimes\mathbf{I}_k\otimes\mathbf{J}_f - \frac{L(m-1)}{lm^2(k-1)}\mathbf{I}_l\otimes\mathbf{J}_{km}, \\
\mathbf{A}^{(2)} &= -\frac{1}{m\Phi}\mathbf{I}_{(L-l)km} + \frac{1}{m^2\Phi}\mathbf{I}_{L-l}\otimes\mathbf{I}_k\otimes\mathbf{J}_m,
\end{aligned}$$

hence

$$\begin{aligned}
\operatorname{Tr}\mathbf{A}^{(1)}\mathbf{V}^{(1)}\mathbf{A}^{(1)}\mathbf{V}^{(1)} &= \frac{Lk(m-1)\Phi - lk(k-1)(m-1)(L-l)}{lm^2(k-1)\Phi^2}\sigma_e^4 \\
&\quad + \frac{2}{lm(k-1)}\sigma_b^2\sigma_e^2 + \frac{1}{l(k-1)}\sigma_e^4, \\
\operatorname{Tr}\mathbf{A}^{(2)}\mathbf{V}^{(2)}\mathbf{A}^{(2)}\mathbf{V}^{(2)} &= \frac{(L-l)k(m-1)}{m^2\Phi^2}\sigma_e^4.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{V}(\hat{\sigma}_b^2) &= 2\operatorname{Tr}\mathbf{AVAV} \\
&= 2\operatorname{Tr}\mathbf{A}^{(1)}\mathbf{V}^{(1)}\mathbf{A}^{(1)}\mathbf{V}^{(1)} + 2\operatorname{Tr}\mathbf{A}^{(2)}\mathbf{V}^{(2)}\mathbf{A}^{(2)}\mathbf{V}^{(2)} \\
&= (6.28). \quad \square
\end{aligned}$$

Theorem 6.3 *If model* (6.26) *is used and no distribution assumption is made, then the variance of $\hat{\sigma}_b^2$ given by* (6.27) *is:*

$$\begin{aligned}
\mathrm{V}(\hat{\sigma}_b^2) &= \frac{2Lk(m-1)}{lm^2(k-1)\left[Lk(m-1)-l(k-1)\right]}\sigma_e^4 + \frac{4}{lm(k-1)}\sigma_b^2\sigma_e^2 \\
&\quad + \frac{2}{l(k-1)}\sigma_b^4 + \frac{1}{lk}\sigma_b^4\gamma_b + \frac{Lk(L-l)(m-1)^2}{lm^3\left[Lk(m-1)-l(k-1)\right]^2}\sigma_e^4\gamma_e \qquad (6.29)
\end{aligned}$$

where γ_b and γ_e are the kurtoses of the random terms.

Proof: From section 1.3.4 we know that without the normality assumption,

$$\mathrm{V}(\hat{\sigma}_b^2) = 2\operatorname{Tr}\mathbf{AVAV} + \operatorname{Tr}\tilde{\mathbf{B}}\Delta_2\tilde{\mathbf{B}},$$

where $\mathbf{B} = \mathbf{U}'\mathbf{AU}$, and

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_b^{(1)} & 0 & \vdots & \mathbf{I} & 0 \\ 0 & \mathbf{U}_b^{(2)} & \vdots & 0 & \mathbf{I} \end{bmatrix},$$

$$\Delta_2 = \begin{bmatrix} \sigma_b^4\gamma_b\mathbf{I}_{Lk} & \\ & \sigma_e^4\gamma_e\mathbf{I}_{Lkm} \end{bmatrix},$$

$$\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{U}_b^{(1)}\widetilde{\mathbf{A}^{(1)}}\mathbf{U}_b^{(1)} & & & \\ & \mathbf{U}_b^{(2)}\widetilde{\mathbf{A}^{(2)}}\mathbf{U}_b^{(2)} & & \\ & & \tilde{\mathbf{A}}^{(1)} & \\ & & & \tilde{\mathbf{A}}^{(2)} \end{bmatrix}.$$

Now

$$\begin{aligned}
\mathbf{U}_b^{(1)}\mathbf{A}^{(1)}\mathbf{U}_b^{(1)} &= \left[\frac{1}{l(k-1)}+\frac{1}{k\Phi}\right]\mathbf{I}_{lk}-\frac{L(m-1)}{l(k-1)\Phi}\mathbf{I}_l\otimes\mathbf{J}_k,\\
\mathbf{U}_b^{(1)}\widetilde{\mathbf{A}^{(1)}}\mathbf{U}_b^{(1)} &= \frac{1}{lk}\mathbf{I}_{lk},\\
\mathbf{U}_b^{(2)}\mathbf{A}^{(2)}\mathbf{U}_b^{(2)} &= 0.
\end{aligned}$$

Therefore,

$$\operatorname{Tr}\tilde{\mathbf{B}}\Delta_2\tilde{\mathbf{B}}=\frac{1}{lk}\sigma_b^4\gamma_b+\frac{Lk(L-l)(m-1)^2}{lm^3\Phi^2}\sigma_e^4\gamma_e,$$

and adding this term to (6.28) we obtain (6.29). □

So far we have given the MINQUE of σ_b^2 and the variance of the estimator for model (6.26). Since the model we should use in general is model (6.18), we ought to know how efficient σ_b^2 is compared to Biemer and Stokes' estimators if model (6.18) is the correct model to use.

Model (6.26) is a special case of model (6.18) when all the strata are homogeneous, for then $\sigma_e^2(\gamma,t)=\sigma_e^2$, $\gamma=1,\ldots,L$, $t=1,\ldots,k$. If in addition, normality is assumed for the distribution of y, then $\hat{\sigma}_b^2$ of (6.27) is the best quadratic unbiased estimator of σ_b^2. Since $\hat{\sigma}_b^2(BS1)$, $\hat{\sigma}_b^2(BS2)$ and $\hat{\sigma}_b^2(BS)$ are all in quadratic forms, hence $\hat{\sigma}_b^2$ is at least as efficient as any of the three estimators.

When the assumption $\sigma_e^2(\gamma,t)=\sigma_e^2$ is not appropriate, we need to compare $\hat{\sigma}_b^2$ with Biemer and Stokes' estimators. In the following we shall argue that the first estimator of Biemer and Stokes, $\hat{\sigma}_b^2(BS1)$, is less preferred than $\hat{\sigma}_b^2$ of (6.27).

Comparing $\hat{\sigma}_b^2(BS1)$ and $\hat{\sigma}_b^2$ yields an interesting question: for a model with complicated random error structure is it better to use a simplified estimator (synthesis-based MINQUE $\hat{\sigma}_b^2(BS1)$) than using a simplified random error structure for the model and the optimal estimator ($\hat{\sigma}_b^2$)? I suspect the answer is negative.

One disadvantage of $\hat{\sigma}_b^2(BS1)$, which is the synthesis-based MINQUE, is that when $\sigma_e^2(\gamma,t)=\sigma_e^2$ model (6.18) then reduces to model (6.26), and $\hat{\sigma}_b^2(BS1)$ is not identical to the MINQUE $\hat{\sigma}_b^2$ for model (6.26). If normality is assumed for y then $\hat{\sigma}_b^2$ is the best quadratic unbiased estimator of σ_b^2 which means that $\hat{\sigma}_b^2$ is more efficient than $\hat{\sigma}_b^2(BS1)$. So $\hat{\sigma}_b^2$ dominates $\hat{\sigma}_b^2(BS1)$ when $\sigma_e^2(\gamma,t)=\sigma_e^2$.

It is necessary to compare the two estimators when $\sigma_e^2(\gamma,t)\neq\sigma_e^2$. Since model (6.18) is a complicated model we use a simple model to demonstrate that in this simple case $\hat{\sigma}_b^2$ is better than $\hat{\sigma}_b^2(BS1)$

Suppose we have a one way random model:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1,2, \quad j = 1,2. \tag{6.30}$$

where $\mathrm{V}(\mathbf{a}) = \sigma_a^2 \mathbf{I}_2 \otimes \mathbf{J}_2$,

$$\mathrm{V}(\mathbf{e}) = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \sigma_2^2 & \\ & & & \sigma_2^2 \end{bmatrix},$$

and $\sigma_1^2 \neq \sigma_2^2$.
The variance covariance matrix of $\mathbf{y}$ is:

$$\mathbf{V} = \sigma_a^2 \mathbf{I}_2 \otimes \mathbf{J}_2 + \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \sigma_2^2 & \\ & & & \sigma_2^2 \end{bmatrix}.$$

The first estimator $\tilde{\sigma}_a^2 = \mathbf{y}'\tilde{\mathbf{A}}_a\mathbf{y}$ is the synthesis-based MINQUE for model (6.21), *i.e.* use $\mathbf{V} = \mathbf{I}$ in the derivation of $\mathbf{R}$.

$$\mathbf{R} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}_4 - \frac{1}{4}\mathbf{J}_4.$$

Now for model (6.30) we know:

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{J}_2 & 0 \\ 0 & \mathbf{J}_2 \end{bmatrix}, \quad \mathbf{V}_2 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \quad \mathbf{V}_3 = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}.$$

Therefore,

$$\mathbf{R}\mathbf{V}_1\mathbf{R} = \mathbf{I}_2 \otimes \mathbf{J}_2 - \frac{1}{2}\mathbf{J}_4,$$

$$\mathbf{R}\mathbf{V}_2\mathbf{R} = \begin{bmatrix} \mathbf{I}_2 - \frac{3}{8}\mathbf{J}_2 & 0 \\ 0 & \frac{1}{8}\mathbf{J}_2 \end{bmatrix},$$

$$\mathbf{R}\mathbf{V}_3\mathbf{R} = \begin{bmatrix} \frac{1}{8}\mathbf{J}_2 & 0 \\ 0 & \mathbf{I}_2 - \frac{3}{8}\mathbf{J}_2 \end{bmatrix}.$$

Let $\Lambda = (\tau_{ij})_{3\times3}$, where $\tau_{ij} = \operatorname{Tr} \mathbf{R}\mathbf{V}_i\mathbf{R}\mathbf{V}_j$, then

$$\Lambda = \begin{bmatrix} 4 & 1 & 1 \\ 1 & \frac{5}{4} & \frac{1}{4} \\ 1 & \frac{1}{4} & \frac{5}{4} \end{bmatrix},$$

$$\Lambda^{-1} = \frac{1}{4}\begin{bmatrix} \frac{3}{2} & -1 & -1 \\ -1 & 4 & 0 \\ -1 & 0 & 4 \end{bmatrix}.$$

So we have $\lambda_1 = \frac{3}{8}$, $\lambda_2 = \lambda_3 = -\frac{1}{4}$. From the MINQUE formula:

$$\begin{aligned} \tilde{\mathbf{A}}_a &= \lambda_1 \mathbf{R}\mathbf{V}_1\mathbf{R} + \lambda_2 \mathbf{R}\mathbf{V}_2\mathbf{R} + \lambda_3 \mathbf{R}\mathbf{V}_3\mathbf{R} \\ &= \begin{bmatrix} -\frac{1}{4}\mathbf{I}_2 + \frac{1}{4}bJ_2 & -\frac{3}{16}\mathbf{J}_2 \\ -\frac{3}{16}\mathbf{J}_2 & -\frac{1}{4}\mathbf{I}_2 + \frac{1}{4}\mathbf{J}_2 \end{bmatrix}. \end{aligned} \tag{6.31}$$

The second estimator, $\hat{\sigma}_a^2 = \mathbf{y}'\hat{\mathbf{A}}_a\mathbf{y}$, is the MINQUE for a simplified model of (6.30):

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1,2, \quad j = 1,2, \tag{6.32}$$

where $\mathrm{V}(\mathbf{a}) = \sigma_a^2 \mathbf{I}_2 \otimes \mathbf{J}_2$, $\mathrm{V}(\mathbf{e}) = \sigma_e^2 \mathbf{I}_4$.

Model (6.32) is equivalent to assuming $\sigma_1^2 = \sigma_2^2$ in model (6.30).

$$\hat{\mathbf{A}}_a = -\frac{1}{4}\mathbf{I}_4 + \frac{3}{8}\mathbf{I}_2 \otimes \mathbf{J}_2 - \frac{1}{8}\mathbf{J}_4. \tag{6.33}$$

Now assume that $\mathbf{y}$ in model (6.30) has a normal distribution with variance covariance matrix $\mathbf{V}$, then

$$\mathrm{V}(\tilde{\sigma}_a^2) = 2\,\mathrm{Tr}\,\tilde{\mathbf{A}}_a\mathbf{V}\tilde{\mathbf{A}}_a\mathbf{V} = 2 \parallel \tilde{\mathbf{A}}_a\mathbf{V} \parallel_E^2 .$$

Using $\tilde{\mathbf{A}}_a$ in (6.31) we have:

$$\tilde{\mathbf{A}}_a\mathbf{V} = \begin{bmatrix} \frac{\sigma_a^2}{4} & \frac{\sigma_a^2+\sigma_1^2}{4} & -\frac{3(2\sigma_a^2+\sigma_2^2)}{16} & -\frac{3(2\sigma_a^2+\sigma_2^2)}{16} \\ \frac{\sigma_a^2+\sigma_1^2}{4} & \frac{\sigma_a^2}{4} & -\frac{3(2\sigma_a^2+\sigma_2^2)}{16} & -\frac{3(2\sigma_a^2+\sigma_2^2)}{16} \\ -\frac{3(2\sigma_a^2+\sigma_1^2)}{16} & -\frac{3(2\sigma_a^2+\sigma_1^2)}{16} & \frac{\sigma_a^2}{4} & \frac{\sigma_a^2+\sigma_2^2}{4} \\ -\frac{3(2\sigma_a^2+\sigma_1^2)}{16} & -\frac{3(2\sigma_a^2+\sigma_1^2)}{16} & \frac{\sigma_a^2+\sigma_2^2}{4} & \frac{\sigma_a^2}{4} \end{bmatrix},$$

and $\mathrm{V}(\hat{\sigma}_a^2) = 2\,\mathrm{Tr}\,\hat{\mathbf{A}}_a\mathbf{V}\hat{\mathbf{A}}_a\mathbf{V} = 2 \parallel \hat{\mathbf{A}}_a\mathbf{V} \parallel_E^2$, using $\hat{\mathbf{A}}_a$ in (6.33):

$$\hat{\mathbf{A}}_a\mathbf{V} = \begin{bmatrix} \frac{\sigma_a^2}{4} & \frac{\sigma_a^2+\sigma_1^2}{4} & -\frac{2\sigma_a^2+\sigma_2^2}{8} & -\frac{2\sigma_a^2+\sigma_2^2}{8} \\ \frac{\sigma_a^2+\sigma_1^2}{4} & \frac{\sigma_a^2}{4} & -\frac{2\sigma_a^2+\sigma_2^2}{8} & -\frac{2\sigma_a^2+\sigma_2^2}{8} \\ -\frac{2\sigma_a^2+\sigma_1^2}{8} & -\frac{2\sigma_a^2+\sigma_1^2}{8} & \frac{\sigma_a^2}{4} & \frac{\sigma_a^2+\sigma_2^2}{4} \\ -\frac{2\sigma_a^2+\sigma_1^2}{8} & -\frac{2\sigma_a^2+\sigma_1^2}{8} & \frac{\sigma_a^2+\sigma_2^2}{4} & \frac{\sigma_a^2}{4} \end{bmatrix}.$$

Since the Euclidean norm $\parallel\ \parallel_E^2$ is the sum of squares of all elements in a matrix and:

$$\frac{3(2\sigma_a^2+\sigma_1^2)}{16} \geq \frac{2\sigma_a^2+\sigma_1^2}{8}, \quad \frac{3(2\sigma_a^2+\sigma_2^2)}{16} \geq \frac{2\sigma_a^2+\sigma_2^2}{8},$$

$\| \tilde{\mathbf{A}}_a \mathbf{V} \|_E^2 \geq \| \hat{\mathbf{A}}_a \mathbf{V} \|_E^2$. Hence

$$\mathrm{V}(\tilde{\sigma}_a^2) \geq \mathrm{V}(\hat{\sigma}_a^2).$$

Straight inequality holds in the above formula when all the variance components are positive.

So in this simple model we have demonstrated that the synthesis-based MINQUE $\tilde{\sigma}_a^2$ for the correct model is less preferred than the MINQUE $\hat{\sigma}_a^2$ for the simplified model by using an equal random error assumption. We suspect that the conclusion holds for more complicated models such as (6.18).

With normality assumption the comparisons between $\hat{\sigma}_b^2(BS)$, $\hat{\sigma}_b^2(BS1)$, $\hat{\sigma}_b^2(BS2)$ and $\hat{\sigma}_b^2$ are presented in the following table:

$\sigma_e^2(\gamma, t) = \sigma_e^2$	$\sigma_e^2(\gamma, t) \neq \sigma_e^2$	
$\hat{\sigma}_b^2 \geq \hat{\sigma}_b^2(BS)$	For model (6.30)	For other models
$\hat{\sigma}_b^2 > \hat{\sigma}_b^2(BS1)$ $\hat{\sigma}_b^2 \geq \hat{\sigma}_b^2(BS2)$	$\hat{\sigma}_b^2 > \hat{\sigma}_b^2(BS1)$	?

$\hat{\sigma}_b^2 \geq \hat{\sigma}_b^2(BS)$ means that $\hat{\sigma}_b^2$ is at least as efficient as $\hat{\sigma}_b^2(BS)$, and $\hat{\sigma}_b^2 > \hat{\sigma}_b^2(BS1)$ means that $\hat{\sigma}_b^2$ is more efficient than $\hat{\sigma}_b^2(BS1)$. The comparisons of the estimators without normality assumption are not available.

To summarize when the strata are roughly homogeneous and the data have a normal distribution $\hat{\sigma}_b^2$ is preferred to Biemer and Stokes' estimators. When the strata are not homogeneous we suspect that $\hat{\sigma}_b^2$ is more efficient than $\hat{\sigma}_b^2(BS1)$ used alone.

6.7 Optimal design for $\hat{\sigma}_b^2$

In this section we consider the optimal design using $\hat{\sigma}_b^2$ as the estimator of σ_b^2 for model (6.26) with a cost constraint and we assume that the data vector $\mathbf{y}$ has a normal distribution. We use the same cost constraint as that considered by Biemer and Stokes.

Assume that the total number of strata, $I = Lk$, is fixed. Also assume that interpenetration of a block containing k strata will increase the usual cost of interviewing by a factor of $\sqrt{k}$. Further assume that in each stratum the increase in cost is fixed, *i.e.* $\Delta_0 = \frac{l}{L}(\sqrt{k} - 1)$ is fixed. We want to find the values of k, the number of interpenetrated interviewers, and l, the number of blocks using the interpenetration scheme, such that $\mathrm{V}(\hat{\sigma}_b^2)$ is a minimum.

Substitute $L = \frac{I}{k}$, $l = \frac{L\Delta_0}{\sqrt{k}-1} = \frac{I\Delta_0}{k(\sqrt{k}-1)}$ into (6.28), then

$$V(\hat{\sigma}_b^2) = \frac{2k(\sqrt{k}-1)}{I\Delta_0(k-1)}\left[\frac{I(m-1)}{m^2\left[I(m-1)-\frac{I}{k(\sqrt{k}-1)}\Delta_0(k-1)\right]}\sigma_e^4 + \frac{2}{m}\sigma_b^2\sigma_e^2 + \sigma_b^4\right]. \tag{6.34}$$

$V(\hat{\sigma}_b^2)$ is in the order of $\sqrt{k}$ and is monotonically increasing from $k = 2$. Hence $V(\hat{\sigma}_b^2)$ achieves a minimum when $k = 2$. Correspondingly, $l = \frac{L\Delta_0}{\sqrt{2}-1}$, which may not be an integer. Now let $k = 2$ in (6.28), then

$$V(\hat{\sigma}_b^2) = \frac{I}{l}\left\{\frac{I(m-1)}{m^2\left[I(m-1)-l\right]}\sigma_e^4 + \frac{2}{m}\sigma_b^2\sigma_e^2 + \sigma_b^4\right\}. \tag{6.35}$$

$V(\hat{\sigma}_b^2)$ in (6.35) is monotonically decreasing when l increases, hence we should choose $l = \left[\frac{L}{\sqrt{2}-1}\Delta_0\right] + 1$ where [ ] denotes the integer part of a real number.

Hence for model (6.26) with normality assumption the pair interpenetration scheme is optimal for $\hat{\sigma}_b^2$ with the specific cost constraint used in this section, and the corresponding optimal choice of the number of blocks being interpenetrated is: $l = \left[\frac{L}{\sqrt{2}-1}\Delta_0\right] + 1$.

For the parameter values used by Biemer and Stokes in their empirical study, $L = 700$, $\Delta_0 = 0.05$, then $l = [84.5] + 1 = 85$, *i.e.* 85 out of 700 blocks should be selected to carry pairwise interpenetrated interviews with this specific cost constraint.

In Biemer and Stokes (1985) they concluded that if $\hat{\sigma}_b^2(BS1)$ is used alone then $k = 2$ is the optimal choice. If $\hat{\sigma}_b^2(BS2)$ is used alone, then k should be chosen as large as possible. They suggested for the composite estimator $\hat{\sigma}_b^2(BS)$ the design problem should be addressed empirically.

6.8 Estimator of the variance of the estimator of the mean

In this section we shall derive an unbiased estimator for the variance of the estimator of the mean using $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$ while $\hat{\sigma}_b^2$ is given by (6.27) and $\hat{\sigma}_e^2$ is given in the following theorem.

Theorem 6.4 *If model (6.26) is used, the MINQUE of σ_e^2 is:*

$$\hat{\sigma}_e^2 = \frac{1}{Lk(m-1)-l(k-1)}\left\{\sum_{\gamma=1}^{l}\sum_{t=1}^{k}\sum_{j=1}^{k}\sum_{s=1}^{f}(y_{\gamma tjs}^{(1)} - \bar{\bar{y}}_{\gamma t..}^{(1)})^2\right.$$

$$+\sum_{\gamma=l+1}^{L}\sum_{t=1}^{k}\sum_{s=1}^{m}(y_{\gamma ts}^{(2)}-\bar{y}_{\gamma t.}^{(2)})^2-m\sum_{\gamma=1}^{l}\sum_{j=1}^{k}(\bar{y}_{\gamma.j.}^{(1)}-\bar{\bar{y}}_{\gamma...}^{(1)})^2\Bigg\}. \tag{6.36}$$

The derivation of $\hat{\sigma}_e^2$ is similar to that of $\hat{\sigma}_b^2$ in theorem 6.1 and hence is omitted.

For model (6.26) we use the least squares estimator to estimate $\boldsymbol{\eta}$ which contains the population mean for each stratum, then

$$\hat{\boldsymbol{\eta}}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}=\bar{\mathbf{y}}, \tag{6.37}$$

where $\bar{\mathbf{y}}=(\bar{y}_{\gamma t})$ and

$$\bar{y}_{\gamma t}=\begin{cases}\bar{\bar{y}}_{\gamma t..}^{(1)}, & \text{if } (\gamma,t) \text{ is in the selected } l \text{ blocks,}\\ \bar{y}_{\gamma t.}^{(2)}, & \text{if } (\gamma,t) \text{ is the remaining } L-l \text{ blocks.}\end{cases}$$

$\hat{\boldsymbol{\eta}}$ is the vector containing the estimated mean of each stratum.

We are interested to draw inference about Π, the population considered in Section 6.2. Suppose $\mathbf{p}'\boldsymbol{\eta}$ is the parameter in Π which we want to estimate. If $\mathbf{p}=\mathbf{1}$, then $\mathbf{p}'\boldsymbol{\eta}$ is the total of Π. If $\mathbf{p}=\frac{1}{\sum_{\gamma=1}^{L}\sum_{t=1}^{k}M_{\gamma t}}\mathbf{1}$, then $\mathbf{p}'\boldsymbol{\eta}$ is the mean of Π.

We use $\widehat{\mathbf{p}'\boldsymbol{\eta}}=\mathbf{p}'\hat{\boldsymbol{\eta}}$ to estimate $\mathbf{p}'\boldsymbol{\eta}$.

From the discussion on $\sigma_e^2(\gamma,t)$ in Section 6.4 we know that sampling errors are included in $\sigma_e^2(\gamma,t)$. The variance of $\widehat{\mathbf{p}'\boldsymbol{\eta}}$ can then be expressed in terms of σ_b^2 and σ_e^2.

$$\begin{aligned}\mathrm{V}(\hat{\boldsymbol{\eta}}|\Pi) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\\ &= \begin{bmatrix}\frac{\sigma_e^2}{m}\mathbf{I}_{lk}+\frac{\sigma_b^2}{k}\mathbf{I}_l\otimes\mathbf{J}_k & 0\\ 0 & \frac{m\sigma_b^2+\sigma_e^2}{m}\mathbf{I}_{(L-l)m}\end{bmatrix},\end{aligned}$$

then

$$\begin{aligned}\mathrm{V}(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi) &= \mathbf{p}'\,\mathrm{V}(\hat{\boldsymbol{\eta}})\mathbf{p}\\ &= \frac{\sigma_e^2}{m}\sum_{\gamma=1}^{l}\sum_{t=1}^{k}p_{\gamma t}^2+\sigma_b^2\sum_{\gamma=l+1}^{L}\sum_{t=1}^{k}p_{\gamma t}^2+\frac{\sigma_b^2}{k}\sum_{\substack{\gamma=1\\ \gamma'=1}}^{l}\sum_{\substack{t=1\\ t'=1}}^{k}p_{\gamma t}p_{\gamma' t'}.\end{aligned}$$

Since $\hat{\sigma}_b^2$ of (6.27) and $\hat{\sigma}_e^2$ of (6.36) are unbiased estimators of σ_b^2 and σ_e^2, respectively, thus

$$\hat{\mathrm{V}}(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi)=\frac{\hat{\sigma}_e^2}{m}\sum_{\gamma=1}^{L}\sum_{t=1}^{k}p_{\gamma t}^2+\frac{\hat{\sigma}_b^2}{k}\sum_{\substack{\gamma=1\\ \gamma'=1}}^{l}\sum_{\substack{t=1\\ t'=1}}^{k}p_{\gamma t}p_{\gamma' t'}+\hat{\sigma}_b^2\sum_{\gamma=l+1}^{L}\sum_{t=1}^{k}p_{\gamma t}^2, \tag{6.38}$$

is unbiased for $V(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi)$. From (6.38) we can see that $\hat{V}(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi)$ has used the estimates of the interviewer's variance and the random error. In the above formulas we used $|\Pi$ to denote that the derivation is conditional on the selection of Π.

As we pointed out in Section 6.2 estimation in this chapter is conditional on the selection of Π. Inference can be made by standard survey estimation to the total population from which Π is selected because there are no more measurement errors involved. As Hartley and Rao (1978) have shown we can use:

$$V(\widehat{\mathbf{p}'\boldsymbol{\eta}}) = \mathop{E}_{\Pi} V(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi) + \mathop{V}_{\Pi} E(\widehat{\mathbf{p}'\boldsymbol{\eta}}|\Pi),$$

where E_Π and V_Π are the expectation and variance over the selection of Π in the original survey design. Hence classical sample survey theory can be used to estimate $V(\widehat{\mathbf{p}'\boldsymbol{\eta}})$ for the total population.

6.9 Conclusions

In this chapter we applied variance components estimation to estimate interviewer's variance in surveys.

We build a model by assuming homogeneity of the variances within strata and derived the MINQUE estimator for the interviewer's variance. If normality is the distribution of the data this MINQUE estimator is the best quadratic unbiased estimator for the interviewer's variance and therefore is at least as efficient as Biemer and Stokes' estimators. We also showed that with a specified cost constraint the pair interpenetration scheme is optimal for the MINQUE estimator and derived the optimal number of blocks taking interpenetrated interviews. Using the MINQUE estimate of interviewer's variance we derived an unbiased estimator for the variance of the estimator of the mean.

When the strata are not homogeneous we suspect that the estimator we derived in section 6.6 is still more efficient than Biemer and Stokes' first estimator. Our suspicion is supported by the investigation on a simple model in Section 6.6. More research is needed to compare the estimators in this case.

Chapter 7

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

7.1 Summary of conclusions

In this thesis we considered quadratic estimators of linear combinations of variance components and the properties of the estimators.

Chapters 2 and 3 investigated the properties of MINQUE, which was proposed for the general variance components model (1.1). MINQUE is unbiased, but it is known that its optimality requires the following two assumptions:

(1) The data have a normal distribution;

(2) The prior values are correct values of the variance components.

MINQUE under these two assumptions gives the best locally quadratic unbiased estimators of variance components. In practice these two assumptions are very restrictive. We often cannot assume that the data collected from a survey or an experiment have a normal distribution. Furthermore, the correct values of the variance components are never known to us. Chapters 2 and 3 are aimed at weakening the two assumptions and hence widening the optimality area of MINQUE.

In Chapter 2 we proved a sufficient condition for the design matrices of the variance components model (1.1) so that the MINQUE will be the best quadratic unbiased estimator without assuming a normal distribution for the data. A class of models fitted to balanced data is shown to satisfy the condition, hence

MINQUE for these models is optimal without assumption (1).

In Chapter 3 we proved a necessary and sufficient condition for MINQUE to be independent of prior values. The condition is simplified for the balanced extended analysis of variance models. If MINQUE achieves optimality and satisfies the condition in Chapter 3, it will be the globally best quadratic unbiased estimator.

We have shown that there are models and designs satisfying the conditions in both Chapters 2 and 3, therefore the MINQUE used for these models is the globally best quadratic unbiased estimator without satisfying assumptions (1) and (2).

Instead of the two well-known assumptions, we can have two alternative assumptions:

(1)' The design matrices of the model satisfy condition (2.14);

(2)' The design matrices of the model satisfy the conditions in Theorem 3.3.

We use the following table to summarize how by using (1)' and/or (2)' as alternative assumptions, the optimality area of MINQUE is widened.

Assumptions	Optimality of MINQUE (quadratic unbiased)	result comes from
(1) (2)	locally best	Rao (1971a,b)
(1)' (2)	locally best	Chapter 2
(1) (2)'	globally best	Chapter 3
(1)' (2)'	globally best	Chapters 2 and 3

Chapters 4 and 5 concentrate on obtaining nonnegative quadratic estimators. After investigating the properties of some existing nonnegative estimators: MINQE, CMINQUE and Hartung's estimator, we proved the nonexistence of a globally minimum biased nonnegative estimator across the parameter space. A modified version of Hartung's estimator, the minimum bias range MINQ estimator, is proposed which has the minimum variance among all the estimators minimizing an upper bound of the bias function. Such an estimator needs prior values. Iterative computing can be used to obtain an estimate independent of prior values.

Numerical and empirical comparisons are presented in Chapter 5. It can be seen that none of the estimators dominates any other throughout the parameter space. Suggestions are made on the use of these estimators.

Chapter 6 applied MINQUE in a complex survey to estimate the interviewer's variance. After setting up the model we applied MINQUE to a simplified model and obtained an estimator for the interviewer's variance. We suspect that our estimator is more efficient than Biemer-Stokes' synthesis-based MINQUE applied to the full model. In the case of homogeneous strata our suspicion is true.

A design problem with a specified cost constraint is solved and an unbiased estimator is given for the variance of the estimator of the population mean.

This thesis has proved that MINQUE can be applied with optimality to more situations than those known before. MINQUE can be modified in various ways to form nonnegative estimators. When used in practice MINQUE is quite efficient compared to some existing estimators.

In the next section we shall give recommendations for some areas where future research is needed.

7.2 Recommendations for future research

There are several problems directly related to the ones considered in this thesis.

(1) In Theorem 3.3 we gave the necessary and sufficient conditions for the MINQUE to be independent of prior values. We have simplified the conditions for balanced E-ANOVA models. When considering the general variance components model (1.1) more work is needed to make the conditions in Theorem 3.3 checkable by computer, so that assumptions (1)' and (2)' can be easily verified.

(2) In Chapters 4 and 5 we have shown that none of the nonnegative estimators considered dominates the others throughout the parameter space. There is a need to set up a reasonable global measure of efficiency, rather than the mean squared error at individual parameter values that we used in Chapters 4 and 5, to compare the estimators. Some estimators may be more favourable than the others under a new measure.

(3) It will be interesting to apply MINQUE to complex surveys with more error sources than the two error sources considered in Chapter 6 and see if the MINQUE is still efficient. The MINQUE algorithm is straightforward to extend to more error sources. The problem with many error sources is that the variance covariance matrix is very complicated if computed algebraically, and has a very large order if computed numerically.

(4) In Section 1.3.5 we have introduced Goldstein's method for the estimation

of variance components. Goldstein has proved that his method is equivalent to the maximum likelihood estimator with a normal distribution assumption. Since the equivalence of maximum likelihood and weighted least squares estimates for a member of the exponential family with one parameter is established, it will be interesting to see if Goldstein's method is equivalent to maximum likelihood estimation for a certain class of distributions. In that case Goldstein's method will be an extension of the maximum likelihood estimation to more distributions.

(5) Herbach (1959) derived the maximum likelihood estimators for the one-way balanced model. More research is necessary to study the likelihood (restricted likelihood) functions for more models and see how to deal with the negative solutions to the likelihood equations.

There are some problem areas where estimation of variance components is concerned.

(1) Optimal designs for the estimation of variance components.

Apart from the result of Mukerjee and Huda (1988) stated in Theorem 1.3 for the unweighted analysis of variance estimator, there have been no results on the optimal designs for the various estimators of variance components, *e.g.* MINQUE, ML and REML.

(2) Prediction of random variables and small area estimation.

In Section 1.2 we discussed some problems associated with the general variance components models such as:

(i) The estimation of $\boldsymbol{\beta}$;

(ii) The estimation of σ_i^2, $i = 1, \ldots, k$;

(iii) Prediction of $\boldsymbol{\xi}_i$, $i = 1, \ldots, k$.

Problem (i) has been widely addressed and problem (ii) is the concern of this thesis. Problem (iii) is closely related to problems (i) and (ii).

Now consider a simpler variance components model than model (1.1):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{d}, \tag{7.1}$$

where $\mathbf{Y}$ is the observed data vector, $\boldsymbol{\beta}$ is the fixed effect parameter, $\mathbf{a}$ and $\mathbf{d}$ are random variables with $\mathrm{E}(\mathbf{a}) = \mathrm{E}(\mathbf{d}) = 0$, $\mathrm{V}(\mathbf{a}) = \mathbf{A}$ and $\mathrm{V}(\mathbf{d}) = \mathbf{D}$.

If the task is to predict $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a}$, Harville (1976) proved that when $\mathbf{A}$ and $\mathbf{D}$ are known and mean squared error is used as the optimality criterion the best linear unbiased estimator of $\mathbf{y}$ is:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{A}(\mathbf{A} + \mathbf{D})^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{7.2}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'(\mathbf{A}+\mathbf{D})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{A}+\mathbf{D})^{-1}\mathbf{Y}. \tag{7.3}$$

In the simplest case of matrices $\mathbf{A}$ and $\mathbf{D}$, $\mathbf{A} = \sigma_a^2\mathbf{I}$, $\mathbf{D} = \sigma_d^2\mathbf{I}$, then $\boldsymbol{\beta}$ is the least square estimator,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{7.4}$$

and

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \frac{\sigma_a^2}{\sigma_a^2+\sigma_d^2}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}). \tag{7.5}$$

The problem is that in practice both σ_a^2 and σ_d^2 are not known. It seems reasonable to estimate σ_a^2 and σ_d^2 first and use $\hat{\sigma}_a^2$ and $\hat{\sigma}_d^2$ in (7.5).

(*). The problem is what kind of estimators $\hat{\sigma}_a^2$ and $\hat{\sigma}_d^2$ in (7.5) will make prediction of $\mathbf{y}$ optimal?

We can use the MINQUE $\hat{\sigma}_a^2$ and $\hat{\sigma}_d^2$ in (7.5), but they cannot make the predictor $\hat{\mathbf{y}}$ in (7.5) optimal, as demonstrated by Peixoto and Harville (1986). One explanation for the failure of MINQUE in this situation is that MINQUE is required to be optimal when estimating linear combinations of the variance components while the variance components appearing in (7.5) are the ratio of the variance components.

So far there has been no answer to the problem (*).

The demand for small area estimation has increased recently. Sample surveys usually provide efficient estimators for the totals of large domain. Small area estimation is desired when estimation is required for subdivision of the population and when the standard errors for the sample survey estimator are unacceptably large for the subdivision. Harville (1985) pointed out that all the existing methods for small area estimation, namely, Bayes approach, empirical Bayes approach and regression model approach, can be derived from the variance components model approach.

Suppose $\mathbf{y}$ is the desired $n \times 1$ vector to be estimated, and $\mathbf{X}$ is the $n \times p$ matrix containing the auxiliary information. We can write the model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a}, \tag{7.6}$$

where $\boldsymbol{\beta}$ is the $p \times 1$ regression coefficient vector to be estimated, $\mathbf{a}$ is the random error term with $\mathrm{E}(\mathbf{a}) = 0$ and $\mathrm{V}(\mathbf{a}) = \mathbf{A}$, $\mathbf{A}$ is an $n \times n$ matrix.

Suppose the observed data vector $\mathbf{Y}$ has a measurement error in $\mathbf{y}$, that is:

$$\mathbf{Y} = \mathbf{y} + \mathbf{d}, \tag{7.7}$$

where $\mathrm{E}(\mathbf{d}) = 0$, $\mathrm{V}(\mathbf{d}) = \mathbf{D}$ and $\mathbf{D}$ is an $n \times n$ matrix.

Now combining (7.6) and (7.7) we obtain a variance components model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{a} + \mathbf{d},$$

which is identical to model (7.1). The small area estimation of $\mathbf{y}$ is then identical to the problem of prediction of random variables, hence an optimal estimator for small area estimator depends on the answer to $(*)$.

Bibliography

[1] Anderson, R. L. (1975) Design and estimators for variance components. *A Survey of Statistical Design and Linear Models* ed. J. N. Srivastava, 1-29. North-Holland Publishing Company.

[2] Anderson, T. W. (1969) Statistical inference for covariance matrices with linear structure. *Proc. Second Internat. Symp. Multivariate Analysis* ed. P. R. Krishnaiah, 55-66. Academic Press, New York.

[3] Arnold, S. F. (1981) *The Theory of Linear Models and Multivariate Analysis*, New York, John Wiley.

[4] Arvesen, J. N. and Layard, M. W. J. (1975) Asymptotically robust tests in unbalanced variance components. *Ann. Statist.* **3** 1122-1134.

[5] Bailey, R. A. (1991) Strata for randomized experiments. *J. R. Statis. Soc.* B **53** 27-78.

[6] Bhapkar, V. P. (1972) On a measure of efficiency of an estimating equation. *Sankhya* A **34** 467-472.

[7] Biemer, P. P. (1978) *The Estimation of Non-Sampling Variance Components in Sample Surveys* Unpublished ph.D thesis, Texas A & M University.

[8] Biemer, P. P. and Stokes, S. L. (1985) Optimal design of interviewer variance experiments in complex surveys. *J. Amer. Statist. Assoc.* **80** 158-166.

[9] Bradley, E. L. (1973) The equivalence of maximum likelihood and weighted least square estimates in the exponential family. *J. Amer. Statist. Assoc.* **68** 199-200.

[10] Brown, K. G. (1976) Asymptotic behavior of MINQUE-type estimators of variance components. *Ann. Statist.* **4** 746-754.

[11] Browne, M. W. (1974) Generalized least squares estimators in the analysis of covariance structures. *South African Statist. J.* **8** 1-24.

[12] Chaubey, Y. P. (1983) A non-negative estimator of variance components closest to MINQUE. *Sankhaya* **45** 201-211.

[13] Cochran, W. G. (1963) *Sampling Techniques* 2nd edition, John Wiley.

[14] Conerly, M. D. and Webster, J. T. (1987) MINQUE for the one-way classification. *Technometrics* **29** 229-236.

[15] Fellegi, I. P. (1974) An improved method of estimating the correlated response variance. *J. Amer. Statist. Assoc.* **69** 496-501.

[16] Fuller, W. A. (1987) *Measurement Error Models* John Wiley, New York.

[17] Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208-1212.

[18] Godambe, V. P. (1976) Conditional likelihood and (unconditional) optimum estimating equation. *Biometrika* **63** 277-284.

[19] Godambe, V. P. and Heyde, C. C. (1987) Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **23** 231-244.

[20] Godambe, V. P. and Thompson, M. E. (1984) Robust estimation through estimating equations. *Biometrika* **71** 115-125.

[21] Goldstein, H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73** 43-56.

[22] Goldstein, H. (1989) Restricted unbiased iterative generalized least-square estimation. *Biometrika* **76** 622-623.

[23] Graham, A. (1981) *Kronecker Products and Matrix Calculus with Applications,* Ellis Hortwood Ltd.

[24] Graybill, F. A. (1954) On quadratic estimates of variance components. *Ann. Math. Statist.* **25** 367-372.

[25] Graybill, F. A. (1983) *Matrices with Applications in Statistics* (2nd edition), Wadsworth International Group.

[26] Graybill, F. A. and Hultquist, R. A. (1961) Theorems concerning Eisenhart's model 2. *Ann. Math. Statist.* **32** 261-269.

[27] Graybill, F. A. and Wortham, A. W. (1956) A note on uniformly best unbiased estimators for variance components. *J. Amer. Statist. Assoc.* **51** 266-268.

[28] Hansen, M. H., Hurwitz, W. N. and Bershad, M. A. (1959) Measurement errors in censuses and surveys. *Bulletin of the Internat. Statist. Inst.* **38** 359-374.

[29] Hartley, H. O. (1981) Estimation and design for non-sampling errors of surveys. *Current Topics in Survey Sampling* (ed. D. Krewski, P. Platek and J. N. K. Rao), Academic Press, 31-46.

[30] Hartley, H. O. and Rao, J. N. K. (1967) Maximum likelihood estimation for mixed analysis of variance models. *Biometrika* **54** 93-103.

[31] Hartley, H. O. and Rao, J. N. K. (1978) Estimation of nonsampling variance components in the sample surveys. *Survey Sampling and Measurement* ed. N. Krishnan Namboodiri, 35-43.

[32] Hartley, H. O., Rao, J. N. K. and LaMotte, L. (1978) A simple synthesis-based method of variance component estimation. *Biometrics* **34** 233-242.

[33] Hartung, J. (1981) Nonnegative minimum biased invariant estimation in variance component models. *Ann. Statist.* **9** 278-292.

[34] Harville, D. A. (1977) Maximum likelihood approaches to variance components estimators and to related problems. *J. Amer. Statist. Assoc.* **72** 320-340.

[35] Harville, D. A. (1985) Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132-138.

[36] Hayman, B. I. (1960) Maximum likelihood estimation of genetic components of variation. *Biometrics* **16** 369-381.

[37] Herbach, L. H. (1959) Properties of type 2 analysis of variance tests. *Ann. Math. Statist.* **30** 939-959.

[38] Holt, D., Smith, T. M. F. and Tomberlin, T. J. (1979) A model based approach to estimation for small subgroup of a population. *J. Amer. Statist. Assoc.* **74** 405-410.

[39] Kackar, R. N. and Harville, D. A. (1984) Approximation for standard errors of estimatiors of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79** 853-861.

[40] Kleffe, J. and Rao, J. N. K. (1986) The existence of asymptotically unbiased nonnegative quadratic estimates of variance components in ANOVA models. *J. Amer. Statist. Assoc.* **81** 692-698.

[41] Kleffe, J. and Seifert, B. (1986) Computation of variance components by the MINQUE method. *J. Multivariate Analysis* **18** 107-116.

[42] Kurkjian, B. and Zelen, M. (1962) A calculus for factorial arrangements. *Ann. Math. Statist.* **33** 600-619.

[43] LaMotte, L. R. (1973a) Quadratic estimation of variance components. *Biometrics* **29** 311-330.

[44] LaMotte, L. R. (1973b) On nonnegative quadratic unbiased estimation of variance components. *J. Amer. Statist. Assoc.* **68** 728-730.

[45] Miller, J. J. (1977) Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746-762.

[46] Morris, C. N. (1983) Parametric empirical bayes inference theory and applications. *J. Amer. Statist. Assoc.* **78** 47-55.

[47] Morton, R. (1981) Efficiency of estimating equations and the use of pivots. *Biometrika* **68** 227-233.

[48] Mukerjee, R. and Huda, S. (1988) Optimal design for the estimation of variance components. *Biometrika* **75** 75-80.

[49] Nelder, J. A. (1965a) The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc.* A **283** 147-162.

[50] Nelder, J. A. (1965b) The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc.* A **283** 163-178.

[51] Patterson, H. D. and Thompson, R. (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* **58** 545-554.

[52] Patterson, H. D. and Thompson, R. (1974) Maximum likelihood estimation of components of variance. *Proc. 8th Internat. Biometric Conference* 197-207.

[53] Peixoto, J. L. and Harville, D. A. (1986) Comparisons of alternative predictors under the balanced one-way model. *J. Amer. Statist. Assoc.* **81** 431-436.

[54] Platek, R., Rao, J. N. K., Sarndal, C. E. and Singh, M. B. (1986) *Small Area Statistics: An International Symposium,* John Wiley, New York.

[55] Pukelsheim, F. (1981) On the existence of unbiased nonnegative estimates of variance covariance components. *Ann. Statist.* **9** 293-299.

[56] Rao, C. R. (1970) Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc* **65** 161-172.

[57] Rao, C. R. (1971a) Estimation of variance and covariance components—MINQUE theory. *J. Multi. Analy.* I 257-275.

[58] Rao, C. R. (1971b) Minimum variance quadratic unbiased estimation of variance components. *J. Multi. Analy.* I 445-456.

[59] Rao, C. R. (1973) *Linear Statistical Inference and its Applications* (2nd edition), John Wiley.

[60] Rao, C. R. (1979) MINQUE theory and its relation to ML and MML estimation of variance components. *Sankhaya* **41** 138-153.

[61] Rao, C. R. and Kleffe, J. (1980) Estimation of variance components. *Handbook of Statistics* **1** 1-40, North-Holland.

[62] Rao, C. R. and Kleffe, J. (1988) *Estimation of Variance Components and Applications*, North-Holland.

[63] Rao, P. S. R. S. and Chaubey, Y. P. (1978) Three modifications of the principle of the MINQUE. *Commun. Statist. -Theor. Meth.* A**7(8)** 767-778.

[64] Royall, R. M. (1976) The linear-square prediction approach to two-stage sampling. *J. Amer. Statist. Assoc.* **71** 657-664.

[65] Scott, A. J. and Smith, T. M. F. (1969) Estimation in Multi-stage surveys. *J. Amer. Statist. Assoc.* **64** 830-840.

[66] Searle, S. R. (1971) *Linear Models*, John Wiley.

[67] Searle, S. R. (1988) Mixed models and unbalanced data: wherefrom, whereat and whereto? *Commun. Statist. -Theor. Meth.* **17** 935-968.

[68] Seely, J. (1971) Quadratic subspaces and completeness. *Ann. Math. Statist.* **42** 710-721.

[69] Seely, J. and EL-Bassiouni, Y. (1983) Applying Wald's variance components Test. *Ann. Statist.* **11** 197-201.

[70] Speed, T. P. (1987) What is an analysis of variance? *Ann. Statist.* **15** 885-910.

[71] Speed, T. P. and Bailey, R. A. (1987) Factorial dispersion models, *Int. Statist. Rev.* **55** 261-277.

[72] Swallow, W. H. and Searle, S. R. (1978) Minimum variance quadratic unbiased estimation (MINQUE) of variance components. *Technometrics* **20** 265-272.

[73] Swallow, W. H. and Monahan, J. F. (1984) Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics* **26** 47-57.

[74] Szatrowski, T. H. (1980) Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariance. *Ann. Statist.* **8** 802-819.

[75] Szatrowski, T. H. and Miller, J. J. (1980) Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. *Ann. Statist.* **8** 811-819.

[76] Thompson, R. (1973) The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **29** 527-550.

[77] Thompson, W. A. Jr. (1962) The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273-289.

[78] Tiao, G. C. and Box, G. E. P. (1967) Bayesian analysis of a three-component hierarchical design model. *Biometrika* **54** 109-125.

[79] Tiao, G. C. and Tan, W. Y. (1965) Bayesian analysis of random-effects models in the analysis of variance, 1. Posterior distribution of variance components. *Biometrika* **52** 37-53.

[80] Tiao, G. C. and Tan, W. Y. (1966) Bayesian analysis of random-effects models in the analysis of variance, 2. effectf of autocorrelated errors. *Biometrika* **53** 477-495.

[81] U. S. Bureau of the Census (1979) *Research and Evaluation Program: Enumerator variance in the 1970 Census*, PHC(E)-13, Washington, D. C.

[82] Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 295-301.

[83] Zelen, M. and Federer, W. (1964) Applications of the calculus for factorial arrangements 2: Two way elimination of heterogeneity. *Ann. Math. Statist.* **35** 658-672.