CANADIAN THESES ON MICROFICHE

THESES CANADIENNES SUR MICROFICHE

I.S.B.N.

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r. 82/08)

Canada

28

0-315-14414-9

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

62323

NAME OF AUTHOR/*NOM DE L'AUTEUR* Christopher Graham McLaren

TITLE OF THESIS/*TITRE DE LA THÈSE* Estimation of Variance Components in Linear Models.

UNIVERSITY/*UNIVERSITÉ* Simon Fraser University

DEGREE FOR WHICH THESIS WAS PRESENTED/*GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* M. Sc.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE DEGRÉ* 1983.

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Professor C. Villegas

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* 27 Jan 1983 SIGNED/*SIGNÉ* [redacted]

PERMANENT ADDRESS/*RÉSIDENCE FIXE* [redacted]

NL-91 (11-79)

ESTIMATION OF VARIANCE COMPONENTS IN LINEAR MODELS

by

Christopher Graham McLaren

B.Sc.(Agric.) University of Natal (S.A.) 1976

B.Sc.(Hons.) University of Cape Town (S.A.) 1977

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

January, 1983

APPROVAL

Name: Christopher Graham McLaren

Degree: Master of Science

Title of Thesis: Estimation of variance components in linear models.

Chairman: B.S. Thomson

C. Villegas
Senior Supervisor

D. Eaves

R. Lockhart

R. Routledge
External Examiner
Associate Professor
Department of Mathematics
Simon Fraser University

Date Approved: Jan 18 83

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis or dissertation (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this thesis for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation:

Estimation of Variance Components in Linear Models

Author: [signature]
(signature)

C. G. McLaren.
(name)

27 / Jan / 83
(date)

ABSTRACT

Linear models containing covariance parameters are classified according to assumed structure. A generalized representation of balanced variance component models is given, Pukelsheim's derived linear model, where the covariance parameters appear as means, is presented.

Modern techniques for estimating variance components are reviewed. Methods considered include the following: Henderson's use of analysis of variance quadratic forms and method of fitting constants; classical optimization by minimum variance (La Motte) and by minimum norm (Rao, C.R.); Hartley and J.N.K. Rao's maximum likelihood, and Patterson and Thompson's marginal maximum likelihood.

Computing techniques are considered and algorithms for Henderson's method III and minimum norm methods are presented. Some computational results are given.

A logical Bayesian analysis of generalized balanced variance component models is developed, which agrees with the approach of Box and Tiao.

ACKNOWLEDGEMENTS

I would like to acknowledge, with thanks, my supervisory committee for their support and interest during the preparation of this thesis.

I extend my most sincere thanks to my wife, Susan, for her patience, support and hard work in typing the thesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

Introduction

Linear models form the basis of the two most extensively used techniques of data analysis: linear regression and the analysis of variance. This prominent position stems from their intuitive appeal, computational and analytical tractability and frequent adequacy for the job at hand.

Linear models seek to express a response variable as a linear combination of effects. These effects are supposed to 'explain' the value of the response in some practical fashion. Complications arise when it is unsuitable, or undesirable, to assume that the effects can be represented by fixed parameters in the model. In this case, it is sometimes possible to view them as realizations of some unobservable random process. The analysis of linear models containing these 'random effects' involves estimating, or at least checking assumptions about, the variances and covariances of the underlying process. These are the variance components.

The objective of this thesis, is to review modern approaches to the estimation of variance components arising in linear models.

The estimation of variance components has two main directions for application. The first is to provide estimates of the covariance matrix of the response variables. This estimate can then be used with the method of generalized least squares, to provide good estimates of location parameters in the linear model for prediction or estimation, (Mitra and Moore, (1973)).

The second direction is to provide a break-down of the covariance of the response into components, having practical significance in their own right. Examples of this latter use are found in industry, where it may be important to identify the factor in a production process which is causing undue variability in a product and in genetic research, where it is necessary to separate variability in a breeding population into components, due to environmental and various genetic factors.

Sahai (1979) observes; 'Before one embarks on a research project in this area, it is imperative to know what has been done previously. So far, there does not exist any publication to serve this purpose.' He goes on to present an extensive bibliography, comprising about eight hundred entries. This, in itself, is an indication of the complexity of and controversy surrounding the subject of variance component estimation.

The technical source of difficulty appears to be in the requirement that the covariance matrix of the response variables be positive definite, or at least non-negative definite. This requirement imposes frequently complex and always complicating constraints on possible values for the variance components.

1.1. The Scope of the Thesis.

Rao, C.R. and Kleffe, (1980) and Sahai (1979), refer to numerous historical applications of and approaches to variance component estimation.

We start by considering various forms of linear models, in which

variance component estimation can be important. The most general linear model is considered first. This model assumes least knowledge of the underlying random processes. Then, because knowledge about these processes can be used in various estimation techniques, more specific models are considered. Finally, a general form of balanced variance component models is developed. This model incorporates important properties of balanced classification models, for which estimation methods are most advanced. Three examples of real and generated data are described. Various estimation techniques are later applied to these examples.

Estimation methods originated with analysis of variance applied to balanced classification models. We first consider Henderson's (1953) extensions of these methods to unbalanced classification models. These are known as the analysis of variance method of estimating variance components and the method of fitting constants. Searle (1968 and 1971), considered these methods in some detail and they remain the most widely used. For classification models, Henderson's (1953) methods are easy to compute, even by hand, and we consider an algorithm for the method of fitting constants, which is applicable to more general structured models. These estimates are computed for the example data sets.

Because of the lack of demonstrable, desirable properties of these estimators, a variety of alternatives, based on different optimality principles, have been developed.

Rao, C.R. (1970, 1971a, b, and 1972), proposed a general method

called minimum norm quadratic estimation. LaMotte (1973 a and b), considered minimum mean square error as a criterion of estimation in normal models. These approaches have been developed and expanded by Kleffe (1976 a, b, 1977 a, b, c and 1980), Rao, P.S.R.S. and Chaubey (1978) and Chaubey (1980). Their general applicability is a major source of appeal for these estimators, but computational difficulties have, so far, limited their use. The adaptation of algorithms developed for a maximal likelihood approach in structured models having a residual error, is considered and applied to the data examples.

Hartley and Rao, J.N.K. (1967), considered the maximum likelihood approach for normal models. This method overcomes the theoretical difficulties associated with the parameter space constraints, which however remain a computational problem.

Severe bias of the maximum likelihood estimates in some models led Patterson and Thompson (1974) to consider maximization of a 'restricted likelihood', which is the likelihood of a maximal invariant to translations in the location parameters. This approach has an appealing justification from the Bayesian point of view, where the posterior distribution is marginalized over the nuisance parameters to yield an 'integrated' or 'marginal' likelihood. The technique is therefore referred to as the marginal maximum likelihood method. In the classical approach of Patterson and Thompson (1974), integration is with respect to ordinary Lebesgue measure, but in the Bayesian context, there is scope for considering alternative measures over the parameter space.

Harville (1977), reviews the maximum likelihood techniques with consideration of various computational methods. The availability of the computational algorithms, (Dixon and Brown (1977)), has led to increased popularity of the technique. A commercially available programme was used to compute the estimates for the example data sets.

There has been limited development of Bayesian methods for variance component estimation. This is possibly due to difficulty in choosing prior distributions for the variance components, which either realistically represent subjective prior knowledge, or conveniently represent a state of ignorance.

The subjective approach has been most generally developed by Rudolph (1976), but the applicability is limited. Kleffe and Pincus (1974a), consider the identification of Bayes estimators in restricted classes of estimating functions, for which the Bayes risk depends only on the first two moments of the prior distribution.

The state of the logical Bayesian approach, using Jeffreys' rule to provide a prior, is presented by Box and Tiao (1973), but is restricted to simple balanced classification models. A different logical Bayesian viewpoint of Villegas (1977a, b and 1980), is used in this thesis, to extend these methods to general balanced variance component models.

There is considerable difficulty in comparing and contrasting the various methods of variance component estimation. One approach, which appears promising in this respect, is to use the 'dispersion-mean correspondence' of Pukelsheim (1976), which relates quadratic functions

in the sample space to linear functions in an isomorphic space. This is used to re-parameterize the model, so that the variance components appear as location parameters. Use of this re-parameterization as a comparative tool, relies on having an explicit expression for quadratic statistics used in the estimation processes. This expression is not yet available for Henderson's (1953) estimation methods. Finding such an expression would be a desirable development.

CHAPTER 2

Linear Models with Variance and Covariance Parameters.

We are interested in models which represent observable response random variables as linear combinations of fixed parameters and unobservable random variables. The main aim is to consider ways of estimating and, if possible, making inferences about the covariance structure of the unobservable random variables.

There are as many forms to the linear model as there are applications. The different forms reflect greater or lesser knowledge about the distribution of the response. This knowledge can often be exploited when making inferences about the covariance structure and so we consider several general classes of linear models which involve covariance parameters.

The model forms are presented, starting with the most general, the one that assumes the least knowledge and progressing to the most specific. This progression is in reverse to the historical development of the methods of analysis.

Searle (1971), Kempthorne (1975), Harville (1977) and Rao, C.R. and Kleffe (1980) give numerous examples and applications of linear models with covariance parameters.

2.1. The General Linear Model with Covariance Parameters.

The general linear model, or Gauss-Markov model, has model equation:

$$\underline{y} = X\underline{\beta} + \underline{e} \qquad (2.1)$$

where:

$\underline{y}$ is an $n \times 1$ observable response vector.

$\underline{\beta}$ is an unknown $p \times 1$ vector of location parameters.

X is a known $n \times p$ matrix of regressors, covariate values or design vectors, called the structure matrix associated with $\underline{\beta}$.

$\underline{e}$ is an $n \times 1$ unobservable random vector.

The parameter space of $\underline{\beta}$, $\Omega_{\underline{\beta}}$, is $\mathbb{R}^p$. X need not have full column rank, but if it does not, there is a re-parameterization of the model to a full rank model, reducing the dimension of $\underline{\beta}$.

The difference between various variance and covariance models is in the assumptions on the distribution of $\underline{e}$. The most general assumptions are:

$$E(\underline{e}) = 0$$

$$\mathrm{Cov}(\underline{e}) = V(\underline{\theta}) = \sum_{i=1}^{q} \theta_i V_i \qquad (2.2)$$

where:

$\underline{\theta} = (\theta_1, \theta_2 \ldots \theta_q)'$ is an unknown q-vector of variance or covariance parameters.

V_i, $i = 1,2 \ldots q$ are known $n \times n$ symmetric matrices.

The parameter space for $\underline{\theta}$, $\Omega_{\underline{\theta}}$ is restricted to an open subset of $\mathbb{R}^q$ such that $V(\underline{\theta})$ is positive definite, (p.d.) for $\underline{\theta} \in \Omega_{\underline{\theta}}$. No assumption is made about the sign of θ_i nor are the V_i necessarily non-negative definite, (n.n.d.)

In general, any covariance matrix may be written in the form (2.2), with $q = n(n+1)/2$. In practical situations, however, q is often small compared to the number of observations, n.

2.1.1. Identifiability

A linear combination, $\underline{h}'\underline{\theta}$ of the covariance parameters is identifiable if and only if $\underline{h}'\underline{\theta}_1 = \underline{h}'\underline{\theta}_2$ whenever $V(\underline{\theta}_1) = V(\underline{\theta}_2)$ for $\underline{\theta}_1, \underline{\theta}_2 \in \Omega_{\underline{\theta}}$. Provided the distribution of the response depends on $\underline{\theta}$ only through $V(\underline{\theta})$, this definition coincides with the notion of identifiability by distribution of Bunke and Bunke (1974).

Pincus (1974) reports the equivalence of identifiability of $\underline{h}'\underline{\theta}$ with the condition:

$$\underline{h} \in R(H) \text{ or equivalently } \underline{h} \in R(H(\underline{s})) \qquad (2.3)$$

where:

$R(H)$ denotes the column space of H

and $H = (tr(V_i V_j))$ $i,j = 1,2\ldots q$

(meaning that the i,j th element of H is the trace of $V_i V_j$).

or $H(\underline{s}) = (tr(V(\underline{s})^{-1} V_i V(\underline{s})^{-1} V_j))$ for $V(\underline{s}) = \sum_i s_i V_i$ p.d.

A proof of this result is as follows:

If $\underline{h} \in R(H)$ there exists $\underline{b}$ such that

$h_i = \sum_j b_j \, tr(V_i V_j) = tr(V_i V(\underline{b}))$. $i = 1,2...q$ where $tr(\,.\,)$ denotes the trace of a matrix.

Thus $\underline{h}'\underline{\theta}_1 = \sum_i \theta_{1i} \, tr(V_i V(\underline{b})) = tr(V(\underline{\theta}_1)V(\underline{b}))$

and $\underline{h}'\underline{\theta}_2 = tr(V(\underline{\theta}_2)V(\underline{b}))$

Clearly if $V(\underline{\theta}_1) = V(\underline{\theta}_2)$ then $\underline{h}'\underline{\theta}_1 = \underline{h}'\underline{\theta}_2$.

On the other hand we note that:

$V(\underline{\theta}_1) = V(\underline{\theta}_2)$ or $V(\underline{\theta}_1 - \underline{\theta}_2) = 0$ if and only if $\underline{t} = \theta_1 - \theta_2$ is orthogonal to all columns of H. This follows because if

$V(\underline{\theta}_1 - \underline{\theta}_2) = V(\underline{t}) = 0$ then $\sum_i t_i \, tr(V_i V_j) = tr(V(\underline{t})V_j) = 0$.

Conversely, if $\underline{t}'H = 0$ or $tr(V(\underline{t})V_j) = 0$, $j = 1,2...q$ then

$\sum_j t_j \, tr(V(\underline{t})V_j) = tr(V(\underline{t})V(\underline{t})) = 0$ and this only occurs if

$V(\underline{t}) = 0$ since $V(\underline{t})$ is symmetric and $tr(AA')$ is the sum of squares of elements of A. Thus if $\underline{h}'\underline{\theta}$ is identifiable, $\underline{h}$ is orthogonal to $\underline{t}$, whenever $\underline{t}$ is orthogonal to $R(H)$ or equivalently $\underline{h} \in R(H)$.

H can be written as TT' where T is $q \times n^2$ with i th row given by $Vec'(V_i)$ (Appendix A.1.).

If $sp\{V_1, V_2...V_q\}$ denotes the span of the set of V_i in the space of $n \times n$ symmetric matrices, and $\dim sp\{V_1...V_q\}$, its dimension, then $rank(H) = rank(T) = \dim sp\{V_1...V_q\}$.

In particular $\underline{\theta}$ is identifiable if and only if $\{V_1, V_2...V_q\}$ is independant or equivalently $rank(H) = q$.

2.1.2. Normal Distribution Assumptions.

Frequently additional distributional assumptions are required for model (2.1). The usual assumption is that $\underline{e}$ has a central normal distribution. In this case we write:

$$\underline{Y} \sim N(X\underline{\beta}, V(\underline{\theta})) \tag{2.4}$$

where: $V(\underline{\theta}) = \sum_i \theta_i V_i$ is p.d. for $\underline{\theta} \in \Omega_{\underline{\theta}}$

The likelihood of $\underline{Y}$ is:

$$\ell(\underline{\beta},\underline{\theta}/\underline{y}) = (2\pi)^{-n/2} \det(V(\underline{\theta}))^{-1/2} \exp\{-(1/2)\|\underline{y} - X\underline{\beta}\|^2_{V(\underline{\theta})}\} \tag{2.5}$$

where: $\|\underline{y} - X\underline{\beta}\|^2_{V(\underline{\theta})} = (\underline{y} - X\underline{\beta})' V(\underline{\theta})^{-1} (\underline{y} - X\underline{\beta})$

and $\det(V(\underline{\theta}))$ is the determinant of $V(\underline{\theta})$.

Assuming X to have full column rank, the likelihood can be factored since:

$$\|\underline{y} - X\underline{\beta}\|^2_{V(\underline{\theta})} = \|\underline{y} - P(\underline{\theta})\underline{y} + X\hat{\underline{\beta}}(\underline{\theta}) - X\underline{\beta}\|^2_{V(\underline{\theta})}$$

$$= \underline{y}' N(\underline{\theta})\underline{y} - 2\underline{y}' N(\underline{\theta}) X(\hat{\underline{\beta}}(\underline{\theta}) - \underline{\beta}) + \|\hat{\underline{\beta}}(\underline{\theta}) - X\underline{\beta}\|^2_{(X'V(\underline{\theta})^{-1}X)^{-1}}$$

and the middle term vanishes.

where: $P(\underline{\theta}) = X(X'V(\underline{\theta})^{-1}X)^{-1}X'V(\underline{\theta})^{-1}$

$N(\underline{\theta}) = V(\underline{\theta})^{-1}(I - P(\underline{\theta}))$

and $\hat{\underline{\beta}}(\underline{\theta}) = (X'V(\underline{\theta})^{-1}X)^{-1}X'V(\underline{\theta})^{-1}\underline{y}$

so that $X\hat{\underline{\beta}}(\underline{\theta}) = P(\underline{\theta})\underline{y}$.

The factorization of (2.5) is then:

$$\ell(\underline{\beta},\underline{\theta}/\underline{y}) = (2\pi)^{-(n-p)/2} \det(V(\underline{\theta}))^{-1/2} \det(X'V(\underline{\theta})^{-1}X)^{-1/2} \exp\{-(1/2)\underline{y}' N(\underline{\theta})\underline{y}\}$$

$$\times (2\pi)^{-p/2} \det((X'V(\underline{\theta})^{-1}X)^{-1})^{-1/2} \exp\{-(1/2)\|\hat{\underline{\beta}}(\underline{\theta}) - \underline{\beta}\|^2_{(X'V(\underline{\theta})^{-1}X)^{-1}}\} \tag{2.6}$$

2.2. Covariance Component Models.

The first major specialization of the general model is to assume a known structure for the random part, $\underline{e}$, of (2.1):

$$\underline{Y} = X\underline{\beta} + U_1\underline{v}_1 + U_2\underline{v}_2 + \ldots + U_q\underline{v}_q \qquad (2.7)$$

where:

$\underline{y}$, X and $\underline{\beta}$ are as in (2.1),

$\underline{v}_i$, for i = 1,2...q are unobservable q_i x 1 random vectors,

U_i, for i = 1,2...q are known structure matrices associated with the $\underline{v}_i$. U_i is n x q_i for i = 1,2..q,

The most general distributional assumptions are:

$$E(\underline{v}_i) = 0 \quad \text{for} \quad i = 1,2\ldots q,$$

$$E(\underline{v}_i\underline{v}_i') = \mathrm{Cov}(\underline{v}_i) = \Sigma_i \quad \text{for } i = 1,2\ldots q,$$

and $$E(\underline{v}_i\underline{v}_j') = 0 \text{ for } i \neq j. \qquad (2.8)$$

The Σ_i matrices are q_i x q_i unknown covariance matrices called the Covariance Components.

Clearly, $\mathrm{Cov}(\underline{y}) = \sum_{i=1} U_i\Sigma_i U_i'$

Model (2.7) may be referred to as the Structured Covariance Component Model to emphasize the structure on $\underline{e}$.

A convenient representation of this model is:

$$\underline{Y} = X\underline{\beta} + U\underline{v}$$

with $E(\underline{v}) = 0$ and $E(\underline{v}\underline{v}') = \Sigma$

so that $\mathrm{Cov}(\underline{y}) = U\Sigma U' \qquad (2.9)$

where:

$U = (U_1, U_2 \ldots U_q)$ is $n \times m$, $m = \Sigma q_i$

$\Sigma = \text{diag}(\Sigma_i, i = 1,2 \ldots q)$ (an $m \times m$ matrix with blocks, Σ_i, on the diagonal and zeros elsewhere.)

The assumption of Normality for model (2.6) stipulates that:

$\underline{v}_i \sim N(0, \Sigma_i) \quad i = 1,2 \ldots q$

or $\underline{y} \sim N(X\underline{\beta}, U\Sigma U') \qquad (2.10)$

The flexibility in model (2.7) with assumptions (2.8) follows from different forms of Σ_i.

One frequent form of model (2.6) has

$\Sigma_i = \Sigma \quad i = 1,2 \ldots q-1$ and $\Sigma_q = \theta I$

then

$$\text{Cov}(\underline{y}) = \sum_{i=1}^{q-1} U_i \Sigma U_i' + \theta U_q U_q' \qquad (2.11)$$

Another frequent modification of (2.7) follows by taking $\Sigma_i = \theta_i I_{qi} \quad i = 1,2 \ldots q-1$ and $\Sigma_q = \Sigma$, then

$$\text{Cov}(\underline{y}) = \sum_{i=1}^{q-1} \theta_i U_i U_i' + U_q \Sigma U_q' \qquad (2.12)$$

Structured Covariance Models may be represented as general linear models with covariance parameters (2.1) by writing each covariance component as a linear combination of symmetric matrices with unknown coefficients. Some techniques, however, may utilise forms like (2.11) to estimate Σ.

2.2.1. Covariance Component Models with a Residual Error.

In structured models having model equation with the form (2.7), it frequently happens that one of the structure matrices, U_i, is the identity matrix. If we assume that $U_q = I_n$, then the model has form:

$$\underline{y} = X\underline{\beta} + U_1\underline{v}_1 + \ldots U_{q-1}\underline{v}_{q-1} + \underline{e} \qquad (2.13)$$

where: $\underline{e}$ is $n \times 1$ and will be called the residual error effect. The distributional assumptions (2.8) hold for (2.13) with $U_q = I$ and $\underline{v}_q = \underline{e}$.

If $\Sigma_q = R$ then:

$$\text{Cov}(\underline{y}) = \sum_i U_i \Sigma_i U_i' + R$$

or

$$\text{Cov}(\underline{y}) = U \Sigma U' + R \qquad (2.14)$$

where

$$U = (U_1, U_2 \ldots U_{q-1})$$

$$\Sigma = \text{diag}(\Sigma_i, \quad i = 1,2 \ldots q-1).$$

If any structure matrix U_i in (2.7) is $n \times n$, with full rank, then the model can be re-parameterized to have form (2.13).

The special forms (2.11) and (2.12) have useful versions with residual errors, ($U_q = I_n$ and $\underline{v}_q = \underline{e}$)

2.2.2. Random Regression Coefficient Models.

One common example of a covariance component model with residual error component is the random regression coefficient model.

$$\underline{y} = X\underline{\beta} + X\underline{v} + \underline{e} \qquad (2.15)$$

with $E(\underline{v}) = E(\underline{e}) = 0$, $E(\underline{v}\underline{e}') = 0$

$E(\underline{v}\,\underline{v}') = \Sigma$ and $E(\underline{e}\,\underline{e}') = R$

so that $Cov(\underline{y}) = X\Sigma X' + R$.

Frequently $R = \theta I$.

2.3 Variance Component Models.

Variance Component Models have the same structure as (2.7), but the covariance components have form $\theta_i I_{qi}$, $i = 1,2...q$. Hence a variance component model is:

$$\underline{y} = X\underline{\beta} + U_1\underline{v}_1 + ..+U_q\underline{v}_q \qquad (2.16)$$

with

$$E(\underline{v}_i) = 0, \quad E(\underline{v}_i\underline{v}_j) = 0 \quad i \neq j$$

$$E(\underline{v}_i\underline{v}_i') = \theta_i I_{qi}, \quad i = 1,2...q$$

so that $Cov(\underline{y}) = \Sigma\theta_i U_i U_i' = \Sigma\theta_i V_i = V(\underline{\theta})$. (2.17)

Alternatively,

$$Cov(\underline{y}) = V(\underline{\theta}) = U D U' \qquad (2.18)$$

where:

$\underline{y}$, X, $\underline{\beta}$, U_i, and $\underline{v}_i$ are as in (2.6)

V_i are defined as U_iU_i', V_i n.n.d.

$U = (U_1, U_2...U_q)$

and $D = diag(\theta_i I_{qi}, i=1,2...q)$.

The θ_i are variance components and model (2.16) may be referred to as a structured variance component model, to emphasise the structure of the random components.

A major difference between variance component models and

general linear models with covariance parameters is that θ_i, i=1,2...q are non-negative.

Hence

$$\Omega_{\underline{\theta}} \subseteq \mathbb{R}^{q+} \text{ such that } V(\underline{\theta}) \text{ p.d. for } \underline{\theta} \in \Omega_{\underline{\theta}} \qquad (2.19)$$

$\Omega_{\underline{\beta}}$ remains as in (2.1)

The normal distribution assumptions frequently included in model (2.16) are:

$$\underline{v}_i \sim N(0, \theta_i I_{qi})$$

or $$\underline{Y} \sim N(X\underline{\beta}, V(\underline{\theta})). \qquad (2.20)$$

2.3.1. Heteroscedastic Regression Models.

One common example of structured variance component models is the heteroscedastic regression model where the response $\underline{Y}$ comprises q subvectors of length qi,

$$\underline{Y} = (\underline{Y}_1', \underline{Y}_2' \cdots \underline{Y}_q')'$$

Each subvector has model:

$$\underline{Y}_i = X_i\underline{\beta} + \underline{v}_i \quad i = 1,2...q \qquad (2.21)$$

$$E(\underline{v}_i) = 0 \quad E(\underline{v}_i\underline{v}_i') = \theta_i I_{qi}$$

or with $X = (X_1', X_2' \ldots X_q')'$ and U_i being the n x qi matrix of zeros, except for the i th block (conforming to the partition of $\underline{y}$) which is I_{qi}, we have

$$\underline{Y} = X\underline{\beta} + U_1\underline{v}_1 + \ldots + U_q\underline{v}_q \qquad (2.22)$$

with $E(\underline{v}_i) = 0$, $E(\underline{v}_i\underline{v}_j') = 0 \quad i \neq j$

and $E(\underline{v}_i\underline{v}_i) = \theta_i I_{qi} \quad i,j = 1,2...q.$

Frequently $X_1 = X_2 = \ldots = X_q$.

2.3.2. Variance Component Models with a Residual Error.

As with covariance component models, if $U_q = I$ then we write $\underline{e}$ for $\underline{v}_q$ and have a variance component model with a residual error effect:

$$\underline{y} = X\underline{\beta} + U_1\underline{v}_1 + \ldots + U_{q-1}\underline{v}_{q-1} + \underline{e} \qquad (2.23)$$

with $\quad Cov(\underline{y}) = \sum_i \theta_i V_i + \theta_q I_n$

or $\quad Cov(\underline{y}) = U\,D\,U' + \theta_q I_n$

where

$$U = (U_1, U_2 \ldots U_{q-1})$$

$$D + \text{diag}\,(\theta_i I_{qi}, \quad i = 1,2 \ldots q-1).$$

Any variance component model having a non-singular structure matrix, say U_i, can be re-parameterized to have the form (2.23).

An important class of structured variance component models having residual errors are the factorial classification models, which include analysis of variance models for fixed, mixed and random effects. These models have structure matrices X and U_i with additional useful properties. The structure matrices are called design matrices and are considered in the next section.

2.4 Classification Models.

In many applications a response vector can be viewed as a collection of samples from a set of classified populations. The populations are classified according to levels of factors, which may be

treatments or attributes of the universe elements.

Linear models representing responses from such classified populations have a parameter representing the overall mean response and parameter sets representing the main effects and interactions associated with fixed effect factors. (Searle (1971) pp 145-159). Where random effect factors are involved (Searle (1971) pp 376-383) the model contains sets of random deviations associated with the random effects, together with a residual error effect representing deviations in the response vector not accounted for through the fixed or random effects.

Classification models are a class of variance component models, where the structure matrices of (2.16) have important properties derived from the classification. The structure matrices are called effect design matrices.

The model equation for classification models has the form:

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \ldots + X_r\underline{\beta}_r + U_1\underline{v}_1 + \ldots + U_{q-1}\underline{v}_{q-1} + \underline{e} \qquad (2.24)$$

where:

$\underline{y}$ is the $n \times 1$ observable response vector.

X_1 is an $n \times 1$ vector of 1,s. X_1 is the mean effect design matrix and may be written as $\underline{1}$. X_1 is always present.

β_1 is the mean effect parameter.

$\underline{\beta}_i$, $i = 2,3\ldots r$ are $p_i \times 1$ vectors of unknown location parameters called fixed effects.

X_i, $i = 2,3...r$ are $n \times p_i$ known design matrices associated with the fixed effects.

$\underline{v}_i$ $i = 1,2...q-1$ are $q_i \times 1$ vectors of unobservable random deviations called random effects.

U_i, $i= 1,2...q-1$ are $n \times q_i$ known design matrices associated with the random effects.

$\underline{e}$ is an $n \times 1$ vector of unobservable random variables, called the residual effect.

$\underline{e}$ is always present.

If $r = 1$ the model is a random effect model. If $q = 1$ it is a fixed effect model and if $r > 1$ and $q > 1$ it is a mixed effect classification model.

The model equation may be written as:

$$\underline{Y} = X\underline{\beta} + U\underline{v} + \underline{e} \qquad (2.25)$$

where:

$$X = (X_1, X_2...X_r), \quad \underline{\beta} = (\beta_1, \underline{\beta}_2'...\underline{\beta}_r')'$$

and $$U = (U_1,....U_{q-1}), \quad \underline{v} = (\underline{v}_1',...\underline{v}_{q-1}')'$$

The basic distributional assumptions associated with model (2.24) are:

$$E(\underline{v}_i) = E(\underline{e}) = 0, \ i=1,2...q-1$$

$$E(\underline{v}_i\underline{v}_j') = 0 \text{ for } i \neq j, \quad E(\underline{v}_i\underline{e}') = 0, \ i=1,2...q-1,$$

$$E(\underline{v}_i\underline{v}_i') = \theta_i I_{qi}, \ i=1,2,..q-1, \text{ and } E(\underline{ee}') = \theta_q I_n \qquad (2.26)$$

As a consequence of these assumptions we see that $\theta_i \geq 0$ all i,

and $\theta_q > 0$ to ensure that the covariance matrix is p.d. Hence

$$\Omega_{\underline{\theta}} \subset R^{q+}. \qquad (2.27)$$

Writing V_i for $U_i U_i'$, we obtain:

$$\text{Cov}(\underline{y}) = V(\underline{\theta}) = \sum_i \theta_i V_i + \theta_q I_n \qquad (2.28)$$

or $$\text{Cov}(\underline{y}) = UDU' + \theta_q I_n \quad \text{from (2.25)} \qquad (2.29)$$

where: $D = \text{diag}(\theta_i I_{qi}, \quad i=i,2\ldots q-1)$

The Normal distribution assumptions are the same as (2.20).

2.4.1. Properties of Effect Design Matrices.

The effect design matrices of (2.24), X_i and U_i, are matrices of zeros and ones. They have exactly one 1 in each row. (Searle (1971) p 166). A consequence of this is that:

$$X_j \underline{1} = \underline{1}, \quad j=1,2\ldots r \qquad (2.30)$$

and $$U_j \underline{1} = \underline{1}, \quad j=1,2\ldots q \qquad (2.31)$$

Also since every row has exactly one 1, the columns within each design matrix are orthogonal.

Each effect in the model represents a sub-classification of the response according to some factorial combination which should be explained by that effect. One column of the effect design matrix is associated with each cell of the sub-classification. Each column has 1,s in positions corresponding to response elements in that cell. Thus if the sub-classification associated with the j,th effect has $n_{i,j}$ observations in the i,th cell, we obtain:

$$X_j' X_j = \text{diag}(n_{i,j}, \quad i=1,2\ldots p_j)$$

or $$U_j' U_j = \text{diag}(n_{i,j}, \quad i=1,2\ldots q_j) \qquad (2.32)$$

A classification design is balanced if and only if

$n_{i,j} = n_j$ for all cells i and each effect j. (2.33)

Since $X_1 = \underline{1}$ we have with (2.30) and (2.31) that

$$R(X_1) \subset R(X_j) \text{ and } R(X_1) \subset R(U_j) \tag{2.34}$$

2.4.2. Appropriateness of Classification Models.

There is considerable controversy over the appropriateness of the model (2.24) and distributional assumptions (2.26) to represent classified data sets.

One aspect of the controversy centers on interpretation of the fixed effects in the light of constraints, which have to be imposed on the parameter estimates to facilitate analysis. (Searle (1971) p 204, Kempthorne (1975) and Urquhart and Weeks (1978)). We are not directly concerned with this problem, as the fixed effect parameters are largely viewed as nuisance parameters for the purposes of variance component estimation. However, one aspect of it spills over to the random effects, associated with interactions involving fixed and random factors. If we assume that the constraints on the parameter estimates also hold for the population values, then the random deviations should also be constrained over their fixed factor indices, (Kempthorne (1975)). This would mean that the assumption of independence within random effect vectors in (2.26) would not hold.

Jennrich and Sampson (1978) suggest a re-parameterization for side-stepping this situation, since many computational procedures rely

on the independence assumption.

2.4.3. Balanced Classification Models.

The property of balance for classification models (2.33) has important consequences for the design matrices.

Specifically in the model (2.24), if the classification is balanced, then the matrices $V_i = U_iU_i'$ have two important properties:

a) $\{V_1,\ldots V_{q-1}, I_n\}$ is a subset of a commutative quadratic subspace of the vector space of symmetric $n \times n$ matrices over the real field. (2.35)

b) $R(X_i)$ $i=1,2\ldots r$ are invariant subspaces of $\mathbb{R}^n$ under the operators V_j, $j = 1,2\ldots q$. (2.36)

The notion of quadratic subspaces and many properties are presented by Seeley (1971). The defining property is:

S is a quadratic subspace if and only if

$$A \in S \Rightarrow A^2 \in S. \tag{2.37}$$

The property of commutativity means that:

$$A, B \in S \rightarrow AB = BA \in S \tag{2.38}$$

The notion of invariance of subspaces of $\mathbb{R}^n$ to linear operators is defined as follows:

B a subspace of $\mathbb{R}^n$ is invariant under A if and only if $\underline{x} \in B \rightarrow A\underline{x} \in B$ (2.39)

The invariance property (2.36) is equivalent to the relationships:

$$P_j V_i = V_i P_j = P_j V_i P_j \quad j = 1,2...r,\ i = 1,2...q \tag{2.40}$$

where: $P_j = X_j (X_j' X_j)^{-1} X_j'$ is the symmetric and idempotent projection operator onto $R(X_j)$.

This equivalence follows by noting that $R(X_j) = R(P_j)$ and P_j is the identity operator on $R(X_j)$. Hence, if $R(P_j)$ is invariant to V_i, then $V_i P_j \in R(P_j)$ so $P_j V_i P_j = V_i P_j$ and the rest follows by symmetry of the matrices. On the other hand, if (2.40) holds, then given $\underline{x} \in R(P_j)$, there exists $\underline{y} \in \mathbb{R}^n$ such that $\underline{x} = P_j \underline{y}$. Hence $V_i \underline{x} = V_i P_j \underline{y} = P_j V_i \underline{y} \in R(P_j)$ so that $R(X_j) = R(P_j)$ is invariant to V_i.

Not only do balanced classification models possess the properties of design matrices 2.4.1, and the properties of balance (2.35) and (2.36), but these properties hold for all the structure matrices, irrespective of the fixed or random nature of the effects. That is if $Z_j = X_j X_j'$, $j=1,2...r$, then Z_j are also elements of the commutative quadratic subspace and the column spaces $R(X_j)$ and $R(U_j)$ are all invariant to all the operators Z_j and V_i.

The properties (2.35) and (2.36) for balanced classification models follow from results given by Searle and Henderson (1979).

We consider an s-factor classification with one observation per cell and n_k, $k=1,2...s$ levels for each factor.

Replication is treated as a factor nested within all others. There will be $n = \prod_{k=1}^{s} n_k$ observations.

Searle and Henderson (1979) report that the matrices $Z_j = X_j X_j'$, $j=1,2...r$ and $V_i = U_i U_i'$, $i=1,2...q$ with $U_q = I_n$ are a subset of:

$$K(s) = \{K_{\underline{i}} : \text{where } \underline{i} = (i_s, i_{s-1}, \ldots i_1) \text{ with } i_j = 1 \text{ or } 0,\ j=1,2..s$$

$$\text{and } K_{\underline{i}} = J_{n_s}^{1-i_s} \otimes J_{n_{s-1}}^{1-i_{s-1}} \otimes \ldots \otimes J_{n_1}^{1-i_1}$$

$$\text{where } J_{n_j}^{1-i_j} \text{ is the } n_j \times n_j \text{ matrix of } 1\text{s if } i_j=0,$$

$$\text{or } I_{n_j} \text{ if } i_j=1\} \qquad (2.41)$$

$K(s)$ has 2^s elements corresponding to the 2^s possible values for the binary index $\underline{i}$. $sp(K(s))$, the subspace of symmetric n x n matrices spanned by the elements of $K(s)$ is a commutative quadratic subspace. This follows from the definition of $K_{\underline{i}}$ since if $K_{\underline{i}}$ and $K_{\underline{j}} \in K(s)$ then:

$$K_{\underline{i}} K_{\underline{j}} = x\, K_{\underline{i}.\underline{j}} \qquad (2.42)$$

where: $x \in \mathbb{R}$ and $\underline{i}.\underline{j}$ is the Boolean addition of the binary vectors $\underline{i}$ and $\underline{j}$. Clearly $K_{\underline{i}.\underline{j}} \in sp(K(s))$.

Furthermore, for $K_{\underline{i}} \in K(s)$, $R(K_{\underline{i}})$ is invariant to all $K_{\underline{j}} \in K(s)$ since

$$K_{\underline{j}} K_{\underline{i}} = K_{\underline{i}} K_{\underline{j}} \qquad (2.43)$$

and the left hand matrix has columns in $R(K_{\underline{i}})$.

Hence $R(X_j)$ is invariant to V_i for i=1,2...q and j=1,2...r, since $Z_j = X_j X_j' = K_{\underline{i}}$ for some $\underline{i}$ and $R(Z_j) = R(X_j)$. (2.44)

An important property of commutative quadratic subspaces is that they have a basis of symmetric, idempotent and mutually orthogonal

matrices (Seely (1971) lemma 6).

Thus, there is a basis $\{Q_1, Q_2, \ldots Q_2 s\}$ for the space sp(K(s)) where $Q_i = Q_i'$, $Q_i Q_i = Q_i$, and $Q_i Q_j = 0$ for i#j. (2.45)

Hence we can write the covariance matrix in terms of this basis since:

$$V_j = \sum_{i=1}^{m} t_{ij} Q_i \quad , \quad j=1,2\ldots q, \quad \text{where } m=2^s. \tag{2.46}$$

The t_{ij} coefficients are known, and in the case of balanced classification models, Searle and Henderson (1979) show that the full m x m matrix $T = (t_{ij})$ of coefficients for expressing the elements of K(s) in terms of symmetric, idempotent and mutually orthogonal matrices is given by:

$$T = \begin{pmatrix} n_1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} n_2 & 0 \\ 1 & 1 \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} n_s & 0 \\ 1 & 1 \end{pmatrix} \tag{2.47}$$

where: $\otimes$ denotes the Kroneker Product (2.76).

Since $V_j \in K(s)$, we can find the t_{ij} of (2.46) from the corresponding rows of T. The covariance matrix $V(\underline{\theta})$ can then be expressed as:

$$V(\underline{\theta}) = \sum_{j=1}^{q} \theta_j V_j = \sum_{j=1}^{q} \theta_j \sum_{i=1}^{m} t_{ij} Q_i = \sum_{i=1}^{m} \omega_i Q_i = V(\underline{\omega}) \tag{2.48}$$

where ω_i is the i th eliment of $\underline{\omega}$ given by:

$$\underline{\omega} = T_1 \underline{\theta} \tag{2.49}$$

and T_1 is the m x q matrix of t_{ij},s obtained from (2.47). T_1 has independent columns because of the independence of V_j, j=1,2...q.

Hence, in the case of balanced classification models we can re-parameterize the model in terms of $\underline{\omega}$ and Q_i, i=1,2...m.

The model becomes:

$$\underline{y} = X\underline{\beta} + \underline{e}$$

with: $$\text{Cov}(\underline{y}) = V(\underline{\omega}) = \sum_{i=1}^{m} \omega_i Q_i \tag{2.50}$$

Where: Q_i are known n x n symmetric matrices satisfying $Q_i^2 = Q_i$,

and $Q_iQ_j = 0$, $i \neq j$ (2.51)

and $R(X)$ is invariant to Q_i, $i=1,2...m$ (2.52)

The parameter space for $\underline{\omega}$, $\Omega_{\underline{\omega}}$, is however subject to rather complicated constraints. If we complete T_1 of (2.49) to a square non-singular matrix by adding $m-q$ independent columns to form $T = [T_1, T_2]$, then:

$$\underline{\omega} = T\begin{pmatrix}\underline{\theta}\\ \underline{0}\end{pmatrix} \tag{2.53}$$

Conversely, if $[B_1, B_2] = T^{-1}$ then

$$B_1\underline{\omega} = \underline{\theta} \tag{2.54}$$

and $$B_2\underline{\omega} = \underline{0} \tag{2.55}$$

$\underline{\theta}$ was restricted (2.27) so that $\underline{\theta} \in R^{q+}$ and hence $\underline{\omega}$ is restricted by:

$$C: \quad B_1\underline{\omega} > \underline{0} \quad \text{and} \quad B_2\underline{\omega} = \underline{0}. \tag{2.56}$$

and $$\Omega_{\underline{\omega}} = \{\underline{\omega} : \underline{\omega} \in R^q \text{ and } \underline{\omega} \text{ satisfies } C\} \tag{2.57}$$

In many applications, (Box and Tiao (1973) Ch 5 and 6), the constraints $B_1\underline{\omega} > \underline{0}$ have the form: $0 < \omega_q \leq \omega_{q-1} \leq \ldots \leq \omega_1$, and the constraints $B_2\underline{\omega} = \underline{0}$ are void.

The advantages of the properties (2.51) and (2.52) of balance are so important that we define a class of variance component models which possess these properties but which are not necessarily classification models.

2.5. Balanced Variance Component Models

Motivated by the properties of balanced classification models, 2.4.3., we consider general models possessing some balance properties which have important consequences for methods of analysis.

$$\underline{y} = X\underline{\beta} + \underline{e} \tag{2.58}$$

$$\text{Cov}(\underline{y}) = V(\underline{\omega}) = \sum_{i=1}^{q} \omega_i Q_i \tag{2.59}$$

and $E(\underline{e}) = \underline{0}$

where: $\underline{y}$, X, $\underline{\beta}$ and $\underline{e}$ are as in (2.1)

Q_i, $i=1,2...q$ are known $n \times n$ symmetric matrices satisfying

$$Q_i^2 = Q_i, \text{ and } Q_iQ_j = 0, \quad i \neq j \tag{2.60}$$

$$R(X) \text{ is invariant to } Q_i, \quad i=1,2...q \tag{2.61}$$

and $\underline{\omega}$ is an unknown q-vector of parameters taking values in $\Omega_{\underline{\omega}}$.

The parameters of interest are some independent linear combinations of $\underline{\omega}$, say $\underline{\theta}$ where:

$$\underline{\theta} = B_1\underline{\omega} \tag{2.62}$$

$\Omega_{\underline{\omega}}$ is a subspace of $\mathbb{R}^{q+}$ restricted by some linear constraints C such that $V(\underline{\omega})$ is p.d.

2.5.1. Consequences of Balance for Variance Component Models.

The properties (2.60) and (2.61) have important consequences for the balanced variance component model (2.58):

a) $PQ_i = Q_iP = PQ_iP$, for $i=1,2...q$ where $P = X(X'X)^{-}X'$ is the projection operator onto $R(X)$. This follows from the definition of invariant subspaces (2.40). (2.63)

Consequently, $Q_iX = XB_i$ for $B_i = (X'X)^{-}X'Q_iX$, $i = 1,2...q$.

b) $NQ_i = Q_iN = NQ_iN$, $i = 1,2...q$, where $N = I - P$ (2.64)

c) If $A = \Sigma a_iQ_i$ then $R(X)$ is invariant to A,

or equivalently, $PA = AP = PAP$. (2.65)

d) If $V(\underline{\omega}) = \Sigma\omega_iQ_i$ with $\omega_i > 0$, then $V(\underline{\omega})^{-1} = (1/\omega_i)Q_i$ (2.66)

This follows from (2.60) and is the basis for the method of computing $V(\underline{\omega})^{-1}$ for balanced classification models which was suggested by Searle and Henderson (1979).

e) $\sum_{i=1}^{q} Q_i = I_n$. This follows since, we see from (2.60) that ΣQ_i is idempotent, and being a sum of idempotent matrices we have that: $n \geq \text{rank}(\Sigma Q_i) = \Sigma\ \text{rank}(Q_i)$.

On the other hand, we know that $V(\underline{\omega}) = \Sigma\omega_iQ_i$ is p.d. for some $\underline{\omega} \in \Omega_{\underline{\omega}}$ and hence that: $n = \text{rank}(V(\underline{\omega})) \leq \Sigma\ \text{rank}(Q_i)$.

Hence ΣQ_i being full rank and idempotent is the identity. (2.67)

f) $V(\underline{\omega})$ has eigen values ω_i with multiplicities m_i, (2.68)

where $m_i = \text{rank}(Q_i)$, $i = 1,2...q$.

From (2.60), $V(\underline{\omega})Q_i = \omega_iQ_i$, $i = 1,2...q$, so ω_i is an eigen value with multiplicity m_i. Also since $\Sigma\ \text{rank}(Q_i) = m_i = n$, by (2.67), we have all the eigen values of $V(\underline{\omega})$.

g) $\det(V(\underline{\omega})) = \prod_{i=1}^{q} \omega_i^{m_i}$ where $m_i = \text{rank}(Q_i)$ (2.69)

h) If X has full column rank, $\det(X'V(\underline{\omega})^{-1}X) = \det(X'X) \prod_{i=1}^{q} \omega_i^{-r_i}$

$= \det(X'X)\det((X'X)^{-1}X'V(\underline{\omega})^{-1}X)$, where $r_i = \text{rank}(PQ_i)$. (2.70)

This follows by observing that if $B_i = (X'X)^{-1}X'Q_iX$, then $(X'X)^{-1}X'V(\underline{\omega})^{-1}XB_i = (1/\omega_i)\ B_i$ and also $\text{rank}(B_i) = \text{tr}(PQ_i) = r_i$.

Also, $(X'X)^{-1}X'V(\underline{\omega})^{-1}X = \Sigma(1/\omega_i)\ B_i$ and hence

$\text{rank}((X'X)^{-1}X'V(\underline{\omega})^{-1}X) \leq \Sigma r_i$.

2.5.2 The Normal Distribution Assumptions for Balanced Variance Component Models.

If in addition to the distributional assumptions (2.59) we assume that

$$\underline{y} \sim N(X\underline{\beta}, V(\underline{\omega})) \tag{2.71}$$

then the likelihood of $\underline{y}$ is given by (2.5).

The properties of balance lead to convenient representations of this likelihood.

Firstly, the cross product terms in the expansion of $\|\underline{y} - P\underline{y} + X\hat{\underline{\beta}} - X\underline{\beta}\|^2_{V(\underline{\omega})}$ involve $(\underline{\beta} - \hat{\underline{\beta}})'X'V(\underline{\omega})^{-1}N\underline{y}$ which vanishes because of the invariance, (2.64) and (2.65).

Where: $\hat{\underline{\beta}} = (X'\bar{X})\bar{X}'\underline{y}$, $P = X(X'\bar{X})\bar{X}'$ and $N = I - P$

Thus: $$\|\underline{y} - X\underline{\beta}\|^2_{V(\underline{\omega})} = \|N\underline{y}\|^2_{V(\underline{\omega})} + \|X\underline{\beta} - X\hat{\underline{\beta}}\|^2_{V(\underline{\omega})} \tag{2.72}$$

Secondly, from (2.64) and (2.66),

$$\|N\underline{y}\|^2_{V(\underline{\omega})} = \sum_{i=1}^{q} (1/\omega_i)\underline{y}'NQ_iN\underline{y} \tag{2.73}$$

Thus with (2.69) we can write the liklihood as:

$$\ell(\underline{\beta},\underline{\omega}/\underline{y}) = (2\pi)^{-n/2} \prod_{i=1}^{q} \omega_i^{-mi/2} \exp\{-\sum_{i=1}^{q} (1/2\omega_i)\underline{y}'NQ_iN\underline{y} - (1/2)\|X(\underline{\beta} - \hat{\underline{\beta}})\|^2_{V(\underline{\omega})}\} \tag{2.74}$$

where $mi = \text{rank}(Q_i)$.

2.6. Derived Models and the Dispersion-Mean Correspondence.

Several techniques for estimating location parameters in linear models lead to estimators with desirable properties. One approach to the estimation of variance components is to re-formulate the model, so that the variance components appear as location

parameters and then to use the known theory of linear estimation to derive estimators with desirable properties.

2.6.1. A General Derived Model.

For the general model with covariance parameters (2.1), we consider the $n^2 x 1$ derived response vector:

$$\tilde{y} = (y - X\beta) \otimes (y - X\beta) \tag{2.75}$$

where

$\otimes$ denotes the Kronecker product defined for matrices A and B in $\mathbb{R}^{pxq}$ and $\mathbb{R}^{rxs}$ respectively as: $A \otimes B = (a_{ij}B)$. $A \otimes B$ is the pr x qs matrix with i,j th block given by $a_{ij}B$ for

$$i=1,2...p, \quad j=1,2...q. \tag{2.76}$$

Important relationships between the Kronecker product and the Vec(.) operator (A.1) are given in Pukelsheim (1976) and Henderson and Searle (1979). These follow from the definition (2.76). For any matrices A,B,C for which the products are defined:

$$\text{a) } Vec(ABC) = (C' \otimes A)Vec(B) \tag{2.77}$$

$$\text{b) } tr(AB) = Vec'(A')Vec(B) \tag{2.78}$$

c) For any vectors x and y,

$$Vec(xy') = y \otimes x \tag{2.79}$$

Hence $\tilde{y} = Vec((y - X\beta)(y - X\beta)')$

and $$E(\tilde{y}) = \text{Vec}(E(\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})')$$

$$= \text{Vec}(V(\underline{\theta}))$$

$$= \sum_{j=1}^{q} \theta_j \text{Vec}(V_j) = \tilde{X}\underline{\theta} \qquad (2.80)$$

Where:

$$\tilde{X} = (\text{Vec}(V_1), \text{Vec}(V_2) \ldots \text{Vec}(V_q)) \text{ is } n^2 \times q.$$

and $$\text{Cov}(\tilde{y}) = \text{Cov}(\text{Vec}(\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})') = F \qquad (2.81)$$

F is the $n^2 \times n^2$ matrix of fourth order moments for the vector of random variables $(\underline{y} - X\underline{\beta})$, with ordering defined by the $\text{Vec}(\ .\ )$ operator.

The derived model is therefore:

$$\tilde{\underline{y}} = \tilde{X}\underline{\theta} + \tilde{\underline{e}} \qquad (2.82)$$

with $E(\tilde{\underline{e}}) = \underline{0}$ and $\text{Cov}(\tilde{\underline{y}}) = F$.

Two difficulties with this derived model are that $\tilde{\underline{y}}$ is not observable since $\underline{\beta}$ is generally unknown, and F depends on both $\underline{\beta}$ and $\underline{\theta}$ and is not usually of full rank. The sample space of $\tilde{\underline{y}}$ is not the complete space $\mathbb{R}^{n^2}$.

2.6.2. The Dispersion-Mean Correspondence.

The advantage of the derived model is that the parameters of interest are now the location parameters. Pukelsheim (1976) observed that linear functions of $\tilde{\underline{y}}$ in the derived model correspond to quadratic functions of $\underline{y}$ in the original model (2.1). This follows from isomorphism properties of the $\text{Vec}(\ .\ )$ operator:
Clearly, to any n^2-vector $\underline{r}$ there corresponds an $n \times n$ matrix R such that $\text{Vec}(R) = \underline{r}$. $\text{Vec}(R')$ contains the same elements as $\underline{r}$

but in a different order. Henderson and Searle (1979) define the $n^2 \times n^2$ Vec-permutation matrix $I_{(n,n)}$ by the relationship:

$$\underline{r} = \text{Vec}(R) = I_{(n,n)}\text{Vec}(R'). \tag{2.83}$$

$I_{(n,n)}$ has n^2 blocks arranged in an $n \times n$ array with each block being $n \times n$, zero except that the i,j th block has a 1 in its j,i th position.

Some properties of the Vec-permutation matrix given by Henderson and Searle (1979) are:

$$\text{a) } I_{(n,n)}I_{(n,n)} = I_{n^2} \tag{2.84}$$

$$\text{b) } I_{(n,n)}\text{Vec}(A) = \text{Vec}(A) \text{ when } A \text{ is symmetric.} \tag{2.85}$$

$$\text{c) } I_{(n,n)}(A \otimes B) = (B \otimes A)I_{(n,n)} \text{ for any A and B.} \tag{2.86}$$

With these results it is easy to demonstrate the correspondence between linear functions in $\tilde{\underline{y}}$ and quadratic functions in $(\underline{y} - X\underline{\beta})$. Pukelsheim calls this the dispersion mean correspondence.

With R and $\underline{r}$ as in (2.83), and noting that $I_{(n,n)}\tilde{\underline{y}} = \tilde{\underline{y}}$

$$\begin{aligned}\underline{r}'\tilde{\underline{y}} &= \text{Vec}'(R)\tilde{\underline{y}} = \text{Vec}'((1/2)R)\tilde{\underline{y}} + \text{Vec}'((1/2)R')I_{(n,n)}\tilde{\underline{y}} \\ &= \text{Vec}'((1/2)(R + R'))\tilde{\underline{y}} = \text{tr}(A(\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})') \\ &= (\underline{y} - X\underline{\beta})'A(\underline{y} - X\underline{\beta})\end{aligned} \tag{2.87}$$

where: $A = (1/2)(R + R')$ is symmetric.

2.6.3 A Derived Model for Invariant Quadratic Estimation.

A device frequently employed to reduce the effect of nuisance parameters is to restrict attention to estimating functions which are invariant to transformations of the nuisance parameters.

In model (2.1) $N\underline{y}$, the projection of the response onto the orthogonal compliment of $R(X)$ is invariant to translations in $\underline{\beta}$.

Hence $N\underline{y} = N(\underline{y} - X\underline{\beta})$ for $\underline{\beta} \in \Omega_{\underline{\beta}}$ since $NX = 0$.

If we define $\tilde{\underline{y}}_N$ by:

$$\tilde{\underline{y}}_N = N\underline{y} \otimes N\underline{y} = N(\underline{y} - X\underline{\beta}) \otimes N(\underline{y} - X\underline{\beta}) = N \otimes N\, \tilde{\underline{y}} \qquad (2.88)$$

then $$E(\tilde{\underline{y}}_N) = N \otimes N\, \tilde{X}\underline{\theta} = \tilde{X}_N\underline{\theta} \qquad (2.89)$$

where $\tilde{X}_N = (\mathrm{Vec}(NV_1N), \mathrm{Vec}(NV_2N), \ldots \mathrm{Vec}(NV_qN))$ using (2.77),

and $$\mathrm{Cov}(\tilde{\underline{y}}_N) = (N \otimes N)\, F\, (N \otimes N) = F_N \qquad (2.90)$$

The derived model for invariant quadratic estimation is:

$$\tilde{\underline{y}}_N = \tilde{X}_N\underline{\theta} + \tilde{\underline{e}}_N \qquad (2.91)$$

with $E(\tilde{\underline{e}}_N) = 0$

and $\mathrm{Cov}(\tilde{\underline{y}}_N) = F_N$

In this derived model linear estimation corresponds to invariant quadratic estimation in the original model. In this case the response $\tilde{\underline{y}}_N$ is observable since it does not depend on $\underline{\beta}$.

2.6.4. A Derived Model for Mean and Dispersion.

Kleffe (1978) considers simultaneous estimation of $\underline{\beta}$ and $\underline{\theta}$ in a model derived from (2.1) by defining $\tilde{\underline{y}}^*$

$$\tilde{\underline{y}}^* = \begin{pmatrix} \underline{y} \otimes \underline{y} \\ \underline{y} \end{pmatrix} \qquad (2.92)$$

$$\underline{y} \otimes \underline{y} = X \otimes X\tilde{\underline{\beta}} + X\underline{\beta} \otimes \underline{e} + \underline{e} \otimes X\underline{\beta} + \underline{e} \otimes \underline{e}$$

hence $$E(\tilde{\underline{y}}^*) = \begin{pmatrix} X \otimes X & 0 \\ 0 & X \end{pmatrix} \begin{pmatrix} \tilde{\underline{\beta}} \\ \underline{\beta} \end{pmatrix} + \begin{pmatrix} \tilde{X} \\ 0 \end{pmatrix} \underline{\theta} \qquad (2.93)$$

where $\tilde{\underline{\beta}} = \underline{\beta} \otimes \underline{\beta}$ and $\tilde{X}$ is as in (2.80)

2.6.5. Covariance Structures in Derived Models.

A major difficulty with linear estimators in the derived models is that their distributions depend on the matrices F and F_N (2.81), (2.90) of fourth order moments. As previously observed, these matrices depend on the unknown parameters and may not have full rank.

The most general model for which explicit forms are available, is the structured variance component model (2.16), with assumptions of independence between and within random components, as well as the assumptions of finite fourth order moments. Pukelsheim (1977), Gnot, Klonecki and Zmyslony (1977), Drygas (1977) and Henderson and Searle (1979), give forms for this general covariance.

For normal models, the symmetry and kurtosis properties simplify the structure considerably.

If $\underset{\sim}{y} \sim N(0,D)$ where $D = \text{diag}(\theta_1\ \theta_2 \ldots \theta_n)$

then $\underset{\sim}{z} = \text{vec}(\underset{\sim}{y}\underset{\sim}{y}') = \underset{\sim}{y} \otimes \underset{\sim}{y} \sim \text{Wishart}(1,D)$

and $E(\underset{\sim}{z}) = \text{vec}(D)$

With the fact that, for normal variables,

$$\text{Cov}(y_i y_j, y_k y_m) = \text{Cov}(y_i y_k)\text{Cov}(y_j y_m) + \text{Cov}(y_i y_m)\text{Cov}(y_j y_k)$$

it is easy to check that:

$$\text{Cov}(\underset{\sim}{z}) = (D \otimes D)(I_{n^2} + I_{(n,n)}) \qquad (2.94)$$

where $I_{(n,n)}$ is the vec-permutation matrix, (2.83).

For $\underline{y} \sim N(X\underline{\beta}, V)$ with V p.d.,

we can find an orthogonal matrix Z such that $Z'VZ = D$,

where D is a diagonal matrix. Then,

$$Z'(\underline{y} - X\underline{\beta}) \sim N(0, D).$$

If $$\underset{\sim}{\underline{z}} = Z'(\underline{y} - X\underline{\beta}) \otimes Z'(\underline{y} - X\underline{\beta}),$$

we have that:

$$\operatorname{Cov}(\underset{\sim}{\underline{z}}) = (D \otimes D)\ (I + I_{(n,n)})$$

and $$\underset{\sim}{\underline{y}} = (\underline{y} - X\underline{\beta}) \otimes (\underline{y} - X\underline{\beta}) = (Z \otimes Z)\underset{\sim}{\underline{z}}.$$

Hence, using (2.86),

$$\begin{aligned}\operatorname{Cov}(\underset{\sim}{\underline{y}}) &= (Z \otimes Z)\ (D \otimes D)\ (I + I_{(n,n)})\ (Z \otimes Z) \\ &= (V \otimes V)\ (I + I_{(n,n)}) \qquad (2.95)\end{aligned}$$

Further, since $\underset{\sim}{\underline{y}}_N = (N \otimes N)\underset{\sim}{\underline{y}}$, we have,

$$\operatorname{Cov}(\underset{\sim}{\underline{y}}_N) = (NVN \otimes NVN)\ (I + I_{(n,n)}). \qquad (2.96)$$

2.7. Example Data Sets from Classification Models.

2.7.1. Data Set 1. Data from a Cattle Breeding Experiment.

As an example of data which is commonly analysed according to unbalanced classification models, we consider data on eight variables, measured on 208 calves during a cattle breeding experiment. The variates are described in Table (2.1). The classifying factors were:

a) Sex of calf — SEX
b) Place of birth — SITE
c) Breed of sire — SBRD
d) Breed of dam — DBRD
e) Sire — SIRE

Table (2.1): Data Variables for Data Set 1.

VARIABLE	NAME	MEAN	STD. ERROR
Mass at birth (kg)	MASB	32.587	4.294
Height at birth (mm)	HTAB	686.89	31.304
Depth at birth (mm)	DPTB	421.82	28.482
Length at birth (mm)	LENB	486.20	36.297
Height at weaning (mm)	HTAW	1041.8	41.317
Depth at weaning (mm)	DPTW	424.12	40.921
Length at weaning (mm)	LENW	954.87	38.486
Mass at weaning (kg)	MASW	198.84	26.320

The model is described in the following table:

Table (2.2) Classification Model for Data Set 1.

EFFECT	DISTRIBUTION	DEGREES OF FREEDOM
MEAN	Fixed	1
SEX	Fixed	1
SITE	Fixed	1
SBRD	Fixed	1
SITE X SBRD	Fixed	1
DBRD	Fixed	1
SITE X DBRD	Fixed	1
SBRD X DBRD	Fixed	1
SIRE (IN SITE X SBRD)	Random	11
SIRE X DBRD	Random	22
RESIDUAL	Random	208

SIRE was nested within SITE X SBRD combinations. The distribution of observations throughout the cells was:

Table (2.3) Number of Observations in Design Cells for Data Set 1.

LEVEL OF: SEX	SITE	SBRD	DBRD	LEVEL OF SIRE 40	61	64	67	34	38	44	23	42	9	27	TOTALS
1	1	1	1	5	5	3	3								16
1	1	1	2	7	2	6	3								18
1	1	2	1					5	3	4					12
1	1	2	2					4	3	4					11
1	2	1	1								8	4			12
1	2	1	2								7	5			12
1	2	2	1										9	4	13
1	2	2	2										3	5	8
2	1	1	1	5	5	7	8								25
2	1	1	2	4	9	4	7								24
2	1	2	1					3	6	5					14
2	1	2	2					4	4	7					15
2	2	1	1								2	3			5
2	2	1	2								4	2			6
2	2	2	1										2	4	6
2	2	2	2										5	6	11
Totals				21	21	20	21	16	16	20	21	14	19	19	208

2.7.2. Data Set 2. Data Generated to Follow the Model for the Cattle Breeding Data.

As a second example, ten variates were generated from the model assumed for the cattle breeding data, described in 2.7.1. The generated data had the same number of observations in the cells of the design as the actual data. (Table (2.3)).

The parameter values used to generate the data are given in the following table:

Table (2.4) Parameter Values for Generated Data, Following the Model for the Cattle Breeding Data.

EFFECT	VALUE	DISTRIBUTION
MEAN	10.0	FIXED
SEX	-2.0	FIXED
SITE	3.0	FIXED
SBRD	5.0	FIXED
SITE X SBRD	1.0	FIXED
DBRD	5.0	FIXED
SITE X DBRD	1.0	FIXED
SBRD X DBRD	4.0	FIXED
SIRE	11 RANDOM LEVELS	N(0, 10)
SIRE X DBRD	22 RANDOM LEVELS	N(0, 3)
RESIDUAL	208 RANDOM LEVELS	N(0, 8)

The sample variances for the random values generated for each variable are given in the following table:

Table (2.5): Sample Variances of Generated Random Effect Values for Data Set 2.

VARIABLE	SAMPLE VARIANCE FOR COMPONENT:		
	SIRE	SIRE X DBRD	RESIDUAL
1	12.778	3.294	7.862
2	7.976	2.759	9.663
3	13.984	1.776	8.959
4	11.786	0.750	8.911
5	7.611	1.994	8.326
6	7.884	1.772	8.504
7	8.665	3.781	7.194
8	16.188	2.410	7.813
9	5.626	2.797	8.211
10	12.795	4.194	7.566

2.7.3. Data Set 3. A Generated Example.

A small and extremely unbalanced data set of 29 observations was generated for ten variables. A three factor mixed effect classification model was used. The model and parameter values are given in Table (2.6). The distribution of observations over the cells of the design is given in Table (2.7) and the sample variances of the random values generated for each variable are given in Table (2.8).

Table (2.6): Model and Parameter Values for Generation of Data Set 3.

EFFECT	DEGREES OF FREEDOM OR NUMBER OF VALUES	DISTRIBUTION
MEAN	1	FIXED WITH VALUE 10.0
F	2	FIXED WITH VALUES 5.0 and -7.0
R1	3	RANDOM $\sim$ N(0,30)
F X R1	9	RANDOM $\sim$ N(0,10)
R2 (NESTED IN F)	6	RANDOM $\sim$ N(0,3)
R1 X R2	14	RANDOM $\sim$ N(0,6)
RESIDUAL	29	RANDOM $\sim$ N(0,12)

Table (2.7): Number of Observations in Design Cells for Data Set 3.

		FACTOR F					
		1		2		3	
FACTOR R2		1	2	1	2	1	2
FACTOR R1	1	3	2	2	0	2	3
	2	1	0	1	3	1	3
	3	1	3	0	2	2	0

Table (2.8): Sample Variances of Generated Random Effect Values for Data Set 3.

VARIABLE	SAMPLE VARIANCE FOR COMPONENT:				
	R1	R2	F X R1	R1 X R2	RESIDUAL
1	14.457	8.328	2.486	7.493	12.762
2	8.477	8.002	6.871	4.249	11.055
3	44.477	15.396	3.054	5.103	15.568
4	8.114	3.546	3.349	3.796	13.095
5	0.551	12.443	1.454	2.118	14.655
6	87.042	16.146	1.208	9.416	9.309
7	14.120	14.272	1.951	5.445	8.911
8	20.921	6.564	1.980	8.268	14.587
9	28.452	11.548	1.880	5.499	8.958
10	19.235	4.562	8.110	5.353	14.158

CHAPTER 3

Estimation Using Quadratic Forms Based on the Structure of the Model

Many methods of variance component estimation are based on the observation that the expectation of the quadratic forms in the response variables are linear functions of the variance components. The fact that they are also quadratic functions of the location parameters, is one major source of difficulty. Methods based on this observation involve computing quadratic forms and equating them to their expectations to produce a set of equations in the unknown parameters.

In classification models, the factorial data structure provides intuitively and computationally appealing quadratic forms as the sums of squares (SSs) one would compute for a fixed effect analysis of variance (ANOVA).

A second set of possible quadratic forms is the set of reductions in residual SS, due to fitting groups of parameters in different orders. These are the SSs for the method of fitting constants (Searle, (1971) p 246).

These methods developed for classification models are extendable to variance component models having a structured random component. The computations are, however, more difficult.

The first attempts at estimation were for balanced models. In this case it is well known that the

ANOVA and fitting constants SSs are identical. Furthermore, in this case many desirable properties result when unbiasedness is the only criterion of estimation.

This relative success led researchers, notably Henderson (1953), to try to use the undemanding unbiasedness criterion with one or other of the likely sets of quadratic forms for unbalanced models. The most outstanding features of the resulting estimators are their relative computational simplicity, their widespread use and their lack of known desirable properties.

3.1. Balanced Variance Component Models.

The balanced variance component model (2.58) has the form:

$$\underline{y} = X\underline{\beta} + \underline{e} \qquad (3.1)$$

with

$$E(\underline{e}) = 0, \ Cov(\underline{y}) = V(\underline{\omega}) = \Sigma\omega_i Q_i \qquad (3.2)$$

where: Q_i, i=1,2...q are known n x n symmetric and idempotent matrices such that $Q_iQ_j = 0$, i # j.

and $R(X)$ is invariant to Q_i, i=1,2...q.

Interest generally centers on some independent set of linear combinations of the parameters, $\underline{\theta} = B_1\underline{\omega}$ and $\underline{\omega}$ is constrained so that $\underline{\theta} > 0$, $V(\underline{\omega})$ is p.d. and $B_2\underline{\omega} = 0$ for some linear combinations independent of those in B_1. (3.3)

3.1.1. Quadratic Forms Based on the Structure of the Model.

The model (3.1)-(3.2) suggests a set of quadratic forms:

$$\underline{y}'Q_i\underline{y}, \quad i=1,2,..q \qquad (3.4)$$

These are also SSs since the symmetry and idempotency of the Q_i implies that they are n.n.d. The expectations of these SSs does however depend on the fixed effect parameters. One intuitive approach to avoiding this complication is to consider quadratic forms in the projection of the response onto the orthogonal complement of R(X):

$$\underline{y}'NQ_iN\underline{y} = \underline{y}'Q_i\underline{y} - \underline{y}'PQ_iP\underline{y} \qquad (3.5)$$

where: $N = I - P$, $P = X(X'X)^{-}X'$ (3.6)

The expectation of any quadratic form is given by:

$$E(\underline{y}'A\underline{y}) = tr(ACov(\underline{y})) + E(\underline{y})'A\,E(\underline{y})$$

(Graybill 1976 p 139).

For models with the features of (2.1) and (2.2) this becomes:

$$E(\underline{y}'A\underline{y}) = \Sigma\theta_i tr(AV_i) + \underline{\beta}'X'AX\underline{\beta} \qquad (3.7)$$

or for balanced models:

$$E(\underline{y}'A\underline{y}) = \Sigma\omega_i tr(AQ_i) + \underline{\beta}'X'AX\underline{\beta} \qquad (3.8)$$

Using (3.8) and (3.5) together with the properties of Q_i, and the fact that $X'N = 0$, we obtain:

$$E(\underline{y}'NQ_iN\underline{y}) = \omega_i tr(NQ_i)$$
$$= \omega_i rank(NQ_i) \qquad (3.9)$$

since NQ_i is idempotent and $NQ_iNQ_j = 0$ for i#j.

Hence an unbiased estimate of $\underline{\omega}$ is $\underline{\hat{\omega}}$ which has components:

$$\hat{\omega}_i = (\underline{y}'NQ_iN\underline{y})/\text{rank}(NQ_i) \qquad (3.10)$$

and an unbiased estimate of $\underline{\theta}$ is:

$$\hat{\underline{\theta}} = B_1\hat{\underline{\omega}} \qquad (3.11)$$

3.1.2. Distributional Properties of the Quadratic Forms.

If we include the Normal distribution assumptions in the balanced variance component model, we can appeal to the following results to obtain some distributional properties of the quadratic forms (3.4) and (3.5).

For random variables $\underline{y}$ distributed $N(\underline{\mu},V)$, with V p.d., the quadratic forms $\underline{y}'A\underline{y}$ and $\underline{y}'B\underline{y}$ with A and B symmetric have the following properties:

a) $\underline{y}'A\underline{y} \sim \chi^2{}'(\text{rank}(A),\underline{\mu}'A\underline{\mu}/2)$ if and only if AV is idempotent. (Searle 1971 p 57). (3.12)

b) $\text{Var}(\underline{y}'A\underline{y}) = 2\text{tr}(AVAV) + 4\underline{\mu}'AVA\underline{\mu}$ (3.13)

(Searle 1971 p 57)

c) $\text{Cov}(\underline{y}'A\underline{y},\underline{y}'B\underline{y}) = 2\text{tr}(AVBV) + 4\underline{\mu}'AVB\underline{\mu}$ (3.14)

(By expanding the joint moment generating function)

d) $\underline{y}'A\underline{y}$ and $\underline{y}'B\underline{y}$ are independent if and only if AVB (or equivalently BVA) is zero (Searle 1971 p 59)

(3.15)

e) $\text{Cov}(\underline{b}'\underline{y},\underline{y}'A\underline{y}) = 2\underline{b}'VA\underline{\mu}$ (Searle 1971 p 56)

(3.16)

For the quadratic forms (3.4) and (3.5) we obtain:

$$(1/\omega_i)\underline{y}'Q_i\underline{y} \sim \chi^2{}'(\text{rank}(Q_i),(1/2)\underline{\beta}'X'Q_iX\underline{\beta}) \qquad (3.17)$$

and, $$(1/\omega_i)\underline{y}'NQ_iN\underline{y} \sim \chi^2(\text{rank}(NQ_i)) \qquad (3.18)$$

This follows by considering the transformed response: $\underline{z} = (1/\sqrt{\omega_i})\underline{y}$ which has $\text{Cov}(\underline{z}) = (1/\omega_i)V(\underline{\omega})$ and

$$\underline{z}'Q_i\underline{z} = (1/\omega_i)\underline{y}'Q_i\underline{y} \qquad (3.19)$$

and $$\underline{z}'NQ_iN\underline{z} = (1/\omega_i)\underline{y}'NQ_iN\underline{y} \qquad (3.20)$$

Also $Q_i\text{Cov}(\underline{z}) = Q_i$ and $NQ_iN\text{Cov}(\underline{z}) = NQ_i$ and both are idempotent so (3.17) and (3.18) follow from (3.12).

Furthermore, for $i \neq j$, $Q_iQ_j = 0$ so that $Q_iV(\underline{\omega})Q_j$ and $NQ_iNV(\underline{\omega})NQ_jN$ are zero from (2.64). Hence if we ignore the constraints on $\Omega_{\underline{\omega}}$ (3.3), we have from (3.15) that the sets (3.4) and (3.5) contain independent Chi-square random variables.

From (3.13) or properties of the central Chi-square distribution, the sampling variances of $\hat{\omega}_i$ (3.10) are: $2\omega_i^2\text{rank}(NQ_i)$

It is thus possible to obtain the sampling variances of the estimates $\hat{\underline{\theta}}$ (3.11) of the parameters of interest. However, the components of $\hat{\underline{\theta}}$ are not independent. Graybill (1976) gives examples of approximate confidence intervals and tests for the components.

3.1.3 Uniform Minimum Variance Unbiased Estimators.

Seely (1971) shows that the statistics $\underline{y}'Q_i\underline{y}$, $i=1,2\ldots q$, and $\hat{\underline{\beta}} = (X'X)^{-}X'\underline{y}$ are jointly complete and sufficient for the normal balanced variance component model. This follows by considering the density of the response in

the form (2.74) and writing,

$$\sum_i (1/\omega_i)\underline{y}'NQ_iN\underline{y} = \sum_i (1/\omega_i)\underline{y}'Q_i\underline{y} - \sum_i (1/\omega_i)\underline{y}'PQ_iP\underline{y}$$

and (3.21)

$$\|X(\underline{\beta} - \hat{\underline{\beta}})\|^2_{V(\underline{\omega})} = \underline{\beta}'X'V(\underline{\omega})^{-1}X\underline{\beta} - 2\underline{\beta}'X'V(\underline{\omega})^{-1}X\hat{\underline{\beta}} + \hat{\underline{\beta}}'X'V(\underline{\omega})^{-1}X\hat{\underline{\beta}} \quad (3.22)$$

but $\hat{\underline{\beta}}'X'V(\underline{\omega})^{-1}X\hat{\underline{\beta}} = \sum_i (1/\omega_i)\underline{y}'PQ_iP\underline{y}$

Hence the density can be written as:

$$h(\underline{\beta},\underline{\omega})\exp\{\sum_{i=1}^{q}(1/2\omega_i)\underline{y}'Q_i\underline{y} + \underline{\beta}'X'V(\underline{\omega})^{-1}X\underline{\beta}\} \quad (3.23)$$

where:

$$h(\underline{\beta},\underline{\omega}) = (2\pi)^{-n/2}\det(V(\underline{\omega}))^{-1/2}\exp(-(1/2)\underline{\beta}'X'V(\underline{\omega})^{-1}X\underline{\beta}) \quad (3.24)$$

If the model is re-parametrized via the transformations:

$$\omega_i^* = (1/2\omega_i) \text{ and } \underline{\beta}^* = X'V(\underline{\omega})^{-1}X\underline{\beta} \quad (3.25)$$

then the transformed parameter space contains an open subset provided $\Omega_{\underline{\omega}} x \Omega_{\underline{\beta}}$ has a non void open set.
In this case $\hat{\underline{\beta}}$ and $\underline{y}'Q_i\underline{y}$ are jointly complete and sufficient according to a result in Lehmann (1959) p 132.

The estimators $\hat{\underline{\omega}}$ and $\hat{\underline{\theta}}$ (3.10) and (3.11) are uniformly minimun variance estimators amongst all unbiased estimators (UMVUE), since they are functions of the complete and sufficient statistics.

3.1.4. Negative Estimates of Variance Components.

The distributions of the estimators $\hat{\omega}_i$ (3.10) are derived without regard for the constraints (3.3)

on the parameter space. There is consequently a positive probability that the $\hat{\omega}_i$ will violate the constraints and lead to estimates $\underline{\hat{\theta}} = B_1\underline{\omega}$ which have negative components.

The principle of unbiasedness and compliance with the constraints appear to be incompatible. One solution to the problem is to truncate the estimators at zero, but then all properties of the sampling distribution of the estimators are lost. There is also the question of whether to change the model and pool quadratic forms when constraint violations occur.

Klotz, Milton and Zacks (1969) compare the unbiased estimators to non-negative estimators in the balanced two-component model. Using squared error loss, they achieve considerable improvement with truncation-type estimators.

3.2. ANOVA Quadratic Forms for Classification Models.

In mixed and random effect classification models (2.24), there is one set of quadratic forms which has considerable appeal for use in variance component estimation. These are the factorial effect SSs computed in the fixed effect analysis of variance. They are easy to compute and in the balanced case, have well-known expectations and distributions.

The expectations are linear functions of the variance components, plus quadratic functions of the fixed effect parameters. It is the appearance of the fixed effect

parameters in the expectation of the SSs which causes difficulty and limits the estimation of variance components by this method. The method is known as the ANOVA method.

We consider some general aspects of the method and then consider special cases and procedures for eliminating the influence of the fixed effects on the variance component estimators. With balanced classification models, the invariance of the fixed effect design matrices to the coefficient matrices of the covariance facilitates the removal of the fixed effect parameters leading to the estimators of 3.1. Henderson (1953) extended the technique to some cases of unbalanced models, and Searle (1968) and (1971) reviewed and extended the procedures.

We consider the classification model (2.24) with basic structures:

$$\underline{y} = X_1\beta_1 + X_2\underline{\beta}_2 + \ldots + X_r\underline{\beta}_r + U_1\underline{v}_1 + \ldots U_{q-1}\underline{v}_{q-1} + \underline{e} \qquad (3.26)$$

or

$$\underline{y} = X\underline{\beta} + U\underline{v} + \underline{e} \qquad (3.27)$$

where:

X_i, $i = 1,2\ldots r$ and U_j, $j = 1,2\ldots q-1$ are the effect design matrices.

$\underline{\beta}_i$ $i=2,3\ldots r$ are the fixed effect parameters

$\underline{v}_j$ $j=1,2\ldots q-1$ are the random effect components.

$\underline{e}$ is the residual error effect.

β_1 is the mean effect and $X_1 = \underline{1}$.

3.2.1. The ANOVA Quadratic Forms.

In classification models the response vector can be partitioned into cells according to subsets of the factors corresponding to each effect in the model. The SSs of the cell means of these sub-classifications are the SSs of the orthogonal projection of the response vector onto the column space of each effect design matrix. The ANOVA SSs are linear contrasts amongst these sub-classification SSs. The contrasts are linear combinations with coefficients summing to zero. In balanced models the ANOVA SSs correspond to repeated projections onto subspaces of lower dimension.

In the model (3.26) we can define the set of projection operators:

$$\{P_k: \text{ For } k=1,2\ldots r \quad P_k = X_k(X_k'X_k)^{-1}X_k', \text{ and for } j=1,2\ldots q \text{ let } k=j+r \text{ and } P_k = U_j(U_j'U_j)^{-1}U_j' \text{ where } U_q = I_n\} \tag{3.28}$$

The sub-classification or projection SS associated with the k th effect is then: $t_k = \underline{y}'P_k\underline{y}$ (3.29)

and we define the vector of these SSs as: $\underline{t}$. (3.30)

Several properties of P_k and t_k, k=1,2..r+q follow from the properties of design matrices, 2.4.1:

a) Since $X_j'X_j = \text{diag}(n_{i,j},\ i=1,2\ldots p_j)$, (2.32), we see that the projection SSs are simply SSs of the sub-classification totals weighted

according to the number of observations in each cell.

b) $P_{r+q} = I_n$ so that $t_{r+q} = \underline{y}'\underline{y}$ (3.31) (3.32)

c) $tr(P_k) = rank(P_k) = p_i$ or q_i, the column size of the effect design matrix. (3.33)

d) $\underline{1}'P_k = \underline{1}'$ and $P_k\underline{1} = \underline{1}$. (3.34)

since, for k=1,2...r, $\underline{1}_n'P_k = \underline{1}_{pk}'X_k'P_k = \underline{1}_{pk}'X_k' = \underline{1}_n'$ by (2.30). Similarly for k=r+1,r+2...r+q, using U_j and (2.31).

e) $\underline{1}'P_k\underline{1} = n$ for k=1,2...r+q. (3.35)

The ANOVA SSs for the random effects are given by q linearly independent contrasts amongst the r+q projection SSs. We define $\underline{r}$ to be the q-vector of ANOVA SSs corresponding to the random effects. Then: $\underline{r} = R\underline{t}$ (3.36)

The matrix R of contrasts has some important properties which can be exploited to estimate the variance components.

a) R is q x r+q with rank q. (3.37)

The independence of the rows of R follows from the relationship between the effects of any factorial classification. They either involve disjoint subsets of the factors and hence their ANOVA SSs involve different components of $\underline{t}$, or there is a distinct hierarchy amongst the effects, higher order effects involving more factors than lower order effects. Thus the ANOVA SSs of the higher order effects involve more components of $\underline{t}$.

b) $R\underline{1} = \underline{0}$ since the rows of R are contrasts. (3.38)

c) The last column of R is zero, except for the (q, r+q) th element, which is 1. (3.39)

This follows because the only ANOVA SS involving $\underline{y}'\underline{y}$, the last component of $\underline{t}$, is the residual or error SS. This SS is the last component of $\underline{r}$.

3.2.2. Expected Values and Estimation Equations.

The Expectation of $\underline{r}$, $E(\underline{r})$ is:

$$E(\underline{r}) = RE(\underline{t}) \tag{3.40}$$

From (3.7),

$$E(\underline{t}) = S\underline{\theta} + \underline{a}(\underline{\beta}) \tag{3.41}$$

where: S is q+rxq with k,j th element given by $tr(P_k U_j U_j') = tr(U_j' P_k U_j)$ (3.42)

and

$\underline{a}(\underline{\beta})$ is a q+r-vector of quadratic functions of the fixed effect parameters.

The k th element is:

$$\sum_{i=1}^{r} \underline{\beta}'_i X_i' P_k X_i \underline{\beta}_i = \underline{\beta}' X' P_k X \underline{\beta} \tag{3.43}$$

Hence the expectation of $\underline{r}$ is:

$$E(\underline{r}) = H_A\underline{\theta} + R\underline{a}(\underline{\beta}) \tag{3.44}$$

where:

$H_A = RS$ is qxq.

The principle of using $E(\underline{r})$ to get unbiased estimates of $\underline{\theta}$ is to find model conditions or transformations, which make the term $R\underline{a}(\underline{\beta})$ vanish, leaving estimation equations for the restricted or transformed model:

$$\underline{r} = H_A\hat{\theta} \qquad (3.45)$$

One feature of (3.44) common to all classification models is that the last equation is automatically free of the fixed effects and all but one of the variance components. This equation therefore, always provides an unbiased estimate of the residual effect component. This is discussed in section 3.2.4.

The rank of the matrix H_A is important for consistency of equations (3.45). This is not discussed in the literature and seems to cause no problem in practice. We assume H_A to have full rank. According to Hartley, Rao and LaMotte (1978), this will follow provided the matrices U_jU_j' are independent and there is no confounding of fixed and random effects. This is related to identifiability (2.3).

Conditions which lead to the elimination of the fixed effect parameters in (3.44), $R\underline{a}(\underline{\beta}) = \underline{0}$, are invariance of the column spaces of the fixed effect design matrices to $V_j = U_jU_j'$ and hence to P_k for $k=r+1,\ldots q$. This leads to identities amongst the elements of $\underline{a}(\underline{\beta})$ and thence to zero elements of $R\underline{a}(\underline{\beta})$ because of the contrast property (3.38) of the rows of R. Because of this, restricted models which possess this invariance or transformations which produce it,

are used to obtain estimating equations like (3.45). Examples are discussed in sections 3.3.6, 3.3.7 and 3.3.8.

Two aspects of equations like (3.45) which make them desirable estimating equations are computational simplicity and efficiency, section 3.2.3, and some distributional properties for the estimators based on them, section 3.2.5. The transformations considered to make the model amenable, generally affect both these properties and one objective is to minimise this effect.

3.2.3. Computational Procedure - Synthesis.

The major advantage of using the ANOVA SSs and equation (3.45) to estimate $\underline{\theta}$ is computational ease. The SSs are easily computed from sub-classification tables.

Hartley (1967), observed that the same computational device was available for calculating elements of S and hence of H_A for any particular model. From (3.42) we can write the k,j th element of S as:

$$tr(U_j' P_k U_j) = \sum_{i=1}^{q_j} \underline{u}_{ji}' P_k \underline{u}_{ji} \qquad (3.46)$$

where $\underline{u}_{ji}$ is the i th column of U_j.
The components of the sum have exactly the same form as components of $\underline{t}$, $t_k = \underline{Y}' P_k \underline{Y}$, and hence can be calculated using the same sub-classifications on $\underline{u}_{ji}$ for S as on $\underline{Y}$ for $\underline{t}$. That is, treat $\underline{u}_{ji}$ as a data vector and calculate the model ANOVA. This method is known as the Method of Synthesis,

Hartley (1967), Rao, J.N.K. (1968).

Computations for the last row and column of S are made trivial when $U_q = I_n$, which is often the case.

(3.47)

3.2.4. The Residual Effect Variance Component.

For all classification models, the last equation of (3.44) is free of all parameters in $\underline{\theta}$ and $\underline{\beta}$ except for θ_q. This is because r_q is the residual SS, after fitting all the fixed effect parameters β, and all the random effects except the last, that is $\underline{v}_1, \underline{v}_2 \cdots \underline{v}_{q-1}$.

Using the model equation in the form (3.27)

$$r_q = \underline{y}' N \underline{y} \tag{3.48}$$

where

$$N = (I - (X,U)((X,U)'(X,U))^{-}(X,U)') \tag{3.49}$$

N is idempotent, symmetric and has the property that $NX_i = X_i'N = NU_j = U_j'N = 0$, for $i=1,2...r$, $j=1,2...q-1$. Hence with (3.7), the last equation of (3.44) is:

$$E(r_q) = \sum_{j=1}^{q} \theta_j \operatorname{tr}(NU_jU_j') + \sum_{i=1}^{r} \underline{\beta}_i' X_i' N X_i \underline{\beta}_i$$

$$= \theta_q \operatorname{tr}(N) = \theta_q \operatorname{rank}(N) \tag{3.50}$$

rank (N) = n-s, where s is the number of independent columns in (X,U). This is usually referred to as the degrees of freedom for the model.

Hence an unbiased estimate of θ_q is

provided by: $\hat{\theta}_q = r_q/(n-s)$ (3.51)

If the normality assumptions are included in the model (2.20), the properties of N, together with (3.12), mean that

$$\hat{\theta}_q \sim \chi^2_{n-s} (\theta_q/(n-s)) \quad (3.52)$$

because $r_q = \theta_q \underline{y}' (N/\theta_q)\underline{y}$

and $(N/\theta_q)V(\underline{\theta}) = N$ which is idempotent.

It follows from (3.13), that the variance of $\hat{\theta}_q$ is:

$$\text{var}(\hat{\theta}_q) = 2\theta_q^2/(n-s) \quad (3.53)$$

Further, because the matrices for the quadratic forms $r_1 \ldots r_{q-1}$ are linear combinations of $P_1 \ldots P_{r+q-1}$ (3.30), (3.31) and (3.36).

and

$$NV(\underline{\theta})P_k = NP_k = 0 \quad k=1,2\ldots r+q-1$$

we have with (3.15) that $\hat{\theta}_q$ is distributed independently of $\hat{\theta}_j$, $j=1,2\ldots q-1$ whenever estimation equations (3.45) are possible. (3.54)

3.2.5. Distributional Properties of ANOVA Estimators.

Assuming that the model is such that the fixed effect parameters vanish from the expectation equations (3.44) we can make some observations about the distribution of the estimators obtained from (3.45) when we include the Normal Distribution assumptions in the model.

$$\underline{\hat{\theta}} = H_A^{-1}\underline{r} = H_A^{-1}R\underline{t} \quad (3.55)$$

Hence:

$$\mathrm{Cov}(\hat{\underline{\theta}}) = H_A^{-1} R \, \mathrm{Cov}(\underline{t}) \, R' H_A^{-1} \qquad (3.56)$$

Elements in $\mathrm{Cov}(\underline{t})$ are obtained from (3.13) and (3.14). Specifically the k,m th element is given by:

$$2 \, \mathrm{tr}(P_k V(\underline{\theta}) P_m V(\underline{\theta})) + 4 \, \underline{\beta}' X' P_k V(\underline{\theta}) P_m X \underline{\beta} \qquad (3.57)$$

Hence we can write $\mathrm{Cov}(\underline{t})$ as:

$$\mathrm{Cov}(\underline{t}) = 2D(\underline{\theta}) + 4A(\underline{\beta}, \underline{\theta}) \qquad (3.58)$$

where the k,m th elements of D and A are in the first and second terms of (3.57) respectively. Thus the elements of $\mathrm{Cov}(\hat{\underline{\theta}})$ depend on the unknown parameters and are difficult to compute; the Method of Synthesis (section 3.2.3) is only available for some components in $D(\underline{\theta})$. The terms in $\underline{\beta}$, however, may vanish from $\mathrm{Cov}(\hat{\underline{\theta}})$ under the same conditions that make them vanish from (3.44).

We have already noted, (3.53 and 3.54), that the last row and column of $\mathrm{Cov}(\hat{\underline{\theta}})$ are zero, except for the last element, the variance of $\hat{\theta}_q$. This, together with the fact that the last column of R is zero, except for the last element (3.39), means that not all of $D(\underline{\theta})$ need be computed to obtain $\mathrm{Cov}(\hat{\underline{\theta}})$.

In addition, apart from $\hat{\theta}_q$, the distributional form of $\hat{\underline{\theta}}$ is unknown, because the matrices of the quadratic forms producing $\hat{\underline{\theta}}$ are not in general idempotent, as they were in the case of the balanced models (3.17).

3.2.6. ANOVA Quadratic Forms for Random Classification Models: Henderson's Method I

Random Effect Classification Models have model equation:

$$\underline{y} = \underline{1}\mu + U_1\underline{v}_1 + \ldots + U_{q-1}\underline{v}_{q-1} + \underline{e} \qquad (3.59)$$

The fixed effect design matrix, $\underline{1}$ is particularly simple and its column space is invariant to $V_j = U_jU_j'$. Components of $\underline{a}(\underline{g})$ in $E(\underline{t})$ of (3.40) are all $n\mu^2$ from (3.34) and hence vanish under the contrasts R to yield estimating equations (3.45) free of the fixed effects.

Unbiased estimates of $\underline{\theta}$ are therefore:

$$\hat{\underline{\theta}} = H_A^{-1}\underline{r} \qquad (3.60)$$

Again, the last equation only involves $\hat{\theta}_q$ and the solution is the same as discussed in section 3.2.4.

The distribution of elements in $\hat{\underline{\theta}}$ are unknown, even with the Normal Distribution assumptions, except for $\hat{\theta}_q$, which is independent of other components (3.54), and $\hat{\theta}_q \sim \chi^2_{n-s}(\theta_q/(n-s))$ (3.52).

For the $\mathrm{Cov}(\hat{\underline{\theta}})$ we need $\mathrm{Cov}(\underline{t})$, which, using (3.57) has k,m th element:

$$2\sum_{i=1}^{q}\sum_{j=1}^{q}\theta_i\theta_j\,\mathrm{tr}(U_j'P_kU_iU_i'P_mU_j) + 4\mu^2\sum_{j=1}^{q}\theta_j\underline{1}'P_kU_jU_j'P_m\underline{1} \qquad (3.61)$$

The second term is constant for all k and m because $\underline{1}'P_k = \underline{1}'$ for k=1,2...q-1 (3.34). Hence the terms involving μ vanish under the contrasts R and using (3.56) and (3.58):

$$\mathrm{Cov}(\hat{\underline{\theta}}) = 2H_A^{-1} R\, D(\underline{\theta}) R' H_A^{-1} \qquad (3.62)$$

Where the k,m th element of $D(\underline{\theta})$ is:

$$\sum_i \sum_j \theta_i \theta_j \mathrm{tr}(U_j' P_k U_i U_i' P_m U_j) \qquad (3.63)$$

The Method of Synthesis, section 3.2.3, can be used for some terms of $D(\underline{\theta})$, Hartley (1967). The independence of $\hat{\theta}_q$ to other elements and the special form of the last column of R, (3.39), can be used to reduce the size of the matrices in (3.63) (Searle (1971) p 433). However the computations are still formidable.

3.2.7. Adjusting for bias in Mixed Effect Classification Models.

The quadratic functions of fixed effect parameters in the expectation of the ANOVA quadratic forms (3.44), makes it difficult to obtain estimators of the variance components which are notbiased by the fixed effects unless they can be eliminated from the equations.

One approach to achieve this, is to transform the data, so that the model does not depend on the fixed effect parameters, or at least on those for which the column space of the effect design matrix is not invariant to the covariance matrix. The method is to provide some estimate of the fixed effect parameters and then produce residuals, based on these estimates. Two approaches are available; estimate the fixed parameters ignoring the random effects, or including the

random effects as if they were fixed, but ignoring their resulting estimates at the adjustment stage.

Searle (1968) gives a general method based on ignoring the random effects and estimating the fixed effect parameters apart from the mean effect.

With the model written as:

$$\underline{y} = \underline{1}\mu + X\underline{\beta} + \sum_{i=1}^{q} U_i \underline{v}_i \qquad (3.64)$$

Estimate $\underline{\beta}$ by $X^{-}\underline{y}$ where X^{-} is a generalized inverse of X then

$$\underline{z} = \underline{y} - XX^{-}\underline{y} = \underline{x}\mu + \Sigma U_i^{*}\underline{v}_i \qquad (3.65)$$

where $\underline{x} = \underline{1} - XX^{-}\underline{1}$, $U_i^{*} = (I - XX^{-})U_i$

and $\underline{\beta}$ vanishes because $X - XX^{-}X = 0$.

Now $\mathrm{Cov}(\underline{z}) = \Sigma\theta_i U_i^{*}U_i^{*}'$ (3.66)

Henderson's method I can be used on the transformed response $\underline{z}$. The SSs are formed using the design model for $\underline{y}$, that is U_i not U_i^{*}, but the expectations must be taken with the transformed model (3.65). Hence,

$$E(\underline{t}) = S_B\underline{\theta} + \underline{b}(\mu) \qquad (3.67)$$

where: $\underline{t}$ is the vector of sub-classification SSs, (3.30).

S_B has i,j th element

$$\mathrm{tr}(P_i U_j^{*} U_j^{*}') \quad i=1,2\ldots q+1 \qquad (3.68)$$

and $P_1 = \underline{1}(\underline{1}'\underline{1})^{-1}\underline{1}'$, $P_i = U_i(U_i'U_i)^{-1}U_i'$

$i=2\ldots q+1$.

also the i th element of $\underline{b}(\mu)$ is:

$$\mu^2 \underline{x}'P_i\underline{x} \qquad i=1,2\ldots q+1 \qquad (3.69)$$

Therefore:

$$E(\underline{r}) = H_B\underline{\theta} + R\underline{b}(\mu) \qquad (3.70)$$

with $H_B = RS_B$. This is a set of linear equations in $\underline{\theta}$ and μ^2, which can be solved for both $\underline{\theta}$ and μ^2. Searle (1971) observes two drawbacks of the method as being non-uniqueness of X^-, different inverses producing different estimates, and computational difficulties because even with the Method of Synthesis , 3.2.3, for elements of S_B, (3.68), columns of U_j^* are required. These are obtained from U_j by the same transformation as $\underline{z}$ from $\underline{y}$. Also U_q^* is no longer I, so the simplifications (3.47) for the last row and column of S are no longer available.

While the non-uniqueness is unfortunate, there is at least one obvious and, in some senses, optimal candidate for X^-, that is $(X'X)^-X'$ which provides the least squares estimator of $\underline{\beta}$. This is invariant to choice of $(X'X)^-$ (Graybill (1976) p 32). Also, the mean effect might as well be adjusted for as well as other fixed effects. In this case, the method is to use the ANOVA SSs to estimate the variance components of the least squares residuals of the fixed effects.

Hartley, Rao and Lamotte (1978), suggest a method based on these residuals, but using different quadratic forms;

the unweighted sub-classification SSs, that is SSs of sub-classification totals ignoring the number of observations per cell. They produce some optimal properties, such as consistency, but arrive at the same solution as C.R. Rao (1971a) (1971b) and (1972), who tackled the sometimes conflicting aspects of invariance to fixed effects and choice of quadratic forms from a different point of view, to be discussed later.

3.2.8. Mixed Models without Interaction between Fixed and Random Effects:- Henderson's Method II

Henderson (1953) proposed adjusting the data for the fixed effect parameters to produce a model independent of them, but without ignoring the random effects or producing the computational difficulties associated with the previous method. Henderson, Searle and Schaeffer (1974) show that the resulting method overcomes the non-uniqueness problem. However, it is only applicable to mixed effect classification models, having no interaction or nesting between fixed and random effects, except for the residual effect. Searle (1971) considers this a serious limitation.

The method is to produce an estimate of $\underline{\beta}$ in (3.64), $L\underline{y}$ such that

a) $XLU_i = 0$ for $i=1,2...q-1$ (3.71)

b) $XL\underline{1}$ is a constant vector (3.72)

and c) $X - XLX$ has identical rows. (3.73)

The model for $\underline{z} = \underline{y} - XL\underline{y}$ then becomes:

$$\underline{z} = \mu^{*}\underline{1} + \sum_{i=1}^{q-1} U_i \underline{v}_i + U_q^{*}\underline{v}_q \qquad (3.74)$$

where:

$$\mu^{*}\underline{1} = \mu\underline{1} - XL\underline{1} + (X-XLX)\underline{\beta}$$

and $$U_q^{*} = U_q - XLU_q = I-XL$$

Now applying Henderson's method 1 to $\underline{z}$, largely overcomes the computational difficulties of the previous method, since only one new U matrix need be evaluated.

The restriction to models having no interaction or nesting between fixed and random effects, except the residual effect, is equivalent to requiring:

$$\text{rank}(X_1 U_1 \ldots U_{q-1}) = \text{rank}(X) + \text{rank}(U_1 \ldots U_{q-1}) - 1 \qquad (3.75)$$

or $$R(X) \cap R(U_1 \ldots U_{q-1}) = R(\underline{1})$$

where X is the fixed effect design matrix excluding the mean effect. Since if this fails to hold, there are some columns of $U_1 \ldots U_{q-1}$, say U and a matrix M, such that $U = XM$. So multiplying $(X - XLX)$ on the right by M,

$$(X - XLX)M = U - XLU = U \qquad (3.76)$$

because $XLU = 0$ by (3.71).

Now $X - XLX$ is required to have identical rows by (3.73) and hence $(X - XLX)M = U$ must have constant columns, but this is not possible for columns of a design matrix, except for the mean effect.

3.2.9. Procedure for Henderson's Method II.

To produce a matrix L satisfying (3.71) a (3.73), we re-write the model (3.64) as:

$$\underline{y} = \underline{1}\mu + X_a\underline{\beta}_a + X_b\underline{\beta}_b + U_a\underline{v}_a + U_b\underline{v}_b + \underline{e} \quad (3.78)$$

where the columns of X and elements of $\underline{\beta}$, and columns of $(U_1, U_2 \ldots U_{q-1})$ and elements of $\underline{v}_1, \underline{v}_2 \ldots \underline{v}_{q-1}$ have been re-ordered and paritioned in such a way that:

a) U_a has full and maximum column rank (3.79)

and b) (X_b, U_a) has full and maximum column rank, (3.80)

Consequences of this are:

a) There exists K, such that $U_b = U_aK$. (3.81)

For otherwise U_b would have columns independent of those of U_a contradicting (3.79).

b) There exists vector $\underline{p}$, such that $U_a\underline{p} = \underline{1}$ (3.82)

For otherwise $\underline{1}$, which is in $R(U_1U_2 \ldots U_{q-1})$, would not be in $R(U_a)$, thereby contradicting (3.79).

c) There exists matrix M and vector $\underline{t}$, such that $X_a = \underline{1}\underline{t}' + X_bM$ (3.83)

This follows because $\underline{1} \in R(U_a)$ (3.82), hence by (3.80) $\underline{1} \notin R(X_b)$ so $(\underline{1}, X_b)$ has full column rank and the columns of $X_a \in R(\underline{1}, X_b)$ for otherwise $\text{rank}(\underline{1}, X_aX_bU_a) > \text{rank}(\underline{1}, X_b) + \text{rank}(U_a) = \text{rank}(X) + \text{rank}(U_1, U_2 \ldots U_{q-1})$ and we have restricted

the model to ensure that this does not occur.

Next writing Z for $(X_a X_b U_a U_b)$, the matrix:

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & G_{11} & G_{12} & 0 \\ 0 & G_{21} & G_{22} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where $$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} X_b'X_b & X_b'U_a \\ U_a'X_b & U_a'U_a \end{bmatrix}^{-1} \qquad (3.84)$$

is a generalized inverse of $Z'Z$ with partitioning conforming to the partitioning of Z. Hence if we treat the random effects as fixed, $GZ'\underline{y}$ is an unbiased estimate of the effect parameters. The part of the estimate for the fixed parameters, $\underline{\beta}_a$ and $\underline{\beta}_b$ is:

$$\begin{bmatrix} 0 \\ G_{11}X_b' + G_{12}U_a' \end{bmatrix} \underline{y} = \begin{bmatrix} 0 \\ L_1 \end{bmatrix} \underline{y} = L\underline{y} \qquad (3.85)$$

Consequences of this choice of G to provide L are:

a) $L_1X_b = I$ from (3.84) (3.86)

b) $L_1U_a = 0$ from (3.84) (3.87)

c) $L_1U_b = L_1U_aK = 0$ from (3.81) & (3.87) (3.88)

d) $L_1\underline{1} = L_1U_a\underline{p} = 0$ from (3.82) & (3.87) (3.89)

e) $L_1X_a = L_1\underline{1}\underline{t}' + L_1X_bM$ from (3.83)

$= M$ from (3.89) and (3.86) (3.90)

L of (3.85) satisfies conditions (3.71) and (3.73) since:

a) $XL(U_aU_b) = X\begin{bmatrix} 0 & 0 \\ L_1U_a & L_1U_b \end{bmatrix} = 0$ (3.91)

from (3.87) and (3.88)

b) $(X_aX_b)L\underline{1} = 0$ from (3.89) (3.92)

c) $(X_aX_b) - (X_aX_b)L(X_aX_b) = (X_aX_b) - (X_bL_1X_a, X_bL_1X_b)$

$= (X_aX_b) - (X_bM, X_b)$ from (3.84) and (3.90)

$= (\underline{1}\underline{t}', 0)$ since $X_a - X_b M = \underline{1}\underline{t}'$ from (3.83) and $(\underline{1}\underline{t}', 0)$ has identical rows. (3.93)

3.2.10. Invariance of Hendersons Method II to Choice of Fixed Effect Estimators.

Searle (1968) and (1971) considered that this method would produce different estimators of the variance components for different choices of L. That this is not so was shown by Henderson, Searle and Schaeffer (1974). The observation rests on re-parameterizing the model (3.78) as:

$$\underline{y} = (X_b, U_a)\begin{pmatrix}\gamma_1 \\ \gamma_2\end{pmatrix} + \underline{e} \qquad (3.94)$$

where $\gamma_1 = M\underline{\beta}_a + \underline{\beta}_b$ using (3.83)

and $\gamma_2 = \mu\underline{p} + \underline{p}\underline{t}'\underline{\beta}_a + \underline{v}_a + K\underline{v}_b$ using (3.83), (3.82) and (3.81).

The linear functions γ_1 and γ_2 of the parameters are estimable because:

$$E(((X_b, U_a)'(X_b, U_a))^{-1}(X_b, U_a)'\underline{y}) = (\gamma_1', \gamma_2')' \qquad (3.95)$$

Hence the same linear functions of estimates of $\underline{\beta}_a$, $\underline{\beta}_b$, $\underline{v}_a$, $\underline{v}_b$ and μ are invariant to the choice of generalized inverse for Z'Z used to solve the normal equations. (Searle (1971) p 181).

Now if $\underline{z}$ is the vector of residuals $\underline{y} - XL\underline{y}$ produced by using L of (3.85), and $\underline{z}^*$ those produced by using L^* a second matrix satisfying (3.71) to (3.73), then:

$$\underline{z} - \underline{z}^* = X_a(\underline{b}_a - \underline{b}_a^*) + X_b(\underline{b}_b - \underline{b}_b^*) \qquad (3.96)$$

where: $\begin{pmatrix} \underline{b}_a \\ \underline{b}_b \end{pmatrix} = L\underline{y}$ and $\begin{pmatrix} \underline{b}^*_a \\ \underline{b}^*_b \end{pmatrix} = L^*\underline{y}$ (3.97)

Hence, $\underline{z} - \underline{z}^* = \underline{1}\underline{t}'(\underline{b}_a - \underline{b}^*_a) + X_b(M\underline{b}_a + \underline{b}_b - \underline{b}^*_b - M\underline{b}^*_a)$

$= \underline{1}\underline{t}'(\underline{b}_a - \underline{b}^*_a) = c\underline{1}$ (3.98)

where $c = \underline{t}'(\underline{b}_a - \underline{b}^*_a)$.

The second term of (3.98) vanishes because $M\underline{b}_a + \underline{b}_b$ is an estimable function of parameter estimates and hence invariant to the particular estimates used.

Now the quadratic forms used in applying Henderson's method I to $\underline{z}$ or $\underline{z}^*$ have matrices P_i of (3.29) which have the property that $P_i\underline{1} = \underline{1}$ and $\underline{1}'P_i = \underline{1}'$ (3.34). Hence,

$$\underline{z}^{*\prime}P_i\underline{z}^* = (\underline{z} - c\underline{1})'P_i(\underline{z} - c\underline{1})$$

$$= \underline{z}'P_i\underline{z} - 2c\underline{1}'\underline{z} + nc^2 \quad (3.99)$$

The ANOVA SSs used to estimate the variance components are contrasts of these SSs and hence terms with c and c^2 vanish.

3.2.11. Computing the Coefficients of the Residual Effect Variance Component in Henderson's Method II

The transformed model for $\underline{z} = \underline{y} - XL\underline{y}$ is given by (3.74) so that:

$$\mathrm{Cov}(\underline{z}) = \sum_{i=1}^{q-1} \theta_i U_i U_i' + \theta_q U^* U^{*\prime} \quad (3.100)$$

and

$$E(\underline{t}) = S_c\underline{\theta} + n\mu^{*2}\underline{1} \quad (3.101)$$

where $\underline{t}$ is the q+1 vector of sub-classification SSs for model (3.74) with components $t_i = \underline{z}'P_i\underline{z}$, where

$P_1 = \underline{1}(\underline{1}'\underline{1})^{-1}\underline{1}$, $P_2 = U_1(U_1'U_1)^{-1}U_1'$, ... $P_{q+1} = I_n$

(Note $P_{q+1} \neq U^*_q(U^{*\prime}_q U^*_q)^{-1}U^{*\prime}_q$) (3.102)

S_c has i,j th element: $tr(P_i U_j U_j')$ for i=1,2 ..q+1, and j=1,2...q-1 and $tr(P_i U_q^* U_q^{*\prime})$ for i=1,2...q+1. (3.103)

Also, $E(\underline{r}) = H_c \underline{\theta}$ (3.104)

where: $\underline{r} = R\underline{t}$, with R being the q x q+1 contrast matrix such that $\underline{r}$ is the vector of random effect ANOVA SSs. $H_c = RS_c$ is a q x q matrix. (3.105)

The computations for $\underline{t}$, $\underline{r}$ and all but the last column of S_c are identical to those for Henderson's method I. For the last column of S_c we have from (3.74) and (3.85):

$$U_q^* = (I - XL) = (I - X_b L_1) \quad (3.106)$$

Hence, $tr(P_i U_q^* U_q^{*\prime}) = tr(U_q^{*\prime} P_i U_q^*)$ for i=1,2...q+1

$$= tr(P_i) - 2tr(X_b L_1 P_i) + tr(P_i X_b L_1 L_1' X_b') \quad (3.107)$$

However, $L_1 P_i = 0$ from (3.87) and (3.89) (3.108)

and $L_1 L_1' = G_{11} X_b' L_1' + G_{12} U_q' L_1' = G_{11}$ (3.109)

from (3.85), (3.86) and (3.87).

Hence, $tr(P_i U_q^* U_q^{*\prime}) = rank(P_i) + tr(P_i W)$ (3.110)

where: $rank(P_i)$ is the number of columns in the associated design matrix (3.111)

and $W = X_b G_{11} X_b'$ (3.112)

W need only be computed once, and used in (3.110) for each element, i=1,2...q+1. The simplification from (3.107) to (3.110) is a result of the particular choice of L (3.85).

3.3 Reductions in Residual SS for Variance Component Models with Residual Error -Henderson's Method III.

A second group of quadratic forms motivated by the model structure and ease of computation, are those representing reductions in residual SS as different groups of the model constants, (fixed effect parameters or random effect values), are added after all other constants have been fitted. The idea stems from fitting constants methods for unbalanced fixed effect models (Searle (1971) p 246 and (1972)) and is therefore sometimes known as the method of fitting constants.

The method overcomes the difficulty of bias through fixed effects by the simple expedient of always fitting the fixed effect parameters first, so that they do not influence the reduction in residual SS. This is equivalent to using least squares residuals from the fixed effect part of the model to estimate the variance components. This is suggested by Searle's (1971) general method of adjusting for bias, section 3.2.7, and again by Hartley, Rao and LaMotte (1978), who use unweighted sub-classification SSs on the residuals.

The difficulty with this method of Fitting Constants is that there are generally more reduction SSs available through fitting the constants in different orders, than are needed to estimate the variance components. The order of fitting does not affect balanced data because the reduction SSs all become the ANOVA SSs, leading to the estimators of

section 3.1. However, for unbalanced data different orders of fitting produce different estimators.

3.3.1. Reduction SSs and their Expectations.

To examine a particular reduction SS we write a structured variance components model with residual error (2.23) in the form:

$$\underline{y} = Z_1\underline{b}_1 + Z_2\underline{b}_2 + \underline{e} \qquad (3.113)$$

or $\underline{y} = Z\underline{b} + \underline{e}$

where:

$Z_1 = (X,U)$

$\underline{b}_1' = (\underline{\beta}', \underline{v}_{11}' \cdots \underline{v}_{1s}') = (\underline{\beta}', \underline{v}')$

$U = (U_{11}, U_{12} \ldots U_{1s}) \quad 0 \leq s < q-1$

$Z_2 = (U_{21}, U_{22} \ldots U_{2t}) \quad t = q-1-s$

$\underline{b}_2' = (\underline{v}_{21}' \cdots \underline{v}_{2t}')$

The residual effect $\underline{e}$ is assumed to have identity design matrix, and $\mathrm{Cov}(\underline{e}) = \theta_q I_n$. Z_1 always contains the fixed effect design matrix and U is the collection of s, complete random effect design matrices for the effects to be fitted first. $(0 \leq s < q-1)$. Z_2 comprises the random effect design matrices for the effects to be fitted last (other than the residual effect). The resulting SS is the reduction in residual SS, due to fitting these effects after those in Z_1, or for fitting $\underline{b}_2$ after $\underline{b}_1$.

We note that after both stages of fitting, the

model contains all the effects of the variance component model. There is a further restriction on the sets of effects in Z_1 and Z_2. This is, that the effects in Z_1 may not be of higher order than those in Z_2. If this occurs, the reduction in residual SS is zero for that effect.(3.114). One classification effect is of higher order than another if the factors represented by the second are a subset of those represented by the first effect.

Model (3.113) with assumptions (2.26) has covariance;

$$\text{Cov}(\underline{y}) = Z\ \text{Cov}(\underline{b})\ Z' + \theta_q I_n \qquad (3.115)$$

$$= U\ D_1 U' + Z_2 D_2 Z_2' + \theta_q I_n \qquad (3.116)$$

where:

$$\text{Cov}(\underline{b}) = \text{diag}(\text{Cov}(\underline{\beta}) = 0,\ \text{Cov}(\underline{v}),\ \text{Cov}(\underline{b}_2))$$

$$D_1 = \text{Cov}(\underline{v}') = E(\underline{v}\ \underline{v}') = \text{Diag}(\theta_{1i} I_{q1i},\ i=1,2\ldots s)$$

and $$D_2 = \text{Cov}(\underline{b}_2) = E(\underline{b}_2 \underline{b}_2') = \text{Diag}(\theta_{2i} I_{q2i}',\ i=1,2\ldots t)$$

also $$E(\underline{b}) = (E(\underline{\beta}'),\ E(\underline{v}'),\ E(\underline{b}_2'))' = (\underline{\beta}', \underline{0}', \underline{0}')' \qquad (3.117)$$

When the model $\underline{y} = Z_1 \hat{\underline{b}}'$ is fitted to the response vector by least squares, the residual SS is smaller than the total SS by an amount $R(\underline{b}_1)$, the SS due to fitting the model,

where: $$R(\underline{b}_1) = \underline{y}' Z_1 (Z_1' Z_1)^{-} Z_1' \underline{y} \qquad (3.118)$$

For the full model $\underline{y} = Z_1 \hat{\underline{b}}_1 + Z_2 \hat{\underline{b}}_2$ the reduction is:

$$R(\underline{b}_1\underline{b}_2) = \underline{y}'(Z_1Z_2)'((Z_1Z_2)'(Z_1Z_2))^{-}(Z_1Z_2)'\underline{y}$$
$$= \underline{y}'Z(Z'Z)^{-}Z'\underline{y} \qquad (3.119)$$

and the reduction due to fitting $\underline{b}_2$ after $\underline{b}_1$ is defined to be:

$$R(\underline{b}_1/\underline{b}_2) = R(\underline{b}_1\underline{b}_2) - R(\underline{b}_1) \qquad (3.120)$$

Taking expectations over the full model (3.113), we have, with (3.7);

$$E(R(\underline{b}_1\underline{b}_2) = tr(Z(Z'Z)^{-}Z'Cov(\underline{y}))$$
$$+ E(\underline{b}')Z'Z(Z'Z)^{-}Z'ZE(\underline{b})$$
$$= tr(Z'Z\ Cov(\underline{b})) + \theta_q\ tr(Z(Z'Z)^{-}Z') + \underline{\beta}'X'X\underline{\beta} \qquad (3.121)$$

where we have used (3.115), the cyclic property of trace, the property of generalized inverses- $AA^{-}A = A$, and (3.117).

Similarly with expectation over the full model,

$$E(R(\underline{b}_1)) = tr(Z'Z_1(Z_1'Z_1)^{-}Z_1'Z(Cov(\underline{b}))+$$
$$+ \theta_q\ tr(Z_1(Z_1'Z_1)^{-}Z_1')$$
$$+ E(\underline{b}')Z'Z_1(Z_1'Z_1)^{-}Z_1'Z\ E(\underline{b}) \qquad (3.122)$$

But using:

$$Z_1(Z_1'Z_1)^{-}Z_1'Z_1 = Z_1 \quad \text{(Searle, (1971) p 20)}$$

$$Z'Z_1(Z_1'Z_1)Z_1'Z = \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_1(Z_1'Z_1)^{-}Z_1'Z_2 \end{bmatrix}$$
$$= Z'Z - \begin{bmatrix} 0 & 0 \\ 0 & Z_2'N_1Z_2 \end{bmatrix} \qquad (3.123)$$

where:

$$N_1 = I - Z_1(Z_1'Z_1)^{-}Z_1' ,$$

Hence,

$$E(R(\underline{b}_1)) = tr(Z'ZCov(\underline{b})) - tr(Z_2'N_1Z_2Cov(\underline{b}_2))$$
$$+ \underline{\beta}'X'X\underline{\beta} - E(\underline{b}_2')Z_2' N_1Z_2 E(\underline{b}_2) \qquad (3.124)$$

The last term is zero, since $\underline{b}_2$ comprises only random effects, $E(\underline{b}_2) = 0$.

Therefore,

$$E(R(\underline{b}_2/\underline{b}_1)) = E(R(\underline{b}_1\ \underline{b}_2)) - E(R(\underline{b}_1))$$
$$= tr(Z_2'N_1Z_2D_2) + \theta_q(rank(Z) - rank(Z_1))$$
$$= \sum_{i=1}^{t} \theta_{2i}\ tr(U_{2i}'\ N_1U_{2i}) + \theta_q(rank(Z)-rank(Z_1)) \qquad (3.125)$$

because $tr(Z(Z'Z)^{-}Z') = rank(Z)$.

Hence, the expected value of the reduction SS is a linear function of the variance components relating to the constraints fitted last. It does not involve the fixed effects or the components for constants fitted first. With judicious partitionings of effects into $\underline{b}_1$ and $\underline{b}_2$, taking care with the restriction (3.114), we can produce a vector $\underline{r}$ of reduction SSs, whose expectations provide a set of estimating equations for the variance components. Given the vector $\underline{r}$ partitioned into effects fitted first and last, we perform a two-stage procedure:

a) Fit $\underline{y} = Z_1\underline{b}_1 + \underline{e}$ and obtain

rank (Z_1), $R(\underline{b}_1)$ and the residuals

$\hat{e}_1 = \underline{y} - Z_1\hat{\underline{b}}_1 = N_1\underline{y}$

b) Fit $\hat{\underline{e}} = Z_2\underline{b}_2 + \underline{e}$ and obtain

rank(Z_1Z_2), $R(\underline{b}_1\ \underline{b}_2)$.

3.3.2. Computational Method - Henderson's Method III.

There are several algorithms available for producing the reduction SSs and the matrices required for coefficients of the expectation (3.125). They are essentially the algorithms for least squares solution of multiple linear regression problems. The method used here is a modification of Gaussian Elimination, the Sweep Operator, as defined by Goodnight (1978b).

Given a set of Normal Equations for model (3.113):

$$A = \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 & Z_1'\underline{y} \\ Z_2'Z_1 & Z_2'Z_2 & Z_2'\underline{y} \\ \underline{y}'Z_1 & \underline{y}'Z_2 & \underline{y}'\underline{y} \end{bmatrix} \qquad (3.126)$$

The Sweep Operator applied to all columns containing $Z_1'Z_1$ in A transforms A to A_1

$$A_1 = \begin{bmatrix} (Z_1'Z_1)^-, & (Z_1'Z)^-Z_1'Z_2, & (Z_1'Z_1)^-Z_1'\underline{y} \\ -Z_2'Z_1(Z_1'Z_1)^-, & Z_2'N_1Z_2, & Z_2'N_1\underline{y} \\ -\underline{y}'Z_1(Z_1'Z_1)^-, & \underline{y}'N_1Z_2, & \underline{y}'N_1\underline{y} \end{bmatrix} \qquad (3.127)$$

where: $N_1 = (I - Z_1(Z_1'Z_1)^-Z_1)$

After this stage of the computation we have $R(\underline{b}_1)$, because $\underline{y}'N_1\underline{y} = \underline{y}'\underline{y} - R(\underline{b}_1)$, and the coefficients for θ_{2i} in (3.125) from the trace of appropriate submatrices of $Z_2'N_1Z_2$. The Sweep operator also counts the rank of the swept columns of A, so we have the rank(Z_1). If sweeping is continued on the columns of A_1 containing $Z_2'N_1Z_2$ the result is:

$$A_2 = \begin{pmatrix} (Z'Z)^-, & (Z'Z)^-Z'\underline{y} \\ -\underline{y}'Z(Z'Z)^-, & \underline{y}'N\underline{y} \end{pmatrix} \qquad (3.128)$$

where: $N = I - Z(Z'Z)^-Z'$ and the partition $Z = (Z_1, Z_2)$ has been collapsed. At this stage we have rank(Z) and $R(\underline{b}_1, \underline{b}_2)$ because $\underline{y}'N\underline{y} = \underline{y}'\underline{y} - R(\underline{b}_1, \underline{b}_2)$. Hence we have the whole equation for the partition $\underline{b}_1, \underline{b}_2$:

$$R(\underline{b}_2/\underline{b}_1) = \sum_{i=1}^{t} \theta_{2i} tr(U_{2i}'N_1U_{2i}) + \theta_q(rank(Z)-rank(Z_1)) \qquad (3.129)$$

The sweep algorithm can be programmed to operate on the lower triangular part of A stored in a one-dimensional array by rows, (Goodnight (1978b)). Although A_1 and A_2 are not symmetric, the upper and lower triangular parts differ at most in sign so that the whole matrix may be recovered from one half.

The columns of A do not need to be ordered in any particular way. The ordering of (3.126) was only for notational convenience. Similarly, the effects selected to be fitted first or last need not be grouped in any particular way. Hence, a convenient way of producing a set of equations

like (3.129) is to specify an order for adding the effects of model (2.23) into the group to be fitted first, while taking account of the restriction (3.114). For example for the model:

$$\underline{y} = X\underline{\beta} + \sum_{j=1}^{q-1} U_j \underline{v}_j + \underline{e} \qquad (3.130)$$

where: $X = (X_1, X_2, \ldots X_r)$ and $\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2', \ldots \underline{\beta}_r')'$

we specify an order: $(k1, k2 \ldots k(q-1))$ (3.131)

which is a permutation of $(1, 2 \ldots (q-1))$, and specifies the order in which the random effect values $(\underline{v}_1, \underline{v}_2 \ldots \underline{v}_{q-1})$ are to be added to the group of effects fitted first. This leads to the set of reduction SSs:

$$\{R(\underline{v}_{k1}, \ldots \underline{v}_{k(q-1)}/\underline{\beta}),\ R(\underline{v}_{k2}, \ldots \underline{v}_{k(q-1)}/\underline{\beta}, \underline{v}_{k1}), \ldots$$
$$R(\underline{v}_{k(q-1)}/\underline{\beta}, \underline{v}_{k1}, \ldots \underline{v}_{k(q-2)}),\ \text{RSS}\}$$

where: RSS is the residual SS after fitting the full model.

An algorithm for computing these SSs is as follows: The columns of the normal equations A of (3.126), corresponding to X are swept first at stage 0. This results in a matrix A_0 with the form of (3.127) from which it is easy to obtain:

$$t_1 = \underline{y}'\underline{y} - R(\underline{\beta}) = \underline{y}' N_0 \underline{y} \qquad (3.132)$$

$$f_1 = \text{rank}(X) = \text{rank}(Z_0) \qquad (3.133)$$

and $$s_{1,i} = \text{tr}(U_{ki}' N_0 U_{ki}) \text{ for } i = 1, 2 \ldots q-1 \qquad (3.134)$$

where: $Z_0 = (X_1, X_2, \ldots X_r)$ and $N_0 = I - Z_0 (Z_0' Z_0)^{-} Z_0'$

Next, a matrix A_1 is produced by sweeping A_0 on the columns corresponding to those of U_{k1} in the normal equations. The process continues producing A_2, A_3, ... A_{q-1}. At the j th

stage, the following statistics are read from A_j:

$$t_{j+1} = \underline{y}'\underline{y} - R(\underline{\beta}_1\underline{\beta}_2\cdots\underline{\beta}_r, \underline{v}_{k1}\cdots\underline{v}_{kj})$$
$$= \underline{y}'N_j\underline{y} \qquad (3.135)$$

$$f_{j+1} = \text{rank}\,(X_1X_2\cdots X_r, U_{k1}\cdots U_{kj})$$
$$= \text{rank}\,(Z_j) \qquad (3.136)$$

$$s_{j+1,i} = \text{tr}(U_{ki}'\, N_j U_{ki}) \quad i=j+1,\ j+2\ldots q-1 \qquad (3.137)$$
$$0 \quad \text{for } i=1,2\ldots j$$

where:

$$Z_j = (X_1X_2\cdots X_rU_{k1}, U_{k2}\cdots U_{kj})$$
$$N_j = I - Z_j(Z_j'Z_j)^{-}Z_j$$

After the final stage, from A_{q-1}, we have;

$$t_q = \underline{y}'\underline{y} - R(\underline{\beta}_1\cdots\underline{\beta}_r, \underline{v}_{k1}\cdots\underline{v}_{k(q-r)})$$

and $$f_q = \text{rank}(X_1X_2\cdots X_v, U_{k1}\cdots U_{k(q-1)}).$$

There are no non-zero values for s_{qi}, like (3.137), because we have fitted the whole model, and $U_{ki}'N_q = 0$ all i.

Let $\underline{r}$ be the q-vector with components

$$r_i = t_q - t_i = R(\underline{v}_{ki}\cdots\underline{v}_{k(q-1)}/\underline{\beta}_1\cdots\underline{\beta}_1\underline{v}_{k1}\cdots\underline{v}_{k(i-1)})$$
$$i=1,2\ldots q-1$$

and $$r_q = t_q = \text{residual SS.} \qquad (3.138)$$

Let R be the obvious qxq matrix of row contrasts, such that:

$$\underline{r} = R\underline{t}$$

Let H_D be the qxq matrix, with i,j th element h_{ij}, where:

$$h_{ij} = s_{ij} \quad i=1,2...q-1,\ j=1,2...q-1$$
$$h_{iq} = f_q \quad \text{for} \quad i=1,2...q-1$$
$$h_{qj} = 0 \quad \text{for } j=1,2...q-1$$
$$h_{qq} = n - f_q \qquad (3.139)$$

(The last row of H_D is the equation for $E(r_q)$, as in (3.50)).

Then the Henderson Method III Estimators are then solutions $\hat{\underline{\theta}}$, to:

$$\underline{r} = H_D\hat{\underline{\theta}} \qquad (3.140)$$

The procedure can produce a variety of estimators, depending on fitting orders (3.131), which are specified as input.

3.3.3. Distributional Properties of the Henderson Method III Estimators.

If the Normal distribution assumptions are included in the model (3.130) we know that r_q and hence $\hat{\theta}_q$, have central chi-square distribution (3.52). Further, since:

$$t_i = \underline{y}'N_{i-1}\underline{y} \quad i=1,2...q,$$
$$r_i = t_q - t_i = \underline{y}'(N_{q-1}-N_{i-1})\underline{y} \quad i=1,2...q-1$$

and:
$$(N_{q-1}-N_{i-1})V(\theta)N_{q-1} = \theta_q(N_{q-1}-N_{i-1})N_{q-1}$$
$$= \theta_q(N_{q-1} - N_{q-1}) = 0$$

we have that r_q and hence $\hat{\theta}_q$ is independent of $r_1 r_2 .. r_{q-1}$.

However, since $N_{i-1}V(\theta)$ for i=1,2...q-1 are not, in general, idempotent (except for balanced models), the distribution of $t_1, t_2 \ldots t_{q-1}$ is unknown. Similarly, for $r_1 \ldots r_{q-1}$.

However, writing

$$\hat{\underline{\theta}} = H_D^{-1} R \underline{t} \tag{3.141}$$

we have

$$\mathrm{Cov}(\hat{\underline{\theta}}) = H_D^{-1} R \mathrm{Cov}(\underline{t}) R' H_D^{-1}, \tag{3.142}$$

and from (3.14),

$$\mathrm{Cov}(t_i, t_j) = 2\mathrm{tr}(N_{i-1}V(\underline{\theta})N_{j-1}V(\underline{\theta})) + 4\underline{\beta}'X'N_{i-1}V(\underline{\theta})N_{j-1}X\underline{\beta} \tag{3.143}$$

However, $X'N_{i-1} = N_{j-1}X = 0$ for any i,j=1,2...q.

So we can write:

$$\mathrm{Cov}(\underline{t}) = 2D(\underline{\theta}) \tag{3.144}$$

and

$$\mathrm{Cov}(\hat{\underline{\theta}}) = 2H_D^{-1} R D(\underline{\theta}) R' H_D^{-1} \tag{3.145}$$

$D(\underline{\theta})$ is symmetric, with lower triangular elements d_{ij} with $i \geq j$ given by:

$$d_{ij} = \mathrm{tr}(N_{i-1}V(\underline{\theta})N_{j-1}V(\underline{\theta})) = \sum_{p=i}^{q-1}\sum_{m=i}^{q-1} \theta_{kp}\,\theta_{km}\,\mathrm{tr}(U'_{km}N_{i-1}U_{kp}U'_{kp}N_{j-1}U_{km}) + \theta_q^2 f_i \tag{3.146}$$

where $\{kj: j=1,2..q-1\}$ is the fitting order specified at (3.131).

because

a) $N_{i-1}N_{j-1} = N_{i-1} - N_{i-1}Z_{j-1}(Z'_{j-1}Z_{j-1})^{-}Z_{j-1}$

$= N_{i-1}$

b) $tr(N_{i-1}N_{j-1}) = rank(N_{i-1}) = rank(Z_{i-1}) = f_i$

c) $N_{i-1}U_{kp} = 0$ for p=1,2...i-1

d) $U'_{km}N_{i-1} = 0$ for m=1,2...i-1

The diagonal elements d_{ii}, and hence variances of t_i are easy to calculate with the algorithm of section 3.3.2., by calculating the SS of elements in the m,p th submatrix of the unswept portion of $A_1 :- (Z'_2 N_{i-1} Z_2)$. That is, the submatrix $U'_{km}N_{i-1}U_{kp}$. This SS is the coefficient of $\theta_{kp}\theta_{km}$ in d_{ii}. However, the off diagonal elements are computationally far more burdensome. In order to get a measure of $Cov(\underline{\theta})$, we would have to substitute $\hat{\underline{\theta}}$ for $\underline{\theta}$ in $D(\underline{\theta})$ of (3.145).

3.4. Henderson's Method III Estimates of Variance Components for Example Data Sets.

Variance component estimates were computed for the example data sets described in 2.7. The method used was the method of fitting constants, Henderson's method III, with the algorithm described in 3.3.2.

3.4.1. Data Set 1.

With the full model as described in Table (2.2), the reductions in residual SS used to compute the variance component estimates were:

a) Reduction due to fitting effect SIRE last,

b) Reduction due to fitting SIRE AND SIRE X DBRD after all other effects,

and c) Residual SS after fitting all effects.

Negative estimates of the SIRE or SIRE X DBRD variance components were obtained for all variates except Length at birth, LENB, which had the following estimated variance components:

a) SIRE 40.726

b) SIRE X DBRD 51.418

c) Residual 759.237

The large number of negative estimates could be a result of an inappropriate model, or the relatively large residual variance which could have swamped the components

However, if the model is reduced by excluding the SIRE X DBRD interaction, several variates still have negative SIRE variance components. The estimates under the reduced model are given in the following table.

Table (3.1): Henderson Method III Estimates for Data Set 1 Under the Reduced Model.

VARIABLE	VARIANCE COMPONENT ESTIMATE FOR:	
	SIRE	RESIDUAL
MASB	0.333	14.611
HTAB	-5.361	823.334
DPTB	-2.825	668.951
LENB	65.454	779.132
HTAW	-54.385	1461.444
DPTW	-45.346	1207.723
LENW	-25.399	2299.078
MASW	24.848	576.397

3.4.2. Data Set 2.

The data generated to follow the model which was assumed for the cattle breeding experiment 2.7.3., was analysed according to the full model of Table (2.2). The Henderson Method III estimates for this data are given in the following table.

Table (3.2): Henderson Method III Variance Component Estimates for Data Set 2.

VARIABLE	VARIANCE COMPONENT ESTIMATE FOR:		
	SIRE	SIRE X DBRD	RESIDUAL
1	14.781	1.921	8.221
2	4.618	2.905	9.275
3	8.510	3.611	8.804
4	9.944	3.061	8.733
5	5.169	4.050	8.404
6	7.661	0.923	8.983
7	8.727	4.644	7.153
8	26.952	2.763	7.940
9	4.430	3.933	7.875
10	18.239	1.427	8.002

The true variance component values were 10.0, 3.0 and 8.0 respectively.

3.4.3. Data Set 3.

The generated data set of 2.7.3. was analysed according to the model of Table (2.6). The reductions in residual SS used for the Henderson Method III estimates were:

a) R1, F X R1, R2, R1 X R2 fitted after MEAN and F,

b) F X R1, R2, R1 X R2 fitted after MEAN, F, and R1,

c) R2, R2 X R1 fitted after MEAN, F, R1, and F X R1,

d) R2 X R1 fitted after MEAN, F, R1, F X R1, and R2,

and e) Residual SS after fitting the full model.

The estimates are given in the following table.

Table (3.3): Henderson Method III Variance Component Estimates for Data Set 3.

VARIABLE	VARIANCE COMPONENT ESTIMATE FOR:				
	R1	R2	F X R1	R1 X R2	RESIDUAL
1	64.184	17.120	-19.575	17.851	15.720
2	23.145	17.392	2.786	0.911	9.889
3	28.554	-3.417	-16.327	24.752	12.556
4	2.299	-6.092	-4.359	9.889	16.429
5	12.140	10.180	-13.023	16.459	16.777
6	76.390	5.798	17.439	-2.486	8.893
7	26.955	17.720	11.606	-4.360	11.311
8	32.975	3.341	-12.974	11.781	20.974
9	68.238	5.955	15.194	7.659	11.330
10	18.698	2.836	19.458	-8.637	19.988

The true variance components were: 30.0, 10.0, 3.0, 6.0, and 12.0 respectively.

CHAPTER 4

Classical Approaches to Optimal Variance Component Estimation.

Discussion of the ANOVA and Fitting Constants methods of variance component estimation illustrates three aspects of the general estimation problem:

a) What criteria should we use to judge between candidate estimators?

b) Can we restrict the class of functions from which the estimators are to be selected, without a high risk of excluding particularly desirable ones?

c) What properties should be required of estimators, and are there any available with these properties?

Two criteria for choice between estimators are used: minimum mean square error and a minimum norm. The choice in the variance component case is a trade-off between extra assumptions on the model and restrictions on the class of estimating functions. This is because variance component estimators have variances or mean square errors, which depend on second to fourth order moments. Hence, not only are the parameters of interest involved, but for non-normal distributions, other unknown parameters may be involved as well.

Rao, C.R. (1971b) assumed finite fourth order moments in the structured model (2.13) and made progress with the criterion of

local minimum variance amongst quadratic unbiased and invariant estimating functions. Apart from the restriction on functions, the local property requires specification of a priori values of variance components and kurtosis parameters.

For normal models, however, local minimum variance and local minimum mean square error provide tractable solutions in general classes of unbiased estimating functions, as well as in some classes of biased, quadratic estimating functions. (LaMotte (1973 a and b)). Fortunately, the second criterion of minimum norm, which does not require normality, produces the same estimators when normality is available.

The minimum norm criterion was introduced by C.R. Rao (1970, 1971a, 1971b and 1972). The motivation is along the lines of linear least squares estimation for location parameters, but applies to quadratic functions. The idea is, given a quadratic estimator $\underline{y}' \Delta \underline{y}$, which is somehow obvious but unobservable, use the estimator $\underline{y}' A \underline{y}$ where A minimizes some norm between A and Δ. This method does not require the normal distribution assumptions, but is restricted to quadratic functions. Like minimum mean square error, it does depend on the variance components and in practice is also a local criterion.

The answer to the second question about restrictions on the class of estimating functions, is forced by the choice between optimality criteria. Restriction to quadratic functions is not severely counter-intuitive, since the parameters of interest are

second order moments. However, there is evidence (4.40) for considering quadratic plus linear functions, or quadratic functions of a transformed response.

The question of properties to be required of candidate estimators is plagued by issues of consistency, as well as trade-offs between desirable features of the estimators. Requirements of unbiasedness, invariance to translations of the location parameters and non-negativity are considered. Conditions for estimability under these restrictions are unfortunately, difficult to test in practice.

4.1. Effects of Unbiasedness Invariance and Non-Negativity on Estimability.

For the general linear model (2.1)

$$\underline{y} = X\underline{\beta} + \underline{e} \qquad (4.1)$$

with $\quad Cov(\underline{y}) = \sum_{i=1}^{q} \theta_i V_i$

we consider estimation of linear functions of the variance parameters, θ_i, given by $\underline{h}'\underline{\theta}$ for $\underline{h} \in \mathbb{R}^q$.

Let Q be the class of quadratic functions of $\underline{y}$, and QL quadratic plus linear functions.

$$Q = \{ \underline{y}' A \underline{y} : A \text{ nxn}, A = A' \} \qquad (4.2)$$

$$QL = \{ \underline{y}' A \underline{y} + \underline{a}'\underline{y}: A \text{ nxn}, A = A' \text{ and } \underline{a} \in \mathbb{R}^n \} \qquad (4.3)$$

4.1.1. Unbiasedness.

By (3.7) we have;

$$E(\underline{y}'A\underline{y} + \underline{a}'\underline{y}) = \sum_{i=1}^{q} \theta_i tr(AV_i) + \underline{\beta}'X'AX\underline{\beta} + \underline{a}'X\underline{\beta} \qquad (4.4)$$

Hence, if $QL(U,\underline{h})$ and $Q(U,\underline{h})$ are the sub classes of QL and Q of functions unbiased for $\underline{h}'\underline{\theta}$ we can write:

$$QL(U,\underline{h}) = \{ \underline{y}'A\underline{y} + \underline{a}'\underline{y} : A = A',\ X'AX = 0,\ \underline{a}'X = 0 \text{ and}$$

$$tr(AV_i) = \underline{h}_i,\ i=1,2\ldots q\} \qquad (4.5)$$

and $$Q(U,\underline{h}) = \{\underline{y}'A\underline{y}: A = A',\ X'AX = 0, \text{ and}$$

$$tr(AV_i) = \underline{h}_i,\ i=1,2\ldots q\} \qquad (4.6)$$

Conditions for estimability, that is, non-emptiness of the classes, were developed in general by Seeley (1970a and b) and more specifically by Pincus (1974), and Kleffe and Pincus (1974 a and b).

Both classes, $QL(U,\underline{h})$ and $Q(U,\underline{h})$ are non-empty if and only if $\underline{h} \in R(H_U)$, where

$$H_U = (tr(V_i N V_j)) \quad i,j=1,2\ldots q \qquad (4.7)$$

and $$N = I - P,\quad P = X(X'X)^{-1}X'$$

An alternative necessary and sufficient condition is:

$$\underline{h} \in R(H_U^*) \text{ where } H_U^* = (tr(V_iV_j - PV_iPV_j)) \qquad (4.8)$$

$i,j=1,2\ldots q.$

This follows from (4.7) because of the fact that

$$R(H_U) = R(H_U^*) \text{ (A.2)}.$$

To establish the result (4.7), we first note that;

H_U^* = TT' where T is q x n^2 with i th row given by $Vec'(V_i - PV_iP)$ (A.1).

Hence,

$$R(H_U) = R(H_U^*) = R(T).$$

Next, if $QL(U,\underline{h})$ or $Q(U,\underline{h})$ are not empty, there exists a symmetric matrix A such that $X'AX = 0$ and $tr(AV_i) = h_i$, i=1,2...q.

Hence,

$$\underline{h} = T\,Vec(A)$$

so that

$$\underline{h} \in R(T) = R(H_U^*) = R(H_U). \tag{4.9}$$

On the other hand, if $\underline{h} \in R(H_U)$, there exists $\underline{b} \neq 0$, such that

$$\underline{h} = H_U^* \underline{b} \tag{4.10}$$

If $A = \sum_{j=1}^{q} (b_j/2)\,(NV_j + V_jN)$ where $N = I - P$

then A is symmetric,

$X'AX = 0$ since $NX = XN = 0$

and
$$tr(AV_i) = \sum_j (b_j/2)(tr(NV_jV_i) + tr(V_jNV_i))$$
$$= \sum_j b_j\, tr(V_iNV_j) = h_i \quad \text{from (4.10)}$$

(using: $tr(AB) = tr(BA)$ and $tr(A) = tr(A')$)

Hence,

$$\underline{y}'A\underline{y} \in Q(U,\underline{h}) \text{ and } \underline{y}'A\underline{y} + \underline{a}'\underline{y} \in QL(U,\underline{h})$$

for any $\underline{a}$ s.t. $\underline{a}'X = 0$. (4.11)

Further equivalent conditions to (4.7) and (4.8) follow from observing

that if B is any p.d. matrix, $R(H_U^*) = R(H_U(B))$ (4.12)

where: $H_U(B) = (tr(B^{-1}(V_i - P(B)V_iP'(B))B^{-1}V_j))$

and $P(B) = X(X'B^{-1}X)^-X'B^{-1}$. (A.3)

In particular, if $\underline{s}$ is a q-vector, such that $V(\underline{s}) = \Sigma s_i V_i$ is p.d.

Then $Q(U,\underline{h})$ and $QL(U,\underline{h})$ are non-empty if and only if

$\underline{h} \in R(H_U(\underline{s}))$ (4.13)

where

$$H_U(\underline{s}) = (tr(V(\underline{s})^{-1}(V_i - P(\underline{s})V_iP(\underline{s})')V(\underline{s})^{-1}V_j))$$

and

$$P(\underline{s}) = X(X'V(\underline{s})^{-1}X)^-X'V(\underline{s})^{-1}$$

Seeley (1970b) gives two corollaries to (4.7): (4.14)

a) θ_k is unbiasedly estimable in Q or QL if and only if there does not exist $\underline{a} \neq 0 \in R^{q-1}$ and Λ symmetric such that

$$V_k = X \Lambda X' + \sum_{i \neq k} a_i V_i$$

b) All components of $\underline{\theta}$ are unbiasedly estimable in Q or QL if and only if $\{V_i \ldots V_q\}$ are independent in the vector space of nxn symmetric matrices and

$$Sp\{V_i, \ldots V_q\} \cap Sp\{X \Lambda X' : \Lambda = \Lambda'\} = \{0\}$$

Comparing condition (4.7) for Estimability of $\underline{h}'\underline{\theta}$ and (2.3) for Identifiability, we note that the two properties do not coincide, since

in general,

$$R(H_U) \subset R(H)$$

Pincus, (1974), shows that in normal models, if $Q(U,\underline{h})$ and $QL(U,\underline{h})$ are empty, then $\underline{h}'\underline{\theta}$ is not unbiasedly estimable by any function of the response.

A class of estimating functions where the bias is unaffected by the fixed effect parameters is $Q(PU)$ where;

$$Q(PU) = \{ \underline{y}'A\,\underline{y} : A=A',\ X'AX = 0 \} \qquad (4.15)$$

4.1.2. Invariance to Translations of the Fixed Effect Parameters.

In the model (4.5) the covariance structure is invariant to translations of $\underline{\beta}$. This provides an intuitive argument for requiring estimators of $\underline{\theta}$ to have the same invariance. (Rao, C.R. (1973) p 343).

$f(\underline{y})$ is invariant if and only if $f(\underline{y} + X\underline{\beta}) = f(\underline{y})$ for all $\underline{\beta} \in \Omega_{\underline{\beta}}$. According to Lehmann((1959)p 215), we need only consider functions of a maximal invariant.

For the linear model (4.5),

$$\underline{z} = N(\underline{s})\underline{y} \qquad (4.16)$$

is a maximal invariant where

$\underline{s} \in \Omega_{\underline{\theta}}$ is s.t. $V(\underline{s}) = \Sigma s_i V_i$ is p.d.

$N(\underline{s}) = V(\underline{s})^{-1}(I - P(\underline{s}))$, $P(\underline{s}) = X(X'V(\underline{s})^{-1}X)^{-}X'V(\underline{s})^{-1}$

Kleffe (1976a) shows this as follows:

a) $N(\underline{s})\ (\underline{y} + X\underline{\beta}) = N(\underline{s})\underline{y}$ all $\underline{\beta} \in R^p$

because $N(\underline{s})X = 0$, so $N(\underline{s})\underline{y}$ is invariant, and

b) Taking $\underline{y}_1 \neq \underline{y}_2$, $(\underline{y}_1, \underline{y}_2 \in R(X)^{\perp}$ without loss of generality) suppose $N(\underline{s})\underline{y}_1 = N(\underline{s})\underline{y}_2$ then $\underline{y}_2 = \underline{y}_1 + X\underline{\beta}$, for otherwise $\underline{y} = \underline{y}_2 - \underline{y}_1 \in R(X)^{\perp}$ and $N(\underline{s})\underline{y} = 0$, which contradicts the fact that rank $(N(\underline{s})) = n - \text{rank } (X)$.

In fact, any function $B\underline{y}$ where $R(B) \perp R(X)$ and $\text{rank}(B) = n - \text{rank } (X)$ is a maximal invariant. In particular, $N(\underline{s})$ can be factored as BB' with B n x (n-rank(x)) having full column rank and then $B'\underline{y}$ is a maximal invariant such that

$$\text{Cov}(B'\ \underline{y}) = B'V(\underline{\theta})\ B \tag{4.17}$$

is non-singular when $V(\underline{\theta})$ is p.d.

If we restrict attention to quadratic estimating functions, Q and quadratic plus linear functions, QL, we can characterise the invariant functions by:

$$QL(I) = \{\underline{y}'A\ \underline{y} + \underline{a}'\underline{y}: A = A',\ AX = 0,\ \underline{a}X = 0\}$$

and

$$Q\ (I) = \{\underline{y}'A\ \underline{y}: A = A',\ AX = 0\} \tag{4.18}$$

since:

$$(\underline{y} + X\underline{\beta})'A(\underline{y} + X\underline{\beta}) + \underline{a}'(\underline{y} + X\underline{\beta}) = \underline{y}'A\underline{y} + \underline{a}'\underline{y}$$

for all $\underline{y}$ and $\underline{\beta}$ if and only if $AX = 0$ and $\underline{a}'X = 0$.

If we include the restfiction of unbiasedness for $\underline{h}'\underline{\theta}$ on these classes, they become:

$$QL(U,I,\underline{h}) = \{\underline{y}'A\underline{y} + \underline{a}'\underline{y}: A = A',\ AX = 0,\ \underline{a}'X = 0 \text{ and } tr(AV_i) = h_i,\ i=1,2...q\} \quad (4.19)$$

and

$$Q(U,I,\underline{h}) = \{\underline{y}'A\underline{y}: A = A',\ AX = 0 \text{ and } tr(AV_i) = h_i,\ i=1,2...q\} \quad (4.20)$$

A necessary and sufficient condition for $Q(U,I,\underline{h})$ or $QL(U,I,\underline{h})$ to be non-empty is that

$$\underline{h} \in R(H_{U,I}) \quad (4.21)$$

where

$$H_{U,I} = (tr(NV_iNV_j)) \quad i,j=1,2...q$$

and $N = I - P$

To show this, first note that $H_{U,I} = TT'$ where T is qxn^2 with i th row $Vec'(NV_iN)$ (A.1).

Secondly, if $Q(U,I,\underline{h})$ or $QL(U,I,\underline{h})$ are non-empty, there exists a symmetric matrix A such that $AX = AP = 0$ and $tr(AV_i) = h_i$, i=1,2..q. Hence, $A = NAN$ and

$$h_i = tr(NANV_i) = tr(NV_iNA)$$
$$= Vec'(NV_iN)\ Vec(A) \text{ for } i=1,2...q.$$

Therefore $\underline{h} = T\ Vec(A)$ so $\underline{h} \in R(T) = R(H_{U,I})$

On the other hand, if $\underline{h} \in R(H_{U,I})$, there exists $\underline{b}$ such that

$$\underline{h} = H_{U,I}\underline{b}$$

Let $A = \Sigma b_j NV_jN$,

then A is symmetric, $AX = 0$

and $tr(AV_i) = \Sigma b_j \, tr(NV_jNV_i) = h_i$.

Hence

$$\underline{y}'A\,\underline{y} \in Q(U,I,\underline{h})$$

Equivalent conditons to (4.21) follow from the fact that for

B p.d., $R(H_{U,I}) = R(H_{U,I}(B))$ (A.4) (4.22)

where

$$H_{U,I}(B) = (tr(N(B)V_iN(B)V_j))$$

and

$$N(B) = \bar{B}^{-1}(I - P(B)),\ P(B) = X(X'\bar{B}^{-1}X)^{-}X'\bar{B}^{-1}.$$

In particular, if $V(\underline{s}) = \Sigma s_i V_i$ is p.d. we write

$$H_{U,I}(\underline{s}) \quad \text{for} \quad H_{U,I}(V(\underline{s}))$$

It is easy to find examples where $Q(U,I,\underline{h})$ is empty and in these situations, unbiasedness and invariance cannot both be required properties. Rao, C.R. (1979) gives the following example:

$$y_1 = \beta_1 + e_1,\ y_2 = \beta_1 + e_2 \qquad (4.23)$$
$$y_3 = \beta_2 + e_3,\ y_4 = \beta_2 + e_4$$

with

$$E(e_ie_j) = 0,\ E(e_1^2) = E(e_3^2) = \theta_1, E(e_2^2) = E(e_4^2) = \theta_2$$

Then symmetric matrices A such that $AX = 0$ imply that $tr(AV_1) = tr(AV_2)$ so that only the linear combination $a(\theta_1+\theta_2)$ is unbiasedly estimable.

Another way to view this is to note that $H_{U,I} = (\frac{1}{2})\ \underline{1}\ \underline{1}'$, while $H_U = I_2$ so that unbiased and invariant estimation is limited, while all functions are unbiasedly estimable.

Where $Q(U,I,\underline{h})$ is empty and unbiasedness is forfeited, Kleffe (1977c) reports that quadratic minimum bias invariant estimation reduces to the class $Q(U,I,\underline{h}^*)$ where $\underline{h}^*$ is the projection of $\underline{h}$ onto $R(H_{U,I})$. Hence, these estimators can be examined with those in $Q(U,I,\underline{h})$.

4.1.3. Non-Negative Variance Component Estimation.

In variance component models (2.16), the variance component parameters, θ_i are non-negative. All classes of estimators considered so far, allow estimators to take negative values, even when the parametric function $\underline{h}'\underline{\theta}$ is positive. The difficulty with non-negative estimation, is that the requirements of non-negativity and unbiasedness are frequently incompatible. If unbiasedness is not required, but the class of estimators is restricted to quadratic forms, which do not take negative values, the class becomes:

$$Q(NND) = \{\underline{y}'A\ \underline{y}:\ A = A'\ A\ n.n.d\} \tag{4.24}$$

and with invariance, but not unbiasedness:

$$Q(I,NND) = \{\underline{y}'A\ \underline{y}:\ A = A'\ A\ n.n.d.\ \ AX = 0\}. \tag{4.25}$$

Because we allow biasedness, these classes always contain estimators for $\underline{h}'\underline{\theta}$. LaMotte (1973a) used locally minimum mean square error to produce optimal estimators in this class for Normal models. Rao, P.S.R.S. and Chaubey (1978) and Rao, C.R. (1979) provide non-negative

estimators in $Q(I,NND)$ and $Q(NND)$, using the locally minimum norm criteria, without the assumptions of Normality. Fortunately, these coincide with LaMotte's (1973a) estimators in the normal case.

LaMotte (1973b), Pukelsheim (1977) and Kleffe (1980) investigated the relationship between unbiasedness and non-negativity.

Firstly, non-negativity and unbiasedness together imply invariance because the unbiasedness condition: $X'AX = 0$ implies $AX = 0$ for A n.n.d.

$$Q(U,NND,\underline{h}) = \{\underline{y}'A\underline{y} : A = A',\ A \text{ n.n.d.},\ AX = 0, \text{ and } tr(AV_i) = h_i,\ i=1,2\ldots q\} \qquad (4.26)$$

and
$$Q(U,NND,\underline{h}) = Q(U,I,NND,\underline{h}) \qquad (4.27)$$

LaMotte (1973b) shows that if V_i, $i=1,2\ldots q$ are n.n.d., and if $\Sigma_{i\neq j} V_i$ is p.d. then θ_j is not $Q(U,NND)$ estimable. Furthermore, if V_j is p.d. then θ_j is the only variance component which is $Q(U,NND)$ estimable. For classification models this means that only the residual error component is non-negatively and unbiasedly estimable.

4.2. Minimum Variance Unbiased Estimation in Normal Models.

If we do not restrict the class of estimating functions but do assume Normality and p.d. covariance structure, then optimal estimators can be found on the basis of local minimum variance, since in general the variances of estimators depend on the parameters. Conditions under which the variance of these estimators is independent of the parameters are precisely the conditions of balance (2.60) and (2.61).

4.2.1. Locally Minimum Variance Unbiased Estimators (LMVUEs).

Kleffe (1977d) developes LMVUEs for Normal models with the form (2.1): $\underline{y} = X\underline{\beta} + \underline{e}$, $\underline{y} \sim N(X\underline{\beta}, V(\underline{\theta}))$ (4.28)

Given $\underline{b}$ and $\underline{s}$, a priori values for $\underline{\beta}$ and $\underline{\theta}$, such that:

$V(\underline{s}) = \sum_{i=1}^{q} s_i V_i$ is p.d., (4.29)

Then a) $\gamma = \underline{h}'\underline{\theta}$ is unbiasedly estimable if and only if

$\underline{h} \in R(H_U(\underline{s}))$

where: $H_U(\underline{s}) = (tr(A_i(\underline{s})V_j))$. (4.13) (4.30)

$A_i(\underline{s}) = V(\underline{s})^{-1}(V_i - P(\underline{s})V_iP(\underline{s})')V(\underline{s})^{-1}$ (4.31)

and $P(\underline{s}) = X(X'V(\underline{s})^{-1}X)^{-}X'V(\underline{s})^{-1}$ (4.32)

and b) If $\underline{h}$ satisfies a) then the LMVUE of γ at $(\underline{b},\underline{s})$ is:

$\hat{\gamma} = \sum \lambda_i(\underline{y} - X\underline{b})'A_i(\underline{s})(\underline{y} - X\underline{b})$ (4.33)

$= \underline{\lambda}'\underline{r}_u(\underline{s})$ (4.34)

$= (\underline{y} - X\underline{b})'A(\underline{s})(\underline{y} - X\underline{b})$ (4.35)

where: $\underline{\lambda} = (\lambda_1, \lambda_2 \ldots \lambda_q)'$ is any solution to

$H_U(\underline{s})\underline{\lambda} = \underline{h}$ (4.36)

$\underline{r}_u(\underline{s}) = ((\underline{y} - X\underline{b})'A_i(\underline{s})(\underline{y} - X\underline{b}),\ i=1,2..q)'$ (4.37)

and $A(\underline{s}) = \Sigma\lambda_i A_i(\underline{s})$ (4.38)

We can write $\hat{\gamma}$ as $\underline{h}'\hat{\underline{\theta}}$ where $\hat{\underline{\theta}}$ is any solution to:

$H_U(\underline{s})\hat{\underline{\theta}} = \underline{r}_u(\underline{s})$ (4.39)

It is interesting to note that the LMVUE turns out to be in $QL(U,\underline{h})$ even though all estimating functions are considered. (4.40)

Establishing part a) rests on the observation that $R(H_U(\underline{s})) = R(H_U)$, (4.13), and that for normal models, if $\underline{h} \notin R(H_U)$ then $\underline{h}'\underline{\theta}$ is not unbiasedly estimable. (Pincus (1974)).

For part b) we note that $\hat{\gamma}$ is unbiased for γ because

$X'A_i(\underline{s})X = 0$, $i=1,2...q$ and

$$tr(A(\underline{s})V(\underline{\theta})) = \sum_i \sum_j \lambda_i tr(A_i(\underline{s})V_j)\theta_j = \underline{\lambda}'H_U'(\underline{s})\underline{\theta} = \gamma$$

Then, if $g(\underline{y})$ is any random variable with zero expectation over $\Omega_{\underline{\beta}} x \Omega_{\underline{\theta}}$ and finite second order moment, then:

$$\int\int g(\underline{y}) \exp\{-(1/2)(\underline{y} - X\underline{\beta})'V(\underline{\theta})^{-1}(\underline{y} - X\underline{\beta})\}\, d\underline{y} = 0 \qquad (4.41)$$

Differentiating (4.41) with respect to θ_i (A.6), we can interchange differentiation and integration to obtain:

$$E_{\underline{\beta},\underline{\theta}}(g(\underline{y})(\underline{y} - X\underline{\beta})'V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}(\underline{y} - X\underline{\beta})) = 0, \quad \text{all } i. \qquad (4.42)$$

hence, $$Cov(g(\underline{y}), (\underline{y}'V(\underline{s})^{-1}V_iV(\underline{s})^{-1}\underline{y} - 2\underline{b}X'V(\underline{s})^{-1}V_iV(\underline{s})^{-1}\underline{y})/\ \underline{b},\underline{s}) = 0$$

$$\text{for } i=1,2...q \qquad (4.43)$$

The second order partial derivatives of (4.41) with respect to $\beta_i\beta_j$, for $i,j=1,2...p$ result in:

$$E_{\underline{\beta},\underline{\theta}}(g(\underline{y})(\underline{y}'V(\underline{\theta})^{-1}x_ix_j'V(\underline{\theta})^{-1}\underline{y} - 2\underline{\beta}'X'V(\underline{\theta})^{-1}x_ix_j'V(\underline{\theta})^{-1}\underline{y}$$
$$+ \underline{\beta}'X'V(\underline{\theta})^{-1}x_ix_j'V(\underline{\theta})^{-1}X\underline{\beta} - x_j'V(\underline{\theta})^{-1}x_i)) = 0$$

where x_i is the i th column of X.

Hence: $$Cov(g(\underline{y}), (\underline{y}'V(\underline{s})^{-1}x_ix_j'V(\underline{s})^{-1}\underline{y} - 2\underline{b}'X'V(\underline{s})^{-1}x_ix_j'V(\underline{s})^{-1}\underline{y})/\ \underline{b},\underline{s}) = 0$$

$$\text{for } i,j = 1,2...p. \qquad (4.44)$$

Now $sp\{x_ix_j'$, $i,j = 1,2...p\} = sp\{XCX' : C \text{ symmetric}\}$ and hence includes the matrices: $P(\underline{s})V_iP(\underline{s})'$ for $i=1,2...q$ and therefore,

$$Cov(g(\underline{y}), (\underline{y}'V(\underline{s})^{-1}P(\underline{s})V_iP(\underline{s})'V(\underline{s})^{-1}\underline{y}$$
$$- 2\underline{b}'X'V(\underline{s})^{-1}P(\underline{s})V_iP(\underline{s})'V(\underline{s})^{-1}\underline{y})/\ \underline{b},\underline{s}) = 0 \qquad (4.45)$$
$$\text{for } i=1,2...q.$$

Subtracting (4.45) from (4.43) gives:

$$\mathrm{Cov}(g(\underline{y}), (\underline{y}'A_i(\underline{s})\underline{y} - 2\underline{b}'X'A_i(\underline{s})\underline{y}) / \underline{b},\underline{s}) = 0$$

or since $X'A_i(\underline{s})X = 0$,

$$\mathrm{Cov}(g(\underline{y}), \hat{\gamma} / \underline{b},\underline{s}) = 0$$

which together with a result on minimum variance unbiased estimation (Rao, C.R. (1973) p 317) means that $\hat{\gamma}$ is LMVUE for γ.

The variance of $\hat{\gamma}$ (4.34) is, from (3.12):

$$\mathrm{Var}(\hat{\gamma} / \underline{\beta},\underline{\theta}) = 2\mathrm{tr}(A(\underline{s})V(\underline{\theta})A(\underline{s})V(\underline{\theta})) + 4\,(\underline{\beta} - \underline{b})'X'A(\underline{s})V(\underline{\theta})A(\underline{s})X(\underline{\beta} - \underline{b}) \qquad (4.46)$$

since $\underline{y} - X\underline{b} = X(\underline{\beta} - \underline{b}) + \underline{e}$.

This attains its minimum when $\underline{\beta} = \underline{b}$ and $\underline{\theta} = \underline{s}$:

$$\mathrm{Var}(\hat{\gamma} / \underline{b},\underline{s}) = 2\mathrm{tr}(A(\underline{s})V(\underline{s})A(\underline{s})V(\underline{s})) = \underline{\lambda}'H_U(\underline{s})\underline{\lambda} = \underline{h}'H_U^{-}(\underline{s})\underline{h} \quad \text{from (4.35)} \qquad (4.47)$$

One modification of (4.34) is to use the best linear unbiased estimator, $\hat{\underline{\beta}} = (X'V(\underline{s})X)^{-}X'\underline{y}$ of $\underline{\beta}$ for the a priori value $\underline{b}$. Then $X\hat{\underline{\beta}} = P(\underline{s})\underline{y}$ and $\hat{\gamma}$ becomes $\hat{\gamma}^*$ given $\underline{s}$:

$$\hat{\gamma}^* = \underline{y}'N(\underline{s})V(\underline{\lambda})N(\underline{s})\underline{y} \qquad (4.48)$$

where $N(\underline{s}) = V(\underline{s})^{-1}(I - P(\underline{s}))$ and $V(\underline{\lambda}) = \Sigma\lambda_i V_i$ with $\underline{\lambda}$ as in (4.36). This is independent of any a priori value for $\underline{\beta}$ but may be biased.

If $H_U(\underline{s})$ is non-singular, and writing $\hat{\underline{\theta}}_0$ for $\underline{s}$ we can establish an iterative scheme:

$$\hat{\underline{\theta}}_i = H_U^{-1}(\hat{\underline{\theta}}_{i-1})r_u(\hat{\underline{\theta}}_{i-1}), \quad \text{and} \quad X\hat{\underline{\beta}}_i = P(\hat{\underline{\theta}}_{i-1})\underline{y} \qquad (4.49)$$

If this scheme converges, the result is the iterated minimum variance unbiased estimator (IMVUE) which is not generally unbiased.

Seely (1971), Rao, C.R. (1971) and Kleffe and Pincus (1974a) examined conditions under which LMVUEs are independent of $\underline{b}$ and $\underline{s}$

and hence UMVUE. The conditions turn out to be equivalent to the conditions of balance in balanced variance component models (2.58), which lead to the ANOVA estimators of 3.1. which are UMVUE. (4.50)

4.2.2. Locally Minimum Variance Invariant and Unbiased Estimators.

If the class of estimating functions is restricted to invariant and unbiased estimators then we need only consider functions of a maximal invariant (4.16). Invariant and unbiased functions have the properties:

$$g(\underline{y} + X\underline{\beta}) = g(\underline{y}) \text{ for all } \underline{\beta} \text{ and } E(g(\underline{y})) = \underline{h}'\underline{\theta} \qquad (4.51)$$

Using a full rank maximal invariant (4.17), Kleffe shows that, given $\underline{s}$, an a priori value for $\underline{\theta}$:

a) The Locally minimum variance unbiased and invariant estimate, (LMVIUE) of $\gamma = \underline{h}'\underline{\theta}$ exists if and only if $\underline{h} \in R(H_{U,I}(\underline{s}))$

where: $H_{U,I}(\underline{s}) = (tr(N(\underline{s})V_i N(\underline{s})V_j))$ (4.22) (4.53)

and b) The LMVIUE of $\gamma = \underline{h}'\underline{\theta}$ is,

$$\hat{\gamma} = \Sigma\lambda_i \underline{y}'N(\underline{s})V_i N(\underline{s})\underline{y} \qquad (4.54)$$

$$= \underline{y}'N(\underline{s})V(\underline{\lambda})N(\underline{s})\underline{y} = \Sigma\lambda_i \underline{z}'V_i\underline{z} \qquad (4.55)$$

$$= \underline{\lambda}'\underline{r}_{U,I}(\underline{s}) \qquad (4.56)$$

or $$= \underline{h}'\hat{\underline{\theta}} \qquad (4.57)$$

where

$\underline{\lambda}$ is any solution to $H_{U,I}(\underline{s})\underline{\lambda} = \underline{h}$

$\underline{r}_{U,I}(\underline{s}) = (\underline{y}'N(\underline{s})V_i N(\underline{s})\underline{y},\ i=1,2\ldots q)'$

and $\hat{\underline{\theta}}$ is any solution to $H_{UI}(\underline{s})\hat{\underline{\theta}} = \underline{r}_{U,I}(\underline{s})$ (4.58)

This result also follows from noting that $R(H_{UI}(\underline{s})) = R(H_{UI})$ (4.22) and applying the LMVUE result (4.33) to the transformed model for $z = N(\underline{s})\underline{y}$ a maximal invariant with respect to translations in $\underline{\beta}$. Putting $\underline{s} = \hat{\underline{\theta}}_o$ and using

$$H_{UI}(\hat{\underline{\theta}}_{i-1})\ \hat{\underline{\theta}}_i = \underline{r}_{U,I}(\hat{\underline{\theta}}_{i-1})$$

we can define iterated MVIUE, IMVIUE (4.59)

4.3. Minimum Norm Quadratic Estimation in General Variance Component Models.

In a series of papers (1970, 1971a, 1971b, 1972), C.R. Rao developed a minimum norm criterion for variance component estimation, starting with heteroscedastic regression models. The main advantage of the method is that it does not require the normal distribution assumptions on the model, but the optimization is only carried out over restricted classes of quadratic estimating functions.

Fortunately the technique results in optimal estimators, which agree with those for Normal models derived through minimum mean square error or maximum likelihood principles, when the classes of estimating functions coincide. (Chaubey 1980). It is argued that the restriction to quadratic forms in the response, $\underline{y}$, or transformed response $\underline{y} - X\underline{b}$, is not severe, because optimal estimators for normal models with other criteria, turn out to be functions of that form.

The minimum norm estimators still suffer from the possibility of producing negative estimators in some classes. Chaubey and Rao, P.S.R.S. (1976), Rao, P.S.R.S. (1977) and Rao, P.S.R.S. and Chaubey (1978),

examined modifications to avoid this. One possibility is to consider biased quadratic estimating functions and in this case, it is possible to incorporate information on the structure of the error term for structured models, (2.16). This does not seem to be possible with other methods of estimation.

The distribution of the minimum norm estimators is unknown. Brown (1976) showed them to be asymptotically normal and consistent. However, he based these results on the assumption that models of increasing size could be regarded as replications of a basic model of small dimension. This assumption is difficult to meet in practice.

4.3.1. Principles of Minimum Norm Quadratic Estimation. (MINQE).

To estimate the function $\gamma = \underline{h}'\underline{\theta}$ with the minimum norm technique, we must first identify a quadratic estimator, $\underline{y}'\Delta\underline{y}$ which would be the obvious choice if the random part of the model were observable. Next, we must choose the observable estimator, $\underline{y}'A\underline{y}$, from some class of quadratic estimators, which minimizes $\|A - \Delta\|$ for some suitable choice of norm, $\| \cdot \|$.

The structured model, (2.16), has form:

$$\underline{y} = X\underline{\beta} + U_1\underline{v}_1 + \ldots + U_q\underline{v}_q = X\underline{\beta} + U\underline{v} \qquad (4.60)$$

with $E(\underline{v}_i) = 0$, $E(\underline{v}_i\underline{v}_j') = 0$ for $i \neq j$,

and $E(\underline{v}_i\underline{v}_i') = \theta_i I_{q_i}$, for $i = 1,2\ldots q$. (4.61)

If the $\underline{v}_i$ were observable, an obvious estimator of θ_i would be: $\hat{\theta}_i = \underline{v}_i'\underline{v}_i/q_i$, $i = 1,2\ldots q$ (4.62)

and hence for $\gamma = \underline{h}'\underline{\theta}$:

$$\hat{\gamma} = \underline{h}'\hat{\underline{\theta}} = \Sigma(h_i/q_i)\underline{v}_i'\underline{v}_i = \underline{v}'D\underline{v} \qquad (4.63)$$

where: $D = \text{diag}((h_i/q_i)I_{q_i}, \quad i = 1,2...q)$

By writing an arbitrary quadratic form in $\underline{y}$ as:

$$\underline{y}'A\underline{y} = \begin{bmatrix} \underline{v} \\ \underline{\beta} \end{bmatrix}' \begin{bmatrix} U'AU & , & U'AX \\ X'AU & , & X'AX \end{bmatrix} \begin{bmatrix} \underline{v} \\ \underline{\beta} \end{bmatrix} \tag{4.64}$$

we see that the quantity to be minimized is:

$$\left\| \begin{matrix} U'AU - D & , & U'AX \\ X'AU & , & X'AX \end{matrix} \right\|^2 \tag{4.65}$$

There is an arbitrariness in the choice of norm. We will show in 4.3.2. and 4.3.5. that if we use the usual Euclidean norm and minimize over the classes $Q(U,I,\underline{h})$ (4.20) and $Q(U,\underline{h})$ (4.6), then the resulting MINQE,s have the same form as the LMVUE,s of 4.2. but having minimum variance at points $\underline{\theta} = \alpha\underline{1}$, for $\alpha > 0$, if the model had been normal.

We may, however, have some prior information about the parameters in the madel, for example, a prior value $\underline{s} \neq \alpha\underline{1}$, for $\underline{\theta}$ and a prior value $\underline{b}$ and precision K for $\underline{\beta}$, (K^{-1} is a prior covariance matrix for $\underline{\beta}$). One way to incorporate this information into the minimum norm method and to produce estimators with good properties at points $\alpha\underline{s} \in \Omega_{\underline{\theta}}$, is to consider the weighted Euclidean norm:

$$\left\| \begin{bmatrix} D(\underline{s})^{\frac{1}{2}} & 0 \\ 0 & K^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} U'AU - D & , & U'AX \\ X'AU & , & X'AX \end{bmatrix} \begin{bmatrix} D(\underline{s})^{\frac{1}{2}} & 0 \\ 0 & K^{\frac{1}{2}} \end{bmatrix} \right\|^2 \tag{4.66}$$

where: $D(\underline{s}) = \text{diag}(\, s_i I_{q_i}, \quad i = 1,2...q)$.

For normal models this norm produces the locally minimum variance estimators at points $\underline{\theta} = \alpha\underline{s}$ in $\Omega_{\underline{\theta}}$.

Use of the weighted norm (4.66) is equivalent to using the unweighted norm on a transformed model:

$$\underline{y}^* = X^*\underline{\beta} + U^*_1\underline{v}^*_1 + \ldots + U^*_q\underline{v}^*_q = X^*\underline{\beta} + U^*\underline{v}^* \tag{4.67}$$

where $\underline{y}^* = \underline{y} - X\underline{b}$, $X^* = XK^{\frac{1}{2}}$, $\beta^* = K^{-\frac{1}{2}}(\underline{\beta} - \underline{b})$,

$U_i^* = \sqrt{s_i}U_i$ and $\underline{v}_i^* = (1/\sqrt{s_i})\underline{v}_i$

The natural estimator of $\gamma = \underline{h}'\underline{\theta}$ becomes

$$\hat{\gamma} = \underline{h}'\underline{\theta} = \Sigma(h_i s_i/q_i)\ \underline{v}_i^{*\prime}\underline{v}_i^* = \underline{v}_i^{*\prime}D\underline{v}_i^* \qquad (4.68)$$

where $D = \text{diag}((h_i s_i/q_i)I_{qi},\ i=1,2\ldots q)$ (4.69)

Then $\underline{y} = \underline{y}^* + X\underline{b} = X^*\underline{\beta} + U^*\underline{v}^*$ and the quantity to be minimized is:

$$\left\| \begin{matrix} U^{*\prime}AU^* - D, & U^{*\prime}AX^* \\ X^{*\prime}AU^* , & X^{*\prime}AX^* \end{matrix} \right\|^2$$

$$= \text{tr}((U^{*\prime}AU^* - D)(U^{*\prime}AU^* - D)) + 2\text{tr}(U^{*\prime}AXKX^{\prime}AU^*) + \text{tr}(X^{\prime}AXKX^{\prime}AXK) \qquad (4.70)$$

which is identical to (4.66).

The obvious estimator in this case reflects the model structure. The MINQE technique is therefore able to utilize the structure where other methods ignore it.

In the general model (2.1), the choice of natural estimator is not as obvious as for the structured model. However, given prior information $\underline{s}$, $\underline{b}$ and K with the transformed model:

$$\underline{y}^* = X^*\underline{\beta}^* + \Lambda(\underline{s})\ \underline{e}^* \qquad (4.71)$$

where

$\underline{y}^*$, X^* and $\underline{\beta}^*$ are as in (4.67),

$\Lambda(\underline{s})$ is a p.d. square root of $V(\underline{s})$

so that $\Lambda(\underline{s})\ \Lambda(\underline{s}) = V(\underline{s})$

and $\underline{e}^* = \Lambda(\underline{s})^{-1}\underline{e}$

If $\underline{e}^*$ were observable and normal, it would have model:

$$\underline{e}^* \sim N(\theta, V^*(\underline{\theta})) \tag{4/72}$$

where $V^*(\underline{\theta}) = \Sigma\theta_i V_i^*$

and $V_i^* = \Lambda(\underline{s})^{-1} V_i \Lambda(\underline{s})^{-1}$

Hence from (4.33) the LMVUE of $\hat{\gamma} = \underline{h}'\underline{\theta}$ given $\underline{s}$ would be;

$$\hat{\gamma} = \underline{e}^{*\prime} T \underline{e}^* \tag{4.73}$$

where $T = \Sigma\lambda_i^* V_i^*$

$\underline{\lambda}^*$ is any solution to $H(\underline{s})\ \underline{\lambda}^* = \underline{h}$

and $H(\underline{s}) = (tr(V^*(\underline{s})^{-1} V_i^* V^*(\underline{s})^{-1} V_j^*))$

$= (tr(V(\underline{s})^{-1} V_i\ V(\underline{s})^{-1} V_j))$

The simplifications over (4.33) being due to the centrality and the fact that $V^*(\underline{s}) = I$.

Similarly to (4.65), we see that the quantity to be minimized is:

$$\left\| \begin{matrix} \Lambda(\underline{s})\ A\ \Lambda(\underline{s}), & \Lambda(\underline{s})\ \Lambda X^* \\ X^{*\prime}\ A\ \Lambda(\underline{s}), & X^{*\prime}\ A\ X^* \end{matrix} \right\|^2$$

$$= tr((\Lambda(\underline{s})A\ \Lambda(\underline{s}) - T)(\Lambda(\underline{s})A\Lambda(\underline{s}) - T)$$

$$+ 2\ tr(\Lambda(\underline{s})A\ X\ K\ X'\Lambda\Lambda(\underline{s})) + tr(X'A\ X\ K\ X'A\ X\ K) \tag{4.74}$$

4.3.2. Minimum Norm in Quadratic, Unbiased and Invariant Estimators - MINQE(U,I).

In the class $Q(U,I,\underline{h})$ (4.20), the conditions:

$$AX = 0 \quad \text{and} \quad tr(AV_i) = h_i, \; i=1,2...q \tag{4.75}$$

simplify the quantity (4.70) or (4.74) to be minimized. In particular, for the general model, (4.74) becomes:

$$tr(V(\underline{s})AV(\underline{s})A) - 2tr(A\Lambda(\underline{s})T\Lambda(\underline{s})) + tr(TT)$$

which is independent of K, and from the definition of T, (4.73), we obtain:

$$\Lambda(\underline{s})T\Lambda(\underline{s}) = \Sigma\lambda_i^* V_i$$

and therefore the second term becomes: $\Sigma\lambda_i^* tr(AV_i) = \underline{h}'\underline{\lambda}^*$ from (4.75), which is independent of A. Hence the quantity to be minimized is:

$$tr(V(\underline{s})AV(\underline{s})A) \tag{4.76}$$

subject to A satisfying (4.75).

Given $\underline{s}$ and $\underline{h}$ such that $Q(U,I,\underline{h})$ is non-empty, we know from (4.22) that there exists $\underline{\lambda}$ satisfying:

$$H_{U,I}(\underline{s})\underline{\lambda} = \underline{h}$$

where $H_{U,I}(\underline{s}) = (tr(N(\underline{s})V_i N(\underline{s})V_j))$.

Under these conditions the optimal choice of A which minimizes (4.76) is

$$A^* = \Sigma\lambda_i N(\underline{s})V_i N(\underline{s}) \tag{4.77}$$

This follows because $\underline{y}'A^*\underline{y}$ is clearly in $Q(U,I,\underline{h})$ and if A is any other matrix satisfying (4.75), then $A = A^* + B$ for some B satisfyig:

$$B \neq 0, \quad B = B', \quad tr(BV_i) = 0, \text{ for } i=1,2...q$$

and $BX = 0$ so that $(I - P(\underline{s}))B = (I - P(\underline{s})')B = B$.

Hence, $\quad tr(V(\underline{s})AV(\underline{s})A) = tr(V(\underline{s})A^*V(\underline{s})A^*) + 2tr(A^*V(\underline{s})BV(\underline{s}))$

$$+ tr(BV(\underline{s})BV(\underline{s})) \qquad (4.78)$$

but $\quad tr(A^*V(\underline{s})BV(\underline{s})) = \Sigma\lambda_i tr(N(\underline{s})V_i N(\underline{s})V(\underline{s})BV(\underline{s}))$

$$= \Sigma\lambda_i tr(BV_i) = 0 \qquad (4.79)$$

because $\quad N(\underline{s})V(\underline{s}) = (I - P(\underline{s})')$.

Since the last term of (4.78) is non negative we have that:

$$tr(V(\underline{s})AV(\underline{s})A) \geq tr(V(\underline{s})A^*V(\underline{s})A^*) \qquad (4.80)$$

and hence, MINQE(U,I) given $\underline{s}$ of $\gamma = \underline{h}'\theta$ is:

$$\hat{\gamma} = \underline{y}'A^*\underline{y} = \Sigma\lambda_i \underline{y}'N(\underline{s})V_i N(\underline{s}) = \underline{\lambda}'\underline{r}_{U,I}(\underline{s}) \qquad (4.81)$$

where: $\underline{\lambda}$ is any solution to $H_{U,I}(\underline{s})\underline{\lambda} = \underline{h}$

$$H_{U,I}(\underline{s}) = (tr(N(\underline{s})V_i N(\underline{s})V_j))$$

and $\quad \underline{r}_{U,I}(\underline{s}) = (\underline{y}'N(\underline{s})V_i N(\underline{s})\underline{y}, \quad i=1,2...q)'$

One common difficulty with MINQE(U,I) is that cases occur where $Q(U,I,\underline{h})$ is empty and then the estimator is not available.

We note that MINQE(U,I) is independent of the model structure. If we had chosen to minimize (4.70) for the structured model, we would choose A^* to minimize:

$$tr(U^*'AU^*U^*'AU^*) = tr(V(\underline{s})AV(\underline{s})A) \qquad (4.82)$$

since $\quad U^*U^*' = \Sigma s_i V_i = V(\underline{s})$.

MINQE(U,I) is also independent of the fixed effect parameters β, as expected for invariant estimators.

Under normality MINQE(U,I) reduces to the LMVIUE (4.54), and as for that estimator we can write (4.81) in the form: $\hat{\gamma} = \underline{h}'\hat{\underline{\theta}}$ where $\hat{\underline{\theta}}$ satisfies:

$$H_{U,I}(\underline{s})\hat{\underline{\theta}} = \underline{r}_{U,I}(\underline{s}) \qquad (4.83)$$

Putting $\hat{\underline{\theta}}_0 = \underline{s}$ we can define an iterative procedure for eliminating the influence of the starting value $\underline{s}$:

$$H_{U,I}(\hat{\underline{\theta}}_{i-1})\hat{\underline{\theta}}_i = \underline{r}_{U,I}(\hat{\underline{\theta}}_{i-1}) \qquad (4.84)$$

If the procedure converges to a solution $\hat{\underline{\theta}} \in \Omega_{\underline{\theta}}$ then $\hat{\underline{\theta}}$ is called iterated MINQE, IMINQE(U,I). Under normality IMINQE(U,I) is also IMVIUE (4.59), and we will see (5.37) that these equations (4.84) are also the marginal maximum liklihood estimators.

Variants of MINQE(U,I) given $\underline{s}$ exist for cases where no prior knowledge of $\underline{s}$ is available. One of these variants is MINQE0(U,I) for models with a residual error, obtained by taking $\underline{s}$ to have zero components, except for the one corresponding to the residual effect, which is taken as unity. (4.85) This estimator has the advantage that it is relatively easy to compute (Goodnight 1978a).

Another variant MINQE1(U,I) is obtained by setting all the elements of $\underline{s}$ equal to $\underline{1}$.

4.3.3. The W-Transformation.

Computing elements of the equation systems (4.83) and (4.84), proves to be a formidable task for general models, since it requires $V(\underline{s})^{-1}$, the inverse of an nxn matrix, where n will typically be large. Even if this was possible, elements such as $tr(N(\underline{s})V_iN(\underline{s})V_j)$ and $\underline{y}'N(\underline{s})V_iN(\underline{s})\underline{y}$ are expensive to compute.

One suggested solution is to restrict $\underline{s}$, so as to make $V(\underline{s})^{-1}$ easy to calculate, MINQE0 of (4.85) is one such solution.

However, the computing problems are considerably eased, even for general $\underline{s}$ in structured variance component models with residual errors (2.23).

$$\underline{y} = X\underline{\beta} + U_1\underline{V}_1 + \ldots + U_{q-1}\underline{V}_{q-1} + \underline{e}$$
$$= X\underline{\beta} + U\underline{V} + \underline{e} \qquad (4.86)$$

so that

$$\mathrm{Cov}(\underline{y}) = V(\underline{\theta}) = UD(\underline{\theta}^*)U' + \theta_q I$$

where

$$\underline{\theta}^* = (\theta_1 \ldots \theta_{q-1})', \quad \theta_i \geq 0, \ \theta_q > 0$$

and

$$D(\underline{\theta}^*) = \mathrm{Diag}(\theta_1 I_{qi}, \ i=1,2\ldots q-1)$$

Writing $\underline{\gamma} + (\theta_i/\theta_q \ i=1,2\ldots q-1)'$ we have:

$$V(\underline{\theta}) = \theta_q H(\underline{\gamma})$$

where $H(\underline{\gamma})$ is not related to $H(\underline{s})$ of (2.3) but,

$$H(\underline{\gamma}) = UD(\underline{\gamma})U' + I \qquad (4.87)$$

It is easy to show (A.5a) that,

$$H^{-1}(\underline{\gamma}) = I_n - UD(\underline{\gamma})(I_m + U'UD(\underline{\gamma}))^{-1}U' \qquad (4.88)$$
$$= I_n - U(D(\underline{\gamma})^{-1} + U'U)^{-1}U'$$

The first form does not require $\gamma_i > 0$ for all i and hence is preferable. Also, since $\theta_q > 0$

$$V(\underline{\theta})^{-1} = (1/\theta_q)H(\underline{\gamma})^{-1} .$$

In this situation, we only require the inverse of $I_m + U'UD$, which is mxm when $m = \sum_{i=1}^{q-1} q_i$ and m is typically much smaller than n.

Furthermore, given $\underline{s}$ such that $V(\underline{s})$ is p.d., and the model (4.86), we let

$$\underline{c} = (s_i/s_q,\ i=1,2\ldots q-1)$$

and

$$W_0 = \begin{bmatrix} U'U & U'X & U'\underline{y} \\ X'U & X'X & X'\underline{y} \\ \underline{y}'U & \underline{y}'X & \underline{y}'\underline{y} \end{bmatrix} = \begin{bmatrix} W_{0,1}, W_{0,2}, W_{0,3} \end{bmatrix} \qquad (4.89)$$

$\underline{c}$ and W_0 are easy to compute and writing H for $H(\underline{c})$ let:

$$W_1 = \begin{bmatrix} U'\bar{H}^{-1}U & U'\bar{H}^{-1}X & U'\bar{H}^{-1}\underline{y} \\ X'\bar{H}^{-1}U & X'\bar{H}^{-1}X & X'\bar{H}^{-1}\underline{y} \\ \underline{y}'\bar{H}^{-1}U & \underline{y}'\bar{H}^{-1}X & \underline{y}'\bar{H}^{-1}\underline{y} \end{bmatrix} \qquad (4.90)$$

Then, by substituting (4.88) into (4.90)

$$W_1 = W_0 - W_{0,1}D(I+U'UD)^{-1}W'_{0,1} \qquad (4.91)$$

where

D is $D(\underline{c}) = \text{diag}(c_i I_{qi},\ i=1,2\ldots q-1)$.

W_1 is easy to compute and provides all the elements required to form $H_{U,I}(\underline{s})$ and $\underline{r}_{U,I}(\underline{s})$ as will be seen in 4.3.4.

This method of forming W_1 is known as the W-Transformation and was suggested by Hemmerle and Hartley (1973). Schemes for computing W_1 from W_0 and D were developed by Thompson (1975), Hemmerle and Lorens (1976), Liu and Senturia (1977), Hemmerle and Downs (1978) and Goodnight and Hemmerle (1979). The latest methods rely on the observation that if a Cholesky or Sweep Operator (Goodnight 1978b) is applied to the m columns of $U'U + \bar{D}^{-1}$ in:

$$\begin{bmatrix} U'U + \bar{D}^{-1} & W_{0,1} \\ W_{0,1}' & W_0 \end{bmatrix} \qquad (4.92)$$

The resulting matrix is:

$$\begin{bmatrix} I & (U'U + \bar{D}^{-1})^{-1} W_{0,1} \\ 0 & W_1 \end{bmatrix} \qquad (4.93)$$

Furthermore, since we are only interested in W_1 and since elements of $U'U+\bar{D}^{-1}$ are easy to obtain from W_0 and $\underline{c}$, we need never form the matrix (4.92), to achieve the transformation to (4.93).

This algorithm can be programmed to take a given $\underline{c}$ and produce the lower triangular part of W_1 in place of that of W_0. The programme can use the Sweep Operator of Goodnight (1978b)and is unaffected by small or zero elements in $\underline{c}$ provided,in place of $(U'U + \bar{D}^{-1})^{-1}$, the algorithm uses $D(I + U'UD)^{-1}$ as in (4.91).

4.3.4. Computing MINQE(U,I) Estimates.

With structured models having a residual error (4.86), we have the W-Transformation of 4.3.3 to produce the matrix W_1 (4,90) from $\underline{c}$ and W_0 (4.89), which is derived from the model structure and the given prior value $\underline{s}$ for $\underline{\theta}$.

Applying the Sweep Operator (Goodnight 1978b)to the columns of W_1, containing $X'\bar{H}^{-1}X$, we get:

$$W_2 = \begin{bmatrix} U'MU & 0 & U'M\underline{y} \\ RX'\bar{H}^{-1}U & I & RX'\bar{H}^{-1}\underline{y} \\ \underline{y}'MU & 0 & \underline{y}'M\underline{y} \end{bmatrix} \qquad (4.94)$$

where

$$M = M(\underline{c}) = H(\underline{c})^{-1}(I - P(\underline{c})) ,$$

$$P(\underline{c}) = R(\underline{c})X'H(\underline{c})^{-1} ,$$

and $$R(\underline{c}) = (X'H(\underline{c})^{-1}X)^{-}$$

$(I - P(\underline{c}))$ is idempotent and $\text{rank}(I - P(\underline{c})) = n - \text{rank}(X)$.

The lower triangular part of W_2 can be formed in place, from W_1. The Sweep Operator of Goodnight (1978b) can be programmed to leave $R(\underline{c})$ instead of I in the central position of W_2 (4.95).

From W_2 we notice that the following quantities are easily calculated. If we consider U partitioned according to the random effect structure matrices;

$$U = (U_1, U_2 \ldots U_{q-1}).$$

Then let $$a_i = \text{tr}(U_i'MU_i) \quad i=1,2\ldots q-1 \tag{4.96}$$

be the trace of the i th block of $U'MU$.

Let $$b_{ij} = \text{tr}(U_j'MU_iU_i'MU_j) \quad i,j=1,2\ldots q-1 \tag{4.97}$$

be the SS of all the elements in the i,j th block of $U'MU$.

(Note: $b_{ij} = b_{ij}$)

and let

$$d_i = \underline{y}'MU_iU_i'M\underline{y} \quad i=1,2\ldots q-1 \tag{4.98}$$

be the SS of elements in the i th block of $U'M\underline{y}$.

These quantities can be computed from W_1, (4.90)

and, together with $\underline{y}'M\,\underline{y}$, are all that are required to form the MINQE(U,I) equations. These equations can be formed and solved to provide the estimates.

In models like (4.80), we have

$$V_i = U_iU_i' \quad i=1,2\ldots q-1, \quad V_q=I$$

and $$N(\underline{s}) = (1/s_q)M(\underline{c}) = (1/s_q)M.$$

Hence, elements of $s_q^2 H_{U,I}(\underline{s})$ are:

$$s_q^2 \operatorname{tr}(N(\underline{s})V_iN(\underline{s})V_j) = \operatorname{tr}(U_j'MU_iU_i'MU_j)$$

$$= b_{ij} \quad i,j=1,2\ldots q-1 \qquad (4.99)$$

For $j=q$, the i,q th element for $i=1,2\ldots q-1$ is:

$$s_q^2\operatorname{tr}(N(\underline{s})V_iN(\underline{s})) = \operatorname{tr}(M\,V_i\,M)$$

$$= \operatorname{tr}(U_i'MMU_i) \qquad (4.100)$$

$$= \operatorname{tr}(U_i'MU_i) - \sum_{j=1}^{q-1} c_j \operatorname{tr}(U_i'MU_jU_j'MU_i) \quad \text{(A.5d)}.$$

$$= a_i - \Sigma c_j b_{ij} \qquad (4.101)$$

Finally, the q,q th element is:

$$s_q^2 \operatorname{tr}(N(\underline{s})N(\underline{s})) = \operatorname{tr}(MM)$$

$$= n - \operatorname{rank}(X) - 2\Sigma c_i \operatorname{tr}(U_i'MU_i) + \sum\sum c_ic_j\operatorname{tr}(U_i'MU_jU_j'MU_i)$$

$$\text{(A.5e)} \qquad (4.102)$$

$$= n - p - 2\Sigma c_i a_i + \sum\sum c_ic_jb_{ij}$$

For elements of $s_q^2 \underline{r}_{U,I}(\underline{s})$,

$$s_q^2 \underline{y}'N(\underline{s})V_iN(\underline{s})\underline{y} = \underline{y}'MU_iV_i'M\underline{y} = d_i \text{ of (4.98)} \quad (4.103)$$

$$i=1,2\ldots q-1$$

and for the q th element,

$$s_q^2 \underline{y}'N(\underline{s})N(\underline{s})\underline{y} = \underline{y}'MM\underline{y} \quad (4.104)$$

$$= \underline{y}'M\underline{y} - \Sigma c_i \underline{y}'MU_iU_i'M\underline{y} = \underline{y}'M\underline{y} + \Sigma c_i d_i \quad (A.5f)$$

Note that if $\underline{s}$ were the correct value of $\underline{\theta}$, the generalized least squares estimator of the fixed effect parameters is:

$$\hat{\beta}(\underline{s}) = (X'V(\underline{s})^{-1}X)^{-}X'V(\underline{s})^{-1}\underline{y}$$

$$= s_qR(\underline{c})X'(1/s_q)H(\underline{c})^{-1}\underline{y}$$

$$= RX'\bar{H}^{-1}\underline{y} \quad (4.105)$$

which is available in W_2 of (4.94).

Also the $\text{Cov}(\underline{\beta}) = (X'V(\underline{s})X)^{-} = s_qR(\underline{c})$ (4.106)

is available in W_2 in view of (4.95).

However $\underline{s}$ is only a prior value for the unknown $\underline{\theta}$. We expect a good estimate of $\underline{\beta}$ to be $\hat{\underline{\beta}}(\hat{\underline{\theta}})$ where $\hat{\underline{\theta}}$ is the solution to the MINQE(U,I) equations (4.83). This value, together with an estimate of $\text{Cov}(\hat{\underline{\beta}}(\hat{\underline{\theta}}))$ becomes available in W_2 if we start the algorithm with $\hat{\underline{\theta}}$ as prior value.

These considerations also indicate how the computing method can be used to produce IMINQE(U,I) estimates (4.84). This algorithm can be used to compute these values, when convergence of (4.84) occurs.

4.3.5. MINQE without Invariance.

For many models it is easy to find cases where $Q(U,I,\underline{h})$ (4.20) is empty. For example, in model (4.23), with $\underline{h} \neq (a,a)$ for $a \in \mathbb{R}$, $Q(U,I,\underline{h})$ is empty. In these situations, conditions of invariance or unbiasedness must be abandoned. In either case, the resulting estimators depend on the fixed effect parameters and prior information about both $\underline{\theta}$ and $\underline{\beta}$ is required. The estimators will have minimum norm properties dependant on the prior information.

For $\underline{\theta}$ a prior value $\underline{s}$, such that $V(\underline{s})$ is p.d., is required, and for $\underline{\beta}$ a prior mean $\underline{b}$ and p.d. dispersion K^{-1} can be incorporated. (4.107)

The mean is used to transform the model according to $\underline{y} - X\underline{b}$. The dispersion is used via its inverse K, which is a precision matrix and is used with the norm (4.70) or (4.74), to weight the contribution of the fixed effect parameters.

If A satisfies the conditions for unbiasedness:

$$X'AX = 0, \quad \mathrm{tr}(AV_i) = h_i \tag{4.108}$$

then the squared norm (4.74) becomes:

$$\mathrm{tr}((\Lambda(\underline{s})A\Lambda(\underline{s}) - T)(\Lambda(\underline{s})A\Lambda(\underline{s}) - T) + 2\mathrm{tr}(\Lambda(\underline{s})AXKX'A\Lambda(\underline{s})) \tag{4.109}$$

Terms involving T in the expansion of (4.109) do not depend on A, as at (4.75) - (4.76), because of the unbiasedness conditions (4.108). Also, because $X'AX = 0$ we can write the terms of (4.109), which do not contain T as:

$$\text{tr}((V(\underline{s}) + XKX')\ A\ (V(\underline{s}) + XKX')A)$$

$$= \text{tr}(RARA) \qquad (4.110)$$

where

$$R = (V(\underline{s}) + XKX')$$

R is p.d. because $V(\underline{s})$ is p.d. and XKX' n.n.d. Hence the quantity to be minimized over A satisfying (4.108) is (4.110).

Pringle (1974) gives an explicit expression for A*, the A matrix which minimizes (4.110), provided $Q(U,\underline{h})$ is non-empty:

$$A^* = \Sigma\lambda_i\ R^{-1}(V_i - P(R)V_iP'(R))\ R^{-1} \qquad (4.111)$$

where

$$P(R) = X(X'R^{-1}X)^{-}X'R^{-1}, \quad R=(V(\underline{s}) + XKX')$$

$\underline{\lambda}$ is any solution of $H_U(R)\underline{\lambda} = \underline{h}$

$$H_U(R) = \text{tr}(R^{-1}(V_i - P(R)V_iP(R'))R^{-1}V_j))$$

which has the same column space as

H_U, H_U^* and $H_U(\underline{s})$ of (4.7 to 4.12).

Clearly $\underline{y}'A^*\underline{y} \in Q(U,\underline{h})$

because $\text{tr}(A^*V_i) = h_i$ (since $H_U(R)\underline{\lambda} = \underline{h}$)

and $X'A^*X = 0$ (since $X'R^{-1}P(R) = X'R^{-1}$

and $P(R\ R^{-1}X = R^{-1}X$.)

If A is any other symmetric matrix satisfying (4.108), then

$A = A^* + B$ with $B \neq 0$ and

$$\text{tr}(BV_i) = \text{tr}(AV_i) - \text{tr}(A^*V_i) = 0$$

and $X'BX = X'A\ X - X'A^*X = 0$

Therefore (4.110) can be written as:

$$tr(RARA) = tr(RA^*RA^*) + 2tr(RA^*RB) + tr(RBRB)$$

but

$$tr(RA^*RB) = \Sigma\lambda_i((V_i - P(R)V_iP(R)')B)$$

$$= \Sigma\lambda_i(tr(V_iB) - tr(V_iP(R)'BP(R))) = 0$$

Also $tr(RBRB) > 0$ since R is p.d. and therefore A^* is optimal.

MINQE(U) given $\underline{s}$, $\underline{b}$ and K is independent of the model structure, (4.67) or (4.71), and of the natural estimator, T or D, but not of the prior precision K in the norm (4.70) or (4.74). K weights the contribution of prior knowledge about $\underline{\beta}$ relative to $\underline{\theta}$ in the minimum norm criterion.

MINQE(U) given $\underline{s}$, $\underline{b}$ and K of $\gamma = \underline{h}'\underline{\theta}$ is:

$$\hat{\gamma} = \Sigma\lambda_i(\underline{y}-X\underline{b})' R^{-1}(V_i-P(R)V_iP(R)')R^{-1}(\underline{y}-X\underline{b})$$

$$= (\underline{y}-X\underline{b})' A^*(\underline{y}-X\underline{b}) \qquad (4.112)$$

where

λ is a solution to $H_U(R)\underline{\lambda} = \underline{h}$,

$R = (V(\underline{s}) + XKX')$

and $H_U(R) = (tr(R^{-1}(V_i - P(R)V_iP(R)')R^{-1}V_j)$.

If $K = 0$, then $R=V(\underline{s})$ and MINQE(U) coincides with LMVUE (4.33) for normal models.

We can write $\hat{\gamma}$ as $\underline{h}'\hat{\underline{\theta}}$ where $\hat{\underline{\theta}}$ is any solution to:

$$H_U(R)\hat{\underline{\theta}} = \underline{r}_U(R) \qquad (4.113)$$

where $\underline{r}_U(R) = (\underline{y} -X\underline{b})'R^{-1}(V_i - P(R)V_iP'(R))R^{-1}(\underline{y} -X\underline{b})$

$i=1,2...q)'$ (4.114)

Starting with $\hat{\underline{\beta}}_0 = \underline{b}$ and $\hat{\underline{\theta}}_0 = \underline{s}$ we can establish an iterative process using (4.113) and

$$\hat{\underline{\beta}}_{i+1} = (X' \, V \, (\underline{\theta}_i)^{-1} X)^{-} X' \, V \, (\underline{\theta}_i)^{-1} \underline{y} \qquad (4.115)$$

If this converges to a value in the parameter space, it is called the iterated MINQE(U), IMINQE(U).

Variants of MINQE(U) are derived by taking K to be $r^2 I$ or equivalently assuming $\underline{\beta}$ to have prior distribution with mean $\underline{b}$ and covariance $(1/r^2)I$. The resulting estimator (4.111) is called r-MINQE(U).

In an attempt to eliminate the influence of the fixed effects on the estimator, ∞-MINQE(U) was defined, (Kleffe 1977b), as the limiting r-MINQE(U), as r tends to infinity when this limit exists. Rao, C.R. and Kleffe (1980), report that ∞-MINQE(U) coincides with MINQE(U,I), (4.81), when the latter exists. (4.116)

4.3.6. MINQE without Unbiasedness.

If we restrict attention to quadratic estimating functions, but do not require unbiasedness for $\gamma = \underline{h}'\underline{\theta}$, we still have three estimator classes of interest, (4.2), (4.15) and (4.18):

$$Q = \{\underline{y}'A\,\underline{y}: \; A \text{ symmetric } (A = A')\}$$

$$Q(PU) = \{\underline{y}'A\,\underline{y}: \; A = A' \text{ and } X'AX = 0\}$$

and $$Q(I) = \{\underline{y}'A\,\underline{y}: \; A = A' \text{ and } AX = 0\}$$

The optimal solution to the minimum norm criterion over these classes, does however, depend on the choice of model structure and the natural estimator, because we no longer have the condition $\mathrm{tr}(AV_i) = h_i, i = 1, 2..q,$

which made terms in the expansion of the norm involving the natural estimators T or D become independent of A.

In effect, without unbiasedness MINQE takes into account the structure of the error term in the model (2.1-6), through the use of this structure in deriving the natural estimator, (4.68). We have, therefore, two cases to consider; One with the structured model (4.67) and natural estimator D of (4.68) and one with general model (4.71) and natural estimator T of (4.73).

However, noting that U* is a square root of $V(\underline{s})$ in the sense that $V(\underline{s}) = U^*U^{*\prime}$ in the same way that $V(\underline{s}) = \Lambda(\underline{s})\Lambda(\underline{s})$, we can write C for U* or $\Lambda(\underline{s})$ and E for D or T in (4.70) and (4.74). Then the norm to be minimized is:

$$\left\| \begin{matrix} C'AC - E, & C'AX \\ X'AC & X'AX \end{matrix} \right\|_K^2$$

$$= \mathrm{tr}(V(\underline{s})AV(\underline{s})A) - \mathrm{tr}(C'ACE) - \mathrm{tr}(EC'AC)$$

$$+ \mathrm{tr}(EE) + \mathrm{tr}(V(\underline{s})AXKX'A) + \mathrm{tr}(X'AV(\underline{s})XK)$$

$$+ \mathrm{tr}(X'AXKX'AXK) \qquad (4.117)$$

This is minimized at A* in the unrestricted class of symmetric matrices if and only if the cross product terms in the expansion, (4.117) vanish for all symmetric B.

A*, which satisfies this condition is given by:

$$A^* = (V(\underline{s}) + XKX')^{-1}CEC'(V(\underline{s}) + XKX')^{-1} \qquad (4.118)$$

We note that:

$$\{A: A = A' \text{ and } X'AX = 0\} = \{B - P(\underline{s})'BP(\underline{s}): B = B'\} \qquad (4.119)$$

and

$$\{A: A = A', AX = 0\} = \{N(\underline{s})BN(\underline{s}): B = B'\} \qquad (4.120)$$

(Kleffe and Pincus(1974b) Lemma 2.4 and (2.15))

where

$$P(\underline{s}) = X(X'V(\underline{s})^{-1}X)^{-}X'V(\underline{s})^{-1}$$

and $$N(\underline{s}) = V(\underline{s})^{-1}(I - P(\underline{s})).$$

Hence, by substituting $A^* + B - P(\underline{s})'BP(\underline{s})$ or $A^* + N(\underline{s})BN(\underline{s})$ for A in the expansion (4.117) and equating the cross product terms to zero, we can obtain expressions for the optimal A^* in the classes $Q(PU)$ and $Q(I)$. These are:

$$A^*(PU) = (V(\underline{s}) + XKX')^{-1}(CEC' - P(\underline{s})CEC'P(\underline{s})'V(\underline{s}) + XKX')^{-1} \quad (4.121)$$

and

$$A^*(I) = N(\underline{s})CEC'N(\underline{s}) \quad (4.122)$$

Then MINQE of $\underline{h}'\underline{\theta}$ given $\underline{s}$, $\underline{b}$ and K in the class Q is:

$$(\underline{y} - X\underline{b})' A^*(\underline{y} - X\underline{b}) \text{ for } A^* \text{ of (4.118)}, \quad (4.123)$$

MINQE(PU) of $\underline{h}'\underline{\theta}$ given $\underline{s}$, $\underline{b}$ and K in the class $Q(PU)$ is:

$$(\underline{y} - X\underline{b})' A^*(PU)(\underline{y} - X\underline{b}) \quad (4.124)$$

and MINQE(I) of $\underline{h}'\underline{\theta}$ given $\underline{s}$, in the class $Q(I)$ is:

$$\underline{y}' A^*(I)\underline{y} \quad (4.125)$$

which is independent of $\underline{b}$ and K.

In the structured model (4.67), with natural estimator $E = D$ of (4.68), we note that;

$$CEC' = U^*DU^{*\prime} = \Sigma d_i V_i \quad (4.126)$$

where $d_i = s_i^2 h_i / q_i$

In the general model, (4.71) with natural estimator T of (4.73),

$$CEC_{\underline{s}}' = \Lambda(\underline{s})T\,\Lambda(\underline{s}) = \Sigma\lambda_i^* V_i \qquad (4.127)$$

where $\underline{\lambda}^*$ is any solution to $H(\underline{s})\underline{\lambda}^* = \underline{h}$

and $H(\underline{s}) = (tr(V^{-1}(\underline{s})V_i V(\underline{s})^{-1}V_j))$

Hence MINQE without unbiasedness incorporates the knowledge of structure of the error term in the model through the choice of natural estimator.

From (4.125) and (4.126) we can obtain an expression for the MINQE(I) of $\underline{\theta}$ in the structured model (2.16) as:

$$\hat{\underline{\theta}} = diag(s_i^2/q_i,\ i=1,2\ldots q)\ \underline{r}_{U,I}(\underline{s}) \qquad (4.128)$$

where

$$\underline{r}_{U,I}(\underline{s}) = (\underline{y}'N(\underline{s})V_i N(\underline{s})\underline{y},\ i=1,2\ldots q)'$$

For the general model (2.1), using (4.121), we can write the MINQE(I) of $\underline{\theta}$ as:

$$\hat{\underline{\theta}} = H(\underline{s})\hat{\underline{\theta}} = \underline{r}_{U,I}(\underline{s}) \qquad (4.129)$$

Both (4.128) and (4.129) suggest an iterative procedure - IMINQE(I) defined by setting $\underline{s} = \hat{\underline{\theta}}_0$. (4.130)

In section 5.1.2, we see that IMINQE(I) defined (4.129) is the maximum likelihood estimator under normality. The difference between (4.128) and (4.129) illustrates a failure on the part of maximum likelihood, to take account of the structure in structured variance component models.

Rao, P.S.R.S. and Chaubey (1978), considered MINQE without unbiasedness as a means of obtaining non-negative estimates of variance components in heteroscedastic regression models. In the more general setting, developed in Chaubey (1980) and Rao, C.R. (1979), MINQE

given $\underline{s}$, $\underline{b}$ and K and MINQE(I) given $\underline{s}$ are non-negative for non-negative functions $\underline{h}'\underline{\theta}$.

4.4. Minimum Norm Estimates of Variance Components for Example Data Sets

Minimum norm estimates of the variance components were calculated for the example data sets, described in 2.7. Estimates with unbiasedness and invariance properties, were calculated using the algorithm of 4.2.4.

4.4.1. Data Set 1.

As with the estimates obtained by Henderson's Method III, 3.4.1., all variates except LENB had negative estimates of variance components, both for MINQE0(U,I) and MINQE1(U,I). For variable LENB the estimates were;

Table (4.1): Minimum Norm Estimates of Variance Components for Variable LENB of Data Set 1.

COMPONENT	MINQE0(U,I)	MINQE1(U,I)	IMINQE(U,I)
SIRE	35.945	34.805	36.646
SIRE X DBRD	60.067	52.901	55.293
RESIDUAL	756.300	759.226	758.485

The IMINQE(U,I) estimates for variable LENB, converged after four iterations of (4.84), starting with the MINQE1(U,I) estimates. Exactly the same values were obtained with the marginal maximum likelihood method. Table (5.1).

4.4.2. Data Set 2.

For the data generated to follow the model for the cattle breeding experiment, better results were obtained. The minimum norm estimates for the ten variables are given in Table (4.2).

Table (4.2). Minimum Norm Estimates of Variance Components for Data Set 2.

VARIABLE	COMPONENT	MINQEO(U,I)	MINQE1(U,I)	IMINQE(U,I)
	SIRE	13.814	15.384	15.425
1	SIRE X DBRD	2.198	1.931	1.935
	RESIDUAL	8.662	8.222	8.221
	SIRE	4.263	5.337	5.260
2	SIRE X DBRD	2.547	2.849	2.858
	RESIDUAL	9.763	9.278	9.280
	SIRE	7.674	9.250	9.200
3	SIRE X DBRD	3.934	3.322	3.411
	RESIDUAL	9.126	8.802	8.797
	SIRE	9.847	9.908	9.910
4	SIRE X DBRD	3.258	3.015	3.034
	RESIDUAL	8.657	8.731	8.730

Table (4.2) contd.

Table (4.2) contd.

VARIABLE	COMPONENT	MINQEO(U,I)	MINQE1(U,I)	IMINQE(U,I)
	SIRE	5.240	5.162	5.154
5	SIRE X DBRD	4.095	3.941	3.967
	RESIDUAL	8.323	8.405	8.403
	SIRE	7.462	7.995	7.885
6	SIRE X DBRD	0.895	0.795	0.918
	RESIDUAL	9.136	8.982	8.968
	SIRE	8.061	8.864	8.871
7	SIRE X DBRD	5.257	4.533	4.550
	RESIDUAL	7.162	7.154	7.152
	SIRE	28.048	25.780	25.732
8	SIRE X DBRD	2.563	2.805	2.792
	RESIDUAL	7.124	7.940	7.942
	SIRE	4.887	4.212	4.231
9	SIRE X DBRD	3.596	4.039	4.016
	RESIDUAL	7.813	7.876	7.877
	SIRE	19.659	17.004	16.915
10	SIRE X DBRD	1.528	1.302	1.390
	RESIDUAL	6.996	8.000	7.993

The IMINQE(U,I) values agreed with the marginal maximum likelihood estimates, computed with programme BMDP3V (Dixon and Brown (1977)).

4.4.3. Data Set 3.

For the small generated data set of 2.7.3, the minimum norm estimates, like the Henderson Method III estimates, 3.4.3, had numerous negative values. The negative values prevented convergence of the iterative scheme (4.84), for IMINQE(U,I).

Table (4.3). Minimum Norm Estimates of Variance Components for Data Set 3.

VARIABLE	ESTIMATOR	VARIANCE COMPONENT FOR:				
		R1	R2	R1 X F	R1 X R2	RESIDUAL
1	MINQE0(U,I)	62.115	51.487	9.239	-13.473	5.107
	MINQE1(U,I)	43.630	25.083	-4.280	5.874	15.757
2	MINQE0(U,I)	22.151	21.841	-5.119	0.814	13.600
	MINQE1(U,I)	27.086	20.094	-3.880	4.211	9.857
3	MINQE0(U,I)	25.945	-10.052	-8.883	38.655	1.179
	MINQE1(U,I)	17.706	-6.745	-7.160	21.048	12.271
4	MINQE0(U,I)	4.132	-9.867	-13.775	20.156	15.011
	MINQE1(U,I)	3.575	-7.982	-14.498	17.834	16.398
5	MINQE0(U,I)	11.115	14.769	-15.583	21.240	11.687
	MINQE1(U,I)	4.559	16.516	-18.561	19.204	15.815
6	MINQE0(U,I)	86.940	18.571	29.259	-25.721	6.245
	MINQE1(U,I)	58.329	16.622	20.904	-6.757	8.964
7	MINQE0(U,I)	30.049	43.450	14.856	-26.841	11.865
	MINQE1(U,I)	18.709	23.302	13.479	-6.207	11.518

Table (4.3) contd.

Table (4.3) contd.

VARIABLE	ESTIMATOR	VARIANCE COMPONENT FOR:				
		R1	R2	R1 X F	R1 X R2	RESIDUAL
8	MINQEO(U,I)	26.608	14.656	17.725	-20.795	25.653
	MINQE1(U,I)	34.005	3.239	-0.552	1.541	21.114
9	MINQEO(U,I)	84.012	14.212	-5.389	6.368	11.182
	MINQE1(U,I)	59.474	15.506	6.891	11.180	11.328
10	MINQEO(U,I)	21.348	13.262	-0.971	-11.094	28.471
	MINQE1(U,I)	28.182	9.791	14.676	-10.453	20.044

CHAPTER 5

Maximum Likelihood Methods of Variance Component Estimation in Normal Models.

Maximum likelihood estimation of variance components was examined by Hartley and Rao, J.N.K. (1967) and reviewed by Harville (1977). Apart from the appeal of the principle, it produces estimators with desirable properties. In particular, the maximum likelihood estimators (MLEs) are constrained to lie in the parameter space and hence positive parameters can only have positive estimates. Thus, at least conceptually, one major problem of other estimating methods is overcome.

When a theoretically tractable and practically applicable notion of a limit of size for variance component models is available, the MLEs also have the desirable property of sufficiency, consistency and limiting normality. Hartley and Rao, J.N.K. (1967) and Miller (1977), used different notions of the limit to obtain these results. Where these limit definitions are applicable to specific models, the ML approach provides some measure of the variance of the estimators through Fisher's information matrix.

Drawbacks of the ML approach are computational difficulties, the fact that the estimators may be unacceptably biased in some cases and the requirement for assuming normality in the model.

The computational problems stem from two sources. Firstly, maximizing the likelihood is a constrained non-linear optimization problem. The constraints are just that the solution must lie in the parameter space. The very feature of ML that allows conceptual solution of the negative estimators problem now becomes a computing problem. Secondly, no closed form solution seems available, except for balanced models and so iterative techniques are required. Further, evaluation of the likelihood requires the inverse of the covariance matrix. This problem has been overcome to some extent for structured models with residual errors through the development of the W-algorithm, section 4.3.3. Jennrich and Sampson (1976) review the computing methods available and have produced a programme, BMDP3V, which works for a commonly used range of models,(Dixon and Brown (1977)).

The problem of unacceptable bias in MLEs for some models can be viewed as a failure of the technique to take account of uncertainty about the fixed effect parameters. Patterson and Thompson (1971 and 1974) and later, Corbeil and Searle (1976a) suggested maximizing the marginal or integrated likelihood of the variance components, MMLE. This technique seems to perform well in practice, Corbeil and Searle (1976b).

Evidence of robustness of the MLEs and MMLEs to non-normality, comes from the observation that cases of

iterated minimum norm techniques produce the same estimating equations for non-normal models as MLEs and MMLEs do for normal cases.

5.1. Maximum Likelihood Estimation in the General Normal Model.

We assume the response in model (2.1) follows the normal distribution:

$$\underline{y} = X\underline{\beta} = \underline{e},\ \mathrm{Cov}(\underline{y}) = \Sigma\theta_i V_i = V(\underline{\theta}) \qquad (5.1)$$

$$\underline{y} \sim N(X\underline{\beta}, V(\underline{\theta})),\ \underline{\beta} \in \Omega_{\underline{\beta}},\ \underline{\theta} \in \Omega_{\underline{\theta}} \qquad (5.2)$$

We also assume $V(\underline{\theta})$ is p.d. for $\underline{\theta} \in \Omega_{\underline{\theta}}$, but no assumption is made about the sign of θ_i in this model.

The log likelihood function of $\underline{\beta}$ and $\underline{\theta}$ for this model is derived from (2.5) as:

$$L(\underline{\beta},\underline{\theta}/\underline{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(V(\underline{\theta})) - \frac{1}{2}(\underline{y}-X\underline{\beta})'V(\underline{\theta})^{-1}(\underline{y}-X\underline{\beta}) \qquad (5.3)$$

and the MLE of $\underline{\beta}$ and $\underline{\theta}$ is the value $(\hat{\underline{\beta}},\hat{\underline{\theta}})$ such that $\hat{\underline{\beta}} \in \Omega_{\underline{\beta}}$ and $\hat{\underline{\theta}} \in \Omega_{\underline{\theta}}$ and

$$L(\hat{\underline{\beta}},\hat{\underline{\theta}}/\underline{y}) \geq L(\underline{\beta},\underline{\theta}/\underline{y}) \quad \text{for all } (\underline{\beta},\underline{\theta}) \in \Omega_{\underline{\beta}} \times \Omega_{\underline{\theta}} \qquad (5.4)$$

5.1.1. The Maximum Likelihood Equations.

The general procedure for obtaining expressions for the MLEs, $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$ of (5.4), is to equate the first order partial derivatives of the log likelihood (5.2) to zero.

This gives the set of maximum likelihood equations: (from A.6h).

$$X'V(\underline{\theta})^{-1} X\underline{\beta} = X'V(\underline{\theta})^{-1}\underline{y} \qquad (5.5)$$

and $$tr(V(\underline{\theta})^{-1}V_i) = (\underline{y}-X\underline{\beta})'V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}(\underline{y}-X\underline{\beta}) \qquad (5.6)$$

$$i=1,2\ldots q$$

When these equations are all satisfied by values $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$, such that $(\hat{\underline{\beta}},\hat{\underline{\theta}}) \in \Omega_{\underline{\beta}} \times \Omega_{\underline{\theta}}$, then $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$ are the MLEs. Attempts to write (5.5) and (5.6) as explicit expressions for $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$ fail. However, writing $\hat{\underline{\beta}}(\underline{\theta})$ as a solution to (5.5) for given $\underline{\theta}$, we have

$$X\hat{\underline{\beta}}(\underline{\theta}) = X(X'V(\underline{\theta})^{-1}X)^{-}X'V(\underline{\theta})^{-1}\underline{y} = P(\underline{\theta})\underline{y} \qquad (5.7)$$

and substituting this for $X\underline{\beta}$ in (5.6), while noting that

$$tr(V(\underline{\theta})^{-1}V_i) = tr(V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}V(\underline{\theta}))$$

$$=\sum_j \theta_j tr(V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}V_j) \qquad (5.8)$$

we obtain;

$$H(\underline{\theta})\underline{\theta} = \underline{r}_{U,I}(\underline{\theta}) \qquad (5.9)$$

where

$$H(\underline{\theta}) = (tr(V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}V_j)) \qquad (5.10)$$

$$\underline{r}_{U,I}(\underline{\theta}) = (\underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y}, \quad i=1,2\ldots q)' \qquad (5.11)$$

Equations (5.7) and (5.9) suggest an iterative procedure for obtaining MLEs, provided the system converges to values in the parameter space.

5.1.2. Properties of the Maximum Likelihood Equations.

The MLE of $\underline{\theta}$, is invariant to translations of $\underline{y}$ by $X\underline{\beta}$ for any $\underline{\beta} \in \Omega_{\underline{\beta}}$, since from (5.11) it is a function of a maximal invariant $N(\underline{\theta})\underline{y}$ (4.16). In fact (5.9) is identical to equation (4.129), which was the basis for defining IMINQE(I) and hence in normal models IMINQE(I) is MLE.

When $\underline{\theta}$ is identifiable, that is, when $\{V_1 \ldots V_q\}$ are independent, 2.1.1, $H(\underline{\theta})$ of (5.10) has full rank, since $H_I(\underline{\theta}) = TT'$, where T, q x n^2 has i th row $Vec'(V_i V(\underline{\theta})^{-1})$ and $\{V_1 \ldots V_q\}$ independent implies $\{V_1 V(\underline{\theta})^{-1} \ldots V_q V(\underline{\theta})^{-1}\}$ is independent and so rank $(H(\underline{\theta}))$ = rank(T) = q. In this case, the iterative procedure for solving (5.9) can be written as:

$$\hat{\theta}_{i+1} = H(\hat{\underline{\theta}}_i)^{-1} \underline{r}_{U,I}(\hat{\underline{\theta}}_i) \qquad (5.12)$$

One difficulty with the ML equations reported in Rao, C.R. (1979) is that even when $\underline{\theta}$ is identifiable in the model (5.1), it may not be in the model for a maximal invariant. For example, in the model (4.23), V_1 and V_2 are independent, but $NV_1N = NV_2N$. In such cases, the likelihood function does not have a maximum and solutions to (5.12) exhibit functional relationships, which may not hold for the true parameter values.

The MLEs are biased since, if (5.9) holds

$$E(\underline{r}_{U,I}(\underline{\theta})) = H_{U,I}(\underline{\theta})\underline{\theta} \neq H(\underline{\theta})\underline{\theta} \qquad (5.13)$$

where

$$H_{U,I}(\underline{\theta}) = (tr(N(\underline{\theta})V_i N(\underline{\theta})V_j)$$

and $$H(\underline{\theta}) = (tr(V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j))$$

If we define an estimator by equating $\underline{r}_{U,I}(\underline{\theta})$ (5.11) to its expected value after the spirit of the ANOVA estimators, we obtain:

$$\underline{r}_{U,I}(\underline{\theta}) = H_{U,I}(\underline{\theta})\underline{\theta} \qquad (5.14)$$

which is identical to (4.84), the defining equation for IMINQE(U,I) and will be seen to be the marginal maximum likelihood equations.

5.1.3. The Information Matrix for MLEs.

One means of getting an estimate of the precision of the MLEs, is via Fisher's Information Matrix, which is the inverse covariance matrix of the limiting distribution of the MLEs. The notion of limits is difficult to formalize in variance component models and it is not known in what circumstances of model or data size the use of the Information Matrix is justified.

Searle (1970) derives the Information Matrix as:

$$I\begin{pmatrix}\underline{\beta}\\ \underline{\theta}\end{pmatrix} = \begin{bmatrix} X'V(\underline{\theta})^{-1}X & 0 \\ 0 & I(\underline{\theta}) \end{bmatrix} \qquad (5.15)$$

where

$$I(\underline{\theta}) = \frac{1}{2}(tr(V(\underline{\theta})V_i V(\underline{\theta})V_j)) = \frac{1}{2}H(\underline{\theta}) \qquad (5.16)$$

This follows from the definition of the information matrix:- the negative expectation of the second order partial derivatives. The derivatives are given in (A.6h).

5.1.4. Computation of MLEs in Structural Models having a Residual Error.

Since closed form solutions to the ML equations are only available for some balanced models, iterative techniques must be used. These techniques require the inversion of the covariance matrix at each step. The computational burden of this requirement limits the practical application of ML to structural models with residual errors, where inversion via (4.88) and the W-algorithm of section 4.3.3. are available.

Several techniques for solving the equations (5.7) and (5.9) are available. The most direct, is to iterate (5.9) and update (5.7), at each cycle. Miller (1979), shows this method to be related to the Fisher Scoring Algorithm, but reports difficulty in imposing the non-negativity constraints and ensuring that, when the system converges to a feasible solution, it is a maximum of the likelihood function.

Hartley and Rao (1967), Hartley and Vaughn (1972) and Hemmerle and Hartley (1973), proposed schemes, while developing the W-Transformation, which are essentially hybrids

of two classical methods of non-linear optimization: the Newton-Raphson and Fisher Scoring Algorithms. Jennrich and Sampson (1976) and Hemmerle and Downs (1978), compare and develop hybrids of these two methods, which appear to work well on many models. Jennrich and Sampson (1978) discuss the combination of the two basic algorithms into a programme BMDP3V (Dixon(1977)). This programme takes advantage of the scoring algorithm's ability to correct poor initial values, and the Newton-Raphson ability to use the quadratic form of the log likelihood to converge quickly from good approximate values.

We consider computation of the components required for both algorithms. The Structured Model with residual error (2.23) has form:

$$\underline{y} = X\underline{\beta} + U_1\underline{v}_1 + U_2\underline{v}_2 + \ldots + U_{q-1}\underline{v}_{q-1} + \underline{e} \quad (5.17)$$

$$= X\underline{\beta} + U\underline{v} + \underline{e} \quad (5.18)$$

where

$$\begin{aligned} \mathrm{Cov}(\underline{y}) &= V(\underline{\theta}) = \theta_q H(\underline{\gamma}) \\ H(\underline{\gamma}) &= UD(\underline{\gamma})U' + I_n \\ D(\underline{\gamma}) &= \mathrm{diag}(\gamma_i I_{qi},\ i=1,2\ldots q-1) \\ \underline{\gamma} &= (\theta_1/\theta_q, \theta_2/\theta_q, \ldots \theta_{q-1}/\theta_q)' \end{aligned} \quad (5.19)$$

If the values of $\underline{\beta}$ and $\underline{\theta}$ at the start of any iteration are $\underline{b}$ and $\underline{s}$; let $\underline{c} = (s_1/s_q, \ldots s_{q-1}/s_q)'$ be the value for $\underline{\gamma}$. As in the

description of the W-Transformation, 4.3.3, we write H and D interchangeably with $H(\underline{c})$ and $D(\underline{c})$.

Starting with $\underline{b}$, $\underline{s}$ and W_0 of (4.89), the W-Transformation produces W_1 (4.90), which is easily transformed to W_2:

$$W_2 = \begin{bmatrix} U'\bar{H}^{-1}U & U'\bar{H}^{-1}X & U'\bar{H}^{-1}\underline{z} \\ X'\bar{H}^{-1}U & X'\bar{H}^{-1}X & X'\bar{H}^{-1}\underline{z} \\ \underline{z}'\bar{H}^{-1}U & \underline{z}\bar{H}^{-1}X & \underline{z}'\bar{H}^{-1}\underline{z} \end{bmatrix} \tag{5.20}$$

where: $\underline{z} = \underline{y} - X\underline{b}$ and the transformation is accomplished via:

$$\underline{z}'\bar{H}^{-1}U = \underline{y}'\bar{H}^{-1}U - \underline{b}'X'\bar{H}^{-1}U$$

$$\underline{z}'\bar{H}^{-1}X = \underline{y}'\bar{H}^{-1}X - \underline{z}'X'\bar{H}^{-1}X$$

and

$$\underline{z}'\bar{H}^{-1}\underline{z} = \underline{y}'\bar{H}^{-1}\underline{y} - 2\underline{b}'X'\bar{H}^{-1}\underline{y} + \underline{b}'X'\bar{H}^{-1}X\underline{b} \tag{5.21}$$

Noting that $V(\underline{\theta}) = \theta_q H(\underline{\gamma})$ and using A.5a) and b), W_2 provides all components necessary to evaluate the log likelihood (5.3) and its first and second order derivatives. Hence the equations for the next estimated values and the information matrix (5.15) can be established from W_2.

We make the following notational definitions for the arrays of first and second order partial derivatives of the log likelihood evaluated at $\underline{b}$ and $\underline{s}$:

NOTATION	ORDER OF DERIVATIVE	DERIVATIVE WITH RESPECT TO:	MATRIX SIZE
$\underline{r}_{\underline{b}}$	1	$\underline{\beta}$	p x 1
$\underline{r}_{\underline{s}}$	1	$\underline{\theta}$	q x 1
$S_{\underline{bb}}$	2	$\underline{\beta},\underline{\beta}$	p x p
$S_{\underline{sb}} = S_{\underline{bs}}$	2	$\underline{\beta},\underline{\theta}$	p x q
$S_{\underline{ss}}$	2	$\underline{\theta},\underline{\theta}$	q x q
$ES_{\underline{ss}}$	Expectation of $S_{\underline{\theta},\underline{\theta}}$		q x q

(5.22)

The forms of the derivatives are in (A.6h) and using the property $\bar{H}^{-1}(\underline{\gamma}) = \theta_q V(\underline{\theta})^{-1}$ A.5 a) and b) and noting that $V_i = U_i U_i'$ and $V_q = I$, we may calculate these components as follows:

$$\underline{r}_{\underline{b}} = (1/s_q)\, X'\bar{H}^{-1}\underline{z} \tag{5.23}$$

$$\underline{r}_{\underline{s}} = (1/2s_q^2)\, (\underline{z}'\bar{H}^{-1}U_i U_i'\bar{H}^{-1}\underline{z} - tr(U_i'\bar{H}^{-1}U_i) \quad i=1,2\ldots q-1$$

and for i=q: $(1/2s_q^2)\, (\underline{z}'\bar{H}^{-1}\underline{z} - \underline{z}'\bar{H}^{-1}UDU'\bar{H}^{-1}\underline{z} - n + tr(U'\bar{H}^{-1}UD))$

(5.24)

$$S_{\underline{bb}} = -(1/s_q)X'\bar{H}^{-1}X \tag{5.25}$$

$$S_{\underline{bs}} = -(1/s_q^2)\, (X'\bar{H}^{-1}U_j U_j'\bar{H}^{-1}\underline{z}) \quad j=1,2\ldots q-1$$

and for j=q: $-(1/s_q^2)\, (X'\bar{H}^{-1}\underline{z} - X'\bar{H}^{-1}UDU'\bar{H}^{-1}\underline{z})$

(5.26)

$$S_{\underline{ss}} = (1/2s_q^2)\ tr(U_i'\bar{H}^{-1}U_jU_j'\bar{H}^{-1}U_i)$$
$$- (1/s_q^3)\ \underline{z}'\bar{H}^{-1}U_iU_i'\bar{H}^{-1}U_jU_j'\bar{H}^{-1}\underline{z} \quad i,j=1,2\ldots q-1$$

and for j=q: $(1/2s_q^2)(tr(U_i'\bar{H}^{-1}U_i) - tr(U_i'\bar{H}^{-1}UDU'\bar{H}^{-1}U_i))$
$$- (1/s_q^3)\ (\underline{z}'\bar{H}^{-1}U_iU_i'\bar{H}^{-1}\underline{z} - \underline{z}'\bar{H}^{-1}U_iU_i'\bar{H}^{-1}UDU'\bar{H}^{-1}\underline{z})$$

and for i=j=q: $(1/2s_q^2)(n-2tr(U'\bar{H}^{-1}UD) + tr(U'\bar{H}^{-1}UDU'\bar{H}^{-1}UD))$
$$- (1/s_q^3)\ (\underline{z}'\bar{H}^{-1}\underline{z} - 2\underline{z}'\bar{H}^{-1}UDU'\bar{H}^{-1}\underline{z} + \underline{z}\bar{H}^{-1}UDU'\bar{H}^{-1}UDU'\bar{H}^{-1}\underline{z}) \qquad (5.27)$$

Finally,

$$ES_{\underline{ss}} = -(1/2s_q^2)\ tr(U_i'\bar{H}^{-1}U_jU_j'\bar{H}^{-1}U_i) \quad i,j=1,2\ldots q-1$$

and for j=q: $= -(1/2s_q^2)(tr(U_i'\bar{H}^{-1}U_i) - tr(U_i'\bar{H}^{-1}UDU'\bar{H}^{-1}U_i))$

$i=1,2\ldots q-1$

and for i=j=q: $= -(1/2s_q^2)(n-2tr(U'\bar{H}^{-1}UD) + tr(U'\bar{H}^{-1}UDU'\bar{H}^{-1}U))$ $\qquad (5.28)$

Having calculated the components of (5.22), for given values of $\underline{\beta}$ and $\underline{\theta}$, $\underline{b}$ and $\underline{s}$, the likelihood can be evaluated and new estimates of the parameters computed so as to increase the likelihood by either of two standard methods.

a) The Newton Raphson Algorithm:

This algorithm computes new estimates $\underline{b}_1$ and $\underline{s}_1$ from $\underline{b}$ and $\underline{s}$, according to:

$$\underline{b}_1 = \underline{b} - \underline{\delta}_{\underline{b}} \text{ and } \underline{s}_1 = \underline{s} - \underline{\delta}_{\underline{s}} \qquad (5.29)$$

where $(\underline{\delta}_{\underline{b}}'\underline{\delta}_{\underline{s}}')'$ is the solution to:

$$\begin{bmatrix} S_{\underline{bb}} & S_{\underline{bs}} \\ S_{\underline{sb}} & S_{\underline{ss}} \end{bmatrix} \begin{bmatrix} \underline{\delta}_{\underline{b}} \\ \underline{\delta}_{\underline{s}} \end{bmatrix} = \begin{bmatrix} \underline{r}_{\underline{b}} \\ \underline{r}_{\underline{s}} \end{bmatrix} \qquad (5.30)$$

b) The Fisher Scoring Algorithm.

This method computes the new estimates $\underline{b}_1$ and $\underline{s}_1$ as:

$$\underline{b}_1 = \underline{b} - \underline{\delta}_{\underline{b}} \quad \text{and} \quad \underline{s}_1 = \underline{S} - \underline{\delta}_{\underline{s}} \tag{5.31}$$

where

$(\underline{\delta}_{\underline{b}}' \underline{\delta}_{\underline{s}})'$ is the solution to:

$$\begin{bmatrix} S_{\underline{bb}} & 0 \\ 0 & ES_{\underline{ss}} \end{bmatrix} \begin{bmatrix} \underline{\delta}_{\underline{b}} \\ \underline{\delta}_{\underline{s}} \end{bmatrix} = \begin{bmatrix} \underline{r}_{\underline{b}} \\ \underline{r}_{\underline{s}} \end{bmatrix} \tag{5.32}$$

5.2. Marginal Maximum Likelihood Estimators in the General Normal Model - MMLE.

The maximum likelihood estimators are biased and do not reduce to the ANOVA estimators in balanced models. Corbeil and Searle (1976b), show that the bias may be considerable in models with numerous fixed effect parameters. In an attempt to remedy these 'deficiencies', Patterson and Thompson (1971) and (1974) suggested maximizing the likelihood of'contrasts' of the response, which have zero expectation. That is, statistics of the data which are invariant to translations in the fixed effect parameters. As in 4.1.2, under the invariance principle, we need only consider maximal invariants $\underline{By}$ where $BX = 0$ and rank B is n-rank(X).

As noted in 5.1.2, the MLEs are also functions of

a maximal invariant and Patterson and Thompson (1974), argue that this indicates that the MLEs can contain no information about $\underline{\theta}$, which is not contained in the likelihood function for a maximal invariant.

This procedure has a Bayesian justification from the point of view of marginalizing a posterior distribution over nuisance parameters, 6.1.4. In the classical approach ordinary Lebesgue measure is used, in a Bayesian approach other measures on the parameter space might appear preferable. Hence we refer to the method as marginal maximum likelihood estimation (MMLE). It is also known as restricted or modified maximum likelihood estimation.

5.2.1. The Marginal Maximum Likelihood Equations.

If $B\underline{y}$ is a maximal invariant (4.17), it has distribution $N(0, BV(\underline{\theta})B')$, since $BX = 0$. Hence the log likelihood of $B\underline{y}$ is: (5.33)

$$L_1(\underline{\theta}/B\underline{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(BV(\underline{\theta})B')) - \frac{1}{2}\underline{y}'B'(BV(\underline{\theta})B')^{-1}B\underline{y}$$

Differentiating with respect to θ_i and setting the resulting expressions to zero, we obtain the MML Equations: (A.6i).

$$\operatorname{tr}((BV(\underline{\theta})B')^{-1}BV_iB') = \underline{y}'B'(BV(\underline{\theta})B')^{-1}BV_iB'(BV(\underline{\theta})B')^{-1}B\underline{y} \qquad (5.34)$$

$i=1,2\ldots q$

These equations are independent of the particular full rank maximal invariant used, since for any B with size $(n-\text{rank}(X))\times n$ having maximal rank and for which $BX = 0$, we have:

$$B'(BV(\underline{\theta})B')^{-1}B = V(\underline{\theta})^{-1}(I-X(X'V(\underline{\theta})^{-1}X)^{-1}X'V(\underline{\theta})^{-1}) = N(\underline{\theta})$$

Searle (1979 p.27). (5.35)

Hence the MML equations are:

$$tr(N(\underline{\theta})V_i) = \underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y} \quad i=1,2...q \qquad (5.36)$$

or, noting that $N(\underline{\theta}) = N(\underline{\theta})V(\underline{\theta})N(\underline{\theta})$ we have:

$$tr(N(\underline{\theta})V_i) = \Sigma\theta_j tr(N(\underline{\theta})V_iN(\underline{\theta})V_j)$$

and the equation can be written:

$$H_{U,I}(\underline{\theta})\underline{\theta} = r_{U,I}(\underline{\theta}) \qquad (5.37)$$

where

$$H_{U,I}(\underline{\theta}) = (tr(V_iN(\underline{\theta})V_jN(\underline{\theta})))$$

and $$r_{U,I}(\underline{\theta}) = (\underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y}, i=1,2...q)' \qquad (5.38)$$

We note that equation (5.37) is just the defining equation for IMINQE(U,I) (4.84) and is unbiased in the sense that $E(r_{U,I}(\underline{\theta})) = H_{U,I}(\underline{\theta})$. The MML equations also reduce to the ANOVA equations for balanced models. (section 5.3).

5.2.2. The Information Matrix for MMLEs.

The Fisher Information Matrix is obtained as the negative expectation of the log likelihood (5.33) and from (A.6i), is:

$$I(\underline{\theta}) = (\tfrac{1}{2}tr(N(\underline{\theta})V_iN(\underline{\theta})V_j)) = \tfrac{1}{2}H_{U,I}(\underline{\theta}) \qquad (5.39)$$

This is the inverse of the assymptotic covariance matrix of the MMLEs. Conditions of size or design for the models, which make $I(\hat{\underline{\theta}})$ a reasonable estimate of $I(\underline{\theta})$ are unknown, as in the case of the information matrix for MLEs (5.15).

5.2.3. Computation of MMLEs in Structural Models having a Residual Error.

Essentially the same problems are associated with computing MMLEs as were encountered with IMINQEs and MLEs. There is no closed form solution to the MMLE equations (5.37), and so iterative procedures must be used. Further, the components of the equations, or evaluation of the log likelihood, require the inversion of the covariance at each step. The computational burden of this last requirement, means that MMLE is not feasible in large general models. Rao, C.R. (1979) quotes some special cases of hierarchical classification models and orthogonal block designs where the inverse is easy to find. The most general case, however, is for structured models having a residual error (2.23), where the inversion form (4.88) and the W-transformation of 4.3.3, are available. (Corbeil and Searle (1976a)).

Direct iterative application of (5.37) does produce the MMLEs, when convergence occurs and the limit is in the parameter space. This method is a simple iteration of the MINQE(U,I) algorithm of 4.3.4. and is recommended by Miller (1979). Problems of slow convergence, due to oscillating values and difficulty in applying the parameter space constraints were encountered, as in Miller (1979). The first of these problems is solved by detecting the occurance of oscillation and taking a new starting value between two oscillations.

Better procedures may be to use the Newton Raphson or Fisher Scoring Algorithms to locate the maximum of the marginal likelihood, in the same way as for MLEs 5.1.4. A combination of the two is used by Jennrich and Sampson (1976) and (1978) in the programme BMDP3V (Dixon and Brown (1977)). All components required to compute the second order partial derivatives (A.6i) are available in W_1 (4.90) from the W-Transformation and W_2 (4.90) obtained from W_1 via the sweep operator.

5.3. Maximum Likelihood Methods for Balanced Variance Component Models.

The balanced variance component model (2.58), has the form:

$$\underline{y} = X\underline{\beta} + \underline{e} \qquad (5.40)$$

$$\text{Cov}(\underline{y}) = \Sigma\, \omega_i Q_i = V(\underline{\omega})$$

where

Q_i, i=1,2...q are symmetric, idempotent and mutually orthogonal,

and $R(X)$ is invariant to Q_i i=1,2...q.

The parameter space $\Omega_{\underline{\omega}}$ is restricted by linear constraints C, and the parameters of interest are linear functions of $\underline{\omega}$, $\underline{\theta} = B_1\underline{\omega}$.

5.3.1. Maximum Likelihood Estimation - MLEs.

From (2.74), the log likelihood of model (5.40), with the normal distribution assumptions, is proportional to:

$$-\|X(\underline{\beta}-\underline{\beta})\|^2_{V(\underline{\omega})} - \sum_{i=1}^{q} (m_i \log(\omega_i) + (1/\omega_i)\underline{y}'NQ_iN\underline{y}) \qquad (5.41)$$

where $m_i = \text{rank}(Q_i)$.

For each $\underline{\omega} \in \Omega_{\underline{\omega}}$ (5.41) is maximized at $\underline{\beta} = \hat{\underline{\beta}} = (X'X)^{-}X'\underline{y}$ which is independent of $\underline{\omega}$.

With $\underline{\beta}$ taking its maximizing value, the first term of (5.41) vanishes and the remainder is maximized for each term of the sum when:

$$\omega_i = \hat{\omega}_i = (\underline{y}'NQ_iN\underline{y})/\text{rank}(Q_i) \qquad (5.42)$$

These maximum liklihood estimators are biased by a factor of:

$$\text{tr}(NQ_i)/\text{tr}(Q_i) = (m_i - \text{tr}(PQ_i))/m_i$$

This bias can be appreciable when there are many fixed effects which results in $\text{tr}(PQ_i)$ being large.

The MLE of $\underline{\theta}$ is $\hat{\underline{\theta}} = B_1\hat{\underline{\omega}}$ (5.43)

where $\hat{\underline{\omega}}$ has eliments defined by (5.42).

5.3.2. Marginal Maximum Liklihood Estimation - MMLE.

If B is n x (n - rank(X)) with full column rank and $B'X = 0$ then:

a) $B'\underline{y}$ is a maximal invariant (4.17).

b) $B'B$ has full rank,

and c) $R(B)$ is the orthogonal compliment to $R(X)$ and therefore $B(B'B)^{-1}B' = N$ and $B'N = B'$, $NB = B$. (5.44)

In the model (5.40) with the normal distribution assumptions, $B'\underline{y} \sim N(0, B'V(\underline{\omega})B)$

Now, $$(B'V(\underline{\omega})B)^{-1} = (B'B)^{-1}B'V(\underline{\omega})^{-1}B(B'B)^{-1} \qquad (5.45)$$

and $$\det(B'V(\underline{\omega})B) = \det(B'B)\det(B'B)^{-1}B'V(\underline{\omega})B)$$

However, $$(B'B)^{-1}B'V(\underline{\omega})B(B'B)^{-1}B'Q_iB = \omega_i(B'B)^{-1}B'Q_iB$$

so that the eigen values of $(B'B)^{-1}B'V(\underline{\omega})B$ are ω_i with multiplicity $\text{rank}((B'B)^{-1}B'Q_iB) = \text{tr}((B'B)^{-1}B'Q_iB) = \text{tr}(Q_iN)$ because of idempotency. Hence,

$$\det(B'V(\underline{\omega})B) = \det(B'B)\prod_{i=1}^{q}\omega_i^{mi-ri}$$

where $$mi - ri = \text{tr}(Q_i) - \text{tr}(PQ_i) = \text{tr}(NQ_i).$$

Thus the log liklihood is proportional to:

$$-\sum_{i=1}^{q}((mi-ri)\log(\omega_i) + (1/\omega_i)\underline{y}'NQ_iN\underline{y}) \qquad (5.46)$$

and this is maximized when:

$$\omega_i = \hat{\omega}_i = \underline{y}'NQ_iN\underline{y}/\text{rank}(NQ_i) \qquad (5.47)$$

which is just the ANOVA estimator (3.10) and hence is the UMVUE.

The same estimators are derived from a logical Bayesian point of view in 6.1.3.

5.4. Maximum Likelihood Estimates of Variance Components for Example Data Sets.

Maximum liklihood and marginal maximum liklihood estimates for the example data sets of 2.7. were computed using the programme BMDP3V (Dixon and Brown (1977)).

5.4.1. Data Set 1.

The SIRE and SIRE X DBRD components were constrained to be zero in all variables except for LENB and MASW which had the following estimates.

Table (5.1): ML and MML Estimates of Variance Components for Variables LENB and MASW of Data Set 1.

VARIABLE	COMPONENT	ML ESTIMATES	MML ESTIMATES
LENB	SIRE	17.482	36.646
	SIRE X DBRD	19.166	55.293
	RESIDUAL	753.831	758.485
MASW	SIRE	4.793	24.871
	SIRE X DBRD	0.	0.
	RESIDUAL	565.151	576.496

The MML estimates for LENB agreed with the IMINQE(U,I) estimates of 4.4.1. The iterative scheme (4.84) did not converge for variable MASW.

5.4.2. Data Set 2.

The ML estimate for the data generated to follow the model assumed for data set 1, were in good agreement with the Henderson Method III estimates 3.4.2, and the MINQE estimates 4.4.2. The MML estimates agreed exactly with the IMINQE(U,I) estimates of Table (4.2).

Table (5.2) ML Estimates of the Variance Components for Data Set 2.

VARIABLE	VARIANCE COMPONENT FOR:		
	SIRE	SIRE X DBRD	RESIDUAL
1	9.684	1.171	8.175
2	3.134	1.806	9.236
3	5.619	2.260	8.746
4	6.127	1.965	8.682
5	3.058	2.653	8.355
6	4.900	0.443	8.912
7	5.398	3.121	7.110
8	16.227	1.793	7.900
9	2.489	2.675	7.835
10	10.671	0.808	7.943

The actual variance component values were 10.0, 3.0 and 8.0 respectively.

5.4.3. Data Set 3.

The ML estimators for the small generated example of 2.7.3, are given in the following table:

Table (5.3) ML and MML Estimates of Variance Components for Data Set 3.

VARIABLE	ESTIMATOR	VARIANCE COMPONENT FOR: R1	R2	R1 X F	R1 X R2	RESIDUAL
1	MML	44.482	22.750	0.0	0.0	16.749
	ML	31.656	11.782	0.0	0.0	16.662
2	MML	20.630	17.817	0.0	5.012	9.425
	ML	13.889	8.967	0.0	4.853	9.432
3	MML	24.136	4.447	2.711	0.0	17.587
	ML	16.711	0.857	0.0	0.0	18.792
4	MML	0.606	0.0	0.0	4.326	16.153
	ML	0.0	0.0	0.0	3.360	14.922
5	MML	0.0	7.504	0.0	15.613	13.063
	ML	0.0	2.023	0.0	14.934	12.805
6	MML	50.959	11.849	13.811	0.0	8.605
	ML	37.602	5.825	8.925	0.0	9.117
7	MML	13.362	18.409	10.477	0.0	10.486
	ML	10.959	10.314	6.174	0.0	10.920
8	MML	28.903	5.869	2.358	0.0	20.895
	ML	19.668	1.977	1.105	0.0	21.056

Table (5.3) contd.

Table (5.3) contd.

VARIABLE	ESTIMATOR	VARIANCE COMPONENT FOR:				
		R1	R2	R1 X F	R1 X R2	RESIDUAL
9	MML	51.049	18.300	6.851	15.289	11.187
	ML	34.268	7.945	0.0	18.785	12.821
10	MML	29.412	2.256	10.888	0.0	17.597
	ML	19.521	0.0	4.446	0.0	18.920

The actual variance component values were: 30.0, 10.0, 3.0, 6.0 and 12.0 respectively.

CHAPTER 6

Bayesian Approaches to Estimation in Normal Variance Component Models.

Bayesian inferences and estimating functions have not been constructed for the covariance parameters of the general linear model, (2.1), even for the normal case.

Lindley and Smith (1972) consider Bayesian influence for the location parameters of the general linear model, assuming the covariance structure to be known. Smith (1973) extends the ideas to models with random effects represented as location parameters having a known covariance structure. No Bayesian methods of estimating this covariance structure are considered. Dempster, Rubin and Tsutakawa (1981) suggest estimating the covariance matrix of random effects by ML or MML and then using the Bayesian approach of Lindley and Smith (1973) to estimate the location parameters. They also consider an algorithm for this purpose, the E-M algorithm.

Bayesian estimation of covariance parameters has only been considered in simple models. Klotz, Milton and Zacks (1969), considered balanced one-way classification models with a quadratic loss function. Zacks(1970) considered the same model with squared error loss, but restricted the class of estimating functions to the set:

$$\{f(\underline{y}) : f(s(\underline{y} + \underline{b})) = s^2 f(\underline{y}),\ s > 0,\ \underline{b} \in R^n\} .$$

The resulting Bayes estimators are termed Bayes Equivariant Estimators.

Tiao and Box (1967), Box and Tiao (1973) and Sahai(1975), consider simple balanced classification models with non-informative

reference priors. Portnoy (1971) considers the use of scale invariant loss functions for the balanced single factor classification model.

Results for the simple models do not generalize well, but do illustrate two approaches to the problem. Firstly, restricting the model, so as to provide tractable results and secondly, restricting the class of candidate estimating functions.

Considering the first approach, that of restricting the model, Rudolph (1976) achieves the most general results by considering a balanced variance component model (2.58) without any restrictions on the variance component parameter space, which is taken to be $\mathbb{R}^{q+}$. Classical approaches yield UMVUEs for these models. Rudolph (1976), uses the conjugate family of Normal-Gamma distributions. Rudolph's technique requires choosing a normal-gamma prior density for transformed values of the parameters of interest. The resulting posterior distribution for the variance components of interest involves an expectation, which must be evaluated for each model. The practicality of the results is severely limited by these difficulties, together with the loss of applicability, due to ignoring the constraints in the parameter space.

We will consider balanced variance component models and derive non-informative prior distributions from the logical Bayesian view developed by Villegas (1977a and b). Like Box and Tiao (1973) we ignore the constraints, while developing the prior and re-introduce them by truncating the posterior distribution. The logical prior turns out to be the same as Jeffreys' prior, used by Box and Tiao (1973), for

special cases of balanced classification models.

The second approach to Bayesian estimation, that of restricting the class of estimating functions and minimizing a Bayes risk, simplifies the problem of choosing a prior distribution in many cases because the risk associated with certain classes of functions, depends only on the first two moments of a prior distribution. Kleffe and Pincus (1974b) and Kleffe (1975) consider the structured variance component model without the assumption of normality. They show that the risk associated with unbiased quadratic estimators under squared error loss, depends only on the first two moments of the prior distribution and the first four moments of the response distribution. No readily applicable forms of the estimators appear to be available.

We will consider the general normal model and some sub-classes of quadratic estimating functions.

6.1. Bayesian Inference in Balanced Variance Component Models.

The balanced variance component model, (2.58), has the form;

$$\underline{y} = X\underline{\beta} + \underline{e} \tag{6.1}$$

$$\text{Cov}(\underline{y}) = V(\underline{\omega}) = \Sigma\omega_i Q_i \tag{6.2}$$

where

Q_i, i=1,2...q are symmetric, idempotent and mutually orthogonal (6.3)

R(X) is invariant to Q_i, i=1,2...q (6.4)

The parameters of interest are linear functions of $\underline{\omega}$, say $\underline{\theta} = B_1\underline{\omega}$. We consider X to have full column rank. If this is not the case, the

original model can be re-parameterized to make X have full rank. The parameter space of $\underline{\omega}$, $\Omega_{\underline{\omega}}$ is a subset of $\mathbb{R}^{q+}$, with linear restrictions $C:\{ B_1\underline{\omega} = \underline{\theta} > 0,\ B_2\underline{\omega} = \underline{0}\}$. In some applications, (Box and Tiao (1973) Ch. 5 and 6), B_2 is void and $B_1\underline{\omega} > 0$ reduces to:

$$0 < \omega_q \leq \omega_{q-1} \leq \ldots \leq \omega_1 \tag{6.5}$$

From (2.74), the likelihood for model (6.1) with the normal distribution assumptions is proportional to:

$$\prod_{i=1}^{a} \omega_i^{-mi} \exp\{ -\sum_{i=1}^{q} (1/2\omega_i)\underline{y}'NQ_iN\underline{y} + \|X(\underline{\beta} - \hat{\underline{\beta}})\|^2_{V(\underline{\omega})}\} \tag{6.6}$$

where $mi = \text{rank}(Q_i)$, $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$

6.1.1. The Multi-Group Structure of Balanced Variance Component Models.

The sample space of model (6.1) is $\mathbb{R}^n$. We consider the set G of transformations of this sample space defined by:

$$g(\underline{\gamma},\underline{\lambda})(\underline{y}) = V(\underline{\lambda})\underline{y} + X\underline{\gamma} \quad \text{for } \underline{\gamma} \in \mathbb{R}^P, \lambda \in \mathbb{R}^{q+} \tag{6.7}$$

where

$$V(\underline{\lambda}) = \Sigma\lambda_i Q_i.$$

$V(\underline{\lambda})\underline{y} = V(\underline{\lambda})\ X\underline{\beta} + V(\underline{\lambda})\underline{e}$ and $R(X)$ being invariant to $V(\underline{\lambda})$, means that $V(\underline{\lambda})X = XB(\underline{\lambda})$ where $B(\underline{\lambda}) = \Sigma\lambda_i(X'X)^{-1}X'Q_iX$ from (2.63) and (2.64).

$$\text{Let } G = \{g\ (\underline{\gamma},\ \underline{\lambda}) :\ \underline{\gamma} \in \mathbb{R}^P,\ \underline{\lambda} \in \mathbb{R}^{q+}\} \tag{6.8}$$

Then for g_1 and g_2 in the set G,

$$g_2 \circ g_1(\underline{y}) = V(\underline{\lambda}_2)\ V(\underline{\lambda}_1)\underline{y} + V(\underline{\lambda}_2)\ X\gamma_1 + X\gamma_2 = g_3(\underline{y}) \tag{6.9}$$

where; $g_3 = g(\gamma_3\ \underline{\lambda}_3)$ with $\gamma_3 = B(\underline{\lambda}_2)\ \gamma_1 + \gamma_2$

and $\underline{\lambda}_3 = ((\lambda_{1i}\lambda_{2i}),\ i=1,2\ldots q)'$

The set G with the binary operation, $g_2 \circ g_1$ defined by (6.9), is a group. Closure follows by observing that $\underline{\gamma}_3 \in \mathbb{R}^p$ and $\underline{\lambda}_3 \in \mathbb{R}^{q+}$, so that $g_3 \in G$. Associativity follows from the mutual orthogonality and idempotency of Q_i, (6.3), which allow $V(\underline{\lambda}_1)$ and $V(\underline{\lambda}_2)$ and therefore, $B(\underline{\lambda}_1)$ and $B(\underline{\lambda}_2)$ to commute. The identity element is $g(\underline{0}, \underline{1})$, since $V(\underline{1}) = I$ by (2.67). The inverse of $g(\underline{\gamma}, \underline{\lambda})$ is $g(-\underline{\gamma}, \underline{\lambda}^{-1})$, where $\underline{\lambda}^{-1} = (1/\lambda_1, 1/\lambda_2, \ldots 1/\lambda_q)' \in \mathbb{R}^{q+}$.

G induces a homomorphic group $\bar{G}$ of transformations of the parameters in the space $\mathbb{R}^p \times \mathbb{R}^{q+}$. This is the parameter space if we ignore the constraints.

$$\bar{G} = \{\bar{g}(\underline{\gamma}, \underline{\lambda}) : \underline{\gamma} \in \mathbb{R}^p, \lambda \in \mathbb{R}^{q+}\} \qquad (6.10)$$

$$\bar{g}(\underline{\gamma}, \underline{\lambda})\ (\underline{\beta}, \underline{\omega}) = (\underline{\beta}^*\underline{\omega}^*)$$

$$\underline{\beta}^* = B(\underline{\lambda})\underline{\beta} + \underline{\gamma} \quad \text{and} \quad \underline{\omega}^* = (\lambda_i^2\, \omega_i,\ i=1,2\ldots q)'$$

The model (6.1) is invariant to G, in the sense of Villegas (1981), meaning that, if $\underline{y}$ has distribution with parameters $\underline{\beta}$ and $\underline{\omega}$, then $g(\underline{\gamma}\ \underline{\lambda})\underline{y}$ has distribution with parameters $\bar{g}(\underline{\gamma}\ \underline{\lambda})\ (\underline{\beta},\underline{\omega})$. (6.11)

This follows from (6.7):

$$\begin{aligned} g(\underline{\gamma}, \underline{\lambda})\underline{y} &= V(\underline{\lambda})\ \underline{y} + X\underline{\gamma} \\ &= V(\underline{\lambda})X\underline{\beta} + X\underline{\gamma} + V(\underline{\lambda})\ \underline{e} \\ &= X(B(\underline{\lambda})\ \underline{\beta} + \underline{\gamma}) + V(\underline{\lambda})\ \underline{e} \\ &= X\underline{\beta}^* + \underline{e}^* \end{aligned}$$

and $\quad \mathrm{Cov}(g(\underline{\gamma}, \underline{\lambda})\underline{y}) = \mathrm{Cov}(\underline{e}^*) = V(\underline{\lambda})V(\underline{\omega})V(\underline{\lambda})$

$$= \sum_i \sum_j \sum_k \lambda_i\lambda_j\omega_k Q_i Q_j Q_k = \sum_k \lambda_k^2\, \omega_k Q_k \quad \text{by} \quad (6.3) \quad = V(\underline{\omega}^*) \qquad (6.12)$$

It is important to note however, that the transformed parameters may violate the constraints (6.5). Some consequences of this will be considered in subsequent sections. However, if we ignore the constraints, model (6.1) is invariant to G and hence a statistical G model (Villegas (1981)).

We are interested in separate inference about $\underline{\beta}$ and $\underline{\omega}$ and so we factor G into subgroups:

$$G = G_{\underline{\omega}} G_{\underline{\beta}} \tag{6.13}$$

where $G_{\underline{\omega}}$ is a direct product of groups G_i, for $i=1,2...q$.

where; $G_i = \{g(0,\underline{\rho}_i) : \underline{\rho}_i = \underline{1}$ except for component i which has value $\lambda_i > 0\}$ (6.14)

and $G_{\underline{\beta}} = \{g(\underline{\gamma},\underline{1}): \underline{\gamma} \in \mathbb{R}^p\}$ (6.15)

$G_{\underline{\beta}}$ is isomorphic to the group of translations of $\underline{\beta}$ in $\mathbb{R}^p$, since $V(\underline{1}) = B(\underline{1}) = I_n$. $G_{\underline{\omega}}$ is isomorphic to the direct product of q groups of multiplication by positive scalers.

G factors, because $G_{\underline{\beta}}$ and G_i are subgroups of G and

$$g(0, \underline{\rho}_1) \times g(0, \underline{\rho}_2) \times \ldots \times g(0, \underline{\rho}_q) = g(0, \underline{\lambda}) \in G_{\underline{\omega}}$$

and $$g(0, \underline{\lambda})\, g(\underline{\gamma},\underline{1}) = g(\underline{\gamma},\underline{\lambda}) \in G \tag{6.16}$$

With the factorization (6.13) and the invariance to G, we have a statistical bi-group model (Villegas (1982)).

6.1.2. The Logical Prior Distribution.

Box and Tiao (1973) consider examples of balanced classification models and derive Jeffreys' prior for $\underline{\beta}$ and $\underline{\omega}$, assuming independence of $\underline{\beta}$ and $\underline{\omega}$ and ignoring the constraints on $\Omega_{\underline{\omega}}$ (6.5). Under the same assumptions, the balanced variance component model (6.1), is a statistical bi-group model, as developed in 6.1.1. Following the principles of logical Bayesian inference for multi-group models (Villegas, 1981 and 1982), we will derive the logical prior, which turns out to be identical to Jeffreys' prior:

$$\pi(\underline{\beta}, \underline{\omega})d\underline{\beta}\, d\underline{\omega} = \prod_{i=1}^{q} (1/\omega_i)\, d\underline{\beta} d\, \underline{\omega} \qquad (6.17)$$

While this prior assigns positive probability to parts of $\mathbb{R}^{q+}$, known to be impossible according to the constraints (6.5), it does provide tractable results and the constraints can be conveniently re-introduced by truncating the prior or posterior distribution and dividing by a normalizing factor as in Box and Tiao((1973) p67).

Two principles are used in deriving logical priors for statistical multi-group models. An invariance principle (Villegas, (1977a)) and a conditionality principle (Villegas (1977b)).

The invariance principle states that the prior, representing ignorance, should be the measure on the parameter space, which is invariant under the action of the group $\overline{G}$, to which the model is also invariant. (6.10). The intuition behind the principle is that ignorance priors should only represent information about the parameters

which is contained in the model. In a group model, this information has been expressed by the invariance of the model to the group of transformations. The prior, therefore, should also be invariant to this group. In our situation, we have ignored information contained in the constraints (6.5), in order to develop the group structure.

When inference is required separately for different subsets of the parameters, anomalies may occur whereby conditional inferences, given values for one parameter subset, differ from inferences obtained by starting with a model having the appropriately reduced parameter set. The conditionality principle states that the logical prior should avoid such anomalies.

In multi-group models a set of reduced models can be obtained by repeated conditioning of the full model on all but one of the parameter subsets of interest. These reduced models are invariant to the appropriate transformation sub-group. Villegas (1982) shows that, in this situation, application of the conditionality principle results in the logical prior being the product of the logical priors for each reduced model.

The first sub-model to consider is (6.1), with $\underline{\omega}$ known. This reduced model has parameter $\underline{\beta}$ and is invariant under $G_{\underline{\beta}}$ (6.15), the group of translations in $\mathbb{R}^p$. This is a location model and the measure invariant under $\overline{G}_{\underline{\beta}}$ is the constant function: $\pi_{\underline{\beta}}(d\underline{\beta}) = d\underline{\beta}$ (6.18)

The other sub-model of interest is obtained by assuming that $\underline{\beta}$ is known. This reduced model is invariant to $G_{\underline{\omega}}$ (6.14), a direct product of q groups of multiplication by positive scalers.

The measure which is invariant to this group, is the product

$$\pi_{\underline{\omega}}(d\underline{\omega}) = \prod_{i=1}^{q} (1/\omega_i)d\omega_i \qquad (6.19)$$

Hence the logical prior for the bi-group model is (6.17), which corresponds to the prior derived through Jeffreys' rule. (Box and Tiao (1973)).

6.1.3. The Posterior Distribution

The joint posterior distribution is obtained as the product of the prior (6.19) and the likelihood (6.6). The constraints (6.5) are reintroduced, either by truncating the prior or the posterior, as is done in the method of Box and Tiao((1973) p 67 and p 279). The posterior differential is proportional to:

$$\prod_{i=1}^{q} \omega_i^{-m_i/2} \exp\{-\underline{y}'NQ_iN\underline{y}/2\omega_i\} (1/\omega_i)d\omega_i$$
$$x \exp\{ -(1/2) \|x(\underline{\beta} - \underline{\beta})\|^2 V(\underline{\omega})\} d\underline{\beta}$$
$$\text{for } 0 < \omega_q \leq \omega_{q-1} \leq \ldots \leq \omega_1 \qquad (6.20)$$

where $m_i = \text{rank}(Q_i) = \text{tr}(Q_i)$.

Hence, conditionally on $\underline{\omega}$, $\underline{\beta}$ has a normal distribution. Integrating over $\Omega_{\underline{\beta}}$ and using (2.70), we obtain the marginal posterior of $\underline{\omega}$ with differential proportional to:

$$\prod_{i=1}^{q} \omega_i^{-(m_i - r_i)/2} \exp\{-\underline{y}'NQ_iN\underline{y}/2\omega_i\}(1/\omega_i)d\omega_i$$
$$\text{for } 0 < \omega_q \leq \omega_{(q-1)} \leq \ldots \leq \omega_1 \qquad (6.21)$$

where: $r_i = \text{rank}(PQ_i) = \text{tr}(PQ_i)$

hence, $m_i - r_i = \text{tr}(Q_i - PQ_i) = \text{rank}(NQ_i)$.

From this joint posterior, we see that, ignoring the constraints, the ω_i are independently distributed as inverse - chi-square random quantities, (Box and Tiao (1973) p 88). That is:

$$\omega_i \sim \underline{y}'NQ_iN\underline{y}\ \chi^{-2}_{\text{rank}(NQ_i)} \quad \text{for } i = 1,2\ldots q \tag{6.22}$$

Ignoring the constraints, the mode of the distribution (6.21) occurs at $\hat{\underline{\omega}}$ with components;

$$\hat{\omega}_i = \underline{y}'NQ_iN\underline{y}/(\text{rank}(NQ_i) + 2) \tag{6.23}$$

The mode of the constrained distribution will occur at the same values, $\hat{\omega}_i$, provided they do not violate the constraints in (6.21), in which case the mode occurs at a boundary of the parameter space.

The point estimator $\hat{\underline{\omega}}$ (6.23) suffers from two disadvantages: Firstly, it is biased and does not coincide with the MMLE (5.46) which is UMVUE (3.10). Secondly, it will not be invariant to transformations in the way that maximum likelihood estimators are. We will generally be interested in transformations of $\underline{\omega}$. The reason for these disadvantages is that we have maximized the derivative of the differential (6.21) with respect to an arbitrary measure, the Lebesque measure. This measure is arbitrary in the sense that it is not an integral part of the model, and is not relatively invariant under power transformations of the parameters. The natural measure for the parameter space $\Omega_{\underline{\omega}}$ is the prior (6.19) and maximizing the derivative of (6.21) with respect to this measure results in estimators which avoid the two difficulties. This derivative is:

$$\prod_{i=1}^{q} \omega_i^{-(m_i-r_i)/2} \exp\{\underline{y}'NQ_iN\underline{y}\} \tag{6.24}$$

which is just the marginal likelihood (5.46). Its maximum is attained

at the MMLE (5.46), which are also the ANOVA estimates (3.10) and are UMVUE.

This provides an interesting justification of the Marginal Maximum Likelihood approach, which can be extended to the general model.

6.1.4. A Bayesian Approach to Marginal Maximum Likelihood in the General Normal Model.

Any prior distribution for $\underline{\beta}$ and $\underline{\theta}$ in the general model (2.1), which specifies a uniform measure for $\underline{\beta}$ and prior independence of $\underline{\beta}$ and $\underline{\theta}$ has the form:

$$\pi(d\underline{\theta})\,d\underline{\beta}$$

If the model is parameterized so that X has full rank and the likelihood factored as in (2.6), then the joint posterior is proportional to:

$$\det(V(\underline{\theta}))^{-\frac{1}{2}}\det(X'V(\underline{\theta})^{-1}X)^{-\frac{1}{2}}\exp\{-(1/2)\underline{y}'N(\underline{\theta})\underline{y}\}\,\pi(d\underline{\theta})$$
$$\times\,(2\pi)^{-p/2}\det((X'(V(\underline{\theta})^{-1}X)^{-1})^{-\frac{1}{2}}\exp\{-(1/2)\|\hat{\underline{\beta}}(\underline{\theta})-\underline{\beta}\|^2_{(X'V(\underline{\theta})^{-1}X)^{-1}}\}\,d\underline{\beta} \tag{6.25}$$

Hence conditionally on $\underline{\theta}$, $\underline{\beta}$ is normally distributed, and therefore the first term of (6.25) is proportional to the marginal differential for $\underline{\theta}$. Maximizing the derivative of this with respect to the prior measure, $\pi(d\underline{\theta})$ provides point estimators of $\underline{\theta}$ even when the form of the prior is unknown. These estimators must satisfy the equations:

$$\mathrm{tr}(-(X'V(\underline{\theta})^{-1}X)X'V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}X) + \mathrm{tr}(V(\underline{\theta})^{-1}V_i) = \underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y}$$

or

$$\mathrm{tr}(N(\underline{\theta})V_i) = \underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y} \quad i=1,2\ldots q \tag{6.26}$$

These are just the marginal maximum likelihood equations (5.36).

6.2. Bayesian Estimators in Restricted Classes of Functions.

The problem of finding applicable prior densities, which give rise to tractable posterior forms, remains unsolved for the general model. One approach that has been considered however, is to investigate the expected risk under squared error loss over classes of estimating functions. Portnoy (1971) considered the one-way balanced model. Harville (1974) showed that for the general normal model, consideration of functions invariant to translations in the fixed effect parameters, meant that the posterior distribution is independent of prior knowledge of the fixed effect parameters. Kleffe and Pincus (1974a) consider the normal variance component model:

$$\underline{y} = X\underline{\beta} + \underline{e} \tag{6.27}$$

$$\text{Cov}(\underline{y}) = V(\underline{\theta}) = \sum_{i=1}^{q} \theta_i V_i \tag{6.28}$$

together with classes of unbiased and invariant quadratic plus linear functions.

In some classes of quadratic plus linear functions, the expected risk under squared error loss depends only on the first and second moments of the prior distributions. We consider the prior information to be summarized by independence of $\underline{\beta}$ and $\underline{\theta}$, together with:

$$E(\underline{\beta}) = \underline{b}, \quad E(\underline{\beta}\,\underline{\beta}') = KK'$$

$$E(\underline{\theta}) = \underline{s}, \quad E(\underline{\theta}\,\underline{\theta}') = MM' \qquad (6.29)$$

where K and M have full column rank.

M can be partitioned by columns $M = (\underline{m}_1, \underline{m}_2 \ldots \underline{m}_m)$.

6.2.1. Estimating Functions with Bayes' Risk Determined by the First Two Moments of the Prior Distribution.

Let $\underline{g}$ be the vector of estimators for r linearly independent functions, $\underline{h}'_i\underline{\theta}$ $i=1,2\ldots r$ of the variance or covariance components. The risk matrix under squared error loss for particular parameter values, $\underline{\beta}$ and $\underline{\theta}$ has i,j th component:-

$$r_{ij} = E_{\underline{y}}(g_i - \underline{h}_i'\underline{\theta})(g_i - \underline{h}_j'\underline{\theta}) \quad i,j=1,2\ldots r \qquad (6.30)$$

where

$E_{\underline{y}}$ denotes expectation over the distribution of $\underline{y}$.

The Bayes risk is the expected risk over the prior distribution of $\underline{\beta}$ and $\underline{\theta}$:

$$E_{\underline{\beta},\underline{\theta}}(r_{ij}) \quad i,j=1,2\ldots r \qquad (6.31)$$

Restricting the estimating functions g_i to be in the class:

$$QL(PU) = \{\underline{y}'A\,\underline{y} + \underline{a}'\underline{y} : A = A',\ X'AX = 0,\ \underline{a}'X = 0\} \qquad (6.32)$$

which have bias independent of the fixed effect parameters, the Bayes' risk is determined by the first two moments of the prior distribution and is independent of the functional form, apart from independence of $\underline{\beta}$ and $\underline{\theta}$. This follows by considering the form of risk for functions in $QL(PU)$.

$$r_{ij} = \mathrm{Cov}(\underline{y}'A_i\underline{y},\ \underline{y}'A_j\underline{y}) + \mathrm{Cov}(\underline{a}_j'\underline{y}\ \underline{y}'\ A_i\underline{y})$$
$$+ \mathrm{Cov}(\underline{a}_i'\ \underline{y}\ \underline{y}'\ A_j\underline{y}) + \mathrm{Cov}(\underline{a}_i'\ \underline{y}\ \underline{y}'\ \underline{a}_j)$$
$$+(\mathrm{tr}(A_iV(\underline{\theta})) - \underline{h}_i'\underline{\theta})\ (\mathrm{tr}(A_jV(\underline{\theta})) - \underline{h}_j'\underline{\theta}) \qquad (6.33)$$

This follows by considering the covariance and product of bias, separately with (3.7) and the properties of A and $\underline{a}$ (6.32). Using (3.14) and (3.15), the forms for the covariances, we obtain:

$$r_{ij} = 2\mathrm{tr}(A_iV(\underline{\theta})A_jV(\underline{\theta})) + 4\underline{\beta}'X'A_iV(\underline{\theta})A_jX\underline{\beta}$$
$$+ 2\ \underline{a}_j'V(\underline{\theta})A_iX\underline{\beta} + 2\underline{a}_i'V(\underline{\theta})A_jX\underline{\beta} + \underline{a}_i'V(\underline{\theta})\underline{a}_j$$
$$+ (\mathrm{tr}(A_iV(\underline{\theta})) - \underline{h}_i'\underline{\theta})\ (\mathrm{tr}(A_jV(\underline{\theta}) - \underline{h}_j'\underline{\theta}) \qquad (6.34)$$

Taking the expectation over $\underline{\beta}$, we note that the second term is $4\mathrm{tr}(A_iV(\underline{\theta})A_jXE_{\underline{\beta}}(\underline{\beta}\ \underline{\beta}')X')$. For expectation over $\underline{\theta}$, we write each term as a sum $\sum_i \sum_j \theta_i\theta_j f(V_i,V_j)$ or $\sum_i \theta_i f(V_i)$ of linear functions of the known matrices V_i and V_j. From (6.29), we note that $E_{\underline{\theta}}\theta_i\theta_j$ can be written as: $\sum_{k=1}^{m} m_{ik}\ m_{jk}$ where $M = (m_{ik})$. Also, $E(\theta_i) = s_i$, $i=1,2\ldots q$.

Hence (6.44) becomes:

$$E_{\underline{\beta},\ \underline{\theta}}(r_{ij}) =$$
$$\sum_{k=1}^{m} (2\mathrm{tr}(A_iV(\underline{m}_k)A_jV(\underline{m}_k)) + (\mathrm{tr}(A_iV(\underline{m}_k) - \underline{h}_i'\underline{m}_k)(\mathrm{tr}(A_jV(\underline{m}_k) - \underline{h}_j'\underline{m}_k))$$
$$+ 4\ \mathrm{tr}(A_iV(\underline{s})A_jXKK'X') + 2\underline{a}_j'V(\underline{s})A_iX\underline{b}$$
$$+ 2\ \underline{a}_i'V(\underline{s})A_jX\underline{b} + \underline{a}_i'V(\underline{s})\underline{a}_j \qquad (6.35)$$

where $\underline{m}_k$ is the k th column of M (6.29).

Although the Bayes' risk (6.35), is independent of the functional form of the prior distribution, there appears to be no workable, or enlightening, solution to the problem of minimizing (6.35) over A_i and A_j, satisfying the conditions of QL(PU) (6.32). The problem is similar to that of obtaining MINQE(PU) (4.124), but we notice that in the Bayes case, we are optimizing over $m + 1$ prior values $\underline{m}_k$, $k=1,2...m$ and $\underline{s}$ of $\underline{\theta}$ and so the resulting estimator should have good minimum norm properties over a wider set of parameter values than the MINQE estimator.

As in the case of MINQE, the functional forms can be restricted to subsets of QL(PU), while maintaining independence from the functional form of the prior. Kleffe and Pincus (1974a) considered Bayes' Quadratic plus linear unbiased estimators BAQLE $(U, \underline{h}_i, i=1,2..r)$. The minimization problem reduces to:

$$\sum_{k=1}^{m} 2\mathrm{tr}(A_iV(\underline{m}_k)A_jV(\underline{m}_k)) + 4\mathrm{tr}(A_iV(\underline{s})A_jXKK'X)$$

$$+ \ 2\underline{a}_j'V(\underline{s})A_iX\underline{b} + 2\underline{a}_i'V(\underline{s})A_jX\underline{b} + \underline{a}_i'V(\underline{s})\underline{a}_j \qquad (6.36)$$

with $A_i, \underline{a}_i$ satisfying requirements of $QL(U,\underline{h}_i)$ and similarly for $A_j\underline{a}_j$. Kleffe and Pincus (1974a) derive conditions for $A_i\underline{a}_i$, $A_j\underline{a}_j$ to exist and be optimal and conclude that it is sufficient to optimize for each function $\underline{h}_i'\underline{\theta}$ independently.

If invariance is required of the estimators then $AX = 0$ and the optimal value of $\underline{a}$ is zero, so the minimization for $BAQE(U,I,\underline{h})$,

the Bayes' quadratic unbiased and invariant estimator, becomes:

$$\sum_{k=1}^{m} \operatorname{tr}(AV(\underline{m}_k)\, AV(\underline{m}_k)) \qquad (6.37)$$

subject to A satisfying conditions for $Q(U,I,\underline{h})$. This will exist provided $Q(U,I,\underline{h})$ is non-empty, (4.21) and minimizing each term of the sum is the MINQE$(U,I,\underline{h})$ problem of 4.3.2.

CHAPTER 7

Linear Estimation of Variance Components in Derived Models.

In 2.6 we observed that quadratic estimation of covariance parameters in the general model (2.1) is equivalent to linear estimation of location parameters in a derived model:

$$\underset{\sim}{y} = \underset{\sim}{X}\underline{\theta} + \underset{\sim}{e} \tag{7.1}$$

with

$$\mathrm{Cov}(\underset{\sim}{y}) = F$$

where

$$\underset{\sim}{y} = \mathrm{Vec}((\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})') = (\underline{y} - X\underline{\beta}) \otimes (\underline{y} - X\underline{\beta})$$

$$\underset{\sim}{X} = (\mathrm{Vec}(V_1), \mathrm{Vec}(V_2)\ldots\mathrm{Vec}(V_q))$$

$$E(\underset{\sim}{e}) = \underline{0}$$

F is the $n^2 \times n^2$ matrix of fourth order central moments.

If we further restrict estimators to invariant quadratic forms in $\underline{y}$, we need only consider functions of a maximal invariant. One such invariant is $N\underline{y}$ where $N = I - X(X'X)X'$ and then we have derived model (2.91):

$$\underset{\sim}{y}_N = \underset{\sim}{X}_N\underline{\theta} + \underset{\sim}{e}_N \tag{7.2}$$

with

$$\mathrm{Cov}(\underset{\sim}{y}_N) = F_N = (N \otimes N)F(N \otimes N)$$

where $\tilde{y}_N = \mathrm{Vec}(Nyy'N) = Ny \otimes Ny$

$= N \otimes N\, y \otimes y = N \otimes N\tilde{y}$

$\tilde{X}_N = N \otimes N\, \tilde{X}$

and $E(\tilde{e}_N) = 0$

As mentioned in 2.6, difficulties with these derived models include dealing with a restricted sample space and complicated covariance structures, F, which, in general, are functions of β and θ.

Using results on linear estimation theory, several researchers have derived quadratic variance component estimators with various optimal properties. Pukelsheim (1976) demonstrates the equivalence of minimum mean square error and minimum norm for quadratic estimators to best linear estimators in the derived model. Pukelsheim (1977), applies the best linear estimation results to the structured variance component model. Gnot, Klonecki and Zmyslony (1977), consider uniform minimum variance estimators amongst quadratic and quadratic plus linear functions. They make extensive use of dispersion mean correspondence. Drygas (1977) and Kleffe (1978) consider best simultaneous estimation of location and covariance parameters.

7.1. Ordinary Least Squares Estimation in Derived Models.

The OLS estimator of θ in model 7.1 is the solution to:

$$\tilde{X}'\tilde{X}\hat{\theta} = \tilde{X}'\tilde{y} \tag{7.3}$$

or

$$H\hat{\theta} = r \tag{7.4}$$

where $H = ((\mathrm{trace}(V_iV_j))) = (\mathrm{Vec}'(V_i)\mathrm{Vec}(V_j)) = \tilde{X}'\tilde{X}$ by (2.78).

It has full rank if θ is identifiable, 2.1.1.

and $\underline{r} = ((\underline{y} - X\underline{\beta})'V_i(\underline{y} - X\underline{\beta}),\ i=1,2\ldots q)' = \tilde{X}'\tilde{\underline{y}}.$

This estimator is not known to have any desirable properties, unless $\underline{\beta}$ were known, when its only property appears to be unbiasedness, unless F is proportioned to I.

With the inclusion of invariance, we can use OLS on model (7.2) and obtain the MINQEO(U,I) (4.85) estimator, since;

$$\tilde{X}_N'\tilde{X}_N = (\text{Vec}'(V_iN)\text{Vec}(NV_j))$$

$$= (\text{tr}(NV_iNV_j)) = H_{U,I} \quad (4.21)$$

and

$$\tilde{X}_N\tilde{\underline{y}}_N = (\underline{y}'NV_iN\underline{y},\ i=1,2\ldots q)' = \underline{r}_{U,I}$$

Hence the OLS equations are;

$$H_{U,I}\,\hat{\underline{\theta}} = \underline{r}_{U,I} \qquad (7.5)$$

with $\hat{\underline{\theta}}$ having minimum norm at the value of $\underline{\theta} \in \Omega_{\underline{\theta}}$, which makes $V(\underline{\theta}) = I$. This value is $(0,0,\ldots 1)'$ for models with a residual error.

7.2. Generalized Least Squares and Best Linear Unbiased Estimation.

For any linear model:

$$\underline{y} = X\underline{\theta} + \underline{e} \qquad (7.6)$$

with $\text{Cov}(\underline{y}) = F.$ (7.7)

A GLS estimate of $\underline{\theta}$ is a solution, $\hat{\underline{\theta}}$ to the equations:

$$X'F^{-}X\hat{\underline{\theta}} = X'F^{-}\underline{y} \qquad (7.8)$$

where F^{-} is a generalized inverse of F.

Zyskind and Martin (1969), give a Generalized Gauss-Markov theorem to the effect that $\hat{\underline{\theta}}$ of (7.8) is the best linear unbiased

estimator of $\underline{\theta}$, provided F^- belongs to a specific class of generalized inverses of F.

In the particular case where $R(X) \subseteq R(F)$, this class includes all generalized inverses. (Zyskind and Martin (1969) Cor.(1.1)). Searle (1979) gives an equivalent condition for $R(X) \subseteq R(F)$ as

$$FF^-X = X. \tag{7.9}$$

7.3. Generalized Least Squares and Maximum Likelihood in Normal Models.

For the normal general linear model with covariance parameters;

$$\underline{y} \sim N(X\underline{\beta},\ V(\underline{\theta})) \quad \text{where} \quad V(\underline{\theta}) = \Sigma\theta_i V_i \tag{7.10}$$

we have seen 2.6.5, that the covariance of the general derived model is:

$$F = (V(\underline{\theta}) \otimes V(\underline{\theta}))\ (I + I_{(n,n)}) \tag{7.11}$$

Using the properties of the Vec-permutation matrices (2.84) to (2.86), it is easy to show that a generalized inverse of F is given by:

$$F^- = (1/4)\ (V^{-1} \otimes V^{-1})(I + I_{(n,n)}) \tag{7.12}$$

where V is written for $V(\underline{\theta})$.

Furthermore,

$$\begin{aligned} FF^-\tilde{X} &= (1/4)(I + I_{(n,n)})(V \otimes V)(V^{-1} \otimes V^{-1})(I + I_{(n,n)})\ \tilde{X} \\ &= (1/2)(I + I_{(n,n)})\ \tilde{X} \\ &= \tilde{X} \quad \text{since} \quad I_{(n,n)}\mathrm{Vec}(V_i) = \mathrm{Vec}(V_i) \quad \text{by (2.85).} \end{aligned} \tag{7.13}$$

and hence the solution, $\hat{\underline{\theta}}$ to the equations:

$$\tilde{X}'F^-\tilde{X}\ \hat{\underline{\theta}} = \tilde{X}'F^-\underline{y} \tag{7.14}$$

is the BLUE of $\underline{\theta}$.

These generalized least squares, (GLS) equations, reduce to the ML equations when we estimate $\underline{\beta}$ by the GLS estimator in the original model:

The equations (7.14) are:

$$(1/4)\ \tilde{X}'(V^{-1} \otimes V^{-1})(I + I_{(n,n)})\tilde{X} = (1/4)\ \tilde{X}'(V' \otimes V')(I + I_{(n,n)})\tilde{y}.$$

But since $I_{(n,n)}\tilde{X} = \tilde{X}$ (2.85), these reduce to;

$$\tilde{X}'(V^{-1} \otimes V^{-1})\tilde{X} = \tilde{X}'(V^{-1} \otimes V^{-1})\mathrm{Vec}((\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})') \tag{7.15}$$

Also,

$$\mathrm{Vec}'(V_i)V^{-1} \otimes V^{-1}\mathrm{Vec}(V_j) = \mathrm{Vec}'(V^{-1}V_iV^{-1})\mathrm{Vec}(V_j) \text{ by (2.77)}$$
$$= \mathrm{tr}(V^{-1}V_iV^{-1}V_j) \text{ by (2.78)} \tag{7.16}$$

and $\mathrm{Vec}'(V_i)(V^{-1} \otimes V^{-1})\mathrm{Vec}((\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})')$

$$= \mathrm{tr}(V^{-1}V_iV^{-1}(\underline{y} - X\underline{\beta})(\underline{y} - X\underline{\beta})')$$
$$= (\underline{y} - X\underline{\beta})'V^{-1}V_iV^{-1}(\underline{y} - X\underline{\beta}) \tag{7.17}$$

Substituting the GLSE of $X\underline{\beta}$ in (7.17), we obtain;

$$X\underline{\beta} = X(X'V(\underline{\theta})^{-1}X)^{-}X'V(\underline{\theta})^{-1}\underline{y} = P(\underline{\theta})\underline{y} \tag{7.18}$$

and

$$(\underline{y} - X\underline{\beta}) = (I - P(\underline{\theta}))\underline{y}.$$

Hence the GLS equations are (7.18) and

$$H(\underline{\theta})\ \underline{\theta} = \underline{r}_{U,I}(\underline{\theta}) \tag{7.19}$$

where

$$H(\underline{\theta}) = (\mathrm{tr}(V(\underline{\theta})^{-1}V_iV(\underline{\theta})^{-1}V_j))\ i,j=1,2\ldots q$$

$$\underline{r}_{U,I}(\underline{\theta}) = (\underline{y}'N(\underline{\theta})V_iN(\underline{\theta})\underline{y},\ i=1,2\ldots q)'$$

and $N(\underline{\theta}) = (I - P'(\underline{\theta}))V(\underline{\theta})^{-1} = V(\underline{\theta})^{-1}(I - P(\underline{\theta}))$

which are exactly the ML equations (5.7) and (5.9). Hence simultaneous BLUE of $\underline{\beta}$ and $\underline{\theta}$ in the derived model is ML estimation in the original normal model.

7.4. Generalized Least Squares Estimators for the Invariant Derived Model under Normality.

In model (7.2), we have that linear estimation corresponds to invariant quadratic estimation in the original model. From (2.96), we have that, under normality,

$$\mathrm{Cov}(\tilde{y}_N) = (NV(\underline{\theta})N \otimes NV(\underline{\theta})N)\ (I + I_{(n,n)}) = F_N$$

where $N = I - X(X'X)^{-}X'$. (7.20)

If $N(\underline{\theta}) = V(\underline{\theta})^{-1}(I - P(\underline{\theta}))$,

then $NN(\underline{\theta}) = N(\underline{\theta})N = N(\underline{\theta})$ (7.21)

so $N(\underline{\theta})$ is a generalized inverse of $NV(\underline{\theta})N$.

Hence, $(1/4)\ (N(\underline{\theta}) \otimes N(\underline{\theta}))\ (I + I_{(n,n)}) = F_N^{-}$ (7.22)

Furthermore, $F_N F_N^{-} X_N = X_N$, so (7.9) is satisfied and the solution to the generalized least squares equations is the BLUE. (7.23)

In the same way as the equations of 7.3. reduced to the ML equations, we have that the GLS equations for the invariant model reduces to:

$$X_N'(N(\underline{\theta}) \otimes N(\underline{\theta}))X_N = X_N'(N(\underline{\theta}) \otimes N(\underline{\theta}))\tilde{y}_N \quad (7.24)$$

or

$$H_{U,I}(\underline{\theta})\ \hat{\underline{\theta}} = \underline{r}_{U,I}(\underline{\theta}) \quad (7.25)$$

where

$$H_{U,I}(\underline{\theta}) = (\mathrm{tr}(N(\underline{\theta})V_i N(\underline{\theta})V_j))$$

$$\underline{r}_{U,I}(\underline{\theta}) = (\underline{y}'NV_i N\underline{y},\ i=1,2\ldots q)'$$

On substituting an a priori value $\underline{s}$, for $\underline{\theta}$ in (7.25), we obtain exactly the MINQE(U,I) equations (4.83). We have already noted that these lead to unbiased and invariant quadratic estimates which, under normality, have locally minimum variance at $\underline{s} \in \Omega_{\underline{\theta}}$ (4.58).

Appendix Mathematical Results.

A.1. The Vec Operator

For a matrix $A = (a_{ij})$ for $i=1,2...m$, $j=1,2...n$, Vec(A) is the mn x 1 vector:

$$(a_{11},a_{21},...a_{m1},a_{12},a_{22},...a_{m2},...a_{1n},a_{2n},..a_{mn})'$$

Properties of the Vec operator and its relationship to the Kronecker product are given in chapter 2, (2.77) - (2.79) and (2.83) - (2.86). Henderson and Searle (1979) discuss these results in detail.

A.2 $R(H_U) = R(H_U^*)$ (4.8)

Given $V_1,V_2,...V_q$ where V_i are n x n symmetric matrices, and X an n x p matrix,

Let $P = X(X'X)^{-}X'$ and $N = I_n - P$, then $NP = PN = 0$

Let $H_U = (tr(V_iNV_j))$ for $i,j=1,2...q$,

$H_U^* = (tr((V_i - PV_iP)V_j))$ for $i,j=1,2...q$,

$\mathcal{B} = sp\{NV_i + NV_i : i=1,2...q\}$

and $\mathcal{B}^* = sp\{V_i - PV_iP : i=1,2...q\}$

Then the following results hold:

a) $Dim(\mathcal{B}) = Dim(\mathcal{B}^*)$

Proof: If $\Sigma a_i(V_i - PV_iP) = (V - PVP) = 0$ $(V = \Sigma a_iV_i)$ then:

$VN - PVPN = VN = 0$ and $NV - NPVP = NV = 0$,

hence, $NV + VN = \Sigma a_i(NV_i + V_iN) = 0$

On the other hand if $\Sigma a_i(NV_i + V_i'N) = 0$

then $NVP + VNP = NVP = VP - PVP = 0$ so $VP = PVP$ and similarly $PV = PVP$.

Hence, firstly:

$NV = V - PV = V - VP = VN$ so that $0 = NV + VN = 2NV$

and secondly,

$V - PVP = V - PV = NV = 0$ or $\Sigma a_i (V_i - PV_iP) = 0$.

b) $\mathbb{B} \cap \mathbb{B}^{*\perp} = \{0\}$

Proof: Suppose $B \in \mathbb{B} \cap \mathbb{B}^{*\perp}$ then $B = \Sigma a_i (NV_i + V_iN)$

$= NV + VN$ for some $V = \Sigma a_i V_i$ and $tr(BA) = 0$

for all $A \in \mathbb{B}^*$.

In particular, $V - PVP \in \mathbb{B}^*$ so $tr((NV + VN)(V - PVP)) =$

$2tr(VNV) = 2tr((VN)(VN)') = 0$ since $NN = N$.

However $tr(AA')$ is the SS of elements of A so if

$tr(AA') = 0$ then $A=0$. Hence $NV = 0 = VN$ and therefore $B = 0$.

c) $S = \mathbb{B} + \mathbb{B}^{*\perp}$ where S is the space of n x n symmetric matrices.

Proof: If $A \in S$, A can be written as $B_1 + B_2$ with $B_1 \in \mathbb{B}^*$ and $B_2 \in \mathbb{B}^{*\perp}$. However $B_1 \in \mathbb{B}$ because $\mathbb{B}$ and $\mathbb{B}^*$ have the same dimension and $\mathbb{B} \cap \mathbb{B}^{*\perp} = \{0\}$. Hence $S = \mathbb{B} + \mathbb{B}^{*\perp}$.

d) $R(H_U) \subseteq R(H_U^*)$

Proof: Suppose $\underline{h} \neq 0$ and $\underline{h} = H_U \underline{a}$ then

$h_i = \Sigma_j a_j tr(V_i N V_j)$ for $i=1,2...q$.

Let $A = \Sigma(a_j/2)(NV_j + V_jN)$

Then $tr(A(V_i - PV_iP)) = h_i$

However, if we write $tr(A(V_i - PV_iP))$ as $Vec'(V_i - PV_iP)Vec(A)$ then we have $\underline{h} = T\,Vec(A)$ where T is the $q \times n^2$ matrix with i th row: $Vec'(V_i - PV_iP)$.

Hence, $\underline{h} \in R(T)$ but since $H_U^* = TT'$, $R(T) = R(H_U^*)$ and so $\underline{h} \in R(H_U^*)$.

e) $R(H_U^*) \subseteq R(H_U)$

Proof: Suppose $h \neq 0$ and $\underline{h} = H_U^*\underline{a} = TT'\underline{a}$.

We have that $T'\underline{a} = \Sigma a_j Vec(V_j - PV_jP) = Vec(\Sigma a_j(V_j - PV_jP))$

$= Vec(A)$ where $A = \Sigma a_j(V_j - PV_jP)$ is symmetric.

Also $h_i = Vec'(V_i - PV_iP)Vec(A) = tr(A(V_i - PV_iP))$

and since $h_i \neq 0$ all i, $A \notin \mathcal{B}^{*\perp}$.

Further, $A \in S$ so $A = B_1 + B_2$ with $B_1 \neq 0$ and $B_1 \in \mathcal{B}$ by c).

Hence, $h_i = tr((B_1 + B_2)(V_i - PV_iP)) = tr(B_1(V_i - PV_iP))$

because $B_2 \in \mathcal{B}^{*\perp}$. Also $B_1 \in \mathcal{B}$ so ther exists $\underline{b}$ such that

$B_1 = \Sigma b_j(NV_j + V_jN)$

Hence $h_i = tr(B_1(V_i - PV_iP)) = 2\Sigma b_j tr(V_iNV_j)$

or $\underline{h} = 2H_U\underline{b}$ and so $R(H_U^*) \subseteq R(H_U)$.

A.3. For B p.d. $R(H_U(B)) = R(H_U^*)$ (4.12)

If $\underline{y}$ is a random variable with model:

$\underline{y} = X\underline{\beta} + \underline{e}$ and $Cov(\underline{y}) = \Sigma\theta_i V_i$

then we know that $\underline{h}'\underline{\theta}$ is unbiasedly estimable by quadratic functions, $\underline{y}'A\underline{y}$ with A symmetric if and only if

$\underline{h} \in R(H_U^*)$ (4.8)

If we let D be the symmetric p.d square root of B then we can consider the model:

$$\underline{y}^* = X^*\underline{\beta} + \underline{e}^* \text{ with } Cov(\underline{y}^*) = \Sigma\theta_i V_i^*$$

where $\underline{y}^* = D^{-1}\underline{y}$, $X^* = D^{-1}X$, $\underline{e}^* = D^{-1}\underline{e}$

and $V_i^* = D^{-1}V_i D^{-1}$.

Clearly, $E(\underline{y}'A\underline{y}) = \underline{h}'\underline{\theta}$ if and only if $E(\underline{y}^{*\prime}DAD\underline{y}^*) = \underline{h}'\underline{\theta}$

In which case $R(H_U^*) = R(Q)$

where $Q = (tr((V_i^* - P^*V_i^*P^*)V_j^*)$, and $P^* = X^*(X^{*\prime}X^*)^{-}X^{*\prime}$

however, with the cyclic property of trace,

$$Q = (tr((B^{-1}(V_i - P(B)V_iP(B)')B^{-1}V_j) = H_U(B)$$

where $P(B) = X(X'B^{-1}X)^{-}X'B^{-1}$

A.4. For B p.d. $R(H_{U,I}(B)) = R(H_{U,I})$.

If $\underline{y}$ is a random variable with model:

$\underline{y} = X\underline{\beta} + \underline{e}$ and $Cov(\underline{y}) = \Sigma\theta_i V_i$,

then we know that $\underline{h}'\underline{\theta}$ is quadratically a d invariantly estimable if and only if $\underline{h} \in R(H_{U,I})$

If we let D be the p.d. square root of B, then we can consider the transformed model:

$$\underline{y}^* = X^*\underline{\beta} + \underline{e}^*, \quad Cov(\underline{y}^*) = \Sigma\theta_i V_i^*$$

where $\underline{y}^* = D^{-1}\underline{y}$, $X^* = D^{-1}X$, $\underline{e}^* = D^{-1}\underline{e}$

and $V_i^* = D^{-1}V_i D^{-1}$

Clearly, $E(\underline{y}'A\underline{y}) = \underline{h}'\underline{\theta}$ if and only if $E(\underline{y}^{*\prime}DAD\underline{y}^*) = \underline{h}'\underline{\theta}$

In which case, $R(H_{U,I}^*) = R(H_{U,I})$

where $H_{U,I}^* = (tr((V_i^* - P^*V_i^*P^*)V_j))$, $P^* = X^*(X^{*\prime}X^*)^{-}X^{*\prime}$

but with the cyclic property of trace, $H_{U,I}^* = H_{U,I}(B)$

A.5.

Let $H = H(\underline{c}) = U\,D(\underline{c})U' + I_n = I_n + UDU'$

where $D(\underline{c}) = D = \text{diag}(c_i I_{qi},\ i=1,2\ldots q-1)$

and $U = (U_1, U_2 \ldots U_{q-1})$.

a) $H^{-1} = I_n - UD(I + U'UD)^{-1}U'$

proof: $HH^{-1} = (I + UDU')(I - UD(I + U'UD)^{-1}U')$

$= I + UDU' - UD(I + U'UD)^{-1}U'$

$\quad - UDU'UD(I + U'UD)^{-1}U'$

$= I + UDU' - UD(I + U'UD)(I + U'UD)^{-1}U' = I.$

(A more general version of this result is available for H with form $R + UDU'$. Searle (1979) p 14).

b) $H^{-1} = I - UDU'H^{-1}$ and $H^{-1}H^{-1} = H^{-1} - H^{-1}UDU'H^{-1}$

Proof: $(I - UDU'H^{-1})H = H - UDU' = I + UDU' - UDU' = I.$

c) Let $P = P(\underline{c}) = X(X'H^{-1}X)^{-}X'H^{-1}$ and $Q = Q(\underline{c}) = I - P(\underline{c})$ Then Q is symmetric, idempotent and $\text{rank}(Q) = n - \text{rank}(X)$

and Let $M = M(\underline{c}) = H^{-1}Q$. Then $MM = MH^{-1}Q = M(I - UDU'H^{-1})Q$

$= MQ - MUDU'H^{-1}Q = M - MUDU'M$ since $M = H^{-1}Q$

and Q is idempotent.

d) $\text{tr}(U_i'MMU_i) = \text{tr}(U_i'MU_i) - \text{tr}(U_i'MUDU'MU_i)$ by c)

However, $UDU' = \Sigma c_j U_j U_j'$ so that:

$\text{tr}(U_i'MMU_i) = \text{tr}(U_i'MU_i) - \Sigma c_j \text{tr}(U_i'MU_j U_j'MU_i)$

e) $\mathrm{tr}(MM) = \mathrm{tr}(M) - \mathrm{tr}(MUDU'M)$ by c)

$= \mathrm{tr}(H^{-1}Q) - \mathrm{tr}(U'MMUD)$

$= \mathrm{tr}(Q) - \mathrm{tr}(UDU'M) - \mathrm{tr}(U'MMUD)$ by b)

$= n - \mathrm{rank}(X) - \mathrm{tr}(U'MUD)$

$- \mathrm{tr}(U'MUD) + \mathrm{tr}(U'MUDU'MUD)$ by c)

$= n - \mathrm{rank}(X) - 2\Sigma c_i \mathrm{tr}(Ui'MU_i)$

$+ \sum_i \sum_j c_i c_j \mathrm{tr}(U_i'MU_jU_j'MU_i)$

because the diagonal blocks of $U'MUD$ are $c_i U_i'MU_i$.

f) $\underline{y}'M\,M\,\underline{y} = \underline{y}'M\,\underline{y} - \underline{y}'M\,UDU'M\,\underline{y}$ by c)

$= \underline{y}'M\,\underline{y} - \Sigma c_j\, \underline{y}'MU_jU_j'M\,\underline{y}.$

g) $\underline{y}'M\,M\,M\,\underline{y} = \underline{y}'M\,\underline{y} - 2\,\underline{y}'MUDU'M\,\underline{y}$

$+ \underline{y}'M\,UDU'M\,U\,DU'M\,\underline{y}.$

A.6. Matrix Differentiation Results.

a) Let $A = (a_{ij}(x))$, $B = (b_{ij}(x))$

Define $\frac{dA}{dx} = (\frac{d}{dx} a_{ij}(x)) = (a^*_{ij}(x))$ with equivalent

definitions of partial derivatives, where x is a vector.

b) $\frac{d}{dx} AB = A \frac{dB}{dx} + \frac{dA}{dx} B$

Proof: . . If $AB = C = (c_{ij}(x))$,

then $c_{ij}(x) = \sum_k a_{ik}(x)\, a_{kj}(x)$

Hence $c_{ij}(x) = \sum_k (a^*_{ik}(x) b_{kj}(x) + a_{ik}(x) b^*_{kj}(x))$

and the result follows.

c) $\frac{dA^{-1}}{dx} = -A^{-1}\frac{dA}{dx}A^{-1}$

Proof: $AA^{-1} = I$ and $\frac{dI}{dx} = 0$

Hence $0 = \frac{dA}{dx}A^{-1} + A\frac{dA}{dx}^{-1}$ by b),

so, $-A^{-1}\frac{dA}{dx}A^{-1} = \frac{dA}{dx}^{-1}$

d) $\frac{d}{dx}\det(A) = \det(A)\ tr(A'^{-1}\frac{dA}{dx})$

Proof: $\det(A) = \sum_P \pm a_{1i_1} a_{2i_2} \ldots a_{ni_n}$

where $P = \{(i_1, i_2 \ldots i_n) : \text{permutations of } (1,2\ldots n)\}$.

Differentiating and grouping the terms of the sum;

$$\frac{d\ detA}{dx} = \sum_P \pm a^*_{1i_1}(a_{2i_2}\ldots a_{ni_n}) \pm a^*_{2i_2}(a_{1i_1}a_{3i_2}\ldots a_{ni_n})$$

$$\pm \ldots$$

$$= \sum_i \sum_j a^*_{ij} A_{ij} \quad \text{where } A_{ij} \text{ is the ij th}$$

cofactor of A.

$$= tr(\frac{dA}{d\theta}\ adj(A'))$$

where adj(A), the classical adjoint, is $(A_{ij})'$.

However, $adj(A) = \det(A).A^{-1}$

so $adj(A') = \det(A).A'^{-1}$ and the result follows.

e) $\frac{d}{dx}tr(A) = tr(\frac{dA}{dx})$

Proof: $\frac{d}{dx}tr(A) = \frac{d}{dx}\sum_i a_{ii}(x) = \sum_i a^*_{ii}(x) = tr\ \frac{dA}{dx}$

f) $\frac{d}{dx} \operatorname{tr}(A^{-1}B) = -\operatorname{tr}(A^{-1}\frac{dA}{dx} A^{-1}B) + \operatorname{tr}(A^{-1}\frac{dB}{dx})$

Proof follows from d), b) and c).

g) For vector $\underline{b} = (b_1(x)\ldots b_m(x))'$ and constant symmetric matrix A with i th column a_i.

$$\frac{d}{dx}\underline{b}'A\underline{b} = 2\frac{d\underline{b}}{dx}'A\underline{b}$$

and $$\frac{d}{dx} A\underline{b} = A\frac{d\underline{b}}{dx}$$

Proof: $\underline{b}'A\underline{b} = \sum_i \sum_j b_i b_j a_{ij}$

Hence, $\frac{d\underline{b}}{dx}'A\underline{b} = \Sigma\ \Sigma b_i^* b_j a_{ij} + b_i b_j^* a_{ij}$

by symmetry $= 2\Sigma\Sigma b_i^* b_j a_{ij} = 2\frac{d\underline{b}}{dx}A\underline{b}$.

In particular,

$\frac{\partial}{\partial\beta_i}\underline{\beta}'X'A X\underline{\beta} = 2x_i' A X\underline{\beta}$ where x_i is the i th column of X.

Hence, $\frac{\partial}{\partial\underline{\beta}}\underline{\beta}'X'A X\underline{\beta} = 2X'A X\underline{\beta}$.

Similarly,

$$\frac{\partial^2}{\partial\beta_i\partial\beta_j}\underline{\beta}'X'AX\underline{\beta} = 2x_i'Ax_j$$

so, $$\frac{\partial^2}{\partial\underline{\beta}^2}\underline{\beta}'X'AX\underline{\beta} = 2X'AX$$

h) Derivatives of the log likelihood.

$$L(\underline{\beta}\ \underline{\theta}/\underline{y}) = -(n/2)\log(2\pi) - (1/2)\log(\det(V(\underline{\theta})) - (1/2)\ (\underline{y} - x\underline{\beta})'\ V(\underline{\theta})^{-1}(\underline{y} - x\underline{\beta})$$

$$\frac{\partial L}{\partial \underline{\beta}} = x'V(\underline{\theta})^{-1}\underline{y} - x'V(\underline{\theta})^{-1}x\underline{\beta} = x'V(\underline{\theta})^{-1}(\underline{y} - x\underline{\beta})$$

$$\frac{\partial L}{\partial \theta_i} = -(1/2)\ tr(V(\underline{\theta})^{-1}V_i) + (1/2)\ (\underline{y} - x\underline{\beta})'\ V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}(\underline{y} - x\underline{\beta})$$

$$i=1,2\ldots q$$

$$\frac{\partial^2 L}{\partial \underline{\beta}^2} = -xV(\underline{\theta})^{-1}x$$

$$\frac{\partial^2 L}{\partial \underline{\beta}\partial\theta_i} = \left(\frac{\partial^2 L}{\partial\theta_i \partial \underline{\beta}}\right)' = -x'V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}(\underline{y} - x\underline{\beta}) \quad i=1,2\ldots q$$

$$\frac{\partial^2 L}{\partial\theta_i \partial\theta_j} = +(1/2)\ tr(V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j)$$

$$-(1/2)\ (\underline{y} - x\underline{\beta})'\ (V(\underline{\theta})^{-1}V_j V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}$$

$$+ V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j V(\underline{\theta})^{-1})\ (\underline{y} - x\underline{\beta})$$

$$= (1/2)tr(V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j) - (\underline{y} - x\underline{\beta})'V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j V(\underline{\theta})^{-1}(\underline{y} - x\underline{\beta})$$

$$i,j=1,2\ldots q$$

$$E\ \frac{\partial^2 L}{\partial \underline{\beta}^2} = -x'V(\underline{\theta})^{-1}x$$

$$E\ \frac{\partial^2 L}{\partial \underline{\beta}\partial\underline{\theta}} = 0$$

$$E\ \frac{\partial^2 L}{\partial\theta_i \partial\theta_j} = -(1/2)tr(V(\underline{\theta})^{-1}V_i V(\underline{\theta})^{-1}V_j) \quad i,j=1,2\ldots q$$

Using (3.7).

i) Derivatives of the log likelihood of a maximal invariant:

$$L_1(\underline{\theta},\ B\underline{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(BV(\underline{\theta})\ B'))$$

$$- \frac{1}{2}\ \underline{y}'B'(BV(\underline{\theta})^{-1}B')^{-1}B\ \underline{y}$$

$$\frac{\partial L_1}{\partial \theta_i} = -\frac{1}{2}\mathrm{tr}((BV(\underline{\theta})B')^{-1}BV_iB')$$

$$+ \frac{1}{2}\ \underline{y}'B'(BV(\underline{\theta})B')^{-1}BV_iB'(BV(\underline{\theta})B')^{-1}B\ \underline{y}$$

$i=1,2\ldots q.$

$$\frac{\partial^2 L_1}{\partial\theta_i\partial\theta_j} = \frac{1}{2}\ \mathrm{tr}((BV(\underline{\theta})B')^{-1}BV_jB'(BV(\underline{\theta})B')^{-1}BV_iB')$$

$$- \underline{y}'B'(BV(\underline{\theta})B')^{-1}BV_jB'(BV(\underline{\theta})B')^{-1}BV_iB'(BV(\underline{\theta})B')B\ \underline{y}$$

$i,j=1,2\ldots q$

$$E\ \frac{\partial^2 L_1}{\partial\theta_i\partial\theta_j} = -\frac{1}{2}\ \mathrm{tr}((BV(\underline{\theta})B')^{-1}BV_jB(BV(\underline{\theta})B')^{-1}BV_iB')$$

$i,j=1,2\ldots q$ Using (3.7).

Note: $B'(BV(\underline{\theta})B')^{-1}B = V(\underline{\theta})^{-1}(I - X(X'V(\underline{\theta})^{-1}X)^{-}V(\underline{\theta})^{-1})$

$= N(\underline{\theta})$

(Searle (1979) p.27).

BIBLIOGRAPHY

1. Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Co., Reading, Massachusetts.

2. Brown, G.B. (1978). Estimation of Variance Components Using Residuals. Journal of the American Statistical Association, 73, p. 141.

3. Brown, K.G. (1976). Asymptotic Behaviour of MINQUE-Type Estimators of Variance Components. Annals of Statistics, 4, p. 746.

4. Bunke, H. and Bunke, O. (1974). Identifiability and Estimability. Mathematische Operationsforschung und Statistik, 5, p. 223.

5. Chaubey, Y.P. (1980). Minimum Norm Quadratic Estimators of Variance Components. Metrica, International Journal of Theoretical and Applied Statistics, 27, p. 255.

6. Chaubey, Y.P. and Rao, P.S.R.S. (1976). Efficiencies of Five Estimators for the Parameters of Two Linear Models with Unequal Variances. Sankhya, Series B, Indian Journal of Statistics, 38, p. 364.

7. Corbeil, R.R. and Searle, S.R. (1976a). Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. Technometrics, 18, p. 31.

8. Corbeil, R.R. and Searle, S.R. (1976b). A Comparison of Variance Component Estimators. Biometrics, 32, p. 779.

9. Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K. (1981). Estimation in Covariance Components Models. Journal of the American Statistical Association, 76, p. 341.

10. Dixon, W.J. and Brown, M.B. Eds. (1977). BMDP Biomedical Computer Programs, P-Series. University of California Press, Berkeley.

11. Drygas, H. (1977). Best Quadratic Unbiased Estimation in Variance-Covariance Component Models. Mathematische Operationsforschung und Statistik, 8, p. 211.

12. Gnot, S., Klonecki, W. and Zmyslony, R. (1977). Uniformly Minimum Variance Unbiased Estimation in Various Classes of Estimators, I. Mathematische Operationsforschung und Statistik, 8, p. 199.

13. Goodnight, J.H. (1978a). Computing MIVQUEO Estimates of Variance Components. SAS Technical Report R-105, SAS Institute, P O Box 8000, Cary, N.C. 27511.

14. Goodnight, J.H. (1978b). The Sweep Operator: Its Importance in Statistical Computing. Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface. North Carolina State University, Raleigh. Genig, T. and Gallart, A.R. Ed.

15. Goodnight, J.H. and Hemmerle, W.J. (1978). A Simplified Algorithm for the W-Transformation in Variance Component Estimation. Technometrics, 21, p. 265.

16. Goodnight, J.H. and Speed, F. (1980). Computing Expected Mean Squares. Biometrics, 36, p. 123.

17. Graybill, F.A. (1976). Theory and Applications of Linear Models. Duxbury Press, North Scltuate, Massachusetts.

18. Hartley, H.O. (1967). Expectation Variances and Covariances of ANOVA Mean Squares by 'Synthesis'. Biometrics, 23, p. 105.

19. Hartley, H.O. and Rao, J.N.K. (1967). Maximum Likelihood Estimation for the Mixed Analysis of Variance Model. Biometrika, 54, p. 93.

20. Hartley, H.O., Rao, J.N.K. and LaMotte, L. (1978). A Simple 'Synthesis'-Based Method of Variance Component Estimation'. Biometrics, 34, p. 233.

21. Hartley, H.O. and Vaughn, W.K. (1972). A Computer Program for the Mixed Analysis of Variance Model, Based on Maximum Likelihood. Statistical Papers in Honour of G.W. Snedecor. T.A. Bankcroft Ed. Iowa State University Press.

22. Harvey, W.R. (1970). Estimation of Variance and Covariance Components in the Mixed Model. Biometrics, 26, p. 485.

23. Harville, D.A. (1974). Bayesian Inference for Variance Components Using only Error Contrasts. Biometrika, 61, p. 383.

24. Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association, 72, p. 320.

25. Hemmerle, W.J. and Downs, B.W. (1978). Mixed Model Algorithms for Estimating Non-Homogeneous Variances. Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface. North Carolina State University, Raleigh. Gallant, A.R. and Gerig, T.M. Ed.

26. Hemmerle, W.J. and Hartley, H.O. (1973). Computing Maximum Likelihood Estimates for Mixed A.O.V. Models Using the W-Transformation. Technometrics, 15, p. 819.

27. Hemmerle, W.J. and Lorens, J.A. (1976). Improved Algorithm for the W-Transformation in Variance Component Estimation. Technometrics, 18, p. 207.

28. Henderson, C.R. (1953). Estimation of Variance and Covariance Components. Biometrics, 9, p. 226.

29. Henderson, C.R., Searle, S.R. and Schaeffer, L.R. (1974). The Invariance and Calculation of Method 2 for Estimating Variance Components. Biometrics, 30, p. 583.

30. Henderson, H.V. and Searle, S.R. (1979). Vec and Vech Operators for Matrices, with some uses in Jacobians and Multivariate Statistics. The Canadian Journal of Statistics, 7, p. 65.

31. Hess, J.L. (1979). Sensitivity of MINQUE with Respect to A Priori Weights. Biometrics, 35, p. 645.

32. Jennrich, R.I. and Sampson, P.F. (1976). Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation. Technometrics, 18, p. 11.

33. Jennrich, R.I. and Sampson, P.F. (1978). Some Problems Faced in Making a Variance Component Algorithm into a General Mixed Model Program. Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface. North Carolina State University, Raleigh. Gallant, A.R. and Gerig, T.M. Ed.

34. Kackar, R.N. and Harville, D.A. (1980). Variance Approximations for Estimators of Fixed and Random Effects in Mixed Linear Models. American Statistical Association, Proceedings of the Statistical Computing Section 1980. p. 52.

35. Kempthorne, O. (1975). Fixed and Mixed Models in the Analysis of Variance. Biometrics, 31, p. 473.

36. Kleffe, J. (1975). Bayes Invariant Quadratic Estimators for Variance Components in Linear Models. Mathematische Operationsforschung und Statistik, 5, p. 753.

37. Kleffe, J. (1976a). A Note on MINQUE for Normal Models. Mathematische Operationsforschung und Statistik, 5, p. 707.

38. Kleffe, J. (1976b). Best Quadratic Unbiased Estimators for Variance Components in Mixed Linear Models. Sankhya, Series B, Indian Journal of Statistics, 38, p. 179.

39. Kleffe, J. (1977a). Invariant Methods for Estimating Variance Components in Mixed Linear Models. Mathematische Operationsforschung und Statistik, 8, p. 233.

40. Kleffe, J. (1977b). A Note on ∞- MINQUE in Variance Covariance Components Models. Mathematische Operationsforschung und Statistik, 8, p. 337.

41. Kleffe, J. (1977c). Optimal Estimation of Variance Components - A Survey. Sankhya, Series B, Indian Journal of Statistics, 39, p. 211.

42. Kleffe, J. (1977d). Best Unbiased Estimators for Variance Components, with Application to the Unbalanced one-way Classification. Biometrical Journal, 19, p. 313.

43. Kleffe, J. (1978). Simultaneous Estimation of Expectation and Covariance Matrix in Linear Models. (with Discussions). Mathematische Operationsforschung und Statistik, 9, p. 443.

44. Kleffe, J. (1980). On Recent Progress of MINQUE Theory - Nonnegative Estimation, Consistency, Asymptotic Normality and Explicite Formulae. Mathematische Operationsforschung und Statistik, 11, p. 563.

45. Kleffe, J. and Pincus, R. (1974a). BAYES and Best Quadratic Unbiased Estimators for Parameters of the Covariance Matrix in a Normal Linear Model. Mathematische Operationsforschung und Statistik, 1, p. 43.

46. Kleffe, J. and Pincus, R. (1974b). BAYES and Best Quadratic Unbiased Estimators for Variance Components and Heteroscedastic Variances in Linear Models. Mathematische Operationsforschung und Statistik, 2, p. 147.

47. Klotz, J.H., Milton, R.C. and Zacks, S. (1969). Mean Square Efficiency of Estimators of Variance Components. Journal of the American Statistical Association, 64, p. 1383.

48. LaMotte, L.R. (1973a). Quadratic Estimation of Variance Components. Biometrics, 29, p. 311 .

49. LaMotte, L.R. (1973b). On Non-Negative Quadratic Unbiased Estimation of Variance Components. Journal of the American Statistical Association, 68, p. 728.

50. Lehmann, E.L. (1959). Testing Statistical Hypotheses. John Wiley, New York.

51. Lindley, D.V. and Smith, A.F.M. (1972). Bayes Estimates for the Linear Model. Journal of the Royal Statistical Society, Series B, 34, p. 1.

52. Liu, L.M. and Senturia, J. (1977). Computation of MINQUE Variance Component Estimates. Journal of the American Statistical Association, 72, p. 867.

53. Miller, J.J. (1977). Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance. Annals of Statistics, 5, p. 746.

54. Miller, J.J. (1979). Maximum Likelihood Estimation of Variance Components - A Monte Carlo Study. Journal of Statistical Computation and Simulation, 8, p. 175.

55. Mitra, S.K. and Moore, B.J. (1973). Gauss-Markov Estimation with an Incorrect Dispersion Matrix. Sankhya, Series A, 35, p. 139.

56. Patterson, H.D. and Thompson, R. (1974). Maximum Likelihood Estimation of Components of Variance. Proceedings of the 8th International Biometric Conference, Constanta, Bucarest, 1974. Corsten, L.C.A. and Postelnicu, T. Ed.

57. Pincus, R. (1974). Estimability of Parameters of the Covariance Matrix and Variance Components. Mathematische Operationsforschung und Statistik, 8, p. 245.

58. Portnoy, S. (1971). Formal Bayes Estimation with Application to a Random Effects Model. Annals of Mathematical Statistics, 42, p. 1379.

59. Pringle, R.M. (1974). Some Results on the Estimation of Variance Components by MINQUE. Journal of the American Statistical Association, 69, p. 987.

60. Pukelsheim, F. (1976). Estimating Variance Components in Linear Models. Journal of Multivariate Analysis, 6, p. 626.

61. Pukelsheim, F. (1977). On HSU's Model in Regression Analysis. Mathematische Operationsforschung und Statistik, 8, p. 323.

62. Rao, C.R. (1970). Estimation of Heteroscedastic Variances in Linear Models. Journal of the American Statistical Association, 65, p. 161.

63. Rao, C.R. (1971a). Estimation of Variance and Covariance Components - MINQUE Theory. Journal of Multivariate Analysis, 1, p. 257.

64. Rao, C.R. (1971b). Minimum Variance Quadratic Unbiased Estimation of Variance Components. Journal of Multivariate Analysis, 1, p. 445.

65. Rao. C.R. (1972). Estimation of Variance and Covariance Components in Linear Models. Journal of the American Statistical Association, 67, p. 112.

66. Rao, C.R. (1973). Linear Statistical Inference and its Applications . Second edition. John Wiley, New York.

67. Rao, C.R. (1979). MINQE Theory and its Relation to Maximum Likelihood and Marginal Maximum Likelihood Estimation of Variance Components. Sankhya, Series B, Indian Journal of Statistics, 41, p. 138.

68. Rao, C.R. and Kleffe, J. (1980). Estimation of Variance Components. Handbook of Statistics, Vol. 1 - Analysis of Variance, p. 1. Krishnaiah, Ed. North-Holland, New York.

69. Rao, J.N.K. (1968). On Expectations, Variances and Covariances of ANOVA Mean Squares by 'Synthesis'. Biometrics, 24, p. 963.

70. Rao, P.S.R.S. (1977). Theory of MINQUE - A Review. Sankyha, Series B, Indian Journal of Statistics, 39, p. 201.

71. Rao, P.S.R.S. and Chaubey, Y.P. (1978). Three Modifications of the Principle of MINQUE. Communications in Statistics, Part A - Theory and Methods, 7, p. 767.

72. Rudolph, P. (1976). Bayesian Estimation in Linear Models under Different Assumptions about the Covariance Structure: Variance Component Models, Equicorrelated Models. Mathematische Operationsforschung und Statistik, 5, p. 649.

73. Sahai, H. (1979). A Bibliography on Variance Components. International Statistical Review, 47, p. 177.

74. Sahai, H. and Ramirez Martinez, A.A. (1975). Some Formal Bayes Estimators of Variance Components in a Crossed Classification Random Effects Model. Australian Journal of Statistics, 17, p. 43.

75. Searle, S.R. (1968). Another Look at Henderson's Methods of Estimating Variance Components. Biometrics, 24, p. 749.

76. Searle, S.R. (1970). Large Sample Variances of Maximum Likelihood Estimators of Variance Components Using Unbalanced Data. Biometrics, 26, p. 505.

77. Searle, S.R. (1971). Linear Models. John Wiley, New York.

78. Searle, S.R. (1972). Using the R-Notation for Reductions in Sums of Squares when Fitting Linear Models. Paper No. BU-417-M in the Biometrics Unit, Cornell University, Ithaca. Presented at the Spring Regional Meetings of ENAR, Ames, Iowa.

79. Searle, S.R. (1974). Prediction, Mixed Models and Variance Components. Reliability and Biometry. Statistical Analysis of Life Length. Society of Industrial and Applied Mathematics, Philadelphia. p. 229. Proschan, F. and Serfling, R.J. Ed.

80. Searle, S.R. (1979). Notes on Variance Component Estimation. Paper BU-673-M, of the Biometrics Unit, Cornell University, Ithaca.

81. Searle, S.R. (1980). The Inscrutable Mixed Model. American Statistical Association, Proceedings of Statistical Computing Section 1980, p. 42.

82. Searle, S.R. and Henderson, H.V. (1979). Dispersion Matrices for Variance Component Models. Journal of the American Statistical Association, 74, p. 465.

83. Seely, J. (1970a). Linear Spaces and Unbiased Estimation. Annals of Statistics, 41, p. 1725.

84. Seely, J. (1970b). Linear Spaces and Unbiased Estimation - Application to the Mixed Linear Model. Annals of Statistics, 41, p. 1735.

85. Seely, J. (1971). Quadratic Subspaces and Completeness. Annals of Mathematical Statistics, 42, p. 710.

86. Smith, A.F.M. (1973). A General Bayesian Linear Model. Journal of the Royal Statistical Society, Series B, 35, p. 67.

87. Speed, F.M. (1979). Choice of Sums of Squares for Estimation of Components of Variance. American Statistical Association, Proceedings of the Statistical Computing Section, 1979. p. 55.

88. Speed, F.M. and Hocking, R.R. (1974). Computation of Expectations, Variances and Covariances of ANOVA Mean Squares. Biometrics, 30, p. 157.

89. Thompson, R. (1975). A Note on the W- Transformation. Technometrics, 17, p. 511.

90. Tiao, G.C. and Box, G.E.P. (1967). Bayesian Analysis of a Three-Component Hierarchical Design Model. Biometrika, 54, p. 109.

91. Urquhart, N.S. and Weeks, D.L. (1978). Linear Models in Messy Data: Some Problems and Alternatives. Biometrics, 34, p. 696.

92. Villegas, C. (1977a). Inner Statistical Inference. Journal of the American Statistical Association, 72, Number 358, Theory and Methods Section.

93. Villegas, C. (1977b). On the Representation of Ignorance. Journal of the American Statistical Association, 72, Number 359, Theory and Methods Section.

94. Villegas, C. (1981). Inner Statistical Inference II. Annals of Statistics, 9, p. 768.

95. Villegas, C. (1982). Personal Communication.

96. Zacks, S. (1970). Bayes Equivariant Estimators of Variance Components. Annals of the Institute of Statistical Mathematics, 22, p. 27.

97. Zyskind, G. and Martin, F.B. (1969). On Best Linear Estimation and a General Gauss-Markov Theorem in Linear Models with Arbitrary Non-Negative Covariance Structure. Society of Industrial and Applied Mathematics Journal on Applied Mathematics, 17, p. 1190.