

TRANSLATION INVARIANT MAXIMUM LIKELIHOOD ESTIMATORS
OF VARIANCE COMPONENTS IN THE MIXED MODEL

R. R. Corbeil and S. R. Searle
Biometrics Unit, Cornell University, Ithaca, N. Y.

BU-496-M

February 1974

Abstract

The maximum likelihood procedure of Hartley and Rao [1967] is modified by adapting a transformation from Patterson and Thompson [1971]. This partitions the likelihood under normality into two parts, one of which is free of fixed effects and so provides estimators of the variance components that are invariant of the fixed effects. A further transformation, adapted from Hemmerle and Hartley [1973], reduces computing requirements to dealing with matrices having order equal to the dimension of the parameter space rather than that of the sample space. These same matrices also occur in the asymptotic sampling variances of the estimators.

1. Introduction

The maximum likelihood procedures derived by Hartley and Rao [1967] lead to simultaneous estimation of both the fixed effects and the variance components that occur in an analysis of variance model involving both fixed and random effects—a model customarily called the mixed model. In particular, estimators of the variance components depend upon the fixed effects. In contrast it has been suggested (e.g. Rao [1971] and La Motte [1973]) that a useful class of estimators of variance components consists of those that are invariant to changes in the fixed effects, i.e. are invariant to changes in translation of the underlying variable. We call these translation invariant estimators. Adaptation of a transformation used by Patterson

and Thompson [1971] leads to a partitioning of the likelihood function into two parts: one part is entirely free of the fixed effects, and maximization of this provides what we call translation invariant maximum likelihood (TIML) estimators of the variance components. These TIML estimators are not only invariant to the fixed effects but also, for balanced data (having equal numbers of observations in the subclasses), they reduce to the familiar analysis of variance estimators for such data, a property not generally possessed by the maximum likelihood estimators of Hartley and Rao [1967].

Adaptation of a transformation described by Hemmerle and Hartley [1973] that simplifies computation of the Hartley-Rao estimators greatly aids the computing of the TIML estimators and also simplifies derivation of their large-sample variances.

Maximizing that portion of the likelihood not used for the TIML estimators provides estimation of the fixed effects.

2. The Model

The model for \underline{y} , a vector of n observations, is specified in terms that closely follow the notation of Hartley and Rao [1967], Hartley and Vaughan [1972] and Hemmerle and Hartley [1973]. We take

$$\underline{y} = \underline{X}\underline{\mu} + \underline{U}_1\underline{b}_1 + \dots + \underline{U}_c\underline{b}_c + \underline{e} \quad (1)$$

where \underline{y} is an n -vector of observations,

\underline{X} is an $n \times k$ matrix of fixed effects for $k \leq n$ and of full column rank,

$\underline{\mu}$ is a k -vector of unknown constants,

\underline{U}_i is an $n \times m_i$ design matrix associated with the i^{th} random factor

with $m_i \leq n$ levels,

\underline{b}_i is an m_i -vector of random variables which are i. i. d. $N(0, \sigma_i^2)$,
 with the \underline{b}_i 's being mutually independent,
 \underline{e} is an n -vector of random variables which are i. i. d. $N(0, \sigma^2)$ and
 independent of the \underline{b}_i 's .

Hence \underline{y} has a multivariate normal distribution with mean and variance

$$E(\underline{y}) = \underline{X}\underline{\mu} \quad \text{and} \quad \text{var}(\underline{y}) = \underline{H}\sigma^2 \quad (2)$$

where

$$\underline{H} = \sum_{i=1}^c \gamma_i \underline{U}_i \underline{U}_i' + \underline{I}_n \quad \text{for} \quad \gamma_i = \sigma_i^2 / \sigma^2 \quad . \quad (3)$$

The symbol $\underline{\mu}$ for fixed effects emphasizes the generality of the model insofar as fixed effects are concerned. $\underline{\mu}$ is a vector of the maximum number of linearly independent estimable functions of the fixed effects. The simplest such vector has as its elements the population means of those of the sub-most cells of the fixed effects factors that contain data. The corresponding \underline{X} of (1) then has a simple form. Define \underline{y} as being the observations ordered so that all those within each sub-most cell of the fixed effects factors follow one another sequentially. If there are k such cells containing data, with the t 'th one having $n_t \neq 0$ observations, then

$$\underline{X} = \begin{bmatrix} \underline{1}_{n_1} & 0 & \dots & 0 \\ 0 & \underline{1}_{n_2} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & & \underline{1}_{n_k} \end{bmatrix} = \sum_{t=1}^k \underline{1}_{n_t} \quad (4)$$

where $\underline{1}_{n_t}$ is a vector of n_t ones and where Σ^+ represents a direct sum of matrices. For example, if the fixed effect factors form a 2-way crossed classification with

the r 'th row effect represented by α_r and the s 'th column effect represented by β_s and the interaction term by γ_{rs} , then $\underline{\mu}$ has elements $(\alpha_r + \beta_s + \gamma_{rs})$ for those of the cells having data in them.

3. The Estimators

The logarithm of the likelihood for $\underline{y} \sim N(\underline{\mu}, H\sigma^2)$ of (2) is

$$\lambda = - \frac{1}{2} n \log 2\pi - \frac{1}{2} n \log \sigma^2 - \frac{1}{2} \log |H| - \frac{1}{2} (\underline{y} - \underline{X}\underline{\mu})' H^{-1} (\underline{y} - \underline{X}\underline{\mu}) / \sigma^2 \quad (5)$$

To partition this into two parts one of which is free of $\underline{\mu}$, Patterson and Thompson [1971] suggest the singular transformation

$$\underline{z} = \begin{bmatrix} \underline{S} \\ \underline{Q} \end{bmatrix} \underline{y} \quad (6)$$

where

$$\underline{Q} = \underline{X}' \underline{H}^{-1}, \quad \text{and} \quad \underline{S} = \underline{I} - \underline{X}(\underline{X}' \underline{X})^{-1} \underline{X}' = \sum_{t=1}^k \left(\underline{I}_{n_t} - n_t^{-1} \underline{J}_{n_t} \right), \quad (7)$$

where \underline{J}_{n_t} is an $n_t \times n_t$ matrix with every element unity. Then, with \underline{S} being symmetric and idempotent and

$$\underline{S} \underline{X} = \underline{0}, \quad (8)$$

\underline{z} has the singular multivariate normal distribution

$$\underline{z} = \begin{pmatrix} \underline{S} \underline{y} \\ \underline{Q} \underline{y} \end{pmatrix} \sim N \left[\begin{pmatrix} \underline{0} \\ \underline{X}' \underline{H}^{-1} \underline{X} \underline{\mu} \end{pmatrix}, \begin{pmatrix} \underline{S} \underline{H} \underline{S} \sigma^2 & \underline{0} \\ \underline{0} & \underline{X}' \underline{H}^{-1} \underline{X} \sigma^2 \end{pmatrix} \right] \quad (9)$$

It is clear from (9) that the distribution of $\underline{\underline{S}}y$ is free of the fixed effects $\underline{\underline{\mu}}$. The likelihood function for $\underline{\underline{S}}y$ therefore forms the basis of our derivation of translation invariant maximum likelihood estimators of the variance components involved in $\underline{\underline{H}}\sigma^2$. However, to avoid the singularity of $\underline{\underline{S}}\underline{\underline{H}}\underline{\underline{S}}$ in (9), arising from the form of $\underline{\underline{S}}$ shown in (7), we use an alternative derived from $\underline{\underline{S}}$ by deleting its n_1 'th, (n_1+n_2) 'th, $(n_1+n_2+n_3)$ 'th, ..., and $(n_1+n_2+\dots+n_k)$ 'th rows. Such a matrix has order $(n-k) \times n$, and denoting it by $\underline{\underline{T}}$, we have

$$\begin{aligned} \underline{\underline{T}} &= \sum_{t=1}^k \left[\begin{pmatrix} \underline{\underline{I}}_{n_t-1} & \underline{\underline{O}}_{1 \times (n_t-1)} \\ \underline{\underline{0}}_{(n_t-1) \times 1} & \underline{\underline{1}}_{n_t-1} \end{pmatrix} - n_t^{-1} \underline{\underline{J}}_{(n_t-1) \times n_t} \right] \\ &= \sum_{t=1}^k \left(\underline{\underline{I}}_{n_t-1} - n_t^{-1} \underline{\underline{J}}_{n_t-1} \left| - n_t^{-1} \underline{\underline{1}}_{n_t-1} \right. \right) \end{aligned} \quad (10)$$

where $\underline{\underline{J}}_{(n_t-1) \times n_t}$ is a matrix of order $(n_t-1) \times n_t$ whose every element is unity. From $\underline{\underline{X}}$ of (4) it is readily seen that

$$\underline{\underline{T}}\underline{\underline{X}} = \underline{\underline{0}} \quad (11)$$

analogous to (8); and by the nature of $\underline{\underline{T}}$ itself, it is easy to show that

$$\underline{\underline{T}}'(\underline{\underline{T}}\underline{\underline{T}}')^{-1}\underline{\underline{T}} = \underline{\underline{S}} \quad (12)$$

As a result of (11), the distribution of $\begin{bmatrix} \underline{\underline{T}}y \\ \underline{\underline{Q}}y \end{bmatrix}$ is just like (9) only with $\underline{\underline{S}}$ replaced by $\underline{\underline{T}}$. Hence the log likelihood of (5) becomes $\lambda = \lambda_1 + \lambda_2$ where

$$\lambda_1 = -\frac{1}{2}(n-k)\log 2\pi - \frac{1}{2}(n-k)\log \sigma^2 - \frac{1}{2}\log |\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'| - \frac{1}{2}\underline{\underline{y}}'\underline{\underline{T}}(\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}')^{-1}\underline{\underline{T}}'\underline{\underline{y}}/\sigma^2 \quad (13)$$

and

$$\lambda_2 = -\frac{1}{2}k\log 2\pi - \frac{1}{2}k\log \sigma^2 - \frac{1}{2}\log |\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{X}}| - \frac{1}{2}(\underline{\underline{y}} - \underline{\underline{X}}\underline{\underline{\mu}})'\underline{\underline{H}}^{-1}\underline{\underline{X}}(\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{X}})^{-1}\underline{\underline{X}}'\underline{\underline{H}}^{-1}(\underline{\underline{y}} - \underline{\underline{X}}\underline{\underline{\mu}})/\sigma^2 \quad (14)$$

The estimators of σ^2 and the γ_i 's that we call translation invariant maximum likelihood (TIML) are, following the method of Patterson and Thompson [1971], those values of σ^2 and the γ_i 's that maximize λ_1 . Differentiation of (13) gives

$$\frac{\partial \lambda_1}{\partial \sigma^2} = -\frac{1}{2}(n-k)/\sigma^2 + \frac{1}{2} \underline{y}' \underline{T}' (\underline{THT}')^{-1} \underline{T} \underline{y} / \sigma^4 \quad (15)$$

and

$$\frac{\partial \lambda_1}{\partial \gamma_i} = -\frac{1}{2} \text{tr}[\underline{U}' \underline{T}' (\underline{THT}')^{-1} \underline{T} \underline{U}_i] + \frac{1}{2} \underline{y}' \underline{T}' (\underline{THT}')^{-1} \underline{T} \underline{U}_i \underline{U}_i' \underline{T}' (\underline{THT}')^{-1} \underline{T} \underline{y} / \sigma^2$$

for $i = 1, 2, \dots, c$. (16)

where $\text{tr}(\underline{A})$ is the trace of a matrix \underline{A} .

Equating (15) and (16) to zero gives the TIML estimators. The resulting equations clearly have no analytic solution and have to be solved numerically. An iterative procedure is to first assign initial values to $\underline{\gamma}' = \{\gamma_1, \dots, \gamma_c\}$ and then (i) solve

$$\hat{\sigma}^2 = \underline{y}' \underline{T}' (\underline{THT}')^{-1} \underline{T} \underline{y} / (n-k) \quad (17)$$

based on (15), and (ii) use the γ -values, and $\hat{\sigma}^2$ from (17), to calculate new γ -values that make (16) closer to zero. Repetition of (i) and (ii), ending at (i), is continued until a desired degree of accuracy is attained.

Although Patterson and Thompson [1971] give a procedure based on Fisher's iterative method for $c = 1$ and suggest how to use it for $c > 1$, the Newton-Rhapson technique is well suited to the problem of finding successive values of $\underline{\gamma}$ that zeroize (16), and has been effectively applied by Hemmerle and Hartley [1973] to similar equations of the Hartley and Rao [1967] maximum likelihood method. We use their application here, first adapting a transformation they use, which simplifies notation and computing procedures.

4. The W-Transformation

The Newton-Rhapson technique for finding values of the elements of \underline{y} that zeroize (16) utilizes the second-order partial derivatives of λ_1 with respect to the Y_i 's. These are, using (16)

$$\frac{\partial^2 \lambda_1}{\partial Y_i \partial Y_j} = \frac{1}{2} \text{tr}[\underline{U}'_i \underline{T}'(\underline{THT}')^{-1} \underline{TU}_j \underline{U}'_j \underline{T}'(\underline{THT}')^{-1} \underline{TU}_i] - \underline{y}'_i \underline{T}'(\underline{THT}')^{-1} \underline{TU}_i \underline{U}'_i \underline{T}'(\underline{THT}')^{-1} \underline{TU}_j \underline{U}'_j \underline{T}'(\underline{THT}')^{-1} \underline{T}_y / \sigma^2 \quad (18)$$

for $i, j = 1, 2, \dots, c$.

The matrix products in (16) and (18) are summarized in the following transformation to \underline{W} suggested by Hemmerle and Hartley [1973].

Define

$$\underline{U} = [\underline{U}_1 \quad \underline{U}_2 \quad \dots \quad \underline{U}_c] \quad (19)$$

and

$$\underline{W} = \{ \underline{W}_{ij} \} \quad \text{for } i, j = 1, 2, \dots, c+1$$

$$= \begin{bmatrix} \underline{U}' \\ \underline{y}' \end{bmatrix} \underline{T}'(\underline{THT}')^{-1} \underline{T} [\underline{U} \quad \underline{y}] \quad (20)$$

so that

$$\underline{W}_{i,j} = \underline{U}'_i \underline{T}'(\underline{THT}')^{-1} \underline{TU}_j, \text{ of order } m_i \times m_j, \text{ for } i, j = 1, \dots, c$$

$$\underline{W}_{i,c+1} = \underline{w}_i = \underline{U}'_i \underline{T}'(\underline{THT}')^{-1} \underline{T}_y, \text{ a vector of order } m_i, \text{ for } i = 1, \dots, c$$

$$\underline{W}_{c+1,c+1} = w = \underline{y}' \underline{T}'(\underline{THT}')^{-1} \underline{T}_y, \text{ a scalar.}$$

Then (16) and (18) are

$$\frac{\partial \lambda_i}{\partial Y_i} = -\frac{1}{2} \text{tr}(W_{ii}) + \frac{1}{2} w_i' w_i / \sigma^2 \quad (21)$$

and

$$\frac{\partial^2 \lambda_i}{\partial Y_i \partial Y_j} = \frac{1}{2} \text{tr}(W_{ij} W_{ij}') - \frac{w_i' W_{ij} w_j}{\sigma^2} \quad \text{for } i, j = 1, \dots, c \quad (22)$$

Additionally, (17) is

$$\hat{\sigma}^2 = w / (n - k) \quad (23)$$

To use the elements of \tilde{W} in (21) — (23) we need, from (20), the inverse of \tilde{THT}' , which has order $n - k$. For many data sets this will be impossibly large for the computing of $(\tilde{THT}')^{-1}$, but the following development reduces the inversion to

that of a matrix of order $m = \sum_{i=1}^c m_i$, the total number of levels of all random effects in the model. Although this itself may also be impossibly large for some data sets, it is always less than $n - k$, frequently much less, and in many instances will be such that the inversion is computable. Defining

$$\tilde{D} = \sum_{i=1}^c \gamma_i I_{i, m_i} \quad (24)$$

and

$$\tilde{Z} = \tilde{U} \tilde{D}^{-\frac{1}{2}}, \quad (25)$$

recalling that γ_i is by definition positive, we have from (3)

$$\tilde{H} = \tilde{I} + \tilde{Z} \tilde{Z}' \quad (26)$$

so that

$$\tilde{THT}' = \tilde{T}\tilde{T}' + \tilde{T}\tilde{Z}\tilde{Z}'\tilde{T}' \quad .$$

Then similar to the well-known result

$$\underline{\underline{H}}^{-1} = (\underline{\underline{I}} + \underline{\underline{ZZ}}')^{-1} = \underline{\underline{I}} - \underline{\underline{Z}}(\underline{\underline{I}} + \underline{\underline{Z}}'\underline{\underline{Z}})^{-1}\underline{\underline{Z}}' \quad (27)$$

we have

$$(\underline{\underline{THT}}')^{-1} = (\underline{\underline{TT}}')^{-1} - (\underline{\underline{TT}}')^{-1}\underline{\underline{TZ}}[\underline{\underline{I}} + \underline{\underline{Z}}'\underline{\underline{T}}'(\underline{\underline{TT}}')^{-1}\underline{\underline{TZ}}]^{-1}\underline{\underline{Z}}'\underline{\underline{T}}'(\underline{\underline{TT}}')^{-1} \quad (28)$$

which, when used in conjunction with (25), leads to $\underline{\underline{T}}'(\underline{\underline{THT}}')^{-1}\underline{\underline{T}}$ of (20) being

$$\underline{\underline{T}}'(\underline{\underline{THT}}')^{-1}\underline{\underline{T}} = \underline{\underline{S}} - \underline{\underline{SUM}}^{-1}\underline{\underline{U}}'\underline{\underline{S}} \quad (29)$$

for

$$\underline{\underline{M}} = \underline{\underline{D}}^{-1} + \underline{\underline{U}}'\underline{\underline{SU}}, \text{ of order } m_i = \sum_{i=1}^c m_i \quad (30)$$

Now define

$$\underline{\underline{W}}_0 = \begin{bmatrix} \underline{\underline{U}}' \\ \underline{\underline{y}} \end{bmatrix} \underline{\underline{S}} \begin{bmatrix} \underline{\underline{U}} & \underline{\underline{y}} \end{bmatrix} = \begin{bmatrix} \underline{\underline{W}}_{00} & \underline{\underline{w}}_0 \\ \underline{\underline{w}}_0' & \underline{\underline{w}}_0 \end{bmatrix} \quad (31)$$

which, because of (12) is $\underline{\underline{W}}$ with $\underline{\underline{H}}$ replaced by $\underline{\underline{I}}$. Then from $\underline{\underline{S}}$ of (7)

$$\underline{\underline{W}}_{00} = \underline{\underline{U}}'\underline{\underline{SU}} = \underline{\underline{U}}'\underline{\underline{U}} - \underline{\underline{U}}'\underline{\underline{X}} \text{diag}\{1/n_1 \dots 1/n_k\}\underline{\underline{X}}'\underline{\underline{U}} \quad (32)$$

where $\underline{\underline{U}}'\underline{\underline{U}}$ is the familiar "coefficient matrix" for the random effects; (i.e., if $\underline{\underline{\mu}}$ were null and the random effects were in fact fixed, the normal equations for them would be $\underline{\underline{U}}'\underline{\underline{U}}\underline{\underline{b}} = \underline{\underline{U}}'\underline{\underline{y}}$). And in (32) a typical sub-matrix in $\underline{\underline{U}}'\underline{\underline{X}}$ is

$\underline{\underline{U}}'\underline{\underline{X}} = \{n_{i(j),t}\}$ for $j = 1, \dots, m_i$ and $t = 1, \dots, k$,
 an $m_i \times k$ matrix whose typical element $n_{i(j),t}$
 is the number of observations in the j 'th level
 of the i 'th random effects factor and the t 'th
 sub-most cell of the fixed effects factors.

Also in (31)

$$\underline{w}_o = \underline{U}'\underline{S}\underline{y} = \{ \underline{U}'\underline{S}\underline{y} \}, m_i \times 1, \text{ for } i = 1, \dots, c .$$

From (7) $\underline{S}\underline{y} = \underline{z}$ is the vector \underline{y} with each observation replaced by its deviation from the cell mean of the sub-most cell of the fixed effects factors in which it occurs :

$$\underline{z} = \underline{S}\underline{y} = \underline{y} - \left(\sum_{t=1}^{k+} n_t^{-1} \underline{J}_{n_t} \right) \underline{y} = \underline{y} - [\{ \bar{y}_{t \cdot n_t} \} \text{ for } t = 1, \dots, k] . \quad (33)$$

Hence

$$\underline{w}_o = \{ \underline{U}_i \underline{z} \} = \{ \text{an } m_i \times 1 \text{ vector of totals of the } z \text{'s, totalled over each level of the } i \text{'th random factor} \} \quad (34)$$

for $i = 1, \dots, c$

and

$$\begin{aligned} w_o &= \underline{y}'\underline{S}\underline{y} = \text{total sum of squares of the } z \text{'s} \\ &= \text{within cell sum of squares of the } y \text{'s for the } k \text{ sub-most} \\ &\quad \text{cells of the fixed effects factors.} \end{aligned} \quad (35)$$

and on using (29) in (20), \underline{W} becomes

$$\begin{aligned} \underline{W} &= \underline{W}_o - \begin{bmatrix} \underline{U}' \\ \underline{y}' \end{bmatrix} \underline{S}^{-1} \underline{U}'\underline{S} \begin{bmatrix} \underline{U} \\ \underline{y} \end{bmatrix} \\ &= \begin{bmatrix} \underline{W}_{oo} & \underline{w}_o \\ \underline{w}'_o & \underline{w}_o \end{bmatrix} - \begin{bmatrix} \underline{W}_{oo} \\ \underline{w}'_o \end{bmatrix} \underline{M}^{-1} \begin{bmatrix} \underline{W}_{oo} & \underline{w}_o \end{bmatrix} \\ &= \begin{bmatrix} \underline{W}_{oo} - \underline{W}_{oo} \underline{M}^{-1} \underline{W}_{oo} & \underline{w}_o - \underline{W}_{oo} \underline{M}^{-1} \underline{w}_o \\ \underline{w}'_o - \underline{w}'_o \underline{M}^{-1} \underline{W}_{oo} & \underline{w}_o - \underline{w}'_o \underline{M}^{-1} \underline{w}_o \end{bmatrix} . \end{aligned} \quad (36)$$

Notice also, from (24), (30) and (32) that

$$\begin{aligned} \underline{\underline{M}} &= \underline{\underline{D}}^{-1} + \underline{\underline{W}}_{oo} \\ &= \sum_{i=1}^{c+} (1/\gamma_i) \underline{\underline{I}}_{m_i} + \underline{\underline{U}}'\underline{\underline{U}} - \underline{\underline{U}}'\underline{\underline{X}} \text{diag}\{1/n_1 \cdots 1/n_k\} \underline{\underline{X}}'\underline{\underline{U}} \end{aligned} \quad (37)$$

In this way $\underline{\underline{W}}$ of (20) for use in (21) — (23) is obtained from (36) using $\underline{\underline{W}}_o$ of (31) — (35) and $\underline{\underline{M}}^{-1}$ from (37). Since $\underline{\underline{M}}^{-1}$ has order m , it is more readily computed than is $(\underline{\underline{Y}}\underline{\underline{Y}}')^{-1}$ of order $n - k$ in (20).

With these expressions, implementation of the iterative solution of equations formed by equating (15) and (16) to zero can be carried out exactly as suggested by Hemmerle and Hartley [1973].

5. Estimation of Fixed Effects

The likelihood (5) has been partitioned into two parts $\lambda_1 + \lambda_2$ in (13) and (14), the first of which has provided the preceding TIML estimators of the variance components. Maximizing the second part, λ_2 , with respect to $\underline{\underline{\mu}}$ provides an estimator

$$\hat{\underline{\underline{\mu}}} = (\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{X}})^{-1}\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{y}} \quad (38)$$

of the fixed effects. Substituting from (27) for $\underline{\underline{H}}^{-1}$ and using (25) for $\underline{\underline{Z}}$ the terms of this expression are

$$\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{X}} = \underline{\underline{X}}'\underline{\underline{X}} - \underline{\underline{X}}'\underline{\underline{U}}(\underline{\underline{D}}^{-1} + \underline{\underline{U}}'\underline{\underline{U}})^{-1}\underline{\underline{U}}'\underline{\underline{X}}$$

and

$$\underline{\underline{X}}'\underline{\underline{H}}^{-1}\underline{\underline{y}} = \underline{\underline{X}}'\underline{\underline{y}} - \underline{\underline{X}}'\underline{\underline{U}}(\underline{\underline{D}}^{-1} + \underline{\underline{U}}'\underline{\underline{U}})^{-1}\underline{\underline{U}}'\underline{\underline{y}}$$

$$\tilde{\tilde{X}}'\tilde{\tilde{X}} = \text{diag}\{n_1, \dots, n_k\}$$

$\tilde{\tilde{X}}'\tilde{\tilde{y}} = \{y_t.\}$ for $t = 1, \dots, k$, a $k \times 1$ vector of the cell totals $y_t.$ of the sub-most cells of the fixed effects factors

$\tilde{\tilde{U}}'\tilde{\tilde{y}}$ is an $m \times 1$ vector of y -totals for each level of the random effects factors

and where $\tilde{\tilde{X}}'\tilde{\tilde{U}}$ is described below (32) and $\tilde{\tilde{D}}^{-1} + \tilde{\tilde{U}}'\tilde{\tilde{U}}$ is, from (37), part of $\tilde{\tilde{M}}$. With the TML estimators of the γ_i used in $\tilde{\tilde{D}}^{-1}$, an estimator of $\tilde{\tilde{\mu}}$ based on those estimators is obtained. If $\tilde{\tilde{H}}$ is the value of $\tilde{\tilde{H}}$ when the TML estimator $\tilde{\tilde{\gamma}}$ is used in place of $\tilde{\tilde{Y}}$ in $\tilde{\tilde{H}}$ the corresponding estimator of $\tilde{\tilde{\mu}}$ will be

$$\tilde{\tilde{\mu}} = (\tilde{\tilde{X}}'\tilde{\tilde{H}}\tilde{\tilde{X}})^{-1}\tilde{\tilde{X}}'\tilde{\tilde{H}}^{-1}\tilde{\tilde{y}} .$$

The true covariance matrix of this estimator is

$$\text{var}(\tilde{\tilde{\mu}}) = (\tilde{\tilde{X}}'\tilde{\tilde{H}}\tilde{\tilde{X}})^{-1}\tilde{\tilde{X}}'\tilde{\tilde{H}}^{-1}\tilde{\tilde{H}}\tilde{\tilde{H}}^{-1}\tilde{\tilde{X}}(\tilde{\tilde{X}}'\tilde{\tilde{H}}\tilde{\tilde{X}})^{-1}\sigma^2$$

and if the TML estimators are used for $\tilde{\tilde{H}}$ and σ^2 this becomes

$$\tilde{\tilde{\text{var}}}(\tilde{\tilde{\mu}}) = (\tilde{\tilde{X}}'\tilde{\tilde{H}}\tilde{\tilde{X}})^{-1}\tilde{\tilde{\sigma}}^2 ,$$

the covariance matrix of $\tilde{\tilde{\mu}}$ using $\tilde{\tilde{H}}\tilde{\tilde{\sigma}}^2$ in place of $\tilde{\tilde{H}}\sigma^2$.

6. Large Sample Variances

The preceding TML estimators of σ^2 and of $\gamma_i = \sigma_i^2/\sigma^2$ for $i = 1, \dots, c$ have been derived from (13) which is the logarithm of the likelihood of $\tilde{\tilde{T}}_y$. They are

therefore the maximum likelihood estimators based on $\underline{\underline{T}}\underline{\underline{y}}$. Define

$$\sigma_0^2 \equiv \sigma^2 \quad (39)$$

and

$$\underline{\underline{\sigma}}^2 = \begin{bmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \vdots \\ \sigma_c^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ \sigma^2 \underline{\underline{\gamma}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \sigma^2 \underline{\underline{I}}_c \end{bmatrix} \begin{bmatrix} \sigma^2 \\ \underline{\underline{\gamma}} \end{bmatrix}. \quad (40)$$

Then, because $\underline{\underline{T}}\underline{\underline{y}} \sim N(0, \underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'\sigma^2)$, the covariance matrix of the large-sample maximum likelihood estimator of $\underline{\underline{\sigma}}^2$ is, from Searle [1971],

$$\text{var } \underline{\underline{\sigma}}^2 = 2\underline{\underline{P}}^{-1} = 2\{p_{ij}\}^{-1} \quad (41)$$

with

$$p_{ij} = \text{tr} \left[\left(\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'\sigma^2 \right)^{-1} \frac{\partial (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'\sigma^2)}{\partial \sigma_i^2} \left(\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'\sigma^2 \right)^{-1} \frac{\partial (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}'\sigma^2)}{\partial \sigma_j^2} \right] \quad (42)$$

for $i, j = 0, 1, 2, \dots, c$.

The elements p_{ij} are readily obtained from

$$\underline{\underline{H}}\sigma^2 = \sum_i \sigma_i^2 \underline{\underline{U}}_i \underline{\underline{U}}_i' + \sigma_{0\sim r}^2 \underline{\underline{I}}$$

of (3). Thus

$$p_{00} = \text{tr} \left[(1/\sigma^2) (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}')^{-1} \underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}' \right]^2 = (n-k)/\sigma^4$$

$$\begin{aligned} p_{0j} &= \text{tr} \left[(1/\sigma^2) (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}')^{-1} \underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}' (1/\sigma^2) (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}')^{-1} \underline{\underline{T}}\underline{\underline{U}}_j \underline{\underline{U}}_j' \underline{\underline{T}}' \right] \\ &= 1/\sigma^4 \text{tr} \left[\underline{\underline{U}}_j' \underline{\underline{T}}' (\underline{\underline{T}}\underline{\underline{H}}\underline{\underline{T}}')^{-1} \underline{\underline{T}}\underline{\underline{U}}_j \right] \\ &= \text{tr}(\underline{\underline{W}}_{jj})/\sigma^4 \text{ from (20), for } j = 1, \dots, c \end{aligned}$$

and

$$\begin{aligned}
 p_{ij} &= \text{tr}[(1/\sigma^2)(\underline{THT}')^{-1} \underline{TU_i U_i' T'} (1/\sigma^2) \underline{THT}' \underline{TU_j U_j' T'}] \\
 &= \text{tr}(\underline{W}_{ij} \underline{W}'_{ij}) / \sigma^4 \quad \text{from (20), for } i, j = 1, \dots, c .
 \end{aligned}$$

Hence

$$\text{var}(\tilde{\sigma}^2) = 2 \begin{bmatrix} (n-k)/\sigma^4 & \{\text{tr}(\underline{W}_{jj})\}' / \sigma^4 \\ \{\text{tr}(\underline{W}_{jj})\} / \sigma^4 & \{\text{tr}(\underline{W}_{ij} \underline{W}'_{ij})\} / \sigma^4 \end{bmatrix}^{-1} \quad (43)$$

for $i, j = 1, \dots, c$.

so that from (40)

$$\text{var} \begin{bmatrix} \tilde{\sigma}^2 \\ \tilde{\gamma} \\ \sim \end{bmatrix} = 2 \begin{bmatrix} (n-k)/\sigma^4 & \{\text{tr}(\underline{W}_{jj})\}' / \sigma^2 \\ \{\text{tr}(\underline{W}_{jj})\} / \sigma^2 & \{\text{tr}(\underline{W}_{ij} \underline{W}'_{ij})\} \end{bmatrix}^{-1} \quad (44)$$

for $i, j = 1, \dots, c$.

These results can also be obtained from second differentials of (15) and (16), by taking expectations and changing their signs. Estimators of (42) and (43) are obtained by using the TIML's in place of σ^2 and γ .

Note in passing that because the matrix in (44) is positive definite, the value of $\text{var}(\tilde{\sigma}^2)$ obtained from (44) will exceed $2\sigma^4/(n-k)$ which is the variance of $\hat{\sigma}^2$ of (17). This result is not unexpected since (17) as it stands assumes γ known whereas (44) is based on estimating σ^2 and the c elements of γ .

References

- Hartley, H. O. and Rao, J. N. K. [1967] Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 54, 93-108.
- Hartley, H. O. and Vaughan, W. K. [1972] A computer program for the mixed analysis of variance model based on maximum likelihood. In Statistical Papers in Honor of George W. Snedecor, Ed. T. A. Bancroft, 129-144, Iowa State University Press, Ames.
- Hemmerle, W. J. and Hartley, H. O. [1973] Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. Technometrics, 15, 819-832.
- LaMotte, L. R. [1973] Quadratic estimation of variance components. Biometrics, 29, 311-330.
- Patterson, H. D. and Thompson, R. [1971] Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.
- Rao, C. R. [1971] Estimation of variance and covariance components -- minque theory. J. of Multivariate Analysis, 1, 257-275.
- Searle, S. R. [1970] Large sample variances of maximum likelihood estimators of variance components using unbalanced data. Biometrics, 26, 505-524.

APPENDIX: The Iterative Procedure

Implementation of the Newton-Rhapson technique follows Hemmerle and Hartley [1973] very closely.

Since by definition the parameter γ_i is positive, negative values are avoided in the iterative process by defining

$$\tau_i = \sqrt{\gamma_i} . \quad (45)$$

We then have

$$\frac{\partial \lambda_1}{\partial \tau_i} = 2\tau_i \frac{\partial \lambda_1}{\partial \gamma_i} \quad (46)$$

and

$$\frac{\partial^2 \lambda_1}{\partial \tau_i \partial \tau_j} = 4\tau_i \tau_j \frac{\partial^2 \lambda_1}{\partial \gamma_i \partial \gamma_j} + 2\delta_{ij} \frac{\partial \lambda_1}{\partial \gamma_i} \quad (47)$$

where δ_{ij} is the Kronecker delta. The iterating is done on the τ_i and at each round γ_i is taken as τ_i^2 . On defining, from (21) and (22),

$$\tilde{f}^*(\tilde{\gamma}) = \left\{ \frac{\partial \lambda_1}{\partial \gamma_i} \right\} = \left\{ -\frac{1}{2} \text{tr}(W_{ii}) + \frac{1}{2} w_{ii}' w_{ii} / \sigma^2 \right\}, \quad c \times 1 \quad (48)$$

$$\tilde{G}^*(\tilde{\gamma}) = \left\{ \frac{\partial^2 \lambda_1}{\partial \gamma_i \partial \gamma_j} \right\} = \left\{ \frac{1}{2} \text{tr}(W_{ij} W_{ij}') - w_{ij}' w_{ij} / \sigma^2 \right\} \quad (49)$$

for $i, j = 1, \dots, c$

and

$$\tilde{D}_1(\tilde{\tau}) = \text{diag}\{\tau_1 \dots \tau_c\}$$

and

$$\tilde{D}_2(\tilde{\gamma}) = \text{diag}\left\{ \frac{\partial \lambda_1}{\partial \gamma_1} \dots \frac{\partial \lambda_1}{\partial \gamma_c} \right\} = \text{diag}\{f_1^*(\tilde{\gamma}) \dots f_c^*(\tilde{\gamma})\} ,$$

we then have from (46) and (47)

$$\underline{f}(\underline{\tau}) = \left\{ \frac{\partial \lambda}{\partial \tau_i} \right\} = 2\underline{D}_1(\underline{\tau})\underline{f}^*(\underline{\gamma}) \quad (50)$$

and

$$\underline{G}(\underline{\tau}) = \left\{ \frac{\partial^2 \lambda}{\partial \tau_i \partial \tau_j} \right\} = 4\underline{D}_1(\underline{\tau})\underline{G}^*(\underline{\gamma})\underline{D}_1(\underline{\tau}) + 2\underline{D}_2(\underline{\gamma}) \quad (51)$$

Suppose $\underline{\gamma}^{(r)}$ is an approximate value of $\underline{\gamma}$ that makes $\underline{f}^*(\underline{\gamma})$ of (48) null and hence (21) zero, and $\sigma^{2(r)}$ is the corresponding value of σ^2 from (23). Let $\underline{\tau}^{(r)}$ be the vector of values $\sqrt{\gamma_i^{(r)}}$. Then the Newton-Rhapson method is based on the approximation

$$\underline{f}(\underline{\tau}^{(r+1)}) + \Delta \underline{\tau} \cong \underline{f}(\underline{\tau}^{(r)}) + \underline{G}(\underline{\tau}^{(r)}) \Delta \underline{\tau}$$

so that if the left-hand side is to be zero, which is what we want,

$$\Delta \underline{\tau} = - \left[\underline{G}(\underline{\tau}^{(r)}) \right]^{-1} \underline{f}(\underline{\tau}^{(r+1)}) \quad (52)$$

and

$$\underline{\tau}^{(r+1)} = \left\{ \tau_i^{(r+1)} \right\} \quad \text{for } \tau_i^{r+1} = |\tau_i^{(r)} + \Delta \tau_i| \quad (53)$$

and

$$\gamma_i^{(r+1)} = \left[\tau_i^{(r+1)} \right]^2, \quad \text{for } i = 1, \dots, c \quad (54)$$

Using (54) in (23) gives $\sigma^{2(r+1)}$, and the approximation can be repeated. The iterative procedure is therefore initially

(i) to calculate W_0 from (31) — (35)

and (ii) to assign initial values $\underline{\gamma}^{(1)}$ to $\underline{\gamma}$,

and then the r'th round of the iteration consists of the following steps.

<u>Step</u>	<u>Calculation</u>	<u>Equation</u>
1	\underline{M} and \underline{M}^{-1} , using $\underline{y}^{(r)}$ for \underline{y}	(37)
2	\underline{W}	(36)
3	$\sigma^{2(r)} = (\underline{w}_0 - \underline{w}_0' \underline{M}^{-1} \underline{w}_0) / (n - k)$	(23) and (36)
4	$\underline{f}^*(\underline{y}^{(r)})$ and $\underline{G}^*(\underline{y}^{(r)})$	(48) and (49)
5	$\underline{f}(\underline{\tau}^{(r)})$ and $\underline{G}(\underline{\tau}^{(r)})$	(50) and (51)
6	$\Delta \underline{\tau}$, $\underline{\tau}^{(r+1)}$, and $\underline{y}^{(r+1)}$	(52), (53) and (54)
7	Return to step 1	

The (r+1)'th iteration ends after step 3, and the procedure is complete whenever sufficient accuracy has been obtained.

The estimators are designed to maximize the likelihood λ_1 given in (13). If $\hat{\sigma}^2$ satisfies (17) then (13) becomes

$$\hat{\lambda}_1 = -\frac{1}{2}(n-1)(\log 2\pi + 1) - \frac{1}{2}[(n-k)\log \hat{\sigma}^2 + \log |\hat{THT}'|] \quad (55)$$

Since the estimates are calculated by iteration it will be of interest to evaluate $\hat{\lambda}_1$ at successive rounds of the iteration. Clearly, from (55) we need only look at

$$(n-k)\log \hat{\sigma}^2 + \log |\hat{THT}'| \quad (56)$$

which can be expected to decrease at each round. To evaluate $|\hat{THT}'|$ we use the determinant of a partitioned matrix.

$$\begin{aligned}
 & \left| \begin{array}{cc} \underline{\underline{M}} & \underline{\underline{U}}' \underline{\underline{T}}' (\underline{\underline{TT}}')^{-1} \\ (\underline{\underline{TT}}')^{-1} \underline{\underline{TU}} & (\underline{\underline{TT}}')^{-1} \end{array} \right| \\
 & \equiv |\underline{\underline{M}}| |(\underline{\underline{TT}}')^{-1} - (\underline{\underline{TT}}')^{-1} \underline{\underline{TU}} \underline{\underline{U}}^{-1} \underline{\underline{U}}' \underline{\underline{T}}' (\underline{\underline{TT}}')^{-1}| \\
 & = |\underline{\underline{M}}| |(\underline{\underline{THT}}')^{-1}| \quad \text{by (28)} \\
 & \equiv |(\underline{\underline{TT}}')^{-1}| |\underline{\underline{M}} - \underline{\underline{U}}' \underline{\underline{T}}' (\underline{\underline{TT}}')^{-1} \underline{\underline{TT}}' (\underline{\underline{TT}}')^{-1} \underline{\underline{TU}}| \\
 & = |(\underline{\underline{TT}}')^{-1}| |\underline{\underline{D}}^{-1}| \quad \text{by (30)} .
 \end{aligned}$$

Hence

$$|\underline{\underline{THT}}'| = |\underline{\underline{M}}| |\underline{\underline{D}}| |\underline{\underline{TT}}'|$$

and so (56) becomes

$$(n-k) \log \hat{\sigma}^2 + \log |\hat{\underline{\underline{M}}}| + \log |\hat{\underline{\underline{D}}}| \tag{57}$$

since $\log |\underline{\underline{TT}}'|$ is constant insofar as successive iterations are concerned. This in turn is

$$(n-k) \log \hat{\sigma}^2 + \sum_{i=1}^c m_i \log \hat{\gamma}_i + \log |\hat{\underline{\underline{M}}}| . \tag{58}$$