THE UNIQUENESS OF METHOD 2 OF ESTIMATING VARIANCE COMPONENTS*

C. R. Henderson[1], S. R. Searle[2] and L. R. Schaeffer[1]

[1]Animal Science Department and [2]Biometrics Unit,
Cornell University, Ithaca, N.Y.

Abstract

Estimating the variance components of a mixed model by Method 2 of Henderson [1953] is shown to be a well-defined procedure so long as there are no interactions between the fixed and random factors of the model.

1. Introduction

Three methods of estimating variance components from unbalanced data are given in Henderson [1953]. Method 2, designed to circumvent biasedness that results from using Method 1 on mixed models, is characterized more generally in Searle [1968] as a special way of executing what is described there as a simplified form of a Generalized Method 2. In that description 2 limitations are discussed: (1) it is suggested that the Generalized Method 2 is not uniquely specified, and (2) it is proven that the simplified form of the Generalized Method 2 can be used only when the model has no interactions between fixed and random effects. From this it is concluded that Henderson's Method 2 suffers from both these limitations.

There is no doubt about limitation (2): Henderson's Method 2 can be used only when the model has no interactions between fixed and random effects. However,

*Paper BU-437-M in the Biometrics Unit, Cornell University, Ithaca, N. Y.

we show here that Henderson's Method 2, even though it is one of the methods that can collectively be called the Generalized Method 2, does not suffer from limitation (1). Indeed it is limitation (2) which causes Henderson's Method 2 not to be affected by limitation (1); i.e., Henderson's Method 2 is well defined. This we now prove.

2. The Model

We follow the same notation as Searle [1968]. Let the equation of the model for the vector of observations $\underline{y}$ be

$$\begin{aligned} \underline{y} &= \mu\underline{1} + \underline{X}\underline{\beta} + \underline{e} \\ &= \mu\underline{1} + \underline{X}_f\underline{\beta}_f + \underline{X}_r\underline{\beta}_r + \underline{e} \quad , \end{aligned} \tag{1}$$

where μ is a general mean, $\underline{1}$ is a vector of 1's, $\underline{\beta}_f$ is the vector of all the fixed effects in the model (other than μ), $\underline{\beta}_r$ is the vector of all the random effects, $\underline{X}_f$ and $\underline{X}_r$ are the corresponding coefficient (design) matrices and $\underline{e}$ is a vector of random error terms. As in the Appendix of Searle [1968] we partition $\underline{X}_f$ and $\underline{X}_r$ as

$$\underline{X}_f = (\underline{F}_1 \quad \underline{F}_2) \qquad \text{and} \qquad \underline{X}_r = (\underline{R}_1 \quad \underline{R}_2) \tag{2}$$

and $\underline{\beta}_f$ and $\underline{\beta}_r$ correspondingly as

$$\underline{\beta}_f = (\underline{\beta}_{f1} \quad \underline{\beta}_{f2}) \qquad \text{and} \qquad \underline{\beta}_r = (\underline{\beta}_{r1} \quad \underline{\beta}_{r2}) \quad .$$

The normal equations for the model (1) are then

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{F}_1 & \mathbf{1}'\mathbf{F}_2 & \mathbf{1}'\mathbf{R}_1 & \mathbf{1}'\mathbf{R}_2 \\ \mathbf{F}_1'\mathbf{1} & \mathbf{F}_1'\mathbf{F}_1 & \mathbf{F}_1'\mathbf{F}_2 & \mathbf{F}_1'\mathbf{R}_1 & \mathbf{F}_1'\mathbf{R}_2 \\ \mathbf{F}_2'\mathbf{1} & \mathbf{F}_2'\mathbf{F}_1 & \mathbf{F}_2'\mathbf{F}_2 & \mathbf{F}_2'\mathbf{R}_1 & \mathbf{F}_2'\mathbf{R}_2 \\ \mathbf{R}_1'\mathbf{1} & \mathbf{R}_1'\mathbf{F}_1 & \mathbf{R}_1'\mathbf{F}_1 & \mathbf{R}_1'\mathbf{R}_1 & \mathbf{R}_1'\mathbf{R}_2 \\ \mathbf{R}_2'\mathbf{1} & \mathbf{R}_2'\mathbf{F}_1 & \mathbf{R}_2'\mathbf{F}_1 & \mathbf{R}_2'\mathbf{R}_1 & \mathbf{R}_2'\mathbf{R}_2 \end{bmatrix} \begin{bmatrix} \mu^o \\ \boldsymbol{\beta}_{f1}^o \\ \boldsymbol{\beta}_{f2}^o \\ \boldsymbol{\beta}_{r1}^o \\ \boldsymbol{\beta}_{r2}^o \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{F}_1'\mathbf{y} \\ \mathbf{F}_2'\mathbf{y} \\ \mathbf{R}_1'\mathbf{y} \\ \mathbf{R}_2'\mathbf{y} \end{bmatrix} . \tag{3}$$

The superscript o on μ^o and the $\boldsymbol{\beta}$-symbols in these equations merely indicates that these are solutions to the normal equations.

The Appendix of Searle [1968] discusses in detail the partitioning in (2), which is done in such a way that $(\mathbf{F}_2 \;\; \mathbf{R}_1)$ has full column rank, equal to the rank of $(\mathbf{1} \;\; \mathbf{X})$. Then, with

$$\begin{bmatrix} \mathbf{F}_2'\mathbf{F}_2 & \mathbf{F}_2'\mathbf{R}_1 \\ \mathbf{R}_1'\mathbf{F}_2 & \mathbf{R}_1'\mathbf{R}_1 \end{bmatrix}^{-1} \equiv \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}' & \mathbf{Q}_{22} \end{bmatrix} \text{ say,} \tag{4}$$

a solution to the normal equations (3) is

$$\begin{bmatrix} \mu^o \\ \boldsymbol{\beta}_{f1}^o \\ \boldsymbol{\beta}_{f2}^o \\ \boldsymbol{\beta}_{r1}^o \\ \boldsymbol{\beta}_{r2}^o \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{11} & \mathbf{Q}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{12}' & \mathbf{Q}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{F}_1'\mathbf{y} \\ \mathbf{F}_2'\mathbf{y} \\ \mathbf{R}_1'\mathbf{y} \\ \mathbf{R}_2'\mathbf{y} \end{bmatrix} . \tag{5}$$

Henderson's Method 2 consists of carrying out Method 1 on

$$
\begin{aligned}
\underset{\sim}{z} &= \underset{\sim}{y} - \underset{\sim}{X}_f \underset{\sim}{\beta}_f^o \\
&= \underset{\sim}{y} - \underset{\sim}{F}_2 \underset{\sim}{\beta}_{f1}^o \ , \quad \text{from (2) and (5)} \ ,
\end{aligned} \tag{6}
$$

where $\beta_f^{o'} = (0 \quad \beta_{f2}^{o'})$ is obtained from (5). Operationally, the crux of the method in making the partitioning (2) has been specified (see Searle [1968]) as choosing F_1 to have as many columns in it as possible and R_2 to have as few as possible. Conversely, F_2 must have as few columns as possible and R_1 as many as possible. We now show first, that this can always be done for mixed models that have no interactions between fixed and random effects, and second, that this leads to Henderson's Method 2 being invariant to whatever solution of the normal equations (5) is used for β_f^o in (6). To establish these results we utilize certain properties of rank and consequent relationships between the matrices involved. These are developed in Section 3 and 4.

3. Considerations of rank

To show that the partitioning (2) can always be done with R_1 having as many columns as possible, let us choose an R_1 of that nature and investigate the consequences. Denote the rank of any matrix A by $r(A)$. Because $(F_2 \quad R_1)$ is to have full column rank, for (4) to exist, so also will F_2 and R_1. And

$$r(F_2 \quad R_1) = r(F_2) + r(R_1) \quad . \tag{7}$$

Furthermore, with R_1 having full column rank it will have as many columns as possible when

$$r(R_1) = r(R_1 \quad R_2) = r(X_2) \quad . \tag{8}$$

Henderson's Method 2 demands that there be no interactions between fixed and random effects. This means that the only relationship between columns of $\underset{-}{X}_f$ and those of $\underset{-}{X}_r$ is that certain groups of columns of $\underset{-}{X}_f$ sum to $\underset{-}{1}$, as do certain groups of columns of $\underset{-}{X}_r$. This represents only one linearly independent relationship between columns of $\underset{-}{X}_f$ and those of $\underset{-}{X}_r$. [Even though there may be several such groups of columns in $\underset{-}{X}_f$ (and in $\underset{-}{X}_r$) the interrelationship of the groups affects just the rank of $\underset{-}{X}_f$ (and $\underset{-}{X}_r$) with only one additional relationship being involved in the rank of $(\underset{-}{X}_f \;\; \underset{-}{X}_r)$.] Hence

$$r(\underset{-}{X}_f \;\; \underset{-}{X}_r) = r(\underset{-}{X}_f) + r(\underset{-}{X}_r) - 1 \quad . \tag{9}$$

Also, because (5) is a solution to (3) the matrices $(\underset{-}{F}_2 \;\; \underset{-}{R}_1)$ and $(\underset{-}{1} \;\; \underset{-}{X})$ have the same rank and so

$$r(\underset{-}{F}_2 \;\; \underset{-}{R}_1) = r(\underset{-}{1} \;\; \underset{-}{X}) = r(\underset{-}{1} \;\; \underset{-}{X}_f \;\; \underset{-}{X}_r) \quad .$$

Since certain columns of $\underset{-}{X}_f$ (and $\underset{-}{X}_r$) sum to $\underset{-}{1}$ this means

$$r(\underset{-}{F}_2 \;\; \underset{-}{R}_1) = r(\underset{-}{X}_f \;\; \underset{-}{X}_r) \quad .$$

Hence from (7)

$$\begin{aligned} r(\underset{-}{F}_2) + r(\underset{-}{R}_1) &= r(\underset{-}{X}_f \;\; \underset{-}{X}_r) \\ &= r(\underset{-}{X}_f) + r(\underset{-}{X}_r) - 1 \;, \quad \text{from (9)} \end{aligned}$$

and so, using (8),

$$r(\underset{-}{F}_2) = r(\underset{-}{X}_f) - 1 \;, \tag{10}$$

or from (2)

$$r(\underset{-}{F}_2) = r(\underset{-}{F}_1 \;\; \underset{-}{F}_2) - 1 \quad .$$

As a condition on $\underset{\sim}{F}_2$, of full column rank, this is easily satisfied. Thus by choosing $\underset{\sim}{R}_1$ to have as many columns as possible, one can also choose $\underset{\sim}{F}_2$ so that $(\underset{\sim}{F}_2 \;\; \underset{\sim}{R}_1)$ has full column rank equal to the rank of $(\underset{\sim}{1} \;\; \underset{\sim}{X})$ and the solution (5) then exists. Hence, so long as there are no interactions between fixed and random effects solution (5) exists.

Three other results are useful. Since $\underset{\sim}{R}_1$ and $\underset{\sim}{X}_r = (\underset{\sim}{R}_1 \;\; \underset{\sim}{R}_2)$ have the same rank, $\underset{\sim}{R}_1$ contains the columns pertaining to <u>all</u> levels of one of the random factors in the model. Denoting these columns by $\underset{\sim}{R}_{12}$ we partition $\underset{\sim}{R}_1$ as in the Appendix of Searle [1968] as

$$\underset{\sim}{R}_1 = (\underset{\sim}{R}_{11} \;\; \underset{\sim}{R}_{12} \;\; \underset{\sim}{R}_{13}) \tag{11}$$

where $\underset{\sim}{R}_{11}$ and/or $\underset{\sim}{R}_{13}$ may be dimensionless, and where

$$\underset{\sim}{R}_{12}\underset{\sim}{1} = \underset{\sim}{1} \quad ; \tag{12}$$

i.e., those columns of $\underset{\sim}{R}_1$ constituting $\underset{\sim}{R}_{12}$ sum to $\underset{\sim}{1}$. Because of this, and because $(\underset{\sim}{F}_2 \;\; \underset{\sim}{R}_1)$ has full column rank, it follows that no columns of $\underset{\sim}{F}_2$ sum to $\underset{\sim}{1}$. Therefore

$$\begin{aligned} r(\underset{\sim}{1} \;\; \underset{\sim}{F}_2) &= r(\underset{\sim}{F}_2) + 1 \\ &= r(\underset{\sim}{X}_f), \qquad \text{from (10)} \\ &= r(\underset{\sim}{1} \;\; \underset{\sim}{X}_f) \\ &= r(\underset{\sim}{1} \;\; \underset{\sim}{F}_1 \;\; \underset{\sim}{F}_2) \quad . \end{aligned} \tag{13}$$

4. Some Matrix Results

The rank relationship in (13) indicates that for some matrix $\underset{\sim}{T}$, $\underset{\sim}{F}_1 = (\underset{\sim}{1} \;\; \underset{\sim}{F}_2)\underset{\sim}{T}$ which can be rewritten as

$$\underset{\sim}{F}_1 = (\underset{\sim}{1} \;\; \underset{\sim}{F}_2) \begin{bmatrix} \underset{\sim}{t}' \\ \underset{\sim}{M} \end{bmatrix} \tag{14}$$

for some row vector $\mathbf{t}'$ and some matrix $\mathbf{M}$. Similarly, (8) indicates that

$$\mathbf{R}_2 = \mathbf{R}_1\mathbf{K}_1 \qquad \text{for some matrix } \mathbf{K}_1 \quad . \tag{15}$$

Then from (14)

$$(\mathbf{1} \quad \mathbf{F}_1) = \left[\mathbf{1} \qquad (\mathbf{1} \quad \mathbf{F}_2)\begin{pmatrix}\mathbf{t}'\\ \mathbf{M}\end{pmatrix}\right]$$

$$= [\mathbf{1} \quad (\mathbf{1}\mathbf{t}' + \mathbf{F}_2\mathbf{M})]$$

$$= \mathbf{F}_2(\mathbf{0} \quad \mathbf{M}) + (\mathbf{1} \quad \mathbf{1}\mathbf{t}') \quad .$$

Now by using (12) we have

$$(\mathbf{1} \quad \mathbf{1}\mathbf{t}') = (\mathbf{R}_{12}\mathbf{1} \quad \mathbf{R}_{12}\mathbf{1}\mathbf{t}') = (\mathbf{R}_{11} \quad \mathbf{R}_{12} \quad \mathbf{R}_{13})\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{1} & \mathbf{1}\mathbf{t}'\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$

and on using (11) and defining

$$\mathbf{K}_3 \equiv \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{1} & \mathbf{1}\mathbf{t}'\\ \mathbf{0} & \mathbf{0}\end{bmatrix} = (\mathbf{u} \quad \mathbf{u}\mathbf{t}') \text{ for } \mathbf{u} \equiv \begin{bmatrix}\mathbf{0}\\ \mathbf{1}\\ \mathbf{0}\end{bmatrix} \tag{17}$$

this becomes

$$(\mathbf{1} \quad \mathbf{1}\mathbf{t}') = \mathbf{R}_1\mathbf{K}_3 \quad . \tag{18}$$

And on defining

$$\mathbf{K}_2 \equiv (\mathbf{0} \quad \mathbf{M}) \tag{19}$$

we see that (16) is

$$(\mathbf{1} \quad \mathbf{F}_1) = \mathbf{F}_2\mathbf{K}_2 + \mathbf{R}_1\mathbf{K}_3 \quad . \tag{20}$$

Note from the left-hand side of (18) that

$$\underline{R}_1\underline{K}_3 \text{ has all its rows the same.} \tag{21}$$

These properties play an important part in what follows.

5. Invariance of the Method

Henderson's Method 2 involves applying Method 1 to $\underline{z} = \underline{y} - \underline{X}_f\underline{\beta}_f^o$ of (6). From Searle [1968, equations (15), (16) and (41)] we have

$$\underline{z} = \mu_1\underline{1} + \underline{X}_r\underline{\beta}_r + (\underline{I} - \underline{X}_f\underline{L})\underline{e} \tag{22}$$

for some scalar μ_1 different from μ, and for

$$\underline{L}' = (\underline{0} \quad \underline{V}') \qquad \text{with} \qquad \underline{V} = \underline{Q}_{11}\underline{F}_2' + \underline{Q}_{12}\underline{R}_1' \quad . \tag{23}$$

On using (2) this gives

$$\underline{z} = \mu_1\underline{1} + \underline{X}_r\underline{\beta}_r + (\underline{I} - \underline{F}_2\underline{V})\underline{e} \ , \tag{24}$$

This is the vector of (adjusted) observations, to which Method 1 is applied. Apart from the term in $\underline{e}$, it has the same model as $\underline{y}$ with the fixed effects excluded.

We note in passing that for $\underline{V}$ of (23) the inverse matrix defined in (4) gives

$$\underline{V}\underline{F}_2 = \underline{I} \qquad \text{and} \qquad \underline{V}\underline{R}_1 = \underline{0} \tag{25}$$

We now show that Method 2 is invariant to whatever solution of the normal equations (3) is used for μ^o and β^o. To do so we derive another $\underline{z}$, call it $\underline{z}^*$, using some solution to (3) other than (5) with $\underline{F}_2$ and $\underline{R}_1$ as defined in Sections 3 and 4. Call this solution $(\mu_1^* \quad \underline{\beta}_1^{*\prime})$. Whatever it may be, it is just some solution different from (5) and so is related to (5), through the familiar properties of solutions of non-full rank equations, as follows.

We have

$$\underline{z}^* = \underline{y} - \underline{X}_f \underline{\beta}_f^* \tag{26}$$

for $\underline{\beta}^{*\prime} = (\underline{\beta}_f^{*\prime} \quad \underline{\beta}_r^{*\prime})$ where $(\mu^* \quad \underline{\beta}^{*\prime})$ is a solution of (3) different from $(\mu^\circ \quad \underline{\beta}^{\circ\prime})$ of (5). Since these are both solutions to the same equations, (3), this new solution can (see, e.g., Searle [1968], Theorem 3, p. 11) be expressed as

$$\begin{bmatrix} \mu^* \\ \underline{\beta}^* \end{bmatrix} = \begin{bmatrix} \mu^\circ \\ \underline{\beta}^\circ \end{bmatrix} + (\underline{H} - \underline{I})\underline{m} \tag{27}$$

where $\underline{m}$ is an arbitrary vector and

$$\underline{H} = \underline{G}\begin{bmatrix} \underline{1}' \\ \underline{X}' \end{bmatrix}(\underline{1} \quad \underline{X}) ,$$

with $\underline{G}$ being the matrix on the right-hand side of (5) and $\begin{bmatrix} \underline{1}' \\ \underline{X}' \end{bmatrix}(\underline{1} \quad \underline{X})$ being the matrix on the left-hand side of (3). Multiplying these gives

$$\underline{H} = \begin{bmatrix} 0 & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{V}\underline{1} & \underline{V}\underline{F}_1 & \underline{V}\underline{F}_2 & \underline{V}\underline{R}_1 & \underline{V}\underline{R}_2 \\ \underline{S}\underline{1} & \underline{S}\underline{F}_1 & \underline{S}\underline{F}_2 & \underline{S}\underline{R}_1 & \underline{S}\underline{R}_2 \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \end{bmatrix},$$

where $\underline{S}$ is

$$\underline{S} \equiv \underline{Q}_{12}'\underline{F}_2' + \underline{Q}_{22}\underline{R}_1' \qquad \text{with} \qquad \underline{S}\underline{F}_2 = \underline{0} \qquad \text{and} \qquad \underline{S}\underline{R}_1 = \underline{I} \tag{28}$$

from (4). With (28), (15) and (25), $\underline{H}$ becomes

$$\underline{H} = \begin{bmatrix} 0 & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{V}(\underline{1} & \underline{F}_1) & \underline{I} & \underline{0} & \underline{0} \\ \underline{S}(\underline{1} & \underline{F}_1) & \underline{0} & \underline{I} & \underline{K}_1 \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \end{bmatrix}$$

On using (19), (20) and (25)

$$\underline{V}(\underline{1} \quad \underline{F}_1) = \underline{V}\underline{F}_2\underline{K}_2 + \underline{V}\underline{R}_1\underline{K}_3 = \underline{K}_2 = (\underline{0} \quad \underline{M})$$

and from (17), (20) and (28)

$$\underline{S}(\underline{1} \quad \underline{F}_1) = \underline{S}\underline{F}_2\underline{K}_2 + \underline{S}\underline{R}_1\underline{K}_3 = \underline{K}_3 = (\underline{u} \quad \underline{u}\underline{t}') \ ,$$

so that substituting these results in (29) and then (27) gives

$$\begin{bmatrix} \mu^* \\ \underline{\beta}^*_{f1} \\ \underline{\beta}^*_{f2} \\ \underline{\beta}^*_{r1} \\ \underline{\beta}^*_{r2} \end{bmatrix} = \begin{bmatrix} \mu^o \\ \underline{\beta}^o_{f1} \\ \underline{\beta}^o_{f2} \\ \underline{\beta}^o_{r1} \\ \underline{\beta}^o_{r2} \end{bmatrix} + \begin{bmatrix} -1 & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & -\underline{I} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{M} & \underline{0} & \underline{0} & \underline{0} \\ \underline{u} & \underline{u}\underline{t}' & \underline{0} & \underline{0} & \underline{K}_1 \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & -\underline{I} \end{bmatrix} \begin{bmatrix} m_0 \\ \underline{m}_1 \\ \underline{m}_2 \\ \underline{m}_3 \\ \underline{m}_4 \end{bmatrix}$$

where m_0 and $\underline{m}_1, \cdots, \underline{m}_4$ are sub-vectors of the arbitrary vector $\underline{m}$ of (27).

Using this solution of (3) in (26) gives

$$\underline{z}^* = \underline{y} - \underline{X}_f \left[\beta^o + \begin{pmatrix} -\underline{m}_1 \\ \underline{M}\underline{m}_1 \end{pmatrix} \right]$$

$$= \underline{y} - \underline{X}_f\underline{\beta}_f^o + \underline{F}_1\underline{m}_1 - \underline{F}_2\underline{M}\underline{m}_1$$

$$= \underline{z} + (\underline{1} \quad \underline{F}_1)\begin{bmatrix} 0 \\ \underline{m}_1 \end{bmatrix} - \underline{F}_2\underline{M}\underline{m}_1 \ ,$$

on using (2); and then using (19) and (20) gives

$$\underline{z}^* = \underline{z} + [\underline{F}_2(\underline{0} \quad \underline{M}) + \underline{R}_1\underline{K}_3]\begin{bmatrix} 0 \\ \underline{m}_1 \end{bmatrix} - \underline{F}_2\underline{M}\underline{m}_1$$

$$= \underline{z} + \underline{R}_1\underline{K}_3\begin{bmatrix} 0 \\ \underline{m}_1 \end{bmatrix} \ .$$

But from (21), $\underline{R}_1\underline{K}_3$ has all its rows the same so that $\underline{R}_1\underline{K}_3\begin{bmatrix} 0 \\ \underline{m}_1 \end{bmatrix} = \tau\underline{1}$ for some scalar τ . Hence

$$\underline{z}^* = \underline{z} + \tau\underline{1} \qquad (30)$$

$$= (\mu_1 + \tau)\underline{1} + \underline{X}_r\underline{\beta}_r + (\underline{I} - \underline{F}_2\underline{V})\underline{e}$$

$$= \mu_1^*\underline{1} + \underline{X}_r\underline{\beta}_r + (\underline{I} - \underline{F}_2\underline{V})\underline{e} \qquad (31)$$

for $\mu_1^* = \mu_1 + \tau$. This is the same form as $\underline{z}$ of (24) except for μ_1^* in place of μ_1 .

Henderson's Method 2 uses Method 1 on $\underline{z} = \underline{y} - \underline{X}_f\underline{\beta}_f^o$. Demonstrating the uniqueness of Method 2 depends upon showing that applying Method 1 to $\underline{z}^*$ yields the same variance components estimates as applying it to $\underline{z}$. That this is so depends on a property of the quadratic forms used in Method 1, namely that when any of the uncorrected sums of squares used in that method (e.g., Searle [1971], equation (35), page 431) are expressed in the form $\underline{y}'\underline{A}\underline{y}$, the symmetric matrix $\underline{A}$ is such that $\underline{A}\underline{1} = \underline{1}$. Differences between two such sums of squares $\underline{y}'\underline{A}\underline{y}$ and $\underline{y}'\underline{B}\underline{y}$ form the basis of Method 1. Calculating these sums of squares for $\underline{z}$ and $\underline{z}^*$ we find from (30) that

$$\underline{z}^{*\prime}\underline{A}\underline{z}^{*} - \underline{z}^{*\prime}\underline{B}\underline{z}^{*} = (\underline{z}' + \tau\underline{1}')(\underline{A} - \underline{B})(\underline{z} + \tau\underline{1})$$

$$= \underline{z}'\underline{A}\underline{z} - \underline{z}'\underline{B}\underline{z}$$

because $\underline{A}\underline{1} = \underline{B}\underline{1} = \underline{1}$. Hence the differences between sums of squares of Method 1 are the same for $\underline{z}^*$ as they are for $\underline{z}$. This means that Henderson's Method 1 using $\underline{z}^*$ is the same as using $\underline{z}$; i.e., that Henderson's Method 2 is independent of whatever solution of (3) is used in (6). Thus we have proved that Henderson's Method 2, which can be used only when there are no interactions between fixed and random effects, is then invariant to whatever solution $\underline{\beta}_f^o$ is obtained from the normal equations (3).

6. The coefficient of σ_e^2

Suppose $\underline{y}'\underline{A}\underline{y}$ is a quadratic used in Method 1. From the model (1) and the familiar expression for the expected value of a quadratic form we know that

$$E(\underline{y}'\underline{A}\underline{y}) \quad \text{contains} \quad \sigma_e^2\,\mathrm{tr}(\underline{A}) \quad . \tag{32}$$

On comparing the models (1) and (24) for $\underline{y}$ and $\underline{z}$ respectively it is clear that both models have the same term $\underline{X}_r\underline{\beta}_r$. Hence, when for Method 2 we use Method 1 on $\underline{z}$, the expected value $E(\underline{z}'\underline{A}\underline{z})$ will contain the same terms in the variance components as does $E(\underline{y}'\underline{A}\underline{y})$, except for the term in σ_e^2 . This similarity is, of course, the objective of Method 2. We now investigate the difference between the σ_e^2-terms in $E(\underline{y}'\underline{A}\underline{y})$ and $E(\underline{z}'\underline{A}\underline{z})$.

Denote the term in σ_e^2 in $E(\underline{z}'\underline{A}\underline{z})$ by $k_A\sigma_e^2$. Then from (24) we have

$$k_A\sigma_e^2 = E[\underline{e}'(\underline{I} - \underline{F}_2\underline{V})'\underline{A}(\underline{I} - \underline{F}_2\underline{V})\underline{e}]$$

$$= \sigma_e^2\,\mathrm{tr}[(\underline{I} - \underline{F}_2\underline{V})'\underline{A}(\underline{I} - \underline{F}_2\underline{V})] \quad . \tag{33}$$

Now whichever of the quadratic forms used in Method 1 $\underline{z}'\underline{A}\underline{z}$ is, it is by the nature of Method 1 also a quadratic form in the totals represented by $\underline{X}_r'\underline{z}$. Hence $\underline{z}'\underline{A}\underline{z} = \underline{z}'\underline{X}_r\underline{A}_1\underline{X}_r'\underline{z}$ for some matrix $\underline{A}_1$. But $\underline{X}_r = (\underline{R}_1 \;\; \underline{R}_2) = \underline{R}_1(\underline{I} \;\; \underline{K}_1)$ by (15) so that for $\underline{A}_2 = (\underline{I} \;\; \underline{K}_1)\underline{A}_1(\underline{I} \;\; \underline{K}_1)'$ we have

$$\underline{z}'\underline{A}\underline{z} = \underline{z}'\underline{R}_1\underline{A}_2\underline{R}_1'\underline{z} \quad \text{with} \quad \underline{A} = \underline{R}_1\underline{A}_2\underline{R}_1' \ . \tag{34}$$

Thus from (33)

$$k_A\sigma_e^2 = \sigma_e^2 \, tr[(\underline{I} - \underline{F}_2\underline{V})'\underline{R}_1\underline{A}_2\underline{R}_1'(\underline{I} - \underline{F}_2\underline{V})]$$

so that

$$k_A = tr[\underline{A}_2(\underline{R}_1'\underline{R}_1 - \underline{R}_1'\underline{F}_2\underline{V}\underline{R}_1 - \underline{R}_1'\underline{V}'\underline{F}_2'\underline{R}_1 + \underline{R}_1'\underline{F}_2\underline{V}\underline{V}'\underline{F}_2'\underline{R}] \ . \tag{35}$$

Now $\underline{V} = \underline{Q}_{11}\underline{F}_2' + \underline{Q}_{12}\underline{R}_1'$, so that $\underline{V}\underline{V}' = \underline{V}\underline{F}_2\underline{Q}_{11} + \underline{V}\underline{R}_1\underline{Q}_{12}' = \underline{Q}_{11}$, by (24). Hence, by also using $\underline{V}\underline{R}_1 = \underline{0}$ from (24) we get from (35)

$$\begin{aligned} k_A &= tr[\underline{A}_2(\underline{R}_1'\underline{R}_1 + \underline{R}_1'\underline{F}_2\underline{Q}_{11}\underline{F}_2'\underline{R}_1)] \\ &= tr[\underline{R}_1\underline{A}_2\underline{R}_1') + tr(\underline{R}_1\underline{A}_2\underline{R}_1'\underline{F}_2\underline{Q}_{11}\underline{F}_2') \end{aligned}$$

and (34) then gives

$$k_A\sigma_e^2 = \sigma_e^2 tr(\underline{A}) + \sigma_e^2 tr(\underline{A}\underline{F}_2\underline{Q}_{11}\underline{F}_2') \ . \tag{36}$$

Comparing (36) with (32) we see that the σ_e^2-term in $E(\underline{z}'\underline{A}\underline{z})$ is the same as that in $E(\underline{y}'\underline{A}\underline{y})$ except for the addition of $\sigma_e^2 tr(\underline{A}\underline{F}_2\underline{Q}_{11}\underline{F}_2')$. Hence, except for this additional term $E(\underline{z}'\underline{A}\underline{z})$ for any Method 1 quadratic form in z is, so far as terms in variance components are concerned, the same as $E(\underline{y}'\underline{A}\underline{y})$.

The additional term in $E(\underline{z}'\underline{A}\underline{z})$

$$\delta_A = \sigma_e^2 tr(\underline{A}\underline{W}) \quad \text{for} \quad \underline{W} = \underline{F}_2\underline{Q}_{11}\underline{F}_2' \tag{37}$$

is a generalization of the result discussed following equation (25) of Searle [1968]. Several characteristics of δ_A are worth noting. First $\underline{W}$ is the same for all $\underline{A}$. This means that for each quadratic $\underline{z}'\underline{A}\underline{z}$ used in the Method 1 analysis of $\underline{z}$, $\underline{W}$ stays the same for all the δ_A-terms. Second, because $\underline{W}$ is symmetric (as is $\underline{A}$ also),

$$\mathrm{tr}(\underline{A}\underline{W}) = \text{sum of element-by-element product of } \underline{A} \text{ and } \underline{W}$$

$$= \Sigma\Sigma a_{ij} w_{ij}$$

where a_{ij} and w_{ij} are typical elements of $\underline{A}$ and $\underline{W}$ respectively. Third, the derivation of $\underline{W}$ in no way depends on the particular properties attributed in Sections 3 and 4 to the partitioning (2). The matrix $\underline{W} = \underline{F}_2\underline{Q}_{11}\underline{F}_2'$ can therefore be defined more generally in terms of that partitioning and the solutions (5) to the normal equations (3). This is given in the summary which follows.

7. A Summary of Method 2

1. Method 2 can be used on any mixed model typified as in (1) by

$$\underline{y} = \mu\underline{1} + \underline{X}_f\underline{\beta}_f + \underline{X}_r\underline{\beta}_r + \underline{e} \qquad (38)$$

provided there are no interactions between the fixed effects, $\underline{\beta}_f$, and the random effects $\underline{\beta}_r$.

2. Temporarily assume the random effects are fixed and solve the normal equations corresponding to (38). Obtain a solution vector in which the solutions for μ and for as many other effects as is necessary are zero. [See equation (5).] In obtaining this solution define $\underline{F}_2$ as being that part of $\underline{X}_f$ corresponding to the fixed effects solutions $\underline{\beta}_{f2}^o$ that were not made zero, and $\underline{Q}_{11}$ as the corresponding sub-matrix of the inverse matrix used in obtaining the solution—see (4) and (5). Define

$$\underline{W} = \underline{F}_2\underline{Q}_{11}\underline{F}_2' \quad .$$

3. Calculate $\underset{\sim}{z} = \underset{\sim}{y} - \underset{\sim}{F}_2 \underset{\sim}{\beta}_{f2}^{o}$.

4. Carry out a Method 1 analysis on $\underset{\sim}{z}$, just as one would on $\underset{\sim}{y}$ were there no fixed effects in its model. For every quadratic form $\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z}$ used in this analysis, the expected value $E(\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z})$ will be identical to $E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y})$ ignoring fixed effects, except for the addition of

$$\delta_A = \sigma_e^2 \mathrm{tr}(\underset{\sim}{A}\underset{\sim}{W}) = \sigma_e^2 \Sigma\Sigma a_{ij} w_{ij}$$

to $E(\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z})$.

The solution for the normal equations suggested in item 2 of this summary is, of course, not the only way of obtaining a solution and proceeding with Method 2 but it is one of the easiest ways both to execute and to describe.

Acknowledgement

The work reported in this paper was partially supported by Grant No. GJ-31746 of the National Science Foundation.

References

Henderson, C. R. [1953] Estimation of variance and covariance components. Biometrics 9: 226-252.

Searle, S. R. [1968] Another look at Henderson's methods of estimating variance components. Biometrics 24: 749-788.

Searle, S. R. [1971] Linear Models, Wiley and Sons, New York.