NORTH-HOLLAND

The Matrix Handling of BLUE and BLUP in the Mixed Linear Model*

Shayle R. Searle
Biometrics Unit
Cornell University
Ithaca, New York 14853

Submitted by Simo Puntanen

ABSTRACT

The mixed model of analysis of variance is a linear model in which some terms that would otherwise be unknown constants are, in fact, unobservable realizations of random variables. Estimation procedures for the constants and for the realized random variables are reviewed, with emphasis on their matrix features. © 1997 Elsevier Science Inc.

1. FIXED-EFFECTS MODELS

1.1. *Basics*

The customary general linear model has model equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y}$ is a vector of N observations, $\mathbf{X}$ is a known matrix [the model matrix, as Kempthorne (1980) calls it], $\boldsymbol{\beta}$ is a vector of p fixed, unknown constants (fixed effects), and $\boldsymbol{\epsilon}$ is a vector of random errors. The latter is defined, *ab initio*, as $\boldsymbol{\epsilon} = \mathbf{y} - E(\mathbf{y}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, so that it has mean zero [$E(\boldsymbol{\epsilon}) = \mathbf{0}$], and we take all elements of $\boldsymbol{\epsilon}$ to be uncorrelated with one another with the same

* Invited paper for the Fourth International Workshop on Matrix Methods for Statistics, Montreal, Quebec, 15–16 July 1995. Paper BU-718 in the Biometrics Unit, Cornell University.

0024-3795/97/$17.00
PII S0024-3795(96)00400-4

variance, σ_ϵ^2, so that the variance-covariance matrix of $\boldsymbol{\epsilon}$ is

$$\text{var}(\boldsymbol{\epsilon}) = \sigma_\epsilon^2 \mathbf{I}_N, \tag{2}$$

with $\mathbf{I}_N$ the identity matrix of order N. When elements of $\boldsymbol{\beta}$ represent effects on y due to factors by which the data are classified (the usual situation for analysis of variance), $\mathbf{X}$ has elements that are 0 or 1 and is called an incidence matrix. But $\boldsymbol{\beta}$ can include regression coefficients for observed covariates corresponding to y, in which case columns of $\mathbf{X}$ contain those observed covariates.

The least-squares equation for estimating $\boldsymbol{\beta}$ of (1) are

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}'\mathbf{y}. \tag{3}$$

When $\mathbf{X}'\mathbf{X}$ is nonsingular, as is usual in regression analysis, the symbol $\hat{\boldsymbol{\beta}}$ is used in place of $\boldsymbol{\beta}^0$ in (3), and then $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. But since, more generally, $\mathbf{X}$ is not of full column rank, and so $\mathbf{X}'\mathbf{X}$ is singular, the notation $\boldsymbol{\beta}^0$ is used to indicate that for each and every generalized inverse of $\mathbf{X}'\mathbf{X}$, namely $(\mathbf{X}'\mathbf{X})^-$ satisfying $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$,

$$\boldsymbol{\beta}^0 = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} \tag{4}$$

is a solution of (3). And although $\boldsymbol{\beta}^0$ clearly depends on the choice of $(\mathbf{X}'\mathbf{X})^-$, the best linear unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ is

$$\mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} = \mathbf{X}\mathbf{X}^+\mathbf{y}, \tag{5}$$

and it does not depend upon the choice of $(\mathbf{X}'\mathbf{X})^-$. ($\mathbf{X}^+$ represents the unique Moore-Penrose inverse of $\mathbf{X}$, satisfying the four conditions $\mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X}$, $\mathbf{X}^+\mathbf{X}\mathbf{X}^+ = \mathbf{X}^+$, and both $\mathbf{X}\mathbf{X}^+$ and $\mathbf{X}^+\mathbf{X}$ symmetric.)

1.2. OLSE and GLSE

The results (4) and (5) are based on least-squares estimation. But if, instead of var($\boldsymbol{\epsilon}$) being $\sigma_\epsilon^2\mathbf{I}_N$ as in (2), we have a more general situation of

$$\text{var}(\boldsymbol{\epsilon}) = \mathbf{V} \tag{6}$$

for some positive definite symmetric matrix $\mathbf{V}$, then in place of the equations (3) we could use the generalized least-squares equations

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{7}$$

with a solution

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{8}$$

This yields the generalized least-squares estimator (GLSE) of $\mathbf{X}\boldsymbol{\beta}$, which, because it is also the best linear unbiased estimator (BLUE) of $\mathbf{X}\boldsymbol{\beta}$, we write as

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \text{GLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{9}$$

In contrast, (5) is often referred to as the ordinary least-squares estimator

$$\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}. \tag{10}$$

Clearly, (9) and (10) are the same when $\mathbf{V} = \sigma^2\mathbf{I}$; and they are also equal for other cases too, as mentioned following (21).

2. MIXED MODELS

All of the preceding discussion is where elements of $\boldsymbol{\beta}$ are deemed to be fixed, unknown constants that one wants to estimate—or at least to estimate estimable functions of them, those that have the form $\mathbf{k}'\boldsymbol{\beta}$ for $\mathbf{k}' = \mathbf{t}'\mathbf{X}$ for any $\mathbf{t}'$. [This is why one concentrates in (9) and (10) on estimating $\mathbf{X}\boldsymbol{\beta}$ rather than $\boldsymbol{\beta}$: all linear combinations of $\mathbf{X}\boldsymbol{\beta}$ are estimable.]

2.1. The Model Equation

But now suppose when var$(\boldsymbol{\epsilon}) = \mathbf{V}$ that we have reason to model $\boldsymbol{\epsilon}$ as

$$\boldsymbol{\epsilon} = \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{11}$$

where $\mathbf{u}$ is a vector of random terms with corresponding matrix $\mathbf{Z}$, and where $\mathbf{e}$ is a vector of random residuals. Then (1) becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}. \tag{12}$$

This is the widely used model equation of the mixed model of analysis of variance.

Before defining mean and variance properties of $\mathbf{u}$ and $\mathbf{e}$, a comparison of $\boldsymbol{\beta}$ and $\mathbf{u}$ is important. $\boldsymbol{\beta}$ is considered as a vector of fixed, unknown constants. $\mathbf{u}$ has been glibly described as a vector of random effects—glibly, because in (12), for a given data vector $\mathbf{y}$, the elements of $\mathbf{u}$ are in fact realized (but unobservable) values of random variables. For example, in the context of genetics (where mixed models are frequently used), suppose the data values in $\mathbf{y}$ are fleece weights of a number of sheep—one fleece per sheep per year, over a number of years. Then whatever sheep have provided the data can, from the genetic viewpoint, be considered as a random sample of sheep from some definable population of sheep. Then each element of $\mathbf{u}$ is a representation of the genetic value of a sheep and so is indeed the unobservable realized value in the data of the random variable "genetic value for wool production of a sheep". Despite this distinction between "realized value" and "random variable", we proceed to use $\mathbf{u}$ to represent both.

2.2. *Means and Variances*

Since we have $E(\boldsymbol{\epsilon}) = \mathbf{0}$, we look at (11) and take

$$E(\mathbf{u}) = \mathbf{0} \quad \text{and} \quad E(\mathbf{e}) = \mathbf{0}. \tag{13}$$

We also define

$$\text{var}(\mathbf{u}) = \mathbf{D}, \qquad \text{var}(\mathbf{e}) = \mathbf{R}, \quad \text{and} \quad \text{cov}(\mathbf{u}, \mathbf{e}') = \mathbf{0}. \tag{14}$$

Then from (12) we have, in contrast to (6),

$$\mathbf{V} = \text{var}(\mathbf{y}) = \text{var}(\boldsymbol{\epsilon}) = \text{var}(\mathbf{Zu} + \mathbf{e}) = \mathbf{ZDZ}' + \mathbf{R}. \tag{15}$$

This is the traditional and much-used mixed model specified by equations (12) through (15). It is a mixture of fixed effects $\boldsymbol{\beta}$ and random effects $\mathbf{u}$.

2.3. *Analysis-of-Variance Modeling*

Rather than the development of the two preceding sections, the mixed model is usually formulated directly in the form of a model equation for a typical analysis-of-variance situation. For example, the model equation for data from a randomized complete block experiment of n observations on each of a treatments in each of b blocks can be taken as

$$y_{ijk} = \mu + t_i + \pi_j + \gamma_{ij} + e_{ijk}, \tag{16}$$

where y_{ijk} is observation k on treatment i in block j. In (16) the μ is a general mean, t_i is the effect of treatment i, π_j is the effect of block j, γ_{ij} is the interaction effect, and e_{ijk} is a random error term. μ and t_i are deemed to be fixed effects, and π_i and γ_{ij} are taken as random effects with the terms of the general model, (12)–(15), being as follows. $\mathbf{y}$ is the vector of data values y_{ijk}, arrayed in lexicon order,

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \mathbf{t} \end{bmatrix}, \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \text{say,} \tag{17}$$

where $\mathbf{t}$, $\boldsymbol{\pi}$, and $\boldsymbol{\gamma}$ are the vectors of the corresponding terms in (16), each with elements in lexicographic order. Then from (14) and (17) we have

$$\mathbf{D} = \operatorname{var}\begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \sigma_\pi^2 \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \sigma_\gamma^2 \mathbf{I}_{ab} \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \sigma_e^2 \mathbf{I}_N \tag{18}$$

for $N = abn$, the total number of observations (length of $\mathbf{y}$). In conformity with $\mathbf{u}$ of (17) we partition $\mathbf{Z}$ as $[\mathbf{Z}_1 \ \ \mathbf{Z}_2]$ and then have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}$$

and

$$\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} = \mathbf{Z}_1\mathbf{Z}_1'\sigma_\pi^2 + \mathbf{Z}_2\mathbf{Z}_2'\sigma_\gamma^2 + \sigma_e^2\mathbf{I}_N.$$

This formulation extends very naturally when there are r random effects factors, be they main effect or interaction factors. Akin to (17) and (18),

$$\mathbf{u} = \{_c\, \mathbf{u}_i\}_{i=1}^r \quad \text{and} \quad \mathbf{D} = \{_d\, \sigma_i^2 \mathbf{I}_{q_i}\}_{i=1}^r, \tag{19}$$

where q_i is the number of levels of the ith random factor that occur in the data, i.e., the order of $\mathbf{u}_i$. Then

$$\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} = \sum_{i=1}^r \mathbf{Z}_i\mathbf{Z}_i'\sigma_i^2 + \mathbf{I}_N\sigma_e^2. \tag{20}$$

On further defining $\mathbf{e} = \mathbf{u}_0$, $\sigma_e^2 = \sigma_0^2$, $N = q_0$, and $\mathbf{I}_N = \mathbf{Z}_0$, we can then write $\mathbf{V}$ even more compactly as

$$\mathbf{V} = \sum_{i=0}^r \mathbf{Z}_i\mathbf{Z}_i'\sigma_i^2. \tag{21}$$

2.4. Estimating Fixed Effects

With $\mathbf{V}$ as in (20), or equivalently (21), we can estimate $\mathbf{X}\boldsymbol{\beta}$ precisely as in (9) or (10):

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \text{GLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

$$\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}.$$

The latter being computationally simpler than the former prompts the question: when are they the same? When $\mathbf{VX} = \mathbf{XF}$ for some $\mathbf{F}$ is one of many equivalent answers (see Puntanen and Styan, 1989); and they are always equal for balanced data. In that case $\mathbf{X}\boldsymbol{\beta}$ (and $\mathbf{V}$) are then each a linear combination of direct products of $\mathbf{I}$-matrices and $\mathbf{1}$-vectors (matrices $\mathbf{11}'$), with $\mathbf{1}$ being a summing vector having all elements unity (see Searle, 1988).

3. RANDOM EFFECTS IN MIXED MODELS

In the genetics example of sheep fleece weights, an underlying economic objective is to breed new generations of sheep from animals that have genetic make-ups associated with high fleece weight. To do this one needs to put a value on the genetic worth of sheep whose fleece weights are currently available as data. This means, in some sense, estimating or predicting for those sheep their unobservable genotype for fleece weight represented by the elements of $\mathbf{u}$ in the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}$. Those elements of $\mathbf{u}$ are realized values of a random variable; and although many sheep may have the same fleece weight, they will not all have the same genotype. Indeed, for a given value of a sheep's average fleece weight, $\tilde{y}$, say, there will be a range of genotypes and some kind of distribution of u-values. Under these circumstances it seems reasonable that for a given $\tilde{y}$ a satisfactory evaluation of the u-value associated with any sheep having average fleece weight $\tilde{y}$ would be $\text{E}(u|\tilde{y})$.

Indeed, this is so. Given that $\mathbf{u}$ represents random variables, suppose we seek the best predictor of $\mathbf{u}$ and call it $\text{BP}(\mathbf{u})$. By this we mean that $\text{BP}(\mathbf{u})$ has minimum mean squared error of prediction, i.e., that $\text{E}\{[\text{BP}(\mathbf{u}) - \mathbf{u}]'\mathbf{A}[\text{BP}(\mathbf{u}) - \mathbf{u}]\}$ is minimized (for any positive definite symmetric $\mathbf{A}$). The result (Searle et al., 1992, Section 7.2a) is

$$\text{BP}(\mathbf{u}) = \text{E}(\mathbf{u}|\mathbf{y}). \tag{22}$$

Note that (22) does not involve normality. Neither does derivation of the best linear predictor BLP($\mathbf{u}$), which is derived by starting from taking

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y} \end{bmatrix} \sim \left[\begin{pmatrix} \boldsymbol{\mu}_U \\ \boldsymbol{\mu}_Y \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{C} \\ \mathbf{C}' & \mathbf{V} \end{pmatrix} \right], \tag{23}$$

again, without imposing normality. Then by seeking BLP($\mathbf{u}$) of the form $\mathbf{a} + \mathbf{B}\mathbf{y}$ we get (*loc. cit.*)

$$\mathrm{BLP}(\mathbf{u}) = \boldsymbol{\mu}_U + \mathbf{C}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y). \tag{24}$$

And if normality is invoked for (22), then (22) becomes identical to (24).

In this formulation, with $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ and $\mathrm{var}(\mathbf{u}) = \mathbf{D}$, the $\mathbf{C}$ of (23) is

$$\mathbf{C} = \mathrm{cov}(\mathbf{u}, \mathbf{y}') = \mathbf{D}\mathbf{Z}'. \tag{25}$$

4. BLUP: BEST LINEAR UNBIASED PREDICTION

In order to use BP($\mathbf{u}$) it is clear from (22) that the joint density function of $\mathbf{u}$ and $\mathbf{y}$ (or at least the conditional density) needs to be known. And for BLP($\mathbf{u}$) of (24) the first and second moments of $\mathbf{u}$ and $\mathbf{y}$ are required. A third, and most popular, procedure is BLUP, best linear unbiased prediction. It requires knowing only second moments of $\mathbf{u}$ and $\mathbf{y}$. In its broadest sense it has wider application than BP or BLP because it estimates not just $\mathbf{u}$ but $\mathbf{u}$ plus linear combinations of $\boldsymbol{\beta}$, namely

$$\mathbf{w} = \mathbf{L}'\boldsymbol{\beta} + \mathbf{u}, \tag{26}$$

where, in order for $\mathbf{L}'\boldsymbol{\beta}$ to be estimable, $\mathbf{L}' = \mathbf{T}'\mathbf{X}$ for some $\mathbf{T}'$. Then

$$\mathrm{BLUP}(\mathbf{w}) = \mathbf{L}'\tilde{\boldsymbol{\beta}} + \mathbf{C}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \tag{27}$$

for $\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ of (9).

The function (27) is best in the sense of minimum mean square of prediction (similar to BP), it is linear in the data, and it is unbiased, in the sense that

$$\mathrm{E}[\mathrm{BLUP}(\mathbf{w})] = \mathbf{L}'\boldsymbol{\beta} + \mathrm{E}(\mathbf{u}). \tag{8}$$

This is a rather special definition of unbiasedness, since the right-hand side of (28) contains $\mathrm{E}(\mathbf{u})$. Unbiasedness is usually defined as the expected value of an estimator equaling the parameter function being estimated; e.g., $\mathrm{E}(\mathbf{X}\boldsymbol{\beta}^0) = \mathbf{X}\boldsymbol{\beta}$. While this can (and does) apply to the $\mathbf{L}'\boldsymbol{\beta}$ part of $\mathbf{w}$ of (26), it cannot be used for $\mathbf{u}$, because $\mathbf{u}$ represents random variables. This is why we have (28). And when $\mathbf{u}$ is taken as having $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, as in (13), the occurrence of $\mathrm{E}(\mathbf{u})$ in (28) is of little concern.

The acronym BLUP describes (27) as a predictor. This, in contrast to "estimator", probably originated from Henderson (1950), who was, I understand (personal communication) rebuked by statisticians after presenting that paper at an Institute of Mathematical Statistics meeting—rebuked because he talked about "estimating" random variables, and so was told that that was not what we did. But Goldberger (1962) used "predictor" in his title, and it has stuck ever since. Robinson (1990) strongly urges the use of "estimator"—but I fear that BLUP trips off the tongue so easily (and BLUE is so well established) so that its more correct acronym, BLUERR (best linear unbiased estimator of realized random variables), would never be accepted.

4.1. Many Derivations of BLUP

Searle et al. (1992, Section 7.4c) give details of a derivation of BLUP($\mathbf{w}$) of (27) that is obtuse and unnecessarily complicated. They also describe five other derivations, which are based, respectively, on two-stage regression, linearity in $\mathbf{y}$, a partitioning of $\mathbf{y}$, Bayes estimation, and Henderson's (Henderson et al., 1959) mixed model equations. There is also Goldberger's (1962) approach of predicting a future observation; and Harville (1990) has yet another viewpoint. Most of these derivations bear little resemblance to the statisticians' straightforward and customary derivation of BLUE($\mathbf{X}\boldsymbol{\beta}$). Recently, however, Searle (19945) showed a derivation which is very similar to that of BLUE and so puts the derivation of BLUP into the mainline of statistical reasoning. An outline of that derivation is as follows. It begins with deriving BLUE($\mathbf{X}\boldsymbol{\beta}$).

4.2. Deriving BLUE

In deriving BLUE($\mathbf{X}\boldsymbol{\beta}$) of (9) we begin by wanting to

$$\text{estimate} \quad \mathbf{t}'\mathbf{X}\boldsymbol{\beta} \text{ for some } \mathbf{t}' \tag{29}$$

by an estimator that is

$$\text{linear in } \mathbf{y}, \text{ i.e., } \boldsymbol{\lambda}'\mathbf{y} \text{ for some } \boldsymbol{\lambda}' \neq \mathbf{0}, \tag{30}$$

and

$$\text{unbiased,} \qquad \text{i.e.,} \quad E(\boldsymbol{\lambda}'\mathbf{y}) = \mathbf{t}'\mathbf{X}\boldsymbol{\beta}. \tag{31}$$

Requiring the latter to be true for all $\boldsymbol{\beta}$ implies

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}'\mathbf{t} \tag{32}$$

and, subject to this, we then

$$\text{choose } \boldsymbol{\lambda} \text{ to minimize } \operatorname{var}(\boldsymbol{\lambda}'\mathbf{y}) = \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda}. \tag{33}$$

This leads to minimizing

$$\theta = \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda} + 2\mathbf{m}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{X}'\mathbf{t})$$

with respect to $\boldsymbol{\lambda}$, where $\mathbf{m}'$ is a vector of Lagrange multipliers. Differentiating θ with respect to $\boldsymbol{\lambda}$ and $\mathbf{m}$ and equating the results to zero, and then letting $\mathbf{t}'$ take on the values of successive rows of $\mathbf{I}_N$, we have, after some matrix manipulation,

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\beta}},$$

which is (9).

Note that (31) and (33) can be rewritten, respectively, as

$$E(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}'\mathbf{X}\boldsymbol{\beta}) = 0. \tag{34}$$

and

$$\text{choose } \boldsymbol{\lambda} \text{ to minimize } \operatorname{var}(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}'\mathbf{X}\boldsymbol{\beta}). \tag{35}$$

(34) is exactly the same as (31); and (35) is equivalent to the left-hand side of (33) because $\mathbf{t}'\mathbf{X}\boldsymbol{\beta}$ is a constant. Yet (34) and (35) are the clue to deriving BLUP, based on designating $\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}'\mathbf{X}\boldsymbol{\beta}$ in (34) and (35) as being "estimation error". For BLUP we use exactly this idea.

4.3. *Deriving BLUP Similarly to BLUE*

In deriving BLUP($\mathbf{w}$) we start by stating what it is that we want to

$$\text{estimate} \quad \mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} + \mathbf{t}_2'\mathbf{u} \quad \text{for non-null } \mathbf{t}_1' \text{ and } \mathbf{t}_2'.$$

The estimator is to be linear in y, as in (30). Also, in the sense of (34) being expected estimation error being zero, we here want

$$E(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} - \mathbf{t}_2'\mathbf{u}) = 0.$$

Since $E(\mathbf{u}) = \mathbf{0}$, this gives

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{X}'\mathbf{t}_1, \tag{36}$$

similar to (32). Then, subject to this we

$$\text{choose } \boldsymbol{\lambda} \text{ to minimize } \operatorname{var}(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} - \mathbf{t}_2'\mathbf{u}), \tag{37}$$

analogous to (3). But now this variance to be minimized is not just $\boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda}$ as in (33), but is, for $\mathbf{u}$ a random variable,

$$\begin{aligned}\operatorname{var}(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} - \mathbf{t}_2'\mathbf{u}) &= \operatorname{var}(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{t}_2'\mathbf{u}) \\ &= \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda} + \mathbf{t}_2'\mathbf{D}\mathbf{t}_2 - 2\boldsymbol{\lambda}'\mathbf{C}'\mathbf{t}_2.\end{aligned}$$

Minimizing this subject to (36) means minimizing

$$\boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda} + \mathbf{t}_2'\mathbf{D}\mathbf{t}_2 - 2\boldsymbol{\lambda}'\mathbf{Z}\mathbf{D}\mathbf{t}_2 + 2\mathbf{m}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{X}'\mathbf{t}_1)$$

with respect to $\boldsymbol{\lambda}$, with $\mathbf{m}$ again being a vector of Lagrange multipliers. This minimization yields

$$\begin{aligned}\boldsymbol{\lambda}'\mathbf{y} &= \mathrm{BLUP}(\mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} + \mathbf{t}_2'\mathbf{u}) \\ &= \mathbf{t}_1'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{t}_2'\mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{y} \\ &= \mathbf{t}_1'\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{t}_2'\mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}).\end{aligned}$$

This is true for any $\mathbf{t}_1'$ and $\mathbf{t}_2'$. Therefore let $\mathbf{t}_1'$ be successive rows of $\mathbf{I}_N$ and take $\mathbf{t}_2' = \mathbf{0}$ and get

$$\mathrm{BLUP}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}).$$

Second, take $\mathbf{t}_1' = \mathbf{0}$ and let $\mathbf{t}_2'$ be successive rows of $\mathbf{I}_q$ (for $\mathbf{u}$ having q elements) and get

$$\mathrm{BLUP}(\mathbf{u}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}).$$

Third, let $\mathbf{t}_1'$ be successive rows of $\mathbf{T}'$ in $\mathbf{L}' = \mathbf{T}'\mathbf{X}$ of $\mathbf{w} = \mathbf{L}'\boldsymbol{\beta} + \mathbf{u}$ in (26); and let $\mathbf{t}_2'$ be successive rows of $\mathbf{I}_q$ and so get (27):

$$\text{BLUP}(\mathbf{w}) = \mathbf{L}'\tilde{\boldsymbol{\beta}} + \mathbf{C}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$$

for $\mathbf{C} = \mathbf{D}\mathbf{Z}'$ of (25).

The clue to this derivation of BLUP is that it is exactly the same as that of BLUE if, in both cases, one states (i) the unbiasedness requirement as being expected estimation error equal to zero; and (ii) the variance to be minimized as being the variance of that same estimation error; e.g., compare (33) and (37).

5. THE MIXED MODEL EQUATIONS (MMES)

Of the numerous derivations of BLUP mentioned in Section 4.1, one of particular interest is that stemming from what have come to be known as the mixed model equations (MMEs) of C. R. Henderson in Henderson et al. (1959). These equations are interesting because of their (i) history, (ii) reducing computational effort, (iii) providing sampling variances of estimators, and (iv) connection with computing maximum-likelihood estimates of variance components. We first summarize these four features and then provide some details.

5.1. *History*

Section 7.1 of Searle et al. (1992) details a question initially asked in a 1940s class offered by A. M. Mood at the then Iowa State College (now University) that was concerned, supposedly, with maximum-likelihood estimation of a student's IQ, given that the student had a test score of 130. And in the version of the question in Mood (1950) there is the tantalizing final sentence "The answer is not 130." This motivated the graduate student C. R. Henderson to what he thought was a maximum-likelihood derivation of BLUP. On assuming $\mathbf{u}$ and $\mathbf{y}$ to be normally distributed as

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{D}\mathbf{Z}' \\ \mathbf{Z}\mathbf{D} & \mathbf{V} \end{pmatrix}\right], \tag{38}$$

in accord with (13), (14), (15), (23), and (25), Henderson maximized

$$\frac{\exp\left\{-\tfrac{1}{2}\left[\mathbf{u}'\mathbf{D}^{-1}\mathbf{u} + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})\right]\right\}}{\left(\sqrt{2\pi}\right)^{N+q}\sqrt{|\mathbf{D}|\,|\mathbf{R}|}}, \tag{39}$$

thinking it was a likelihood. But it is, of course, the joint density function $f(\mathbf{u}, \mathbf{y}) = f(\mathbf{y}|\mathbf{u})f(\mathbf{u})$, whereas the likelihood function is

$$\frac{\exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\}}{(\sqrt{2\pi})^N|\mathbf{V}|}.$$

But in maximizing (39) with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$ we get

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{D}^{-1}+\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \tag{40}$$

These are Henderson's mixed model equations (MMEs). Their solution, as shown in Section 5.5.i, is

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad \text{with} \quad \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\tilde{\boldsymbol{\beta}} \tag{41}$$

and

$$\tilde{\mathbf{u}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}) = \text{BLUP}(\mathbf{u}). \tag{42}$$

5.2. *Computational Economy*

The expressions for $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$, as they stand, require $\mathbf{V}^{-1}$, a matrix of order N, the number of observations. But the MMEs (40) have order $p + q$, where p and q are, respectively, the numbers of fixed and random effects occurring in the data. Any in many situations $p + q \ll N$, so that solving (40) requires much less computational effort than does calculating the inverse of $\mathbf{V}$. This is particularly so when, as is often the case, $\mathbf{R}$ is diagonal, particularly for the common situation of $\mathbf{R} = \sigma_e^2\mathbf{I}_N$. Also, $\mathbf{D}$ being diagonal [as it often is, as in (19)] further contributes to the ease of solving (40).

5.3. *Sampling Variances*

There is at least one generalized inverse of the matrix on the left-hand side of (40) which provides basic expressions for sampling variances of estimators. They are as follows: First,

$$\text{var}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}, \quad \text{so that} \quad \text{var}[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}', \tag{43}$$

the latter being invariant to $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$, whereas var$(\tilde{\boldsymbol{\beta}})$ is not. Here the result for var$(\hat{\boldsymbol{\beta}})$ holds only if $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ is a symmetric, reflexive generalized inverse. Second,

$$\text{var}[\text{BLUP}(\mathbf{u}) - \mathbf{u}] = \mathbf{D} - \mathbf{DZ'PZD}$$

for

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}. \tag{44}$$

Note that this is the sampling variance not of BLUP(**u**), but of its prediction error BLUP(**u**) − **u**. And we also get

$$-\text{cov}(\tilde{\boldsymbol{\beta}}, \mathbf{u}') = \text{cov}[\tilde{\boldsymbol{\beta}}, (\tilde{\mathbf{u}} - \mathbf{u}')] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZD} \tag{45}$$

for $\tilde{\mathbf{u}}$ = BLUP(**u**) of (42). The first equality in (45) arises because $\text{cov}(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{u}}') = \mathbf{0}$; and of course, (45) is invariant to $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ only when adapted by having $\mathbf{X}\tilde{\boldsymbol{\beta}}$ in place of $\tilde{\boldsymbol{\beta}}$.

5.4. *Maximum-Likelihood Estimation of Variance Components*

For unbalanced data, on the basis of normality assumptions, the equations for calculating maximum-likelihood estimators of variance components are very complicated and have to be solved numerically—see Hartley and Rao (1967). As noted by Patterson and Thompson (1971) and by Henderson (1973), iterative procedures for solving those equations can be established in terms of sums of squares $\tilde{\mathbf{u}}'\tilde{\mathbf{u}}$ for $\tilde{\mathbf{u}}$ = BLUP(**u**). Deriving these procedures, which include submatrices of $(\mathbf{I} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZD})^{-1}$, is tedious: details are available in Searle et al. (1992, Sections 7.6c, d, and e).

5.5. *Some Technical Details*

i. Solutions. The general MMEs are, as in (40),

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

Regarding these as two (vector) equations, it is easily seen that the second of them gives

$$\begin{aligned} \tilde{\mathbf{u}} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}}) \\ &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \end{aligned} \tag{46}$$

and then the first is

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}.$$

This, because

$$\mathbf{V}^{-1} = (\mathbf{ZDZ}' + \mathbf{R})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}, \quad (47)$$

reduces to

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

thus giving $\tilde{\boldsymbol{\beta}}$ of (41). Also

$$\begin{aligned}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})\mathbf{DZ}' &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZDZ}' + \mathbf{Z}' = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{ZDZ}' + \mathbf{R})\\ &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{V}.\end{aligned}$$

Therefore

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1} = \mathbf{DZ}'\mathbf{V}^{-1}, \quad (48)$$

and so $\tilde{\mathbf{u}}$ of (46) is (42):

$$\tilde{\mathbf{u}} = \mathbf{DZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}).$$

ii. Special Cases. A common form for $\mathbf{R}$ is $\mathbf{R} = \sigma_e^2\mathbf{I}_N$. This simplifies the MMEs to

$$\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma_e^2\mathbf{D}^{-1}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\beta}}\\ \tilde{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix}. \quad (49)$$

When $\mathbf{D}$ is diagonal of the form $\mathbf{D} = \{_d\ \sigma_i^2\mathbf{I}_{q_i}\}$ of (19), writing $\lambda_i = \sigma_e^2/\sigma_i^2$ simplifies the equations further to

$$\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \{_d\ \lambda_i\mathbf{I}_{q_i}\}_{i=1}^{r}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\beta}}\\ \tilde{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix}. \quad (50)$$

And if there is only one random-effects factor in the model (i.e., $r = 1$), the equations become

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \{_d\, n_t + \lambda_1\}_{t=1}^{q_1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \tag{51}$$

where n_t is the number of observations in the tth level of the random-effects factor.

Finally, when dealing with the one-way classification random model, having the model equation

$$y_{tj} = \mu + \alpha_t + e_{tj}$$

for $j = 1, \ldots, n_t$ and $t = 1, \ldots, q_1 = a$, and with $\lambda_1 = \sigma_e^2/\sigma_\alpha^2$ and $\mathbf{X} = \mathbf{1}_N$ and $\mathbf{Z} = \{_d \mathbf{1}_{n_i}\}$, then (51) becomes

$$\begin{bmatrix} N & \{_r\, n_t\} \\ \{_c\, n_t\} & \{_d\, n_t + \lambda_1\} \end{bmatrix} \begin{bmatrix} \tilde{\mu} \\ \tilde{\alpha} \end{bmatrix} = \begin{bmatrix} y_{\cdot} \\ y_{t\cdot} \end{bmatrix}.$$

The solutions to these equations simplify to

$$\tilde{\mu} = \sum_{t=1}^{a} \frac{n_t \bar{y}_{t\cdot}}{\sigma_e^2 + n_t\sigma_\alpha^2} \Bigg/ \sum_{t=1}^{a} \frac{n_t}{\sigma_e^2 + n_t\sigma_\alpha^2}$$

and

$$\tilde{\alpha}_t = \frac{n_t\sigma_\alpha^2}{\sigma_e^2 + n_t\sigma_\alpha^2}(\bar{y}_{t\cdot} - \tilde{\mu}),$$

so giving

$$\text{BLUP}(\mu + \alpha_t) = \tilde{\mu} + \tilde{\alpha}_t = \theta_t\bar{y}_{t\cdot} + (1 - \theta_t)\tilde{\mu}$$

for

$$\theta_t = \frac{n_t\sigma_\alpha^2}{\sigma_e^2 + n_t\sigma_\alpha^2} < 1.$$

This kind of result is, of course, in keeping with the concepts of "regressino toward the mean" and of Stein estimation. A particular example of it familiar to geneticists uses their parameter $h = 4\sigma_\alpha^2/(\sigma_e^2 + \sigma_\alpha^2)$. This puts

BLUP($\mu + \alpha_t$) in the form

$$\text{BLUP}(\mu + \alpha_t) = \tilde{\mu} + \frac{n_t h}{4 + (n_t - 1)h}(\bar{y}_{t\cdot} - \tilde{\mu}).$$

iii. Sampling Variances. On writing the partitioned coefficient matrix of the MMEs as

$$\mathbf{L} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{T} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{S} \end{bmatrix},$$

thus defining $\mathbf{T}$, $\mathbf{Q}$, and $\mathbf{S}$ as submatrices of $\mathbf{L}$, a standard expression for a generalized inverse of $\mathbf{L}$ is

$$\mathbf{L}^{-} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{S}^{-1}\mathbf{Q}' \end{bmatrix}(\mathbf{T} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{Q}')^{-}\begin{bmatrix} \mathbf{I} & -\mathbf{Q}\mathbf{S}^{-1} \end{bmatrix}. \quad (52)$$

Then

$$\begin{aligned} \mathbf{T} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{Q}' &= \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} \\ &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \end{aligned}$$

from (47); and with the help of (48)

$$\mathbf{S}^{-1}\mathbf{Q}' = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}. \quad (53)$$

Therefore from (52)

$$\mathbf{L}^{-} = \begin{bmatrix} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} & -(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D} \\ -\mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} & \mathbf{S}^{-1} + \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D} \end{bmatrix}.$$

The lower right-hand matrix, call it $\mathbf{B}$, simplifies as follows:

$$\begin{aligned} \mathbf{B} &= \mathbf{S}^{-1} + \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D} \\ &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1} + \mathbf{D}\mathbf{Z}'(\mathbf{V}^{-1} - \mathbf{P})\mathbf{Z}\mathbf{D} \\ &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1} \\ &\quad + \mathbf{D}\mathbf{Z}'\left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}\mathbf{R}'\right]\mathbf{Z}\mathbf{D} - \mathbf{D}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{D}. \quad (54) \end{aligned}$$

Writing $\mathbf{A}$ for $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ gives

$$\begin{aligned}\mathbf{B} &= (\mathbf{A}+\mathbf{D}^{-1})^{-1} + \mathbf{DAD} - \mathbf{DA}(\mathbf{A}+\mathbf{D}^{-1})^{-1}\mathbf{AD} - \mathbf{DZ}'\mathbf{PZD}\\ &= (\mathbf{A}+\mathbf{D}^{-1})^{-1} + \mathbf{DA}(\mathbf{A}+\mathbf{D}^{-1})^{-1}(\mathbf{A}+\mathbf{D}^{-1}-\mathbf{A})\mathbf{D} - \mathbf{DZ}'\mathbf{PZD}\\ &= (\mathbf{I}+\mathbf{DA})(\mathbf{A}+\mathbf{D}^{-1})^{-1} - \mathbf{DZ}'\mathbf{PZD}\\ &= \mathbf{D} - \mathbf{DZ}'\mathbf{PZD}. \end{aligned} \tag{55}$$

Thus

$$\mathbf{L}^{-} = \begin{bmatrix} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} & -(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZD} \\ -\mathbf{DZ}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} & \mathbf{D}-\mathbf{DZ}'\mathbf{PZD} \end{bmatrix}, \tag{56}$$

and this, as indicated by (43), (44), and (45), is

$$\mathbf{L}^{-} = \begin{bmatrix} \operatorname{var}(\tilde{\boldsymbol{\beta}}) & -\operatorname{cov}(\tilde{\boldsymbol{\beta}},\mathbf{u}') = \operatorname{cov}[\tilde{\boldsymbol{\beta}},(\tilde{\mathbf{u}}-\mathbf{u})'] \\ -\operatorname{cov}(\mathbf{u},\tilde{\boldsymbol{\beta}}') = \operatorname{cov}[(\tilde{\mathbf{u}}-\mathbf{u}),\boldsymbol{\beta}'] & \operatorname{var}(\tilde{\mathbf{u}}-\mathbf{u}) \end{bmatrix}, \tag{57}$$

as we now establish.

Basic variances and covariances are

$$\operatorname{var}(\mathbf{y}) = \mathbf{V}, \qquad \operatorname{var}(\mathbf{u}) = \mathbf{D}, \quad \text{and} \quad \operatorname{cov}(\mathbf{y},\mathbf{u}') = \mathbf{ZD}.$$

Then with $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ of (41), and assuming $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ to be a symmetric reflexive generalized inverse (which can always be achieved), we get

$$\operatorname{var}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}.$$

And from (42), with $\mathbf{P}$ of (44) having $\mathbf{PVP} = \mathbf{P}$ and $\mathbf{PX} = \mathbf{0}$

$$\tilde{\mathbf{u}} = \mathbf{DZ}'\mathbf{Py}, \qquad \text{so that} \quad \operatorname{var}(\tilde{\mathbf{u}}) = \mathbf{DZ}'\mathbf{PZD},$$

and

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{u}}') = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{VPZD} = \mathbf{0},$$

because $\mathbf{X'P} = \mathbf{0}$, since $\mathbf{PX} = \mathbf{0}$. But

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}}, \mathbf{u}') = (\mathbf{X'V}^{-1}\mathbf{X})^{-}\mathbf{X'V}^{-1}\mathbf{ZD} \quad \text{and} \quad \operatorname{cov}(\tilde{\mathbf{u}}, \mathbf{u}') = \mathbf{DZ'PZD}.$$

Hence

$$\operatorname{var}(\tilde{\mathbf{u}} - \mathbf{u}) = \mathbf{DZ'PZD} + \mathbf{D} - 2\mathbf{DZ'PZD} = \mathbf{D} - \mathbf{DZ'PZD},$$

and

$$\operatorname{cov}\left[\tilde{\boldsymbol{\beta}}, (\tilde{\mathbf{u}} - \mathbf{u})'\right] = -\operatorname{cov}(\tilde{\boldsymbol{\beta}}, \mathbf{u}') = -(\mathbf{X'V}^{-1}\mathbf{X})^{-}\mathbf{X'V}^{-1}\mathbf{ZD}.$$

Thus is (57) confirmed.

It is to be emphasized that the terms $\operatorname{var}(\tilde{\boldsymbol{\beta}})$ and $\operatorname{cov}[\tilde{\boldsymbol{\beta}}, (\tilde{\mathbf{u}} - \mathbf{u})']$ are not invariant to the choice of $(\mathbf{X'V}^{-1}\mathbf{X})^{-}$, and should only be used in the form $\operatorname{var}(\mathbf{X}\tilde{\boldsymbol{\beta}})$ and $\operatorname{cov}[\mathbf{X}\tilde{\boldsymbol{\beta}}, (\tilde{\mathbf{u}} - \mathbf{u})']$ which do have the invariance property.

iv. Maximum Likelihood for σ^2s. The model for (19) and (20) is that in $\mathbf{u}' = [\mathbf{u}_1' \ \mathbf{u}_2' \ \cdots \ \mathbf{u}_i' \ \cdots \ \mathbf{u}_r']$ each $\mathbf{u}_i$ is a vector of the effects of all q_i levels of the ith random factor that occur in the data; and $\operatorname{var}(\mathbf{u}_i) = \sigma_i^2\mathbf{I}_{q_i}$. The maximum-likelihood equations for estimating those variance components can be written (Searle et al., 1992, Section 7.6c) as

$$\sigma_e^2 = \frac{\mathbf{y}'\left(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{Z}\tilde{\mathbf{u}}\right)}{N}$$

and

$$\operatorname{tr}\left(\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\right) = \mathbf{y'PZ}_i\mathbf{Z}_i'\mathbf{Py} \qquad \text{for} \quad i = 1, \ldots, r. \tag{58}$$

On defining $\mathbf{W}_{ii}$ as the (i, i)th submatrix of $\mathbf{W} = (\mathbf{I} + \mathbf{Z'R}^{-1}\mathbf{ZD})^{-1}$ it can be shown that

$$\operatorname{tr}\left(\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\right) = \frac{q_i - \operatorname{tr}(\mathbf{W}_{ii})}{\sigma_i^2}$$

and

$$\mathbf{y'PZ}_i\mathbf{Z}_i'\mathbf{Py} = \tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_i/\sigma_i^4,$$

and these lead to the iterative procedure

$$\sigma_i^{2(m+1)} = \frac{\tilde{\mathbf{u}}_i'^{(m)}\tilde{\mathbf{u}}_i^{(m)} + \sigma_i^{2(m)}\,\mathrm{tr}\left(\mathbf{W}_{ii}^{(m)}\right)}{q_i}$$

or

$$\sigma_i^{2(m+1)} = \frac{\tilde{\mathbf{u}}_i'^{(m)}\tilde{\mathbf{u}}_i^{(m)}}{q_i - \mathrm{tr}\left(\mathbf{W}_{ii}^{(m)}\right)}$$

and

$$\sigma_e^{2(m+1)} = \frac{\mathbf{y}'\left(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}^{(m)} - \mathbf{Z}\mathbf{u}^{(m)}\right)}{N}.$$

With $\mathbf{W}_{ii}$ replaced by $\mathbf{T}_{ii}$, the (i, i)th submatrix of $(\mathbf{I} + \mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{D})^{-1}$ where $\mathbf{S} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{R}^{-1}$, and N replaced by $N - r(\mathbf{X})$, the above procedure yields restricted maximum-likelihood estimates of the variance components. Searle et al. (1992) has full details, including derivation of information matrices in terms of $\mathbf{W}$ and $\mathbf{T}$.

5.6. *A Conditional Distribution*

The function in (39) is $f(\mathbf{u}, \mathbf{y})$ derived as $f(\mathbf{y}|\mathbf{u})f(\mathbf{u})$. It is a standard result that when

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}\right]$$

one has

$$\mathbf{x}_1|\mathbf{x}_2 \sim N\left[\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),\quad \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\right].$$

Applying this directly to (38) gives

$$\mathbf{u}|\mathbf{y} \sim N\left[\mathbf{0} + \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),\quad \mathbf{D} - \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D}\right].$$

Thus

$$\mathrm{E}(\mathbf{u}|\mathbf{y}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

which has $\boldsymbol{\beta}$ where $\tilde{\mathbf{u}} = \text{BLUP}(\mathbf{u})$ has $\tilde{\boldsymbol{\beta}}$. And

$$\text{var}(\mathbf{u}|\mathbf{y}) = \mathbf{D} - \mathbf{DZ}'\mathbf{V}^{-1}\mathbf{ZD}, \tag{59}$$

which has $\mathbf{V}^{-1}$ where $\tilde{\mathbf{u}}$ has $\mathbf{P}$. It can also be noted, using (48), that (59) is

$$\begin{aligned}\text{var}(\mathbf{u}|\mathbf{y}) &= \mathbf{D} - (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZD} \\ &= \mathbf{D} - (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} - \mathbf{D}^{-1})\mathbf{D} \\ &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}.\end{aligned}$$

From (24) we see that $E(\mathbf{u}|\mathbf{y})$ is BLP($\mathbf{u}$).

5.7. *BLUPs That Add to Zero*

In models that have μ in their model equation [e.g., (16)] the vector $\boldsymbol{\beta}$ and matrix $\mathbf{X}$ will take forms $[\mu \ \boldsymbol{\beta}'_0]'$ and $[\mathbf{1}_N \ \mathbf{X}_0]$ for some $\boldsymbol{\beta}_0$ and $\mathbf{X}_0$, respectively. Then, on confining attention to the customary case when $\mathbf{R}$ and $\mathbf{D}$ are diagonal, as in (18) and (19), when the MMEs then have the form of (50), the first of those equations will be

$$\mathbf{1}'_N\mathbf{1}_N\tilde{\mu} + \mathbf{1}'_N\mathbf{X}_0\tilde{\boldsymbol{\beta}}_0 + \mathbf{1}'_N\sum_{j=1}^{r}\mathbf{Z}_j\tilde{\mathbf{u}}_j = \mathbf{1}'\mathbf{y}, \tag{60}$$

and the ith set of equations coming from the last row of (50) will be

$$\mathbf{Z}'_i\mathbf{1}_N\tilde{\mu} + \mathbf{Z}'_i\mathbf{X}_0\tilde{\boldsymbol{\beta}}_0 + \mathbf{Z}'_i\sum_{j=1}^{r}\mathbf{Z}_j\tilde{\mathbf{u}}_j + \lambda_i\tilde{\mathbf{u}}_i = \mathbf{Z}'_i\mathbf{y}. \tag{61}$$

Now it frequently is the case that $\mathbf{Z}$ is an incidence matrix, with each $\mathbf{Z}_i$ being one also; and each $\mathbf{Z}_i$ corresponds to all q_i effects (occurring in the data) of the ith random-effects factor. Then every row of $\mathbf{Z}_i$ is null except for one element being 1. Therefore

$$\mathbf{Z}_i\mathbf{1}_{q_i} = \mathbf{1}_N. \tag{62}$$

Now premultiply (61) by $\mathbf{1}'_{q_i}$ and use the transpose of (62), and subtract the result from (60). This gives $\lambda_i\mathbf{1}'_{q_i}\tilde{\mathbf{u}}_i = 0$; i.e, if $\tilde{\mathbf{u}}_{it}$ is the tth element of $\tilde{\mathbf{u}}_i$,

$$\sum_{t=1}^{q_i}\tilde{u}_{it} = 0.$$

Thus has been established the result that for each random factor the sum of the BLUPs of its effects is zero. And this, be it noted, has nothing to do with a restriction that is often used for fixed effects—e.g., in the fixed-effects model $y_{tj} = \mu + \alpha_t + e_{tj}$, that $\Sigma_t \alpha_t = 0$. That is not used when the α_i's are taken as random effects, and yet $\Sigma_t \tilde{\alpha}_t = \Sigma_t$ BLUP$(\alpha_t) = 0$, as is evident from $\tilde{\alpha}_t$ following (51).

6. VARIANCE COMPONENTS

The reader will notice that save for the few sentences in Sections 5.4 and 5.5.iv, no discussion is given of estimating the variance components σ_e^2 and σ_i^2 for $i = 1, \ldots, r$ occurring in $\mathbf{D}$ of (19) and hence in $\mathbf{V}$, $\mathbf{V}^{-1}$, and $\mathbf{P}$. And, of course, estimates are needed in order to get numerical values for $\mathbf{X}\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$ and their sampling variances.

REFERENCES

Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model, *J. Amer. Statist. Assoc.* 57:369–375.

Hartley, H. O., and Rao, J. N. K., 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54:93–108.

Harville, D. A. 1990. Comment on Robinson (1990), *Statist. Sci.* 6:35–39.

Henderson, C. R. 1950. Estimation of genetic parameters (Abstract), *Ann. Math. Statist.* 21:309–310.

Henderson, C. R. 1973. Maximum likelihood estimation of variance components. Unpublished manuscript, Department of Animal Science, Cornell University, Ithaca, N.Y., 14850.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. N. 1959. Estimation of environmental and genetic trends from records subject to culling, *Biometrics* 13:192–218.

Kempthorne, O. 1980. The term design matrix, *Amer. Statist.* 34:249.

Mood, A. M. 1950. *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Patterson, H. D., and Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554.

Puntanen, S. and Styan, G. P. H. 1989. On the equality of the ordinary least squares estimator and the best linear unbiased estimator, *Amer. Statist.* 43:153–164.

Robinson, G. K. 1990. The estimation of random effects, *Statist. Sci.* 6:15–32.

Searle, S. R. 1988. Best linear unbiased estimation in mixed models of the analysis of variance, in *Probability and Statistics: Essays in Honor of Franklin A. Graybill* (J. Srivastava, Ed.), North-Holland, Amsterdam, pp. 233–241.

Searle, S. R. 1994. A derivation of BLUP which is similar to that of BLUE, Tech. Rep. BU-1261-M, Biometrics Unit, Cornell Univ., Ithaca, N.Y.

Searle, S. R., Casella, G., and McCulloch, C. E. 1992. *Variance Components*, Wiley, New York.

Received 19 August 1995; final manuscript accepted 8 May 1996