A Derivation of BLUP That is Similar to That of BLUE

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

BU-1261-M

August 1994

A Derivation of BLUP That is Similar to That of BLUE

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y. 14853

BU-1261-M August 1994

ABSTRACT

A simple derivation of best linear unbiased prediction (BLUP) is shown to be very similar to that of best linear unbiased estimation (BLUE). Indeed, although BLUP is usually concerned with random effects, BLUP of fixed effects is the same as BLUE.

1. INTRODUCTION

We deal with the familiar mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

where

$\mathbf{y}$ is a vector of data;

$\boldsymbol{\beta}$ is a vector of fixed, unknown constants;

$\mathbf{X}$ is a known matrix, often an incidence matrix (of elements 0 or 1), although it can include columns of observed covariates;

$\mathbf{u}$ is a vector of unobservable random effects;

$\mathbf{Z}$ is a known incidence matrix;

$\mathbf{e}$ is a vector of random error terms.

The first and second moments attributed to $\mathbf{u}$ and $\mathbf{e}$ are

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \right] \tag{2}$$

so that

$$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} .$$

Of the numerous estimation procedures applicable to (1), the most common two are ordinary least squares estimation (OLSE) and best linear unbiased estimation (BLUE) of the fixed effects $\boldsymbol{\beta}$. To circumvent problems of estimability, attention is confined to the vector $\mathbf{X}\boldsymbol{\beta}$, or to linear combinations of its elements, and we have

$$\text{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \tag{3}$$

and

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{4}$$

where, for any matrix $\mathbf{A}$ say, $\mathbf{A}^{-}$ is a generalized inverse satisfying $\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$. For notational convenience we write

$$\boldsymbol{\beta}^{o} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{5}$$

and so have

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}^{o} \; . \tag{6}$$

Another estimation procedure used with (1) is best linear prediction (BLUP) of $\mathbf{u}$:

$$\text{BLUP}(\mathbf{u}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{o}) \; . \tag{7}$$

Derivation of BLUE($\mathbf{X}\boldsymbol{\beta}$) is well known; and there are numerous derivations of BLUP($\mathbf{u}$), as in Robinson (1991) and Searle *et al.* (1992). Between them they present derivations based on (i) Goldberger's (1962) prediction of a future observation, (ii) Henderson's (1959) mixed model equations, (iii) a Bayes model, (iv) a two-stage regression-like approach, and (v) a partitioning of $\mathbf{y}$ into two orthogonal parts. And Harville (1990) has yet another approach.

Most of these derivations involve some quite heady matrix manipulations. Moreover, connection of deriving BLUP($\mathbf{u}$) to deriving BLUE($\mathbf{X}\boldsymbol{\beta}$) can appear tenuous. BLUP($\mathbf{u}$) of (7) is, in appearance, so different from BLUE($\mathbf{X}\boldsymbol{\beta}$) of (6) that many are those who feel that BLUP and BLUE are quite different concepts. It is the purpose of this note to show that this is not so: that BLUP can be derived in precisely the same way as BLUE and, indeed, the derivation of BLUP yields

$$\text{BLUP}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}^{o} = \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) \; .$$

Moreover, the derivation seems to be about as simple and direct as one can imagine.

2. DERIVING BLUE($\mathbf{X}\beta$)

Although deriving BLUE($\mathbf{X}\beta$) is well known, a familiar derivation is provided here as a basis for deriving BLUP. Begin by defining what is to be estimated: $\mathbf{t}'\mathbf{X}\beta$ for any row vector $\mathbf{t}' \neq \mathbf{0}$. We want the estimator to be

$$\text{(i)} \quad \text{linear in } \mathbf{y}\text{:} \ \lambda'\mathbf{y} \ \text{ for } \ \lambda' \neq \mathbf{0}\ ; \tag{8}$$

$$\text{(ii)} \quad \text{unbiased: } \mathrm{E}(\lambda'\mathbf{y}) = \mathbf{t}'\mathbf{X}\beta\ , \tag{9}$$

which implies

$$\lambda'\mathbf{X}\beta = \mathbf{t}'\mathbf{X}\beta \quad \forall\ \beta, \tag{10}$$

i.e.,

$$\lambda'\mathbf{X} = \mathbf{t}'\mathbf{X}, \quad \text{or} \quad \mathbf{X}'\lambda = \mathbf{X}'\mathbf{t}\ ; \tag{11}$$

and (iii) of minimum variance, i.e.,

$$\mathrm{var}(\lambda'\mathbf{y}) = \lambda'\mathbf{V}\lambda \text{ is to be minimized w.r.t. } \lambda\ . \tag{12}$$

To achieve (12) subject to (11) we minimize

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

where $\mathbf{m}'$ is a vector of Lagrange multipliers. To do this

$$\partial\theta/\partial\lambda = \mathbf{0} \quad \text{gives} \quad 2\mathbf{V}\lambda + 2\mathbf{X}\mathbf{m} = \mathbf{0},\ \text{ i.e., } \ \lambda = -\mathbf{V}^{-1}\mathbf{X}\mathbf{m}\ . \tag{13}$$

$$\partial\theta/\partial\mathbf{m} = \mathbf{0} \quad \text{gives} \quad \mathbf{X}'\lambda = \mathbf{X}'\mathbf{t}\ . \tag{14}$$

Then (13) and (14) give $\mathbf{m} = -(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{t}$ and so using this in (13) yields

$$\lambda = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{t}$$

and so

$$\lambda'\mathbf{y} = \mathrm{BLUE}(\mathbf{t}'\mathbf{X}\beta) = \mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\ . \tag{15}$$

Letting $\mathbf{t}'$ be the successive rows of an identity matrix yields the result in (6)

$$\mathrm{BLUE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}\beta^{o}\ .$$

3. DERIVING BLUP

Before deriving BLUP along the same lines as the preceding derivation of BLUE, we make one small change to that derivation. Instead of (9) we write it equivalently as

$$\mathrm{E}(\lambda'\mathbf{y} - \mathbf{t}'\mathbf{X}\beta) = \mathbf{0}\ , \tag{16}$$

and similarly rewrite (12) as

$$\mathrm{var}(\lambda'\mathbf{y} - \mathbf{t}'\mathbf{X}\beta) \text{ is to be minimized w.r.t. } \lambda\ . \tag{17}$$

Both of these statements involve $\lambda'y - t'X\beta$ which can be called the estimation error, the difference between the estimator $\lambda'y$ and what we seek to estimate, namely $t'X\beta$. In adapting the preceding derivation of BLUE to this concept of estimation error (or prediction error) we derive BLUP.

Begin by wanting to estimate $t'X\beta + s'u$, for any $[t' \quad s'] \neq 0$. We want the estimator to be

$$\text{(i) linear in y: } \lambda'y \quad \text{for} \quad \lambda' \neq 0;$$

$$\text{(ii) unbiased: } E[\lambda'y - (t'X\beta + s'u)] = 0\,. \tag{18}$$

This is the adaptation of (16), with $\lambda'y - (t'X\beta + s'u)$ being the prediction error; and with $E(u) = 0$, as in (2), equation (18) gives

$$\lambda'X\beta - t'X\beta = 0 \quad \forall\ \beta$$

$$\Rightarrow \quad \lambda'X = t'X \quad \text{or} \quad X'\lambda = X't\,, \tag{19}$$

exactly as in (11). But note that this is so only because $E(u) = 0$ of (2). Were $E(u) \neq 0$, equation (18) would yield something different from (19).

The final thing we want for the estimator is that it be of minimum variance. And here, the adaptation of (17) is that we want to

$$\text{minimize } \operatorname{var}[\lambda'y - (t'\beta + s'u)] = \lambda'V\lambda + s'Ds - 2\lambda'ZDs\,. \tag{20}$$

To achieve (20) subject to (19) we minimize

$$\theta = \lambda'V\lambda + s'Ds - 2\lambda'ZDs + 2m'(X'\lambda - X't)\,.$$

To do this

$$\partial\theta/\partial\lambda = 0 \quad \text{gives} \quad 2V\lambda - 2ZDs + 2Xm = 0$$

$$\Rightarrow \quad \lambda = -V^{-1}Xm + V^{-1}ZDs\,, \tag{21}$$

$$\partial\theta/\partial m = 0 \quad \text{gives} \quad X'\lambda = X't\,. \tag{22}$$

Then, just like (13) and (14), so do (21) and (22) give $m = -(X'V^{-1}X)^{-}(X't - X'V^{-1}ZDs)$ and using this in (21) gives

$$\lambda = V^{-1}X(X'V^{-1}X)^{-}X't + [I - V^{-1}X(X'V^{-1}X)^{-}X']\,V^{-1}ZDs$$

and so

$$\begin{aligned}\lambda'y &= \text{BLUP}(t'X\beta + s'u)\\ &= t'X(X'V^{-1}X)^{-}X'V^{-1}y + s'DZ'V^{-1}[I - X(X'V^{-1}X)^{-}X'V^{-1}]\,y\\ &= t'X\beta^{o} + s'DZ'V^{-1}(y - X\beta^{o})\,.\end{aligned} \tag{23}$$

The form of this result, and the method of obtaining it, are similar to the work of Goldberger (1962), except that he was concerned with predicting a future observation, whereas (23) predicts any linear combination of elements of $\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{u}$.

In particular, if in (23) we let $\mathbf{t}'$ be successive rows of an identity matrix and $\mathbf{s}'$ be null, we get

$$\text{BLUP}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}^o = \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) \ . \tag{24}$$

Similarly, if in (23) we take $\mathbf{t}'$ as null, and let $\mathbf{s}'$ be rows of an identity matrix we get

$$\text{BLUP}(\mathbf{u}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^o) \ . \tag{25}$$

This, as Speed (1990) indicates, is an uncomplicated expression for BLUP($\mathbf{u}$); it is also available in Searle *et al.* (1992, pages 270 and 274).

Thus we see that the adaptation of considering expectation (or prediction) error provides a very straightforward derivation of BLUP, giving (23), (24) and (25) as the BLUPs of $\mathbf{t}'\mathbf{X}\boldsymbol{\beta} + \mathbf{s}'\mathbf{u}$, of $\mathbf{X}\boldsymbol{\beta}$ and of $\mathbf{u}$.

REFERENCES

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

Harville, D.A. (1990). Comment on Robinson (1990). *Statistical Science* **6**, 35-39.

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.

Robinson, G.K. (1990). The estimation of random effects. *Statististical Science* **6**, 15-32.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*, Wiley & Sons, N. Y.

Speed, T.M. (1990). Comment on Robinson (1990). *Statististical Science* **6**, 42-44.