ALTERNATIVE DERIVATIONS OF BEST LINEAR UNBIASED PREDICTION (BLUP) IN THE MIXED MODEL

BU-861-M

by

October, 1984

Shayle R. Searle and George Casella[1]
Biometrics Unit, Cornell University, Ithaca, New York

Abstract

A new insightful derivation of Henderson's (1972) BLUP procedure is offered; Bulmer's (1980) predictor is generalized; and its relationship to its own sampling variance is shown to greatly simplify establishing its equality to BLUP.

We deal with what is known as the mixed linear model denoted by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{1}$$

where, for E representing expectation over repeated sampling, $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\alpha}$, $\mathbf{X}$ is an incidence matrix, $\boldsymbol{\alpha}$ is a vector of fixed effects parameters, $\mathbf{Z}$ is an incidence matrix, and $\mathbf{b}$ is a vector of random variables representing the random effects in the model; and $\mathbf{e}$ is a vector of residual error terms. Estimation problems in this model include those of estimating $\boldsymbol{\alpha}$, and of predicting $\mathbf{b}$. The different descriptions "estimating $\boldsymbol{\alpha}$" and "predicting $\mathbf{b}$" are used to distinguish fixed effects from random effects.

1/ Research supported by National Science Foundation Grant No.MCS81-02541.

1. Henderson's BLUP

A well-known procedure for estimating α and predicting b, especially in the context of evaluating the relative genetic merit of bulls used in artificial insemination in the dairy industry, is the BLUP (best, linear unbiased prediction) of Henderson (1972), wherein the predictor of b is taken as

$$\tilde{b} = \mathrm{cov}(b,\ y')V^{-1}(y - X\alpha^{o}) \tag{2a}$$

for

$$\alpha^{o} = (X'V^{-1}X)^{-}X'V^{-1}y \tag{2b}$$

where $\mathrm{cov}(b,\ y')$ is the covariance matrix of b with y', and the dispersion matrix of y is

$$\mathrm{var}(y) = V\ , \tag{3}$$

taken as being positive definite. Derivation of (2a) and (2b) is obtained by Henderson (1972) on the basis of making normality assumptions about b and e in (1), and then maximizing a density function — as discussed in Henderson, Kempthorne, Searle and von Krosigk (1959).

2. Bulmer's two-stage procedure

Bulmer (1980) suggests a two-stage approach to the problem of predicting b: first, form a vector of the records y corrected for the fixed effects (in the genetic context, corrected for the environmental effects):

$$w = y - X\alpha^{o}\ , \tag{4}$$

where α^{o} is as in (2b), and $X\alpha^{o}$ is the best linear unbiased estimator

(BLUE) of $\underset{\sim}{X}\underset{\sim}{\alpha}$. Then, under normality assumptions, $\underset{\sim}{b}$ is predicted by the intuitively appealing regression estimator $E(\underset{\sim}{b}|\underset{\sim}{w})$,

$$\hat{\underset{\sim}{b}} = \text{cov}(\underset{\sim}{b}, \underset{\sim}{w}')[\text{var}(\underset{\sim}{w})]^{-1}\underset{\sim}{w} ,$$

which is well-known to be optimal among the class of predictors based on $\underset{\sim}{w}$.

Gianola and Goffinet (1982) show that Bulmer's $\hat{\underset{\sim}{b}}$ is identical to Henderson's $\tilde{\underset{\sim}{b}}$, and include a discussion by Bulmer in which the equivalence is gladly acknowledged. In showing this equivalence, Gianola and Goffinet note that the elements of $\underset{\sim}{w}$ are not linearly independent (indeed, $\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{w} = \underset{\sim}{0}$), so that the dispersion matrix of $\underset{\sim}{w}$,

$$\underset{\sim}{V}_w = \text{var}(\underset{\sim}{w}) , \tag{5}$$

is singular. Hence their general form of Bulmer's predictor is

$$\underset{\sim}{b}^* = \text{cov}(\underset{\sim}{b}, \underset{\sim}{w}')\underset{\sim}{V}_w^-\underset{\sim}{w} \tag{6}$$

where $\underset{\sim}{V}_w^-$ is a generalized inverse of $\underset{\sim}{V}_w$ satisfying $\underset{\sim}{V}_w\underset{\sim}{V}_w^-\underset{\sim}{V}_w = \underset{\sim}{V}_w$.

3. Equivalence of the BLUP and Bulmer procedures

In showing the equality of $\underset{\sim}{b}^*$ of (6) to $\tilde{\underset{\sim}{b}}$ of (2a), Gianola and Goffinet (1982) indicate that from (4) and (2b) one can easily derive $\underset{\sim}{V}_w$ of (5) as

$$\underset{\sim}{V}_w = \underset{\sim}{V} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^-\underset{\sim}{X}' . \tag{7}$$

They then observe that "it is easy to show that"

$$\underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{V}_w = \underset{\sim}{V}_w , \tag{8}$$

and hence, they write, "Replacing $\underset{\sim}{V}_w^-$ in (6) by $\underset{\sim}{V}^{-1}$ yields (2a)." All this is correct. But the development both of (8) and of (2a) from (6) can be achieved more directly than might be apparent from Gianola and Goffinet's presentation. Indeed, after replacing $\underset{\sim}{V}_w^-$ in (6) they give no indication as to why cov($\underset{\sim}{b}$, $\underset{\sim}{w}'$) in (6) can be apparently replaced by cov($\underset{\sim}{b}$,$\underset{\sim}{y}'$) — an indication that seems to be much needed since these two covariance matrices are not equal.

In passing, observe that (7) is more general than Gianola and Goffinet's form of $\underset{\sim}{V}_w$ which has the regular inverse, $(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}$, rather than $(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^-$ as in (7), because they define $\underset{\sim}{X}$ as having full column rank. That limitation on $\underset{\sim}{X}$ is not assumed here. Nevertheless, (7) is both symmetric and invariant to whatever generalized inverse of $\underset{\sim}{X}\underset{\sim}{V}^{-1}\underset{\sim}{X}$ is used for $(\underset{\sim}{X}\underset{\sim}{V}^{-1}\underset{\sim}{X})^-$. This is so because, since $\underset{\sim}{V}$ is positive definite there exists a non-singular $\underset{\sim}{K}$ such that $\underset{\sim}{V}^{-1} = \underset{\sim}{K}'\underset{\sim}{K}$, and so

$$\begin{aligned}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^-\underset{\sim}{X}' &= \underset{\sim}{K}^{-1}\underset{\sim}{K}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{K}'\underset{\sim}{K}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{K}\underset{\sim}{K}'^{-1}\\ &= \underset{\sim}{K}^{-1}\underset{\sim}{T}(\underset{\sim}{T}'\underset{\sim}{T})^-\underset{\sim}{T}'\underset{\sim}{K}'^{-1} \quad \text{for } \underset{\sim}{T} = \underset{\sim}{K}\underset{\sim}{X}\ ;\end{aligned}$$

and it is well known [e.g., Searle (1982), page 221] for any $\underset{\sim}{T}$ that $\underset{\sim}{T}(\underset{\sim}{T}'\underset{\sim}{T})^-\underset{\sim}{T}'$ is both symmetric and invariant to the choice of $(\underset{\sim}{T}'\underset{\sim}{T})^-$; also $\underset{\sim}{T}(\underset{\sim}{T}'\underset{\sim}{T})^-\underset{\sim}{T}'\underset{\sim}{T} = \underset{\sim}{T}$.

The verity of (8) is more readily established indirectly from (7) than directly, by observing from (2b), (3), (4) and (7) that

$$\underset{\sim}{w} = \underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{y}\ . \tag{9}$$

Hence

$$\text{var}(\underset{\sim}{w}) = \underset{\sim}{V}_w = \underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{V}\underset{\sim}{V}^{-1}\underset{\sim}{V}_w = \underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\ ,$$

which is (8).

Now we show the steps that lead to

$$\text{cov}(\underset{\sim}{b},\ \underset{\sim}{w}')\underset{\sim}{V}^{-1}\underset{\sim}{w} = \text{cov}(\underset{\sim}{b},\ \underset{\sim}{y}')\underset{\sim}{V}^{-1}\underset{\sim}{w}\ .$$

Using $\underset{\sim}{V}^{-1}$ for $\underset{\sim}{V}_w^-$ in (6) gives

$$\begin{aligned}\underset{\sim}{b}^* &= \text{cov}(\underset{\sim}{b}, \underset{\sim}{w}')\underset{\sim}{V}^{-1}\underset{\sim}{w} \\ &= \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}'\underset{\sim}{V}^{-1}\underset{\sim}{V}_w)\underset{\sim}{V}^{-1}\underset{\sim}{w} \\ &= \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{y}, \text{ on using (9)} \\ &= \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\underset{\sim}{V}^{-1}\underset{\sim}{y}, \text{ from (8)} \\ &= \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}\underset{\sim}{w}, \text{ using (9) again.}\end{aligned}$$

Thus is the replacement of $\text{cov}(\underset{\sim}{b}, \underset{\sim}{w}')$ by $\text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')$ established, even though these covariance matrices are not equal: on defining $\text{var}(\underset{\sim}{b}) = \underset{\sim}{D}$, and $\text{cov}(\underset{\sim}{b}, \underset{\sim}{e}') = \underset{\sim}{0}$ as is usual, we have $\text{cov}(\underset{\sim}{b}, \underset{\sim}{y}') = \underset{\sim}{D}\underset{\sim}{Z}'$, and (9) then gives $\text{cov}(\underset{\sim}{b}, \underset{\sim}{w}') = \underset{\sim}{D}\underset{\sim}{Z}'\underset{\sim}{V}^{-1}\underset{\sim}{V}_w$.

Finally, of course, we have

$$\underset{\sim}{b}^* = \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}\underset{\sim}{w} = \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}^o) = \tilde{\underset{\sim}{b}} \qquad (10)$$

of (2a).

4. A direct derivation of BLUP

At first it might seem that a difference between the methods of Henderson and Bulmer is that in Bulmer's method one adjusts the data for the fixed effects before predicting $\underset{\sim}{b}$, while in Henderson's method the fixed and random effects are handled simultaneously. However, if one's interest is in unbiased linear prediction of $\underset{\sim}{b}$, then it makes no difference whatsoever whether the data are adjusted for the fixed effects or not. A linear predictor, $\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{y}$, of $\underset{\sim}{b}$ is unbiased if and only if $\underset{\sim}{a} = \underset{\sim}{0}$ and $\underset{\sim}{B}\underset{\sim}{X} = \underset{\sim}{0}$, i.e.,

$$E(\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{y}) = \underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{X}\underset{\sim}{\alpha} = E\underset{\sim}{u} = \underset{\sim}{0} \iff \underset{\sim}{a} = \underset{\sim}{0} \text{ and } \underset{\sim}{B}\underset{\sim}{X} = \underset{\sim}{0} .$$

Using Lemma 2 of the appendix, it follows that if $\mathbf{BX} = \mathbf{0}$, then we can write

$$\mathbf{B} = \mathbf{L}(\mathbf{I} - \mathbf{H}) \quad \text{for} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} , \tag{11}$$

for some matrix $\mathbf{L}$. Thus our predictor is $\mathbf{By} = \mathbf{L}(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{L}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}^{o})$ for some $\mathbf{L}$. Hence prediction of $\mathbf{b}$ is based on $\mathbf{y}$ only through the quantity $\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}^{o}$, showing that the data need not be adjusted for the fixed effects; the condition of unbiasedness does this adjustment automatically.

It is instructive, and statistically illuminating, to carry this argument further. Multiplying the model (1) by first $\mathbf{H}$ and then $(\mathbf{I} - \mathbf{H})$ gives us two models:

$$\mathbf{Hy} = \mathbf{HX}\boldsymbol{\alpha} + \mathbf{HZb} + \mathbf{He} \tag{12a}$$

and

$$(\mathbf{I} - \mathbf{H})\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{Zb} + (\mathbf{I} - \mathbf{H})\mathbf{e} , \tag{12b}$$

since $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$. Note that $\mathbf{y} = \mathbf{Hy} + (\mathbf{I} - \mathbf{H})\mathbf{y}$, and since by the nature of $\mathbf{H}$,

$$\mathbf{HV} = \mathbf{VH}' \text{ and } \mathbf{H}^2 = \mathbf{H}, \quad \text{so that} \quad \mathbf{HV}(\mathbf{I} - \mathbf{H})' = \mathbf{0} , \tag{13}$$

we can observe that the two model equations in (12) represent a split of model equation (1) into two uncorrelated pieces. One of them then provides information on $\boldsymbol{\alpha}$ and the other on $\mathbf{b}$.

From (12a) the BLUE of $\boldsymbol{\alpha}$ is

$$\dot{\boldsymbol{\alpha}} = [\mathbf{X}'\mathbf{H}'(\mathbf{HVH}')^{-}\mathbf{X}'\mathbf{H}']^{-}(\mathbf{HVH}')^{-}\mathbf{Hy} . \tag{14}$$

Using (13) we find that a generalized inverse of $\mathbf{HVH}'$ is $\mathbf{H}'\mathbf{V}^{-1}\mathbf{H}$, and along with $\mathbf{HX} = \mathbf{X}$ it is then straightforward to show that (14) reduces to the usual BLUE

$$\dot{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^{o} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} , \tag{15}$$

of (2b). Similarly, from (12b), the best linear predictor of $\underset{\sim}{b}$ is given by

$$\dot{\underset{\sim}{b}} = \text{cov}[\underset{\sim}{b}, \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H})][\text{var}((\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y})]^{-}(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y} . \qquad (16)$$

Since $(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y} = \underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}^o = \underset{\sim}{w}$ of (4), it is clear that (16) is the same as (6) which in (10) we have already shown is the BLUP $\tilde{\underset{\sim}{b}}$. The same algebra reduces (16) to $\tilde{\underset{\sim}{b}}$.

A variant of this derivation is to determine $\underset{\sim}{L}$ of $\underset{\sim}{B} = \underset{\sim}{L}(\underset{\sim}{I} - \underset{\sim}{H})$ of (11) by deciding that $\underset{\sim}{B}\underset{\sim}{y}$ shall be the predictor of $\underset{\sim}{b}$ and that it will also be the best linear predictor of $\underset{\sim}{b}$ based on $\underset{\sim}{B}\underset{\sim}{y}$. Thus $\underset{\sim}{L}$ is determined from the equation

$$\text{cov}[\underset{\sim}{b}, (\underset{\sim}{B}\underset{\sim}{y})'][\text{var}(\underset{\sim}{B}\underset{\sim}{y})]^{-}\underset{\sim}{B}\underset{\sim}{y} = \underset{\sim}{B}\underset{\sim}{y} ,$$

which is to be true for all $\underset{\sim}{y}$. Writing $\underset{\sim}{C}$ for $\text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')$ and replacing $\underset{\sim}{B}$ by $\underset{\sim}{L}(\underset{\sim}{I} - \underset{\sim}{H})$ we therefore determine $\underset{\sim}{L}$ from

$$\underset{\sim}{C}(\underset{\sim}{I} - \underset{\sim}{H}')\underset{\sim}{L}'[\underset{\sim}{L}(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{V}(\underset{\sim}{I} - \underset{\sim}{H}')\underset{\sim}{L}']^{-}\underset{\sim}{L}(\underset{\sim}{I} - \underset{\sim}{H}) = \underset{\sim}{L}(\underset{\sim}{I} - \underset{\sim}{H}) . \qquad (17)$$

Using (13) again, together with $\underset{\sim}{V}_w = (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{V}$, and post-multiplying (17) by $\underset{\sim}{V}\underset{\sim}{L}'$, gives

$$\underset{\sim}{C}\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\underset{\sim}{L}'(\underset{\sim}{L}\underset{\sim}{V}_w\underset{\sim}{L}')^{-}\underset{\sim}{L}\underset{\sim}{V}_w\underset{\sim}{L}' = \underset{\sim}{L}\underset{\sim}{V}_w\underset{\sim}{L}'$$

and hence

$$\underset{\sim}{C}\underset{\sim}{V}^{-1}\underset{\sim}{V}_w\underset{\sim}{L}' = \underset{\sim}{L}\underset{\sim}{V}_w\underset{\sim}{L}'$$

which is satisfied by $\underset{\sim}{L} = \underset{\sim}{C}\underset{\sim}{V}^{-1}$. Hence

$$\tilde{\underset{\sim}{b}} = \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y} = \text{cov}(\underset{\sim}{b}, \underset{\sim}{y}')\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}^o)$$

as before.

Appendix

Lemma 1 $\mathbf{BX} = \mathbf{0}$ if and only if $\mathbf{B} = \mathbf{K}(\mathbf{I} - \mathbf{M})$ for any $\mathbf{L}$ and for $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$.

Proof $\mathbf{MX} = \mathbf{X}$, so that if $\mathbf{B} = \mathbf{K}(\mathbf{I} - \mathbf{M})$, $\mathbf{BX} = \mathbf{X} - \mathbf{X} = \mathbf{0}$.

If $\mathbf{BX} = \mathbf{0}$, $\mathbf{BM} = \mathbf{0}$, i.e., $\mathbf{MB}' = \mathbf{0}$. Then, on solving this matrix equation for $\mathbf{B}'$ as an extension of the solving of vector equations (e.g., Searle, 1982, Theorem 4, p. 237), we get

$$\mathbf{B}' = (\mathbf{I} - \mathbf{M}^{-}\mathbf{M})\mathbf{K}'$$

for arbitrary $\mathbf{K}$. But $\mathbf{M}$ is idempotent, so $\mathbf{M}^{-}$ can be taken as $\mathbf{M}$, and then

$$\mathbf{B}' = (\mathbf{I} - \mathbf{M}^2)\mathbf{K}' = (\mathbf{I} - \mathbf{M})\mathbf{K}'$$

i.e.,

$$\mathbf{B} = \mathbf{K}(\mathbf{I} - \mathbf{M}) \text{ for arbitrary } \mathbf{K} .$$

Lemma 2 $\mathbf{B} = \mathbf{L}(\mathbf{I} - \mathbf{H})$ for arbitrary $\mathbf{L}$ and for $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}$.

Proof $\mathbf{BX} = \mathbf{0}$ implies $\mathbf{BV}^{\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X} = \mathbf{0}$ for positive definite $\mathbf{V}$. Therefore, from the lemma

$$\begin{aligned}\mathbf{BV}^{\frac{1}{2}} &= \mathbf{K}[\mathbf{I} - \mathbf{V}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-\frac{1}{2}}] \\ &= \mathbf{KV}^{-\frac{1}{2}}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{V}^{\frac{1}{2}} ;\end{aligned}$$

i.e.,

$$\mathbf{B} = \mathbf{L}(\mathbf{I} - \mathbf{H}) \text{ for arbitrary } \mathbf{L}, \text{ in place of } \mathbf{KV}^{-\frac{1}{2}} .$$

References

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics.* Oxford University Press.

Gianola, D. and Goffinet, B. (1982). Sire evaluation with best linear unbiased predictors. *Biometrics* 38, 1085-1088.

Henderson, C. R. (1972). Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, 10-41. Champaign, Illinois: American Society of Animal Science and American Dairy Science Association.

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* 15, 192-218.

Searle, Shayle R. (1982). *Matrix Algebra Useful for Statistics.* Wiley, New York.