Comments on best linear unbiased prediction (BLUP) as used in beef and dairy production improvement plans[1]

Shayle R. Searle[2]

Biometrics Unit, Cornell University, Ithaca, N.Y., U.S.A.

and

Institut für Mathematik, Universität Augsburg,

D-8900 Augsburg, Federal Republic of Germany

BU-874-M   April 1985

Abstract

The selection of superior animals for breeding for improved production of beef and dairy animals has for many years been based on the procedure known as BLUP: best, linear, unbiased prediction. Well-known properties of BLUP are its unbiasedness, its linearity in the observations, its minimum variance, its maximum correlation with the genetic value being predicted, and its maximizing of the probability of correctly ranking pairs of animals. Initial development was through maximizing a function similar to a likelihood, but alternative derivations now available that provide more insight into understanding its use are (1) as a Bayes estimator, (2) as a regression estimator from corrected records, and (3) as a predictor that is invariant to fixed effects. These and other derivations are briefly reviewed.

1. Introduction

Increasing agricultural production of meat and dairy foodstuffs relies upon production improvement programs that utilize both improving environmental factors and improving the genetic level of the species concerned. This demands assessing the relative genetic merit of individual animals available for being the sires and dams of each succeeding generation of animals that will yield meat and/or milk. This assessment of breeding values is usually based on what is called the best linear unbiased predictor (BLUP). Although originally developed from a kind of maximum likelihood derivation, using what are known as the mixed model equations (Henderson, 1972), BLUP can now also be interpreted from several other viewpoints. (I) It is a Bayes estimator (Lindley and Smith, 1972, and Dempfe, 1977). (II) It is a regression predictor, based on an adjusted form of the data (Bulmer, 1980, and Gianola and Goffinet, 1982). (III) It is a best predictor that is invariant to fixed effects (Searle and Casella, 1984). Features of these and other derivations are summarized.

2. Example

The simplest example is that of having single records on n_i daughters of sire i, for sires labeled i = 1 through i = a. Let y_{ij} be the record of the j'th daughter from the i'th sire, for j = 1,...,n_i and i = 1,2,...,a. A usual model equation for y_{ij} is

$$y_{ij} = \mu + s_i + e_{ij} \tag{1}$$

where μ is a general mean, s_i is a (random) sire effect and e_{ij} is a random error term. Usual assumptions on the s_i and e_{ij} are that they have zero means and variances σ_s^2 and σ_e^2 respectively, for all i and j, and <u>all</u> covariances zero. Under these conditions the best linear unbiased estimator of μ is

$$\hat{\mu} = \sum_{i=1}^{a} \frac{n_i \bar{y}_{i\cdot}}{\sigma_e^2 + n_i \sigma_s^2} \Bigg/ \sum_{i=1}^{a} \frac{n_i}{\sigma_e^2 + n_i \sigma_s^2} \quad , \tag{2}$$

and the best linear unbiased predictor (BLUP) of the sire's daughter average

$$d_i = \mu + s_i, \tag{3}$$

is

$$\tilde{d}_i = \hat{\mu} + \tilde{s}_i = \hat{\mu} + \frac{n_i \sigma_s^2}{\sigma_e^2 + n_i \sigma_s^2} (\bar{y}_{i\cdot} - \hat{\mu}) \quad . \tag{4}$$

Using $h^2 = 4\sigma_s^2/(\sigma_s^2 + \sigma_e^2)$, the parameter familiar to geneticists as heritability (in the narrow sense), reduces (4) to the immediately-recognizable estimated breeding value used in animal selection programs, namely

$$\tilde{d}_i = \hat{\mu} + \frac{n_i h^2}{4+(n_i-1)h^2}\ (\bar{y}_{i.} - \hat{\mu})\ . \tag{5}$$

Well-known properties of $\tilde{d}_i$ are

    (i) $\tilde{d}_i$ is a linear function of the observations;

    (ii) $\tilde{d}_i$ is unbiased for $E(d_i)$;

among all linear unbiased predictors of d_i

    (iii) $\tilde{d}_i$ has smallest variance;

    (iv) $\tilde{d}_i$ has the largest correlation with d_i;

    (v)   ranking sires i and j on the basis of $\tilde{d}_i$ and $\tilde{d}_j$

maximizes the probability of their being correctly ranked.

Development of these and other variance and covariance properties is shown in full detail in Searle (1974).

3. The General Case

The general form of mixed model, of which (1) is a simple example, is

$$y = X\beta + Zu + e, \tag{6}$$

where y is a vector of observations, β is a vector of fixed effects with X its incidence matrix, u is a vector of random effects with Z being its incidence matrix, and e is a vector of random error terms. Means and dispersion matrices of u and e are assumed to be

$$E\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \text{var}\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & R \end{bmatrix} \tag{7}$$

so that

$$V = \text{var}(y) = ZDZ' + R. \tag{8}$$

Then a generalized least squares (GLS) estimator of β is

$$\hat{\beta} = (X'V^{-1}X)^-X'V^{-1}y \tag{9}$$

and the best linear unbiased predictor (BLUP) of $k'\beta + r'u$ for $k' = t'X$ for some t' is $k'\hat{\beta} + r'u$ where

$$\tilde{u} = DZ'V^{-1}(y - X\hat{\beta}) \tag{10}$$

is the BLUP of $\tilde{u}$. Although $\check{d}_i$ of (4) is an example of $k'\hat{\beta} + r'u$ (as detailed in the appendix) we concentrate on $\tilde{u}$ of (10).

4. Derivations

There are at least seven approaches to deriving the BLUP of $\tilde{u}$ of (10), each of which contributes something to our understanding of it. We briefly describe each of the seven.

4.1 Mixed model equations

Henderson (1963), on making normality assumptions about u and e of (9), maximized a density function (Henderson, et al, 1959) and obtained what have come to be called the mixed model equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + D^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} . \tag{11}$$

That $\hat{\beta}$ and $\tilde{u}$ derived from these equations are identical to those of (9) and (10) depends on the matrix identity

$$(ZDZ' + R)^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}, \tag{12}$$

as shown in Henderson et al (1959), a matrix procedure that is reviewed and extended in Henderson and Searle (1981).

Deriving $\tilde{u}$ by means of (11) was for some time called maximum likelihood; this is nearly so, were it not that (11) comes from maximizing a density function rather than a likelihood function. The density function is

$$\frac{\exp - \tfrac{1}{2}(y-X\beta-Zu)'R^{-1}(y-X\beta-Zu)}{(2\pi)^{\frac{1}{2}n}|R|^{\frac{1}{2}}} \cdot \frac{\exp - \tfrac{1}{2}u'D^{-1}u}{(2\pi)^{\frac{1}{2}q}|D|^{\frac{1}{2}}} \tag{13}$$

where the orders of y and u are n and q respectively; the likelihood function is

$$\frac{\exp - \tfrac{1}{2}(y-X\beta)'V^{-1}(y-X\beta)}{(2\pi)^{\frac{1}{2}n}|V|^{\frac{1}{2}}} . \tag{14}$$

4.2 Estimated conditional mean

On assuming normality for u and e in (7), and then considering

$$\begin{bmatrix} y \\ u \end{bmatrix} \sim N \begin{bmatrix} \begin{pmatrix} X\beta \\ 0 \end{pmatrix}, & \begin{pmatrix} V & ZD \\ DZ' & D \end{pmatrix} \end{bmatrix} , \tag{15}$$

standard results from multivariate normal distribution theory show that the conditional mean of u given y is $E(u|y) = DZ'V^{-1}(y-X\beta)$ so that $\tilde{u}$ of (10) is

$$\tilde{u} = \widehat{E(u|y)};\qquad\qquad (16)$$

i.e., $\tilde{u}$ is the maximum likelihood estimator of the conditional mean $E(u|y)$.

This is a useful way of thinking about $\tilde{u}$. For the given set of records, y, we have $\tilde{u}$ as an optimum estimator of the mean of u, given those records.

4.3 Best linear unbiased predictor

$\lambda_i'y$ is a linear combination of the observations in y. On seeking to derive a λ_i such that $\lambda_i'y$ is unbiased for u_i in the sense of having $E(\lambda_i'y) = E(u_i)$, and such that $E(\lambda_i'y - u_i)^2$ is minimized, we find that λ_i turns out to be such that $\hat{u}_i = \lambda_i'y$ are the elements of $\tilde{u}$ of (12). This derivation is given in Henderson (1973). It is to be noted that the property of being "best" means "minimum mean square" not over all possible predictors but over only those that are linear combinations of the observations, and which are unbiased in the sense already used, namely have expected value equal to $E(u_i)$.

4.4 Bayes estimation

Lindley and Smith (1972), using normality for y and assuming normal priors for the parameters (in our case β and u), derive a Bayes estimator of u from the mean of the posterior distribution of the parameters. Dempfle (1977) shows details corresponding to (6) and (7). As examples, Lindley and Smith's elementary case, (1), is (10) with every $n_i = 1$, and $\mu^* + \alpha^*$ of their (20) is (10) with every $n_i = b$.

4.5 A weighted mean, in the Bayes context

Both Lindley and Smith (1972) and Dempfle (1977) suggest that the Bayes estimator is a weighted mean of two estimators of $[\beta'\ u']$. One is the generalized least squares estimator obtained from y, for a given β and u; and the other is the expected value of $[\beta'\ u']$ in their prior distribution. Combining these two means weighted by the inverses of their dispersion matrices leads to the Bayes estimator of u, which is $\tilde{u}$ of (10). Again, specific details are to be found in Dempfle (1977).

4.6 Regression on corrected records

The suggestion made by Bulmer (1980) is that of regression on corrected records. First, form a vector of the records y corrected for the fixed effects (in the animal breeding context, corrected for the environmental effects):

$$w = y - X\hat{\beta}\qquad\qquad (17)$$

for $\hat{\beta}$ being the generalized least squares estimator of β given in (11). Then, under normality assumptions, u is predicted by the intuitively appealing regression estimator

$$u^* = cov(u,w') \left[var(w)\right]^{-1}w. \qquad (18)$$

That $u^* = \tilde{u}$ is shown by Gianola and Goffinet (1982), wherein, in the discussion, the equivalence is gladly acknowledged by Bulmer. The paper also points out, though, that var(w) in u^* is singular and so the regular inverse there must be replaced by a generalized inverse for which, conveniently, V^{-1} happens to be suitable.

4.7 Unbiased prediction

Consider $a + By$, a vector of non-homogenous functions of elements of y. It is unbiased for u in the sense already defined, of having expected value equal to $0 = E(u)$, if and only if $a = 0$ and $BX = 0$. Moreover, even if $E(u) \neq 0$ but is invariant to β, the condition $BX = 0$ must still be satisfied. And, as shown in Searle and Casella (1984), $BX = 0$ if and only if $B = L\left[I - X(X'TX)^{-}X'T\right]$ for any L and for any positive definite matrix T. Using $L = I$ and $T = V^{-1}$ we then see that it is unbiasedness alone which brings the correction for fixed effects that motivated Bulmer's two-stage approach; and it also makes $\tilde{u} = By$ invariant to those effects. Indeed, this is true for any T used in B; and it is the minimum variance property that leads to using $T = V^{-1}$.

5. Summary

Section 2 lists, for an example, some well-known properties of the BLUP procedures. They also apply more generally to (elements of) $\tilde{u} = DZ'V^{-1}(y-X\hat{\beta})$ of (12). In addition, the various derivations just considered give the following properties.

1      $\hat{\beta}$ and $\tilde{u}$ maximize a distribution function.

2      $\tilde{u}$ is the estimate of the conditional mean $E(u|y)$.

3      $\tilde{u}$ is, of course, BLUP.

4      $\tilde{u}$ is a Bayes estimator.

5      $\tilde{u}$ is a weighted mean in the Bayes context.

6      $\tilde{u}$ is the regression predictior of u based on records corrected for fixed effects.

7      $\tilde{u}$ is the minimum variance version of the only form of predictor that can be invariant to the fixed effects β.

6. Appendix

Notation in (4) is simplified by using s for σ_s^2 and e for σ_e^2. Let I_{n_i} be an identity matrix of order n_i, and let $\underset{\sim}{1}_{n_i}$ and J_{n_i} be a vector, and a square matrix respectively, of order n_i with every element unity; and let $\Delta\{A_i\}$ be a block diagonal matrix with matrices A_i on its diagonal, for $i = 1,\ldots,a$. We show that $\hat{\beta}$ of (9) is $\hat{\mu}$ of (2) and that $\tilde{u}$ of (10) is the vector of terms $\tilde{s}_i = d_i - \hat{\mu}$ in (5).

For the model (1), the terms in the general model (6)-(8) are

$$X = \underset{\sim}{1}_{n.}, \quad \beta = \mu, \quad Z = \Delta\{\underset{\sim}{1}_{n_i}\}, \quad \text{and} \quad u = [s_i \ s_2 \ \ldots \ s_a]',$$

$$D = sI_a \quad \text{and} \quad V = \Delta\{eI_{n_i} + sJ_{n_i}\}$$

and

$$V^{-1} = \Delta\{I_{n_i} - w_i J_{n_i}\}/e$$

for

$$w_i = s/(e + n_i s) \quad \text{and} \quad 1 - n_i w_i = e/(e + n_i s).$$

Making these substitutions in (9) gives

$$\hat{\beta} = \hat{\mu} = \frac{(\underset{\sim}{1}'_{n.} - [n_1 w_1 \underset{\sim}{1}'_{n_1} \quad n_2 w_2 \underset{\sim}{1}'_{n_2} \quad \ldots \quad n_a w_a \underset{\sim}{1}'_{n_a}])y/e}{\sum\limits_{i=1}^{a} (n_i - n_i^2 w_i)/e}$$

$$= \frac{\sum\limits_{i=1}^{a} (1 - n_i w_i)y_{i.}}{\sum\limits_{i=1}^{a} n_i(1 - n_i w_i)} = \sum\limits_{i=1}^{a} \frac{n_i \bar{y}_{i.}}{\sigma_e^2 + n_i \sigma_s^2} \Big/ \sum\limits_{i=1}^{a} \frac{n_i}{\sigma_e^2 + n_i \sigma_s^2}$$

as in (2); and similarly, (10) gives

$$u = s\Delta\{\underset{\sim}{1}'_{n_i}(1 - n_i w_i)\} (y - \underset{\sim}{1}_{n.}\hat{\mu}) / e,$$

i.e., for $i = 1,\ldots,a$

$$\{\tilde{u}_i\} = \left\{ \frac{n_i \sigma_s^2}{\sigma_e^2 + n_i \sigma_s^2} (\bar{y}_{i.} - \hat{\mu}) \right\} = \tilde{d}_i - \hat{\mu} \text{ of (4).}$$

References

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

Dempfle, L. (1977). Relation entre BLUP (best linear unbiased prediction) et estimateurs Bayesiens. Annals de Genetique et de Selection Animale 9, 27-32.

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

Henderson, C.R. (1963). Selection index and expected genetic advance. Statistical Genetics in Plant Breeding 141-163, National Academy of Sciences, National Research Council Publication No. 982, Washington D.C.

Henderson, C.R. (1973). Sire evaluation and genetic trends. 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.

Henderson, H.V., and Searle, S.R. (1981). On deriving the inverse of a sum of matrices. SIAM Review 23, 53-60.

Lindley, D.V., and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). J. Royal. Stat. Soc. (B) 34, 1-42.

Searle, S.R. (1974). Prediction, mixed models and variance components. In Reliability and Biometry, eds. F. Proschnan and R.J. Serfling, Philadelphia, Pennsylvania: Society of Industrial and Applied Mathematics, 229-266.

Searle, S.R., and Casella, G. (1984). Alternative derivations of best linear unbiased prediction (BLUP) in the mixed model. Technical Report BU-861-M, Biometrics Unit, Cornell University.