PREDICTION OF BREEDING VALUES WHEN VARIANCES ARE NOT KNOWN

D.Gianola, J.L.Foulley and R.L.Fernando

Department of Animal Sciences
University of Illinois at Urbana Champaign
U.S.A

SUMMARY

The joint distribution of breeding values and of records usually depends on unknown parameters such as means, variances and covariances in the case of the multivariate normal distribution. If the objective of the analysis is to make selection decisions, these parameters should be considered as "nuisances". If the values of the parameters are unknown, the state of uncertainty can be represented by a prior probability distribution. This is then combined with the information contributed by the data to form a posterior distribution from which the needed predictors are calculated after integrating out the "nuisances". Prediction under alternative states of knowledge is discussed in this paper and the corresponding solutions presented. It is shown that when the dispersion structure is unknown, modal estimators of variance parameters should be considered. Because a Bayesian framework is adopted, the estimates so obtained are necessarily non-negative. If prior knowledge about means and variances is completely vague and the distribution is multivariate normal, the "optimal" predictors in the sense of maximizing the expected merit of the selected candidates are those obtained by using the "mixed model equations" with the unknown variances replaced by restricted maximum likelihood estimates. This leads to empirical Bayes predictors of breeding values.

INTRODUCTION

The problem of improvement by selection can be stated as follows: it is wished to elicit favorable genetic change in a "merit" function presumably related to economic return by retaining "superior" breeding animals and discarding "inferior" ones. Merit, e.g., breeding value or a future performance, is usually unobservable so culling decisions must be based on data available on the candidates themselves or on their relatives. The joint distribution of merits and of data usually depends on unknown parameters. In the multivariate normal distribution, these are means, variances and covariances. These must be estimated from the data at hand or, more generally, from a combination of data and pertinent prior information. What predictors of merit should be used when parameters are unknown?. For simplicity and for reasons of space we restrict attention to the multivariate normal distribution and to simple models. The general principles used apply to other distributions and models although the technical details differ. A Bayesian framework is used throughout. Zellner (1971) and Box and Tiao (1973) have reviewed foundations of Bayesian statistics. See Gianola and Fernando (1986) for some applications of Bayesian inference to animal breeding.

GENERAL FRAMEWORK

Model and assumptions

Suppose the data **y**,an n x 1 vector,are suitably described by the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad [1]$$

where $\boldsymbol{\beta}$ and **u** are p x 1 and q x 1 vectors,respectively,**X** and **Z** are known matrices and **e** is an independent residual. Assume,without loss of generality, that rank(**X**)= p. The vector $\boldsymbol{\beta}$ can include elements such as age of dam or herd-year effects which are regarded as "nuisance" parameters when the main objective is to predict breeding values. The vector **u** may consist of producing abilities or breeding values. Define"merit" as a linear function of **u** which in some sense depicts economic returns accruing from breeding. For example,the function **Mu**,for some matrix **M**,is the classical "aggregate genetic value" of selection index theory(Smith,1936;Hazel,1943).

The random process in [1] is a two-stage one.Prior to the realization of **y**, $\boldsymbol{\beta}$ and **u** follow a conceptual (prior) joint distribution. Assume temporarily that

$$\boldsymbol{\beta} \sim N(\boldsymbol{\alpha}, \Gamma\sigma^2_{\beta}) \quad , \quad \mathbf{u} \sim N(\boldsymbol{\emptyset}, \mathbf{A}\sigma^2_u) \qquad [2]$$

are independent. Above, **A** is the additive relationship matrix and σ^2_u is proportional to the additive genetic variance; observe that the distribution of **u** depends on this last parameter. When the variances in [2] are known,the joint density of $\boldsymbol{\beta}$ and **u** can be written as

$$f(\boldsymbol{\beta},\mathbf{u}\,|\sigma^2_{\beta},\sigma^2_u) = f(\boldsymbol{\beta}|\sigma^2_{\beta}) \cdot f(\mathbf{u}\,|\sigma^2_u)$$

$$\propto f(\boldsymbol{\beta}|\sigma^2_{\beta}) \cdot (\sigma^2_u)^{-q/2} \cdot \exp\left[-\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}/2\sigma^2_u\right] \qquad [3]$$

If $\sigma^2_{\beta} \rightarrow \infty$,the distribution of $\boldsymbol{\beta}$ becomes flat and all such vectors tend to be equally likely. This implies vague prior knowledge about $\boldsymbol{\beta}$ or,from a classical viewpoint,that this is a "fixed" vector. Thus,[3] is strictly proportional to the distribution of **u** in [2] above when prior knowledge about $\boldsymbol{\beta}$ is diffuse. If the variance of **u** is unknown,a prior distribution for this parameter would be needed but we assume in this paper that this distribution is also "flat",so as to represent complete ignorance about this variance.

The second stage relates to the realization of **y**. Given $\boldsymbol{\beta}$,**u** and σ^2_u from the first stage distribution,$\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}$ in [1] is fixed prior to the realization of the data. Thus,**e** is a discrepancy due to second stage sampling. The model for this stage,assuming normality,is

$$\mathbf{y}|\boldsymbol{\beta},\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}, \mathbf{R}\sigma^2_e) \qquad [4]$$

where **R** is a known matrix and σ^2_e is the variance of the residuals **e**. This

distribution or likelihood is

$$f(\mathbf{y}|\boldsymbol{\beta},\mathbf{u},\sigma_e^2) \propto (\sigma_e^2)^{-n/2} \exp[-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})/2\sigma_e^2] \qquad [5]$$

which is independent of the variance of u. If σ_e^2 is unknown, uncertainty can be introduced via another prior distribution, and we take here a flat prior to represent complete ignorance about this parameter.

Remembering that flat prior distributions have been taken for all parameters except **u**, the posterior distribution of all unknowns is given by Bayes theorem (Box and Tiao, 1973)

$$f(\boldsymbol{\beta},\mathbf{u},\sigma_u^2,\sigma_e^2|\mathbf{y}) \propto [5] \times [3] \qquad [6]$$

with $-\infty < \beta_i < \infty$ (i=1,...p), $-\infty < u_j < \infty$ (j=1,...q), $\sigma_u^2 \geq 0$ and $\sigma_e^2 > 0$. This distribution contains all available information about the unknown parameters and provides a point of departure for constructing predictors of merit when the variances are unknown.

Choosing the predictor

Cochran(1951), Bulmer(1980), Goffinet(1983), Goffinet and Elsen(1984) and Fernando and Gianola(1986) considered predictors that maximize expected merit in a selected group of individuals. Suppose there are q candidates for selection and that $k<q$ are needed for breeding. If **u** were observable, one would choose its largest k elements. Because this is not the case, it is intuitively appealing to calculate expectations conditionally on **y**, and to retain the k individuals with the largest conditional means. Cochran(1951) showed that selection upon conditional means maximizes expected merit in a series of trials where a proportion α is selected, on average. For this to hold, the joint distribution of merit and of records has to be identical and independent from candidate to candidate. The other authors showed that these restrictive assumptions are not needed when selecting a fixed number k out of m available items. In this case, selection upon conditional means maximizes expected merit in the selected sample irrespective of the form of the joint distribution. Henderson(1973), Searle(1974) and Harville(1985) have shown that over repeated sampling of **y**, the conditional mean is an unbiased predictor of merit and that minimizes mean squared prediction error. Thus, conditional means are appealing in animal breeding applications. In the next section we consider prediction under several alternative states of knowledge.

PREDICTION UNDER ALTERNATIVE STATES OF KNOWLEDGE

Known fixed effects and variances

Suppose one wishes to predict **u** from **y** in [1]. with $\boldsymbol{\beta}$, σ_u^2 and σ_e^2 known. The conditional mean would be calculated from the distribution

$$f(\mathbf{u}|\boldsymbol{\beta},\sigma_u^2,\sigma_e^2,\mathbf{y}) \qquad [7]$$

to obtain as predictor under multivariate normality

$$\hat{\mathbf{u}} = E(\mathbf{u}|\boldsymbol{\beta},\text{variances},\mathbf{y}) = \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \qquad [8]$$

where $\mathbf{C}'=\text{Cov}(\mathbf{u},\mathbf{y}')$ and $\mathbf{V}=\text{Var}(\mathbf{y})$. The posterior distribution [7] is normal with parameters

$$\mathbf{u}|\boldsymbol{\beta},\text{variances},\mathbf{y} \sim \mathbf{N}(\hat{\mathbf{u}},\ \mathbf{A}\sigma_u^2-\mathbf{C}'\mathbf{V}^{-1}\mathbf{C}) \qquad [9]$$

Putting in [8] $\mathbf{B} = \mathbf{V}^{-1}\mathbf{C}$,it is seen at once that $\hat{\mathbf{u}}$ is a selection index predictor. Because this predictor is derived from [7],the fact that selection indexes depend on exact knowledge of means,variances and covariances is highlighted. It is unrealistic to assume in practice that the values of all these parameters are known. A possibility would be to replace them by estimates obtained in some manner. Unfortunately,selection index theory does not guide on how these estimates should be chosen. Clearly,if the means and the variances are estimated from the same body of data from which the predictions are made,the distribution is no longer [7]. It would be incorrect to put any $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\sigma_u^2=\hat{\sigma}_u^2$,$\sigma_e^2=\hat{\sigma}_e^2$, and use [7] under the pretense that these are the "true" parameters. Any inference based on [7] using estimated parameters would ignore the "error" of the estimates.

Unknown fixed effects and known variances

The posterior distribution is now

$$f(\mathbf{u},\boldsymbol{\beta}|\text{variances},\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta},\mathbf{u},\sigma_e^2)\ .\ f(\mathbf{u}|\sigma_u^2)\ .\ f(\boldsymbol{\beta}) \qquad [10]$$

remembering that the prior distribution of $\boldsymbol{\beta}$ is flat. Because this vector is a "nuisance",we integrate it out of [10]. In other words,uncertainty about $\boldsymbol{\beta}$ is taken into account by marginalizing the above posterior distribution. Thus

$$f(\mathbf{u}|\text{variances},\mathbf{y}) \propto \int [10]\ d\boldsymbol{\beta} \qquad [11]$$

where the integration is over the p-space of $\boldsymbol{\beta}$. From [11] and [8] it follows that the predictor is

$$\hat{\mathbf{u}}^0 = E_{\boldsymbol{\beta}}(\hat{\mathbf{u}}) = \mathbf{C}'\mathbf{V}^{-1}[\mathbf{y}-\mathbf{X}\,E_{\boldsymbol{\beta}}(\boldsymbol{\beta})] \qquad [12]$$

where the expectation is taken with respect to $f(\boldsymbol{\beta}|\text{variances},\mathbf{y})$.The predictor in [12] is thus a weighted average of selection index predictions using the marginal posterior distribution of $\boldsymbol{\beta}$ (given the variances) as the weight function. Equivalently,[12] takes into account the fact that $\boldsymbol{\beta}$ is not known but estimated from the data,with the uncertainty taken into account via the marginal posterior distribution of $\boldsymbol{\beta}$. In order to obtain this posterior distribution,observe in [1] that

$$\mathbf{y}|\boldsymbol{\beta} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta},\mathbf{V}) \qquad [13]$$

with $\mathbf{V}=\mathbf{Z}\mathbf{A}\mathbf{Z}'\sigma_u^2+\mathbf{R}\sigma_e^2$.Hence,and because the prior distribution of $\boldsymbol{\beta}$ is flat:

$$f(\boldsymbol{\beta}|\text{variances},\mathbf{y}) \propto \exp[-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2] \quad [14]$$

Letting $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, one can write

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) \quad [15]$$

where it should be noted that only the second part of the expression depends on $\boldsymbol{\beta}$. Using [15] in [14] and remembering that the only variable in this posterior distribution is $\boldsymbol{\beta}$, one can write:

$$f(\boldsymbol{\beta}|\text{variances},\mathbf{y}) \propto \exp[-(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})/2) \quad [16]$$

This is in the form of the multivariate normal distribution

$$\boldsymbol{\beta}|\text{variances},\mathbf{y} \sim \mathbf{N}[\hat{\boldsymbol{\beta}}, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}] \quad [17]$$

Thus, the posterior distribution of $\boldsymbol{\beta}$ when the variances are known and when prior knowledge about this vector is vague is centered at the best linear unbiased estimator of $\boldsymbol{\beta}$ (Searle, 1971). We can now evaluate [12] to obtain the predictor

$$\overset{0}{\mathbf{u}} = \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \quad [18]$$

which is the best linear unbiased predictor or BLUP of u (Henderson, 1973). Without giving the details, the posterior distribution of $\mathbf{u}$ is

$$\mathbf{u}|\text{variances},\mathbf{y} \propto \mathbf{N}[\overset{0}{\mathbf{u}}, (\mathbf{Z}'\mathbf{M}\mathbf{Z}+\mathbf{A}\alpha)^{-1}\sigma_e^2] \quad [19]$$

where $\mathbf{M}$ is the projection matrix $\mathbf{R}^{-1}-\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}$, and α is the ratio between the variance of the residuals and the variance of u. The distribution in [19] is a function of the unknown variances. Unfortunately, these parameters are not always known. In practice, one could replace the variances by estimates obtained in some manner using a combination of data with prior knowledge. However, the theory of best linear unbiased prediction does not answer how these estimates should be obtained. It is clear that if

[18] above is evaluated at, say, $\hat{\alpha}$, a function of the data, then the predictor is no longer linear nor necessarily best in the sense of Henderson (1973). However, [18] remains unbiased provided that certain conditions are met (Kackar and Harville, 1981). While BLUP depends on exact knowledge of the variances, it is an improvement over selection indexes, where uncertainty on $\boldsymbol{\beta}$ is ignored.

Unknown fixed effects and variances known to proportionality

Suppose now that there is certainty with respect to the value of α, but $\boldsymbol{\beta}$ and the variance of the residuals are unknown; this would include the case where heritability is known. The joint posterior density of the unknowns is

$$f(\mathbf{u},\boldsymbol{\beta},\sigma_e^2|\alpha,\mathbf{y}) \quad [20]$$

Mathematically, this has the same form of [10] because a flat prior is taken for the residual variance. *Statistically*, the residual variance is a random

variable in [20] but a constant in [10]. In order to take into account uncertainty about $\boldsymbol{\beta}$ and the residual variance,these variables are integrated out of [20]. The predictor is calculated by succesive integration of nuisance parameters as

$$\tilde{\mathbf{u}} = \underset{\sigma_e^2}{E_2}\left\{ E\left[E\left(\mathbf{u}|\boldsymbol{\beta},\alpha,\sigma_e^2,\mathbf{y}\right) |\alpha,\sigma_e^2,\mathbf{y} \right] |\alpha,\mathbf{y}\right\}$$

$$= \underset{\sigma_e^2}{\text{Average}}[\text{BLUP}] = \underset{\sigma_e^2}{\text{Average}}\{\underset{\boldsymbol{\beta}\ \mathbf{u}}{\text{Average}}[\text{Selection index}]\} \qquad [21]$$

The predictor $\tilde{\mathbf{u}}$ is a weighted average of BLUP predictions ,using the posterior density $f(\sigma_e^2|\alpha,\mathbf{y})$ as weight function. Equivalently,it is a weighted average of selection index evaluations using $f(\boldsymbol{\beta},$residual variance $|\mathbf{y})$ as weighting function. Because the BLUP predictor depends on α but not on the residual variance (Henderson,1973,1977;Thompson,1979),it follows that.$\tilde{\mathbf{u}}$=BLUP($\mathbf{u}$). Hence, BLUPis the predictor of choice when the fixed effects and the residual variance are unknown.

While the distributions $\mathbf{u}|\alpha,\sigma_e^2,\mathbf{y}$ in [19] and $\mathbf{u}|\alpha,\mathbf{y}$ have the same mean,they do not have the same variance. Intuitively,some information should be used to remove uncertainty about the residual variance so one would expect the predictions stemming from [19] to me more precise that those based on [20]. In fact,it can be shown (Zellner,1971;Box and Tiao,1973)) that the distribution of $\mathbf{u}$ given α and $\mathbf{y}$,i.e.,with the residual variance integrated out, is a multivariate-t distribution with mean equal to the BLUP predictor,and variance as in [19] with the residual variance evaluated at

$$\hat{\sigma}_e^2=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}_*^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})/(n-p) \qquad [22]$$

where $\mathbf{V}_*=\mathbf{V}/$residual variance,and $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$. The marginal and conditional distributions of elements of $\mathbf{u}$ also follow univariate or multivariate t distributions. Because in animal breeding applications n-p is large,one can assume that the distribution is normal as in [19], using [22] or expressions easier to compute in lieu of the residual variance.

Unknown fixed effects and variance components

The joint posterior distribution of all unknowns in [6] is explicitly

$$f(\mathbf{u},\boldsymbol{\beta},\sigma_u^2,\sigma_e^2|\mathbf{y}) \propto (\sigma_e^2)^{-n'/2}.(\sigma_u^2)^{-q'/2} \cdot$$

$$\cdot \exp\left\{ -\tfrac{1}{2}\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})/\sigma_e^2 + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}/\sigma_u^2 \right] \right\} \qquad [23]$$

with the same restrictions as in [6]. The predictor would be

$$E(\mathbf{u}|\mathbf{y}) = \underset{\mathbf{u}}{E}\left\{ \underset{v}{E}[\underset{\boldsymbol{\beta}}{E}(\mathbf{u}|\mathbf{y},\boldsymbol{\beta},\sigma^2_u,\sigma^2_e) \; |\mathbf{y},\sigma^2_u,\sigma^2_e] \; |\mathbf{y} \right\} \qquad [24]$$

where v denotes the variances. As in [21],the predictor is obtained upon succesive integration of "nuisance" parameters,these being the fixed effects and the variance components. Equivalently,by interchange of the order of integration,the predictor is a weighted average of BLUP predictions,and the weighting function is the marginal density of the variance components. The necessary integrations leading to [24] are technically complex so we consider several approximations. These involve taking the mode of different posterior distributions rather than the mean. The approximations presented below follow an increasing order of desirability related to the extent to which [23] is marginalized with respect to the nuisance parameters (O'Hagan,1976).

1) Joint maximization with respect to all unknowns

The procedure involves finding the mode of the joint posterior density [23] without formally integrating out any of the nuisance parameters. The **u** component of this mode is then used as an approximation to $E(\mathbf{u}|\mathbf{y})$ in [24]. The values of $\mathbf{u},\boldsymbol{\beta}$ and of the variances maximizing [23] are the maximum a posteriori (MAP) estimates of the corresponding unknowns (Beck and Arnold,1977). MAP can be regarded as an extension of estimation by maximum likelihood as the estimates obtained are the "most likely" values of the unknowns given data and prior knowledge. Because [23] is asymptotically normal (Zellner,1971),the **u**-component of the mode would tend towards $E(\mathbf{u}|\mathbf{y})$ as the amount of information on the unknowns increases. This is so because in the multivariate normal distribution the mode is equal to the mean and the individual elements of the vector of joint means give directly the corresponding marginal means.

The first derivatives of [23] with respect to the unknowns are needed to find the MAP estimates. We have

$$\frac{\delta}{\delta\boldsymbol{\beta}}[f(\mathbf{u},\boldsymbol{\beta},\sigma^2_u,\sigma^2_e|\mathbf{y})] = \frac{\delta}{\delta\boldsymbol{\beta}}[f(\mathbf{u}\,\boldsymbol{\beta}|\sigma^2_u,\sigma^2_e,\mathbf{y}).f(\sigma^2_u,\sigma^2_e|\mathbf{y})]$$

$$\propto \frac{\delta}{\delta\boldsymbol{\beta}}[f(\mathbf{u},\boldsymbol{\beta}|\text{variances},\mathbf{y}] \qquad [25A]$$

because the marginal posterior density of the variances does not depend on $\boldsymbol{\beta}$. Likewise,

$$\frac{\delta}{\delta\mathbf{u}}[f(\mathbf{u}\,\boldsymbol{\beta},\text{variances}|\mathbf{y})] \propto \frac{\delta}{\delta\mathbf{u}}[f(\mathbf{u},\boldsymbol{\beta}|\text{variances},\mathbf{y})] \qquad [25B]$$

Further

$$\frac{\delta}{\delta\sigma^2_e}[f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y})] = \frac{\delta}{\delta\sigma^2_e}[f(\mathbf{u},\boldsymbol{\beta}|\mathbf{y}).f(\text{variances}|\mathbf{u},\boldsymbol{\beta},\mathbf{y})]$$

$$\propto \frac{\delta}{\delta\sigma_e^2}[f(\text{variances}|\boldsymbol{\beta},\mathbf{u},\mathbf{y})] \qquad [25C]$$

and

$$\frac{\delta}{\delta\sigma_u^2}[f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y})] \propto \frac{\delta}{\delta\sigma_u^2}[f(\text{variances}|\boldsymbol{\beta},\mathbf{u},\mathbf{y})] \qquad [25D]$$

In order to find the MAP estimates,[25A]-[25D] are equated to **0**. Observe that [25A] and [25B] involve densities corresponding to the state of knowledge where **u** and $\boldsymbol{\beta}$ are unknown but the variances are known. From results of Henderson et al (1959),Ronningen (1971) and Dempfle(1977),the values of **u** and $\boldsymbol{\beta}$ satisfying simultaneously [25A]=**0** and [25B]=**0** can be found by solving the mixed model equations of Henderson (1973)

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z}+\mathbf{A}^{-1}\alpha^{[k]} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{[k]} \\ \mathbf{u}^{[k]} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix} \qquad [26]$$

with $\alpha^{[k]}$ being the ratio of variances evaluated at their "current"value. This is obtained by maximization of [23] as if **u** and $\boldsymbol{\beta}$ were known,as equations [25C] and [25D] indicate. Differentiating [23] with respect to the variances yields

$$\sigma_e^{2[k+1]} = \hat{\mathbf{e}}^{[k]\prime}\mathbf{R}^{-1}\hat{\mathbf{e}}^{[k]}/n \qquad [27]$$

and

$$\sigma_u^{2[k+1]} = \mathbf{u}^{[k]\prime}\mathbf{A}^{-1}\mathbf{u}^{[k]}/q \qquad [28]$$

where $\hat{\mathbf{e}}^{[k]}$ *is the current value of the residual vector in* [1]. Equations [26],[27] and [28] define a double-iterative scheme which can be described as follows:

i) Choose starting values for the variance components and use them to solve [26];

ii) using the values of **u** *and* $\boldsymbol{\beta}$ *so obtained,update the variance components using* [27] *and* [28];

iii) return to [26] *and repeat as needed until* $\boldsymbol{\beta}$ *and* **u** *stabilize.*

If the algorithm converges to a non-trivial solution,the values obtained give the MAP of the unknowns. Observe that [27] and [28] guarantee non-negativity of the estimated variance components. The algorithm does not involve elements of the inverse of the coefficient matrix in [26],which implies that the procedure can be applied to large problems,as this system of equations can be solved by iteration without great difficulty. The expressions in [27] and [28] parallel the "estimators" of variance components derived by Lindley and Smith (1972) for two-way cross-classified random models; these authors,however,used an informative prior distribution for the variance components,as opposed to the

flat priors employed here. Lindley and Smith(1972) asserted that if a flat prior is used for the variance of **u**,then [28] would converge to 0. It can be verified numerically and analytically that this is not always the case albeit in many applications this variance does go to 0. This can happen in sire evaluation models when progeny group sizes are small or more generally,when α is large. The problem seems to be related to the fact that "many" parameters are estimated simultaneously so there is little information in the data about each of them. Thompson (1980) gave conditions under which the procedure produces non-zero estimates of the variance of the u's in one-way models. Harville (1977) conjectured that the problem may stem from "dependencies". The procedure needs further study as it is computationally feasible in very large models. Extensions to the multivariate domain would make the joint estimation of (co)variance components and breeding values possible in large data sets.

2) *Marginal maximization with respect to* **u** *and the variances*

We now take into account uncertainty about $\boldsymbol{\beta}$ by integrating it out of [23]. This involves working with the joint posterior density $f^{\star}=f(\mathbf{u},\text{variances}|\mathbf{y})$.Maximization of $f^{\star}$ with respect to the unknowns gives the corresponding MAP estimates and the **u** component of this joint posterior mode would be a closer approximation to [24] than the one presented in the preceding section. Putting $\boldsymbol{\gamma}'=[\mathbf{u}',\sigma_u^2,\sigma_e^2]$,we need to satisfy

$$\frac{\delta}{\delta\boldsymbol{\gamma}} \ln f^{\star} = \mathbf{0} \qquad [29]$$

Write

$$\frac{\delta}{\delta\boldsymbol{\gamma}} \ln f^{\star} = (f^{\star})^{-1} \frac{\delta}{\delta\boldsymbol{\gamma}} f^{\star}$$

$$= (f^{\star})^{-1} \frac{\delta}{\delta\boldsymbol{\gamma}} \left[\int_{\boldsymbol{\beta}} f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y})\, d\boldsymbol{\beta} \right] \qquad [30]$$

Putting $f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y}) = f(\boldsymbol{\beta}|\mathbf{u},\text{variances},\mathbf{y}).f^{\star}$,equation [30] can be expressed as

$$\frac{\delta}{\delta\boldsymbol{\gamma}} \ln f^{\star} = \int_{\boldsymbol{\beta}} \left[\frac{\delta}{\delta\boldsymbol{\gamma}} \ln f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y}) \right] f(\boldsymbol{\beta}|\text{variances},\mathbf{y})\, d\boldsymbol{\beta}$$

$$= E\left[\frac{\delta}{\delta\boldsymbol{\gamma}} \ln f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y}) \right] \qquad [31]$$

where the expectation is taken with respect to

$$\boldsymbol{\beta}|\mathbf{u},\text{variances},\mathbf{y} \sim \mathbf{N}\left[(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u}),(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\sigma_e^2\right] \qquad [32]$$

From [23]

$$\frac{\delta}{\delta\mathbf{u}} \ln f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y}) = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})/\sigma_e^2 - \mathbf{A}^{-1}\mathbf{u}/\sigma_u^2 \qquad [33A]$$

$$\frac{\delta}{\delta\sigma_u^2} \ln f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y}) = -q/2\sigma_u^2 + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}/2\,\sigma_u^4 \qquad [33B]$$

and

$$\frac{\delta}{\delta\sigma_e^2} \ln f(\mathbf{u},\boldsymbol{\beta},\text{variances}\,|\mathbf{y}) = -n/2\sigma_e^2 + \mathbf{e}'\mathbf{R}^{-1}\mathbf{e}/2\sigma_e^4 \qquad [33C]$$

Taking the expectation of [33A] with respect to the distribution in [32] and setting to 0 gives

$$\left(\mathbf{Z}'\mathbf{M}\mathbf{Z}+\mathbf{A}^{-1}\alpha^{[k]}\right)\mathbf{u}^{[k]}=\mathbf{Z}'\mathbf{M}\mathbf{y} \qquad [34A]$$

These are the mixed model equations of [26] after "absorption" of $\boldsymbol{\beta}$ and evaluated at the "current" value of the variance ratio. The equation for the variance of the u's follows directly from [33B]

$$\sigma_u^{2[k+1]} = \mathbf{u}'^{[k]}\mathbf{A}^{-1}\mathbf{u}^{[k]}/q \qquad [34B]$$

The expectation of [33C] with respect to [32] involves

$$E[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})] = p\sigma_e^2+(\mathbf{y}-\mathbf{Z}\mathbf{u})'\mathbf{M}^*\mathbf{R}^{-1}\mathbf{M}^*(\mathbf{y}-\mathbf{Z}\mathbf{u})$$

where $\mathbf{M}^*=\mathbf{R}\mathbf{M}$. Using this result when setting the expectation of [33C] to 0 gives:

$$\sigma_e^{2[k+1]}=(\mathbf{y}-\mathbf{Z}\mathbf{u}^{[k]})'\mathbf{M}^*\mathbf{R}^{-1}\mathbf{M}^*(\mathbf{y}-\mathbf{Z}\mathbf{u})/(n-p) \qquad [34C]$$

It can be shown that the numerator of [34C] can be written as $\hat{\mathbf{e}}'^{[k]}\mathbf{R}^{-1}\hat{\mathbf{e}}^{[k]}$. Iteration as in the previous section but with equations [34A]-[34C] yields an algorithm to obtain the MAP estimates of $\mathbf{u}$ and of the variances after integrating $\boldsymbol{\beta}$ out of [23]. Again, expressions [34B] and [34C] guarantee non-negativity of the estimated variance components. The algorithm does not involve elements of the inverse of the coefficient matrix in [34A] so it can be applied, at least potentially, to large problems. Extensions to the multivariate situation are straightforward. Because the main computational difficulty is the "absorption" of $\boldsymbol{\beta}$ into $\mathbf{u}$ to obtain [34A], it may be more efficient to solve [26] directly by an iterative procedure. Equation [34B] has the same form of [28], arising in MAP estimation by "joint maximization", so the problems presented by the estimator of Lindley and Smith (1972) are probably also encountered in this method. On the other hand, the expression for the residual variance in [34C] has n-p in the denominator instead of n as in [27]. In this sense, the method takes into account "losses in degrees of freedom" resulting from "estimation" of $\boldsymbol{\beta}$ (Patterson and Thompson, 1971; Harville, 1977). In the Bayesian view, n-p appears because $\boldsymbol{\beta}$ is integrated out of [23]. Because joint and marginal maximization as described in this paper are based on posterior densities subject to the non-negativity constraints for the variances (see [6]), these procedures utilize all "information" contained in $\mathbf{y}$. This would also be true when working with the posterior densities $f(\boldsymbol{\beta},\text{variances}|\mathbf{y})$ and $f(\text{variances}|\mathbf{y})$. When flat priors are used, these two densities lead to maximum likelihood and restricted maximum likelihood estimators of variance components, respectively (Harville, 1974, 1977).

3) *Approximate integration of the variances*

The conditional expectation in [24] can also be written as

$$E(\mathbf{u}|\mathbf{y})=\int_{\mathbf{u}} \mathbf{u}\,[\int_0^\infty \int_0^\infty f(\mathbf{u}|\text{variances},\mathbf{y}).f(\text{variances}|\mathbf{y})\, d\sigma_e^2\, d\sigma_u^2]\, \mathbf{du} \quad [35]$$

and we note that the expression inside the brackets is $E[f(\mathbf{u}|\text{variances},\mathbf{y})]$,taken over the marginal posterior distribution of the variances. This latter distribution gives the plausibility of values taken by the residual variance and the variance of the u's,given the data.If this density is reasonably peaked,which occurs when there is a large amount of information about the unknown variances in the data,most of the density is at the mode(Zellner,1971;Box and Tiao,1973). If this condition is met,one can write

$$E[f(\mathbf{u}|\text{variances},\mathbf{y}) \simeq f(\mathbf{u}|\sigma_u^2=\sigma_u^{\star 2},\sigma_e^2=\sigma_e^{\star 2},\mathbf{y}) \quad [36]$$

where $\sigma_u^{\star 2}$ and $\sigma_e^{\star 2}$ are the two components of the mode of $f(\text{variances}|\mathbf{y})$. Using [36] in [35] gives

$$E(\mathbf{u}|\mathbf{y}) \simeq E(\mathbf{u}|\sigma_u^2=\sigma_u^{\star 2},\sigma_e^2=\sigma_e^{\star 2},\mathbf{y}) \quad [37]$$

This result indicates that the variances should be estimates by maximization of $f(\text{variances}|\mathbf{y})$,and the predictor obtained by calculating the mean of the conditional distribution [36],which is multivariate normal as stated in [19]. The problem is then solved using results obtained in the section for unknown fixed effects and known variance components,taking α at the modal values of the posterior density of the variance components. The predictor obtained belongs to the class of Empirical Bayes estimators (Vinod and Ullah,1981;Judge et al 1985) as the variance of the prior distribution of **u** is obtained from the data as opposed to being actually "prior".

Put $\boldsymbol{\lambda}'=[\sigma_u^2,\sigma_e^2]$ and $f^+=f(\text{variances}|\mathbf{y})$. In order to maximize f^+,we need to satisfy:

$$\frac{\delta}{\delta\boldsymbol{\lambda}} ln\, f^+ = 0 \quad [38]$$

Using a result similar to the one leading to [31]

$$\frac{\delta}{\delta\boldsymbol{\lambda}} ln\, f^+ = \int_{\boldsymbol{\beta}} \int_{\mathbf{u}} [\frac{\delta}{\delta\boldsymbol{\lambda}} ln\, f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y})]\, f(\mathbf{u},\boldsymbol{\beta}|\text{variances},\mathbf{y})\, \mathbf{du}\, d\boldsymbol{\beta}$$

$$=E[\frac{\delta}{\delta\boldsymbol{\lambda}} ln\, f(\mathbf{u},\boldsymbol{\beta},\text{variances}|\mathbf{y})] \quad [39]$$

where the expectation is taken with respect to $f(\mathbf{u},\boldsymbol{\beta}|\text{variances}|\mathbf{y})$. Evaluating these expectations and setting to zero to satisfy [38] gives

$$\sigma_u^{2[k+1]} = \{[\hat{\mathbf{u}}'\mathbf{A}^{-1}\hat{\mathbf{u}}][k]+tr(\mathbf{A}^{-1}\mathbf{C}^{[k]})\sigma_e^{2[k]}\}/q \quad [40]$$

and

$$\sigma_e^{2[k+1]} = \left\{ [\hat{\mathbf{e}}'\mathbf{R}^{-1}\hat{\mathbf{e}}]^{[k]} + (p+q-tr\mathbf{A}^{-1}\mathbf{C}^{[k]}\alpha^{[k]})\sigma_e^{2[k]} \right\}/n \qquad [41]$$

where [k] indicates iterate number and $\mathbf{C}^{[k]}$ is the q x q lower sub-matrix of the inverse of the mixed model equations evaluated at the current value of α. Equations [40] and [41] in conjunction with [26] define an iterative scheme. Once the variances stabilize,[26] is solved once more to obtain the necessary predictions. The main difficulty of this procedure is the computation of the matrix **C**; in practice,it may be necessary to approximate the traces needed in [40] and [41] and Harville(1977) has suggested some possibilities.

We note that [40] and [41] are expressions arising in the EM algorithm (Dempster et al.,1977) when a restricted likelihood is maximized(Patterson and Thompson,1971);similar equations were described by Henderson(1984). This is not surprising as Harville(1974,1977) showed that restricted maximum likelihood corresponds to Bayesian estimates obtained by maximization of f(variances|**y**) when flat priors are used for the variances and for the fixed effects. This was the approach followed in this section of the paper. It should be observed that [40] and [41] are "natural" expressions derived directly from the posterior distribution without invoking numerical "trickery". Thus,the estimates so obtained would be non-negative as they are based on a posterior distribution which would return with probability equal to zero any negative value. Wu(1983) discussed numerical aspects of the EM algorithm. Slow convergence has been reported by Thompson(1979),Meyer (1985),and Thompson and Meyer(1985). These authors advocate algorithms based on second differentials but they warn about the non-null probability of obtaining estimates outside of the parameter space. This is a disturbing property of such algorithms,especially when employed in multivariate cases.

CONCLUSIONS

The theory presented in this paper indicates that under normality and in the absence of prior information about the dispersion parameters,breeding values should be predicted using BLUP methodology,with the unknown variances replaced by their corresponding REML estimates obtained from the data from which predictions are to be made. The resulting predictors are not BLUP but yield the closest possible approximation to E(**u**|**y**),as uncertainty about the values of fixed effects is taken into account,and the variances are approximately integrated out.The results dismiss quadratic unbiased estimators and point to statistics obtained from maximization of posterior densities or of likelihood in the classical sense when flat priors are employed Several issues which are not dealt with here for reasons of space include prediction using data from selected individuals,specification of informative prior distributions for the unknown variance parameters,and non-normal settings such as when major genes segregate in the population or when the variables are categorical. It is felt,however,that the Bayesian paradigm gives a completely general framework to adress hereto unsolved statistical problems in animal breeding.

LITERATURE CITED

BOX, G.E.P and TIAO,G.C. 1973. Bayesian inference in statistical analysis. Addison-Wesley,Reading,Massachusetts.

BECK,J.V. and ARNOLD,K.J. 1977. Parameter estimation in engineering and science. J.Wiley and Sons,New York

BULMER,M.G. 1980. The mathematical theory of quantitative genetics. Clarendon Press,Oxford.

COCHRAN, W.G. 1951. Improvement by means of selection. Proc.Second Berkeley Symp. Math. Stat. and Prob.,449-470.University of California Press, Berkeley.

DEMPFLE, L. 1977. Relation entre BLUP (best linear unbiased prediction) et estimateurs Bayesiens. Ann.Genet.Sel.Anim. 9,27-32.

DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm Journal of the Royal Stat. Soc. B 39,1-38.

FERNANDO, R.L. and GIANOLA,D. 1986. Optimal properties of the conditional mean as a selection criterion. Submitted.

GIANOLA, D. and FERNANDO, R.L. 1986. Bayesian methods in animal breeding theory. J.Anim.Sci. (in press)

GOFFINET,B. 1983. Selection on selected records. Genet.Sel.Evol.15,91-98.

GOFFINET, B. and ELSEN, J.M. 1984. Critere optimal de selection : quelques resultats generaux. Genet. Sel. Evol. 16,307-318.

HARVILLE, D.A. 1974. Bayesian inference for variance components using only error contrasts. Biometrika 61,383-385.

HARVILLE,D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. J.Amer.Stat.Assoc.72,320-338.

HARVILLE,D.A. 1985. Decomposition of prediction error. J.Amer.Stat.Assoc. 80,132-138.

HAZEL, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 28,476-490.

HENDERSON, C.R. 1973. Sire evaluation and genetic trends pp.10-41 In Proc.Anim.Breed.and Genet.Symp. in Honor of Dr.J.L.Lush. American Society of Animal Science and American Dairy Science Association,Champaign,Illinois

HENDERSON,C.R. 1977. Prediction of future records.pp 615-638. In Proc. Int.Conf.Quant.Genet. E.Pollak,O.Kempthorne and T.B.Bailey Jr.,Editors. Iowa State University Press,Ames.

HENDERSON,C.R. 1984. ANOVA,MIVQUE,REML,and ML algorithms for estimation of variances and covariances.pp 257-280. In "Statistics: an appraisal" H.A.David and H.T.David,Editors. Iowa State University Press,Ames.

HENDERSON,C.R., KEMPTHORNE,O., SEARLE,S.R. and von KROSIGK,C.N. 1959. Estimation of genetic and environmental trends from records subject to culling. Biometrics 13,192-218.

JUDGE, G.C., GRIFFITHS, W.E., HILL,R.C., LUTKEPOL,H. and LEE,T.C. 1985. The theory and practice of econometrics. 2nd Ed. J.Wiley and Sons, New York.

KACKAR, R.N. and HARVILLE, D.A. 1981. Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Comm.Stat ,Theory and Methods A10,1249-1261.

LINDLEY, D.V. and SMITH, A.F.M. 1972. Bayes estimates for the linear model. J.Royal Stat. Soc. B 34,1-18

MEYER, K. 1985. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41,153-165.

O'HAGAN, A. 1976. On posterior joint and marginal modes. Biometrika 63, 329-333.

PATTERSON ,H.D. and THOMPSON,R. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58,545-554.

RONNINGEN, K. 1971. Some properties of the selection index derived by "Henderson's Mixed Model Method". Z.Tierz.Zuchtungsbiol. 88,186-193.

VINOD,H.D and ULLAH.A. 1981. Recent advances in regression methods. Marcel Dekker,New York.

SEARLE, S.R. 1971. Linear Models. J.Wiley and Sons,New York.

SEARLE,S.R. 1974. Prediction,mixed models and variance components. In Proc. of a Conference on Reliability and Biometry. F.Proschan and R.J.Serfling,Editors.S.I.A.M.,Philadelphia,Pennsylvania.

SMITH, H.F. 1936. A discriminant function for plant selection. Ann.Eugen. 7, 240-250

THOMPSON, R. 1979. Sire evaluation. Biometrics 35,339-353

THOMPSON, R. 1980. Maximum likelihood estimation of variance components. Math.Operationsforsch. Statist. 11,545-561.

THOMPSON, R. and MEYER, K. 1985. Theoretical aspects in the estimation of breeding values for multi-trait selection. Mimeo.,20pp. 36th Annual Meeting of EAAP,Kallithea,Greece.

WU, C.F.J. 1983. On the convergence properties of the EM algorithm. Ann. Stat.11,95-103.

ZELLNER, A. 1971 An introduction to Bayesian inference in econometrics. J.Wiley and Sons,New York.