Theor Appl Genet (1986) 72:671–677

On selection criteria and estimation of parameters when the variance is heterogeneous

D. Gianola

Department of Animal Sciences, 126 Animal Sciences Laboratory, 1207 W. Gregory Drive,
University of Illinois, Urbana, IL 61801, USA

Received November 5, 1985; Accepted March 15, 1986
Communicated by D. van Vleck

Summary. Procedures for ranking candidates for selection and for estimating genetic and environmental parameters when variances are heterogeneous are discussed. The best linear unbiased predictor (BLUP) accounts automatically for heterogeneous variance provided that the covariance structure is known and that the assumptions of the model hold. Under multivariate normality BLUP allowing for heterogeneous variance maximizes expected genetic progress. Examples of application of BLUP to selection when residual or genetic variances are heterogeneous are given. Restricted maximum likelihood estimation of heterogeneous variances and covariances via the expectation-maximization algorithm is presented.

Key words: Heterogeneous variance – BLUP – Selection

Introduction

Hill (1984) considered methods of accounting for heterogeneous variance when constructing selection criteria in simple settings and suggested that observations should be scaled using estimated standard deviations. In general, homogeneity of variance or covariance is not a requirement in a best linear unbiased prediction (BLUP) analysis. This is illustrated by multiple trait BLUP (e.g. Henderson and Quaas 1976) where the covariance structure is clearly heterogeneous across traits. Also, Henderson (1984) has dealt to some extent with sire and cow evaluation when variances are unequal.

The objective of this expository paper is to show that if the records are modeled correctly, i.e., that the assumptions hold, BLUP automatically scales the observations. Further, if multivariate normality is tenable, BLUP with the correct model maximizes genetic progress in certain settings. Estimation of variances when heteroscedasticity is present is discussed.

Theory

Suppose we are interested in predicting a random vector **u**. This vector can represent breeding values, producing abilities or transmitting abilities for one or more traits (Henderson 1973) for a set of animals with or without records (Henderson 1977).

Let the relationship between **u** and the records (**y**) be linear and with the form

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

where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$ contains the mean values for each of the records, β is usually but not necessarily a vector of "nuisance" parameters (e.g., herd-year-season effects), **X** and **Z** are known incidence matrices, and **e** is a vector of random residuals, e.g., environmental, non-systematic, effects. Take $\mathrm{E}(\mathbf{u}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, and assume that the variance-covariance matrix of **u** and **e** is completely general

$$\mathrm{Var}\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{T}\\ \mathbf{T}' & \mathbf{R}\end{bmatrix}. \tag{2}$$

Note that with this formulation, each element of **u** or **e** can have different variance. Furthermore, **T**, the covariance matrix between breeding values, say, and environmental effects can be non-null; Schaeffer and Henderson (1983) have dealt with this problem. From (1) and (2)

$$\begin{aligned}\mathrm{Var}(\mathbf{y}) &= \mathrm{Var}(\mathbf{Z}\,\mathbf{u} + \mathbf{e})\\ &= \mathbf{Z}\,\mathbf{G}\,\mathbf{Z}' + \mathbf{Z}\,\mathbf{T} + \mathbf{T}'\mathbf{Z}' + \mathbf{R} = \mathbf{V}.\end{aligned} \tag{3}$$

Because **V** is a non-singular variance-covariance matrix, it can be decomposed as $\mathbf{V} \times \mathbf{P}\,\mathbf{P}'$, with **P** non-singular (Graybill 1976).

Consider now the linear transformation of (1)

$$\mathbf{P}^{-1}\mathbf{y} = \mathbf{P}^{-1}\mathbf{X}\beta + \mathbf{P}^{-1}\mathbf{Z}\,\mathbf{u} + \mathbf{P}^{-1}\mathbf{e} \tag{4}$$

or, equivalently,

$$y^* = X^* \beta + Z^* u + e^* \tag{5}$$

where $y^* = P^{-1} y$, $X^* = P^{-1} X$, $Z^* = P^{-1} Z$ and $e^* = P^{-1} e$. Clearly

$$\begin{aligned} \mathrm{Var}(y^*) &= \mathrm{Var}(Z^* u + e^*) \\ &= P^{-1} \mathrm{Var}(Z\,u + e)\, P'^{-1} \\ &= P^{-1} V P'^{-1} = I\,. \end{aligned} \tag{6}$$

Hence, each transformed record has the same variance, which is equal to 1.

Bulmer (1980) has shown that the expected "merit" of a fixed number of selected candidates is maximized by selecting upon the conditional mean of merit given the data, regardless of the distribution. Hence, if **u** in (1) is the vector of merits, then we would select with E (**u** | **y**). Under linearity or multivariate normality, and with known variances and covariances:

$$\begin{aligned} E(u \mid y) &= \mathrm{Cov}(u, y')\, V^{-1} (y - X\beta) \\ &= (G Z' + T)\, V^{-1} (y - X\beta) \\ &= [G Z' (P')^{-1} + T (P')^{-1}]\, (P^{-1} y - P^{-1} X \beta) \\ &= \mathrm{Cov}(u, y^{*\prime})\, (y^* - X^* \beta) \\ &= E(u \mid y^*)\,. \end{aligned} \tag{7}$$

The above shows that the optimal ranking rules, **E** (**u** | **y**), and **E** (**u** | **y***) are identical. In other words, the standardization or scaling of the records is done automatically. In BLUP, β of (7) is replaced by a solution to

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

or to

$$(X^{*\prime} X^*)\, \tilde{\beta} = X^{*\prime} y^*$$

which lead to the best linear unbiased estimator (BLUE) of $X\beta$ and to the best linear unbiased predictor of **u** (Henderson 1973). The BLUP predictor

$$u = \mathrm{Cov}(u, y')\, V^{-1} (y - X\tilde{\beta}) \tag{8}$$

maximizes expected genetic progress in the class of translation invariant linear functions of the records (Goffinet 1983; Goffinet and Elsen 1984; Fernando 1984). Furthermore, it is unbiased by selection or mating decisions that do not depend on β provided that all records used for such decisions have been employed in the analysis (Henderson 1975a, 1982; Gianola and Fernando 1985; Henderson 1985). The expression in (8) is usually tedious to compute and the mixed model equations of Henderson (1973) provide a more efficient algorithm usually amenable to an iterative implementation.

When heterogeneity of (co)variance is only with respect to the distribution of the residuals **e**, one can write $R = R^{.5} R^{.5}$, so the transformation becomes

$$\begin{aligned} y^{**} = R^{-.5} y &= R^{-.5} X \beta + R^{-.5} Z\, u + R^{-.5} e \\ &= X^{**} \beta + Z^{**} u + e^{**}\,. \end{aligned} \tag{9}$$

In the univariate case, the elements of **y**** are usually observations divided by their corresponding residual standard deviations. Clearly, Var (**e****) = **I**. Further

$$\begin{aligned} E(u \mid y^{**}) &= (G Z' R^{-.5} + T R^{-.5})\, R^{.5} V^{-1} R^{.5} (R^{-.5} y - R^{-.5} X \beta) \\ &= (G Z' + T)\, V^{-1} (y - X\beta) \\ &= E(u \mid y) \end{aligned} \tag{10}$$

Table 1. Hypothetical data on milk production records of daughters of three artificial insemination sires producing in two herds [a]

Herds	Sires		
	1	2	3
1	y_{111}, y_{112}	y_{121}	–
2	y_{211}	–	y_{231}, y_{232}

[a] y_{ijk} is a record on the *kth* daughter of the *jth* sire producing in the *ith* herd

as one would anticipate from the more general result given in (7). The BLUE of β and the BLUP of **u** can be obtained directly (when **T** = **0**) by solving the mixed model equations:

$$\begin{bmatrix} X^{**\prime} X^{**} & X^{**\prime} Z^{**} \\ Z^{**\prime} X^{**} & Z^{**\prime} Z^{**} + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^{**\prime}\ y^{**} \\ Z^{**\prime}\ y^{**} \end{bmatrix}.$$

Applications

1 Heterogeneous residual variance

In order to illustrate the computation of BLUP when the variance is heterogeneous, consider the hypothetical data in Table 1. These are milk production records of daughters of 3 sires producing in 2 herds. Sire 3 does not have progeny in herd 1, and sire 2 does not have records in herd 2. Let the linear model describing the records be

$$y_{ijk} = H_i + s_j + e_{ijk}, \quad i = 1, 2;\ j = 1, 2, 3;\ k = 0, 1, 2 \tag{11}$$

with $E(y_{ijk}) = H_i$, $s_j \sim \mathrm{ID}(0, \sigma_s^2)$, $e_{ijk} \sim \mathrm{ID}(0, \sigma_{e_i}^2)$ and $\mathrm{Cov}(s_j, e_{ijk}) = 0$ for all such pairs; ID stands for independently distributed. Note that the residual variance differs from herd to herd. Using the notation of (1) and (2), we have

$$\begin{aligned} \beta' &= [H_1, H_2] \\ u' &= [s_1, s_2, s_3] \\ G &= I_{3\times 3}\, \sigma_s^2 \end{aligned}$$

where $I_{3\times 3}$ is an identity matrix. Further

$$R = \mathrm{Var}\begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{211} \\ e_{231} \\ e_{232} \end{bmatrix} = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{e_1}^2 & 0 & 0 & 0 & 0 \\ & & \sigma_{e_1}^2 & 0 & 0 & 0 \\ & & & \sigma_{e_2}^2 & 0 & 0 \\ & & & & \sigma_{e_2}^2 & 0 \\ \text{SYMMETRIC} & & & & & \sigma_{e_2}^2 \end{bmatrix}$$

and **T** = **0**. Using the data in Table 1, and putting $\gamma_i = \sigma_{e_i}^{-2}$ (i = 1, 2) and $k = \sigma_s^{-2}$, direct application of the mixed model equations yields

$$\begin{pmatrix} 3\gamma_1 & 0 & 2\gamma_1 & \gamma_1 & 0 \\ & 3\gamma_2 & \gamma_2 & 0 & 2\gamma_2 \\ & & 2\gamma_1+\gamma_2+k & 0 & 0 \\ & & & \gamma_1+k & 0 \\ \textbf{SYMMETRIC} & & & & 2\gamma_2+k \end{pmatrix} \begin{pmatrix} \hat{H}_1 \\ \hat{H}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{pmatrix} = \begin{pmatrix} \sum_j \sum_k y_{1jk}\gamma_1 \\ \sum_j \sum_k y_{2jk}\gamma_2 \\ \sum_i \sum_k y_{i1k}\gamma_i \\ \sum_i \sum_k y_{i2k}\gamma_i \\ \sum_i \sum_k y_{i3k}\gamma_i \end{pmatrix} \qquad (12)$$

Solving the above system gives the BLUE of herd effects ($\hat{H}_i$; $i = 1, 2$) and the BLUP of the transmitting abilities of the sires ($\hat{s}_j$; $j = 1, \ldots, 3$). Following Henderson (1973, 1975a), the inverse of the coefficient matrix in (12) gives

$$\mathrm{Var}\begin{pmatrix} \hat{H}_1 \\ \hat{H}_2 \\ \hat{s}_1 - s_1 \\ \hat{s}_2 - s_2 \\ \hat{s}_3 - s_3 \end{pmatrix}.$$

After some algebra

$$\hat{s}_j = \frac{\sum_i \sum_k \left[\left(\frac{h_i^2}{4-h_i^2}\right)(y_{ijk} - \hat{H}_i)\right]}{1 + \sum_i n_{ij}\left(\frac{h_i^2}{4-h_i^2}\right)} \qquad (13)$$

where $h_i^2 = 4/(1 + k/\gamma_i)$. Observe that when $h_i^2 = h^2$ for all i, (13) becomes

$$\hat{s}_j = \frac{h^2\left(n_{.j}\bar{y}_{.j} - \sum_i n_{ij}\hat{H}_i\right)}{4 + (n_{.j} - 1)h^2} \qquad (14)$$

where $n_{.j} = \sum_i n_{ij}$, and $\bar{y}_{.j}$ is the average of the daughters of sire i. Finally, when all herd effects are equal ($\hat{H}_i = \hat{\mu}$, say, for all i), equation (14) yields

$$\hat{s}_j = \frac{n_{.j}h^2}{4 + (n_{.j} - 1)h^2}(\bar{y}_{.j} - \hat{\mu}) \qquad (15)$$

which is the BLUP evaluation of unrelated sires via progeny testing in a one-way model. Therefore, the solution to (12) reduces to (13) when residual variance is heterogeneous and herd effects differ, to (14) when residual variance is homogeneous and herd effects differ, and to (15) when there is homogeneity with respect to herd effects and residual variance.

2 Heterogeneous residual and sire variances

Consider again the data in Table 1 and suppose that the variances of residuals and of transmitting abilities are heterogeneous with respect to herds, i.e., $\sigma^2_{s_1} \neq \sigma^2_{s_2}$ and $\sigma^2_{e_1} \neq \sigma^2_{e_2}$. Because the distribution of breeding values now differs from herd to herd, the model in (11) needs to be amended as follows

$$y_{ijk} = H_i + s_{ij} + e_{ijk} \qquad (16)$$

with $E(y_{ijk}) = H_i$, $s_{ij} \sim (0, \sigma^2_{s_i})$ is the transmitting ability of sire j when a daughter produces in herd i, $e_{ijk} \sim (0, \sigma^2_{e_i})$ and $\mathrm{Cov}(s_{ij}, e_{ijk}) = 0$ for all such pairs. The variance-covariance matrix of transmitting abilities is now

$$\mathrm{Var}\begin{pmatrix} s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \\ s_{23} \end{pmatrix} = \begin{pmatrix} \sigma^2_{s_1} & 0 & 0 & \varrho_{12}\sigma_{s_1}\sigma_{s_2} & 0 & 0 \\ 0 & \sigma^2_{s_1} & 0 & 0 & \varrho_{12}\sigma_{s_1}\sigma_{s_2} & 0 \\ 0 & 0 & \sigma^2_{s_1} & 0 & 0 & \varrho_{12}\sigma_{s_1}\sigma_{s_2} \\ & & & \sigma^2_{s_2} & 0 & 0 \\ & & & & \sigma^2_{s_2} & 0 \\ \text{SYMMETRIC} & & & & & \sigma^2_{s_2} \end{pmatrix} \qquad (17)$$

where ϱ_{12} is the correlation between the breeding values of the same sire in different herds. We have then

$$\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{I}_{3\times 3} \qquad (18)$$

where $\otimes$ is the Kronecker product operator,

$$\mathbf{G}_0 = \begin{bmatrix} \sigma^2_{s_1} & \sigma_{s_{12}} \\ \sigma_{s_{12}} & \sigma^2_{s_2} \end{bmatrix} \qquad (19)$$

and $\sigma_{s_{12}} = \varrho_{12}\sigma_{s_1}\sigma_{s_2}$ is the covariance between breeding values of the same sire in different herds. Clearly, ϱ_{12} is a measure of genotype × environment interaction, and this parameter can be formally included in the model and estimated as described in the last section of this paper. Now from (18), $\mathbf{G}^{-1} = \mathbf{G}_0^{-1} \otimes \mathbf{I}$, so

$$\mathbf{G}^{-1} = \begin{pmatrix} g^{11} & 0 & 0 & g^{12} & 0 & 0 \\ 0 & g^{11} & 0 & 0 & g^{12} & 0 \\ 0 & 0 & g^{11} & 0 & 0 & g^{12} \\ g^{12} & 0 & 0 & g^{22} & 0 & 0 \\ 0 & g^{12} & 0 & 0 & g^{22} & 0 \\ 0 & 0 & g^{12} & 0 & 0 & g^{22} \end{pmatrix}$$

where g^{11}, g^{12}, g^{22} are elements of $\mathbf{G}_0^{-1}$. The mixed model equations are then

$$\begin{bmatrix} 3\gamma_1 & 0 & 2\gamma_1 & \gamma_1 & 0 & 0 & 0 & 0 \\ & 3\gamma_2 & 0 & 0 & 0 & \gamma_2 & 0 & 2\gamma_2 \\ & & 2\gamma_1+g^{11} & 0 & 0 & g^{12} & 0 & 0 \\ & & & \gamma_1+g^{11} & 0 & 0 & g^{12} & 0 \\ & & & & g^{11} & 0 & 0 & g^{12} \\ & & & & & \gamma_2+g^{22} & 0 & 0 \\ & & & & & & g^{22} & 0 \\ & & & \text{SYMMETRIC} & & & & 2\gamma_2+g^{22} \end{bmatrix} \begin{bmatrix} \hat{H}_1 \\ \hat{H}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \\ \hat{s}_{23} \end{bmatrix} = \begin{bmatrix} \sum\limits_j \sum\limits_k y_{1jk}\,\gamma_1 \\ \sum\limits_j \sum\limits_k y_{2jk}\,\gamma_2 \\ \sum\limits_k y_{11k}\,\gamma_1 \\ \sum\limits_k y_{12k}\,\gamma_1 \\ 0 \\ \sum\limits_k y_{21k}\,\gamma_2 \\ 0 \\ \sum\limits_k y_{23k}\,\gamma_2 \end{bmatrix} \tag{20}$$

Solving the above equations gives the BLUE of herd effects and the BLUP of transmitting abilities. The diagonal elements of the inverse of the coefficient matrix in (20) give the variances of estimates of herd effects and of prediction errors.

Consider the equations pertaining to sires that do not have daughters in the *ijth* cell. For example, the equation for the BLUP of the transmitting ability of sire 3 in herd 1 is

$$g^{11}\,\hat{s}_{13} + g^{12}\,\hat{s}_{23} = 0\,.$$

Thus

$$\hat{s}^{13} = -\,g^{12}\,\hat{s}_{23}/g^{11}$$

which illustrates clearly that the evaluation $\hat{s}_{13}$ is a function of the evaluation of the breeding value of sire 3 in herd 2. Incorporating relationships among sires (Henderson 1975b, 1976) does not pose conceptual difficulty. In this case, the variance-covariance matrix of the transmitting abilities becomes

$$\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$$

where $\mathbf{A}$ is a matrix of additive relationships. Hence, $\mathbf{G}^{-1} = \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}$.

In some circumstances, e.g., Hill (1984), one may wish to assume that $\varrho_{12} = 1$, i.e., absence of genotype x environment interaction. If $\varrho_{12} = 1$, then $\mathbf{G}_0$ is singular. This means that the general model in (16) is overparameterized when the correlation between breeding values of the same sire in different herds is perfect. Observe that $\varrho = 1$ implies a perfect linear relationship between the breeding values of sire j in herds i and i′ so

$$s_{i'j} = b_{i'i} \cdot s_{ij}$$

with $b_{i'i} = \sigma_{s_{i'}}/\sigma_{s_i}$. Now, let $k_i = \sigma_{s_i}^{-2}$ and rewrite (16) as

$$\begin{aligned} y^*_{ijk} &= (H_i + s_{ij} + e_{ijk})\sqrt{k_i} \\ &= H^*_i + s^*_j + e^*_{ijk}\,. \end{aligned} \tag{22}$$

Observe in (22) that subscript i (indexing herd) does not appear in the sire effect. The reason is that $s^*_j = s_{ij}/\sigma_{s_i} \sim (0, 1)$, so the variance of transmitting abilities in (22) no longer depends on the herd. Further

$$\mathrm{Cov}\left[\frac{s_{ij}}{\sigma_{s_i}}, \frac{s_{i'j}}{\sigma_{s_{i'}}}\right] = \mathrm{Var}\,(s^*_j) = 1\,. \tag{23}$$

Also,

$$e^*_{ijk} \sim \left[0, \frac{4-h^2_i}{h^2_i}\right] \tag{24}$$

where h^2_i is heritability in herd i. In vector notation

$$\mathbf{s}^* \sim (\mathbf{0}, \mathbf{A})$$

and

$$\mathbf{e}^* \sim (\mathbf{0}, \mathbf{R}^*)$$

where $\mathbf{R}^*$ is an $n \times n$ diagonal matrix with typical element $(4-h^2_i)/h^2_i$ for observations in the *ith* herd. For the data in Table 1, the mixed model equations to obtain BLUE of H^*_i and BLUP of s^*_j are

$$\begin{bmatrix} 3w_1 & 0 & 2w_1 & w_1 & 0 \\ & 3w_2 & w_2 & 0 & 2w_2 \\ & & 2w_1+w_2+a^{11} & a^{12} & a^{13} \\ & & & w_1+a^{22} & a^{23} \\ & \text{SYMMETRIC} & & & 2w_2+a^{33} \end{bmatrix} \begin{bmatrix} \hat{H}^*_1 \\ \hat{H}^*_2 \\ \hat{s}^*_1 \\ \hat{s}^*_2 \\ \hat{s}^*_3 \end{bmatrix} = \begin{bmatrix} \sum\limits_j \sum\limits_k y_{1jk}\,w_1 \\ \sum\limits_j \sum\limits_k y_{2jk}\,w_2 \\ \sum\limits_i \sum\limits_k y_{i1k}\,w_i \\ \sum\limits_i \sum\limits_k y_{i2k}\,w_i \\ \sum\limits_i \sum\limits_k y_{i3k}\,w_i \end{bmatrix} \tag{25}$$

where $w_i = h_i^2/(4-h_i^2)$, and a^{ij} is a typical element of $\mathbf{A}^{-1}$. Then, by linear invariance

$$\mathrm{BLUE}\,(H_i) = \sigma_{s_i} \cdot \mathrm{BLUE}\,(H_i^*) = \sigma_{s_i} \cdot \hat{H}_i^*$$

and

$$\mathrm{BLUP}\,(s_{ij}) = \sigma_{s_i} \cdot \mathrm{BLUP}\,(s_j^*) = \sigma_{s_i} \cdot \hat{s}_j^* \,.$$

Unknown covariance structure

The implementation of BLUP depends on knowledge of the needed variances and covariances, as illustrated by the preceding examples. In many instances, the only available information on variances and covariances is the data used to predict transmitting abilities in which case all parameters need to be estimated simultaneously. Gianola and Fernando (1986) suggest that in the absence of prior knowledge about variances and covariances, and when the likelihood function is reasonably "peaked", estimation should be by restricted maximum likelihood (REML, Patterson and Thompson 1971). These estimates would be used in the mixed model equations of Henderson (1973) to obtain the necessary predictions. This result applies strictly to multivariate normal data as REML is not defined as yet for other distributions. In the context of heterogeneous variance, if $\varrho \neq 1$, the problem is essentially a multiple-trait one. Hence, estimators of $\mathbf{G}_0$ should yield a "permissible" (Verdooren 1980) or "coherent" matrix (Foulley and Ollivier 1985). Otherwise, an absurd ranking of candidates for selection can be obtained (Hill and Thompson 1978).

The estimation of heterogeneous variances by REML in multiple trait settings was discussed by Schaeffer et al. (1978). We describe here an algebraically simpler procedure based on the EM algorithm for restricted maximum likelihood (Dempster et al. 1977), and consider the case where $\varrho \neq 1$. Suppose there are H sources (e.g., herds) of heterogeneity for the residual variance and for the variance of breeding values. Let the linear model describing the records in the *ith* "source" be

$$\mathbf{y}_i = \mathbf{X}_i \beta_i + \mathbf{Z}_i \mathbf{u}_1 + \mathbf{e}_i \,; \quad i = 1, \ldots, H \tag{26}$$

where β_i, $\mathbf{u}_i$ and $\mathbf{e}_i$ are the fixed effects, breeding values and residuals in the *ith* source, respectively. We assume

$$\begin{aligned}
&\mathrm{Cov}\,(\mathbf{u}_i, \mathbf{u}_j') = \mathbf{A}_{ij}\,\sigma_{u_{ij}}\,; && (i,j) = 1, \ldots, H \\
&\mathrm{Cov}\,(\mathbf{u}_i, \mathbf{e}_j') = \mathbf{0}\,; && (i,j) = 1, \ldots, H \\
&\mathrm{Cov}\,(\mathbf{e}_i, \mathbf{e}_j') = \mathbf{0}\,; && i \neq j \\
&\mathrm{Cov}\,(\mathbf{e}_i, \mathbf{e}_i') = \mathbf{I}\,\sigma_{e_i}^2; && i = 1, \ldots, H\,.
\end{aligned}$$

It is convenient (Henderson 1984) to include all potential breeding values in the *ith* "source", as done in the example in the preceding section. For example, with H = 2.

$$\mathrm{Var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\,\sigma_{u_1}^2 & \mathbf{A}_{12}\,\sigma_{u_{12}} \\ \mathbf{A}_{21}\,\sigma_{u_{12}} & \mathbf{A}_{22}\,\sigma_{u_2}^2 \end{bmatrix} \tag{27}$$

with inverse

$$\mathbf{G}^{-1} = \begin{bmatrix} g^{11}\,\mathbf{A}^{11} & g^{12}\,\mathbf{A}^{12} \\ g^{12}\,\mathbf{A}^{21} & g^{22}\,\mathbf{A}^{22} \end{bmatrix} \tag{28}$$

where g^{11}, g^{12} and g^{22} are as before.

The mixed model equations (H = 2) are then

$$\begin{pmatrix} \mathbf{X}_1'\,\mathbf{X}_1\,\gamma_1 & \mathbf{0} & \mathbf{X}_1'\,\mathbf{Z}_1\,\gamma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\,\mathbf{X}_2\,\gamma_2 & \mathbf{0} & \mathbf{X}_2'\,\mathbf{Z}_2\,\gamma_2 \\ \mathbf{Z}_1'\,\mathbf{X}_1\,\gamma_1 & \mathbf{0} & \gamma_1\,\mathbf{Z}_1'\,\mathbf{Z}_1 + \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{0} & \mathbf{Z}_2'\,\mathbf{X}_2\,\gamma_2 & \mathbf{G}^{21} & \gamma_2\,\mathbf{Z}_2'\,\mathbf{Z}_2 + \mathbf{G}^{22} \end{pmatrix} \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\mathbf{u}}_1 \\ \tilde{\mathbf{u}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\,\mathbf{y}_1\,\gamma_1 \\ \mathbf{X}_2'\,\mathbf{y}_2\,\gamma_2 \\ \mathbf{Z}_1'\,\mathbf{y}_1\,\gamma_1 \\ \mathbf{Z}_2'\,\mathbf{y}_2\,\gamma_2 \end{pmatrix} \tag{29}$$

Let a regular or symmetric generalized inverse of the coefficient matrix in (29) be

$$\begin{pmatrix} \underset{\sim}{\mathbf{C}}_{\beta_1\beta_1} & \underset{\sim}{\mathbf{C}}_{\beta_1\beta_2} & \underset{\sim}{\mathbf{C}}_{\beta_1 u_1} & \underset{\sim}{\mathbf{C}}_{\beta_1 u_2} \\ & \underset{\sim}{\mathbf{C}}_{\beta_2\beta_2} & \underset{\sim}{\mathbf{C}}_{\beta_2 u_1} & \underset{\sim}{\mathbf{C}}_{\beta_2 u_2} \\ & & \underset{\sim}{\mathbf{C}}_{u_1 u_1} & \underset{\sim}{\mathbf{C}}_{u_1 u_2} \\ \text{SYMMETRIC} & & & \underset{\sim}{\mathbf{C}}_{u_2 u_2} \end{pmatrix}$$

In the EM algorithm, there is a distinction between "complete" and "incomplete" data. The "incomplete" data are the observables

$$\mathbf{y}' = [\mathbf{y}_1', \mathbf{y}_2'] \,. \tag{30}$$

The "complete" data are given by the vector

$$\mathbf{z}' = [\mathbf{y}_1', \mathbf{u}_1', \mathbf{e}_1', \mathbf{y}_2', \mathbf{u}_2', \mathbf{e}_2'] \,. \tag{31}$$

The parameter vector ($\boldsymbol{\theta}$) consists of the unknown variances and covariances:

$$\boldsymbol{\theta}' = [\sigma_{u_1}^2, \sigma_{u_2}^2, \sigma_{u_{12}}, \sigma_{e_1}^2, \sigma_{e_2}^2] \,. \tag{32}$$

If $\mathbf{z}$ were observed, the sufficient statistics for the above parameters would be given by the vector

$$\mathbf{t} = \begin{pmatrix} \mathbf{u}_1'\,\mathbf{A}^{11}\,\mathbf{u}_1 \\ \mathbf{u}_2'\,\mathbf{A}^{22}\,\mathbf{u}_2 \\ \mathbf{u}_1'\,\mathbf{A}^{12}\,\mathbf{u}_2 \\ \mathbf{e}_1'\,\mathbf{e}_1 \\ \mathbf{e}_2'\,\mathbf{e}_2 \end{pmatrix} . \tag{33}$$

The expected value of the sufficient statistics for the complete data is

$$\mathbf{E}(\mathbf{t}\mid\boldsymbol{\theta}) = \begin{bmatrix} q\,\sigma^2_{u_1} \\ q\,\sigma^2_{u_2} \\ \sigma_{u_{12}}\mathrm{tr}\,(\mathbf{A}^{12}\mathbf{A}_{21}) \\ n_1\,\sigma^2_{e_1} \\ n_2\,\sigma^2_{e_2} \end{bmatrix} \tag{34}$$

where q is the number of breeding values to be estimated in each of the sources, and n_1 (n_2) is the number of records in source 1 (2).

Unfortunately, (33) is not observed. The E-step of the EM algorithm consists of finding the conditional expectation of the vector of sufficient statistics given the data $\mathbf{y}_1$ and $\mathbf{y}_2$. After algebra, we have

$$\mathrm{E}\,(\mathbf{u}_1'\mathbf{A}^{11}\mathbf{u}_1\mid\mathbf{y}_1,\mathbf{y}_2,\boldsymbol{\theta}) = \tilde{\mathbf{u}}'\mathbf{A}^{11}\tilde{\mathbf{u}}_1 + \mathrm{tr}\,(\mathbf{A}^{11}\mathbf{C}_{u_1u_1}) \tag{35}$$

$$\mathrm{E}\,(\mathbf{u}_2'\mathbf{A}^{22}\mathbf{u}_2\mid\mathbf{y}_1,\mathbf{y}_2,\boldsymbol{\theta}) = \tilde{\mathbf{u}}_2'\mathbf{A}^{22}\tilde{\mathbf{u}}_2 + \mathrm{tr}\,(\mathbf{A}^{22}\mathbf{C}_{u_2u_2}) \tag{36}$$

$$\mathrm{E}\,(\mathbf{u}_1'\mathbf{A}^{12}\mathbf{u}_2\mid\mathbf{y}_1,\mathbf{y}_2,\boldsymbol{\theta}) = \tilde{\mathbf{u}}_1'\mathbf{A}^{12}\tilde{\mathbf{u}}_2 + \mathrm{tr}\,(\mathbf{A}^{12}\mathbf{C}_{u_2u_1}) \tag{37}$$

Further

$$\mathrm{E}\,(\mathbf{e}_1'\mathbf{e}_1\mid\mathbf{y}_1,\mathbf{y}_2,\boldsymbol{\theta}) = \hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1 + \mathrm{tr}\,(\mathbf{B}_{11}) \tag{38}$$

$$\mathrm{E}\,(\mathbf{e}_2'\mathbf{e}_2\mid\mathbf{y}_1,\mathbf{y}_2,\boldsymbol{\theta}) = \hat{\mathbf{e}}_2'\hat{\mathbf{e}}_2 + \mathrm{tr}\,(\mathbf{B}_{22}) \tag{39}$$

where $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\tilde{\beta}_i - \mathbf{Z}_i\tilde{\mathbf{u}}_i$, and

$$\mathbf{B}_{ii} = \mathrm{tr}\,(\mathbf{C}_{\beta_i\beta_i}\mathbf{X}_i'\mathbf{X}_i + 2\,\mathbf{C}_{\beta_i u_i}\mathbf{Z}_i'\mathbf{X}_i + \mathbf{C}_{u_iu_i}\mathbf{Z}_i'\mathbf{Z}_i) \quad i = 1, 2\,.$$

Finally, the M-step of the algorithm consists of equating (34) to (35)–(39), and iterating with

$$\hat{\sigma}^{2\,[k+1]}_{u_1} = \frac{[\tilde{\mathbf{u}}_1'\mathbf{A}^{11}\tilde{\mathbf{u}}_1 + \mathrm{tr}\,(\mathbf{A}^{11}\mathbf{C}_{u_1u_1})]^{[k]}}{q} \tag{40}$$

$$\hat{\sigma}^{2\,[k+1]}_{u_2} = \frac{[\tilde{\mathbf{u}}_2'\mathbf{A}^{22}\tilde{\mathbf{u}}_2 + \mathrm{tr}\,(\mathbf{A}^{22}\mathbf{C}_{u_2u_2})]^{[k]}}{q} \tag{41}$$

$$\hat{\sigma}^{2\,[k+1]}_{u_{12}} = \frac{[\tilde{\mathbf{u}}_1'\mathbf{A}^{12}\tilde{\mathbf{u}}_2 + \mathrm{tr}\,(\mathbf{A}^{12}\mathbf{C}_{u_2u_1})]^{[k]}}{q} \tag{42}$$

$$\hat{\sigma}^{2\,[k+1]}_{e_1} = \frac{[\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1 + \mathrm{tr}\,(\mathbf{B}_{11})]^{[k]}}{n_1} \tag{43}$$

$$\hat{\sigma}^{2\,[k+1]}_{e_2} = \frac{[\hat{\mathbf{e}}_2'\hat{\mathbf{e}}_2 + \mathrm{tr}\,(\mathbf{B}_{22})]^{[k]}}{n_2} \tag{44}$$

where [k] indicates iterate number. The algorithm can be slow to converge as pointed out by Dempster et al. (1977). However, the EM algorithm is relatively easy to derive and guarantees that the eigenvalues of the estimated variance-covariance matrix of the breeding values are positive provided that the starting values for iteration are appropriately chosen (Henderson 1984). In general, this would insure "permissibility" of $\mathbf{G}_0$.

Clearly, the above computations are seldom feasible in data sets as large as the ones arising in national sire evaluation programs because of the size of the matrices involved. If "herds" are the source of heterogeneity, a large number of variances and covariances would need to be estimated. For example, with H = 300, 600 variances, and 44,850 potential covariances would need to be estimated. Further, individual parameters would not be well estimated as there would be little information on each of them. A possible shortcut would be to stratify herds into groups as has been done by, e.g., Hill et al. (1983), and then apply the procedures described here. The methodology yields estimates of the magnitude of "source" × genotype interactions. For example, the correlation between breeding values in different "sources" is estimated as $\varrho_{12} = \hat{\sigma}_{u_{12}}/(\hat{\sigma}_{u_1}\cdot\hat{\sigma}_{u_2})$.

Henderson (1984) and Hill (1984) suggested that the estimated variances could be regressed towards an average. For example, the estimator of the variance of breeding values in the *ith* herd given by Hill (1984) is a weighted average of a "prior" variance ($\bar{\sigma}^2_u$), and of the variance of breeding values in the *ith* herd estimated from the data ($\hat{\sigma}^2_{u_i}$). The weights are reciprocals of the variance of the "prior" variance and of the variance of the estimated variance component, so we have

$$\tilde{\sigma}^2_{u_i} = \left[\frac{\hat{\sigma}^2_{u_i}}{\mathrm{Var}\,(\hat{\sigma}^2_{u_i})} + \frac{\bar{\sigma}^2_u}{\mathrm{Var}\,(\bar{\sigma}^2_u)}\right] \Big/ [1/\mathrm{Var}\,(\hat{\sigma}^2_{u_i}) + 1/\mathrm{Var}\,(\bar{\sigma}^2_u)]\,.$$

In fact, this is the Bayesian solution for pooling prior and data based estimates of means in a normal distribution (e.g., Box and Tiao 1973). However, the weights chosen may not be optimal for pooling variance components. Further, the "correct" weights may be difficult or impossible to calculate in the data sets usually arising in animal breeding practice. However, the concept is appealing as prior knowledge is more "helpful" when the information is scant than when it is plentiful. In order to render the method suggested by Henderson (1984) and Hill (1984) completely Bayesian, it would be necessary to regard the variances in different sources as random variables with an exchangeable prior distribution (Lindley and Smith 1972). This prior distribution would be combined with the likelihood of the data to form a "posterior" distribution. Then, estimates of variances would be obtained from the posterior distribution following the Bayesian strategy.

Note that in the absence of prior information, e.g., $\mathrm{Var}\,(\bar{\sigma}^2_u) \to \infty$, the pooled estimate given above is precisely $\hat{\sigma}^2_{u_i}$, which would be the REML estimate described in this paper.

Acknowledgements. I am grateful to R. L. Fernando and to C. R. Henderson for comments. S. D. Kachman and A. L. Carriquiry are thanked for conveying their understanding of the EM algorithm to me.

References

Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Addison-Wesley, Reading Mass

Bulmer MG (1980) The mathematical theory of quantitative genetics. Clarendon Press, Oxford

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc, Ser B 39:1–38

Fernando RL (1984) Selection and assortative mating. PhD Thesis, University of Illinois, Urbana

Foulley JL, Ollivier L (1985) Criteres de coherence des parametres servant à ètablir un indice de selection. In: 36th Annu Meet EAAP. Kallithea, Greece

Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. J Anim Sci 63:217–244

Goffinet B (1983) Selection on selected records. Genet Sel Evol 15:91–97

Goffinet B, Elsen JM (1984) Critère optimal de selection: quelques résultats génèraux. Genet Sel Evol 16:249–394

Graybill FA (1976) Theory and application of the linear model. Duxbury Press, North Scituate Mass

Henderson CR (1973) Sire evaluation and genetic trends. In: Proc Anim Breed Genet Symp in Honor of JL Lush. Am Soc Anim Sci, Am Dairy Sci Asso, Champaign Ill, pp 10–41

Henderson CR (1975a) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

Henderson CR (1975b) Use of relationships among sires to increase accuracy of sire evaluation. J Dairy Sci 58: 1731–1738

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83

Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives' records. J Anim Sci 43:1188–1197

Henderson CR (1977) Best linear unbiased prediction of breeding values not in the model for records. J Dairy Sci 60:783–787

Henderson CR (1982) Best linear unbiased prediction in populations that have undergone selection. In: Barton RA, Smith WC (eds) Proc World Congr Sheep and Beef Cattle Breeding. Dunmore Press, Palmerston North, pp 191–201

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph (Canada)

Henderson CR (1985) Accounting for potential biases in predictions of breeding values with assortative mating. In: Abstracts 18th Meet Midwestern Sect Am Soc Anim Sci, p 79

Hill WG, Thompson R (1978) Probabilities of non-positive definite between-group or genetic covariance matrices. Biometrics 34:429–439

Hill WG, Edwards MR, Ahmed MKA, Thompson R (1983) Heritability of milk yield and composition at different levels and variability of production. Anim Prod 36:59–68

Hill WG (1984) On selection among groups with heterogeneous variance. Anim Prod 39:473–477

Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. J R Stat Soc, Ser B 34:1–41

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58: 545–554

Schaeffer LR, Wilton JW, Thompson R (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. Biometrics 34:199–208

Schaeffer LR, Henderson CR (1983) Best linear unbiased prediction when error vector is correlated with other random vectors in the model. Genet Sel Evol 15:395–400

Verdooren LR (1980) On estimation of variance components. Stat Neerl 34:83–106