Animal Model for Genetic Evaluation of Multibreed Data

J. W. Arnold3, J. K. Bertrand, and L. L. Benyshek

Department of Animal and Dairy Science, The University of Georgia, Athens 30602

ABSTRACT: Extension of beef cattle genetic evaluation procedures to multibreed data sets is proposed as a way to allow inclusion of crossbred animals into current analyses and to provide comparisons between purebred animals of different breeds. Previous papers dealing with multibreed BLUP have proposed sire or sire-maternal grandsire models. Because current models used in the beef industry are predominantly of the reduced animal model form, models were developed for animal model and reduced animal model mixed-model evaluations that would account for fixed and random additive genetic effects, along with fixed and random nonadditive genetic effects for populations with heterogeneous means and variances.

Key Words: Mixed Model Methods, Genetic Evaluation, Best Linear Unbiased Prediction, Crossbreeding, Beef Cattle

Introduction

Mixed-model procedures for genetic evaluation of beef cattle have become widely accepted as a selection tool by both purebred and commercial producers. A need has consequently developed for evaluation procedures to include cattle of diverse genetic composition. This need may include evaluations for composite breeds, for synthetic breeds in development, for breeds allowing a grading-up process to purebred status, and for commercial crossbred cattle. Genetic comparisons between breeds would also be valuable in development of breeding objectives and strategies for multiple-breed management systems. Although Elzo and Bradford (1985) and Elzo and Famula (1985) presented sire-maternal grandsire models for genetic evaluations involving heterogeneous variances, most U.S. beef evaluation programs currently employ the reduced animal model (RAM) for their evaluations (Benyshek et al., 1988; Benyshek and Bertrand, 1990).

The objective of this study is to present the animal model and RAM formulations for mixed-model procedures to evaluate simultaneously animals of diverse genetic composition using heterogeneous genetic and environmental (colvariances. These models will allow simultaneous analysis of data from multiple purebred populations as well as from any resulting crossbred progeny.

Model Development

Most programs presently in use for the genetic evaluation of U.S. beef cattle populations require that all animals in the evaluation be members of the same overall population. A significant outgrowth of this requirement is the assumption that additive genetic and environmental (colvariances are homogeneous throughout the entire population within each trait in the analysis. This affords the advantage of simplifying the expression of the mixed-model equations (MME) by augmenting all random components in the evaluation with the same basic set of variances and covariances. This assumption, although largely justified by the

1This study was partially funded by the Georgia Agric. Exp. Sta. and was part of Hatch Project 546.
2The authors express their appreciation to the reviewers of this paper for their constructive suggestions and detailed reviews.
Accepted April 8, 1991.
Accepted July 1, 1992.

Published December 11, 2014
MULTIBREED GENETIC EVALUATION

structure of the populations currently being evaluated, is not a necessary condition, and mixed-model theory can accommodate the evaluation of a population with heterogeneous variances if the (co)variance structure of that population is known (Henderson, 1984).

A given phenotypic record is typically represented in the single-breed animal model as a combination of fixed contemporary group effects, additive genetic effects, and random error. In the multiple-breed model, this is expanded to include nonadditive genetic effects, giving:

\[ y = Xb + Zu + Wh + e \]

where \( X \) is an incidence matrix relating the fixed contemporary group effects \((b)\) to the vector of observations \((y)\), \( Z \) is an incidence matrix relating the vector of total additive genetic values \((u)\) to \( y \), \( W \) is an incidence matrix relating the vector of total nonadditive genetic values \((h)\) to \( y \), and \( e \) is a vector of random residual effects. An initial modification to express [1] more precisely is the partitioning of \( u \) into fixed and random components. Following the logic of Quaas and Pollak (1981) and Famula et al. (1983), the total additive genetic value of an individual \((u)\) may be expressed as

\[ u = Qg + a \]

where \( Q \) is a matrix relating fractions of breed group effects to the animal, with these fractional contributions proportional to the breed composition of the animal, \( g \) is a vector of fixed additive breed group effects, and \( a \) is a vector of random additive genetic effects.

Throughout the paper, the term breed group \((g)\) will be used to refer to the fixed additive genetic effect that is intrinsic to the purebred animals of a particular breed in comparison to the other breeds in the evaluation. Animals of mixed-breed ancestry receive a linear combination of breed-group effects proportional to their breed composition. In some situations, it may also be necessary to assign additional fixed group effects to account for differences in genetic base and/or time period. That has not been done here in order to maintain clarity of presentation.

Similarly, the total nonadditive genetic (heterosis) effect \((h)\) will be partitioned into fixed and random components and modeled as

\[ h = Sd + Td \]

where \( S \) is a known incidence matrix relating the vector of fixed heterosis effects \((d)\) to parents of the individual making the record and \( T \) is an incidence matrix relating the vector of random heterosis effects \((d)\) to the vector of observations.

The vector of fixed heterosis effects \((d)\) represents the fixed effects of particular sire breed \( \times \) dam breed interactions \((SB \times DB)\). There are alternative ways to partition the heterosis effects (Gregory et al., 1978; Dillard et al., 1980). There will be one element in \( d \) for each \( SB \times DB \) combination. For example, the effect of mating Breed 2 with Breed 3 will be \( d_{23} \) (it will be assumed that \( d_{23} = d_{32} \)). The magnitude of the nonzero values in \( S \) will relate to the expected heterozygosity in the progeny due to the breed composition of the parents. For example, in the sample data set presented in Table 1, Animal 22 is the progeny of a sire that is an \( F_1 \) cross between Breeds 1 and 2 mated with a purebred dam of Breed 1. This animal (22) would be expected to display 50% of the maximum heterosis resulting from this breed combination, and therefore .5 is the corresponding value in \( S \).

The random component of heterosis \((h)\) is a deviation from the average \( SB \times DB \) effect \((d)\) and may represent a combination of sire \( \times DB \) and dam \( \times SB \) interactions. This effect recognizes that a male or female parent has a genetic composition that, when in combination with a particular breed, will contribute to individual heterosis in the progeny to a greater or lesser extent than another parent of the same breed. The non-zero elements in the columns of \( T \) will relate records from crossbred animals to their parents. This is necessary because heterosis is a genotypic effect (as

Table 1. Example data set to illustrate structure of incidence matrices and vectors

<table>
<thead>
<tr>
<th>Breed</th>
<th>CG*</th>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>11</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>13</td>
<td>9</td>
<td>11</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>14</td>
<td>9</td>
<td>12</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>16</td>
<td>10</td>
<td>12</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>17</td>
<td>1</td>
<td>11</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>18</td>
<td>1</td>
<td>13</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>19</td>
<td>3</td>
<td>12</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>20</td>
<td>9</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>21</td>
<td>10</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>22</td>
<td>17</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>23</td>
<td>17</td>
<td>18</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Contemporary group.
opposed to a gametic effect), and the progeny exhibiting the heterosis will not transmit it to the next generation in the same way. As with $S$, the non-zero values in $T$ will correspond to the expected heterozygosity in the progeny.

The possibility of predicting individual values for random heterosis is problematic at best and may be impractical in realistic beef cattle populations. A sire $\times$ DB component might be estimated for some widely used sires, but prediction of specific dam $\times$ SB values would be much less likely. Although the model is presented as above for completeness, initial applications of this model to existing cattle populations will most likely operate under the assumption that a fixed component of heterosis is sufficient.

Substitution for $u$ and $h$ into [1] yields the multiple-breed model:

$$y = Xb + ZQg + Za + WSd + WT\delta + e$$  \[2\]

where $y$ is a vector of observations, $X$ is an incidence matrix relating the vector of fixed contemporary group effects ($b$) to $y$, $ZQ$ is an incidence matrix relating the vector of fixed additive breed group effects ($g$) to $y$, $Z$ is an incidence matrix relating the vector of random additive genetic effects ($a$) to $y$, $WS$ is an incidence matrix relating the vector of fixed heterosis effects ($d$) to $y$, $WT$ is an incidence matrix relating the vector of random heterosis effects ($\delta$) to $y$, and $e$ is a vector of random residual effects.

$$E(y) = Xb + ZQg + WSd$$

$$V \begin{bmatrix} a & \delta \\ \delta & e \end{bmatrix} = \begin{bmatrix} G & 0 & 0 \\ 0 & H & 0 \\ 0 & 0 & R \end{bmatrix}$$

It should be noted that $W$ is a diagonal matrix with ones on the diagonals corresponding to crossbred animals. It will therefore be seen that $WSd = Sd$ and $WT\delta = T\delta$. The structure of $G$ is discussed in the Appendix. In the MME, coefficients for diagonal blocks of the equations for $\delta$ are augmented by adding the inverse of the covariance matrix of the $\delta (H^{-1})$. If a zero correlation is assumed for interaction effects for individual sires and dams, $H^{-1}$ will be diagonal and $\delta$ will be predicted only for parents having crossbred progeny with records. In a procedure similar to that described by Henderson (1976), Elzo (1990b) has shown how $H^{-1}$ may be constructed to include relationships for the nonadditive effects. Incorporation of this procedure into the MME would allow prediction of $\delta$ for nonparents. The matrix $R$ represents the covariance structure among the residuals for animals with records.

The matrix representation of [2] is similar to that presented by Famula et al. (1983). Other than the addition of nonadditive effects, one distinction is that in [2], breed group is a cross-classified, fixed effect common to all animals in a breed and is not used to account for selection of base animals or time trend within a breed. The presence of heterogeneous base populations within a breed may make the inclusion of within-breed grouping strategies desirable (Wiggans et al., 1988). A fixed effect for within-breed genetic grouping could be added to the model in addition to the groups as described here. In the present paper an animal model using the inverse of the numerator relationship matrix ($A^{-1}$) will be assumed to minimize the need for a within-breed grouping strategy (Henderson, 1975). In addition, the present model makes no assumption of homogeneity of variances across breeds.

The MME for Model [2] are as follows:

$$X'R^{-1}X \quad X'R^{-1}ZQ \quad X'R^{-1}Z \quad X'R^{-1}WS \quad X'R^{-1}WT \quad X'R^{-1}y$$

$$Q'Z'R^{-1}ZQ \quad Q'Z'R^{-1}Z \quad Q'Z'R^{-1}WS \quad Q'Z'R^{-1}WT \quad Q'Z'R^{-1}y$$

$$Z'R^{-1}Z \quad Z'R^{-1}WS \quad Z'R^{-1}WT \quad Z'R^{-1}y$$

$$S'W'R^{-1}WS \quad S'W'R^{-1}WT \quad S'W'R^{-1}x$$

$$T'W'R^{-1}WT \quad T'W'R^{-1}y$$

$$R^{-1}$$

$$H^{-1}$$

$\delta = G^{-1}$

$\bar{\delta}$

$\bar{\delta}$

$T'W'R^{-1}y$

$Q'Z'R^{-1}y$

$S'W'R^{-1}y$

Specification of Matrices

Setting up Model [2] and the corresponding MME [3] is a matter of specifying an appropriate structure for the various matrices and vectors involved. To that end, arrange the vector of observations ($y$) as follows:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{1p} \\ y_{1n} \\ y_{2p} \\ y_{2n} \\ \vdots \\ y_{xp} \\ y_{xn} \end{bmatrix}$$
In this section, matrices and vectors are represented in the form $M_{ij}$, where $i=\{1,2,x\}$ corresponds to animals of breed 1, 2 or crossbred animals, respectively, and $j=\{p,n\}$ is the portion of the matrix or vector corresponding to records of parents and nonparents, respectively. The other matrices and vectors of Model [2] are similarly organized:

$$X = \begin{bmatrix} X_{11} \\ X_{21} \\ X_{x1} \end{bmatrix}; \quad Z = \begin{bmatrix} Z_{11} & 0 & 0 \\ 0 & Z_{21} & 0 \\ 0 & 0 & Z_{x1} \end{bmatrix} = \begin{bmatrix} Z_{1p} & 0 & 0 \\ 0 & I_{1n} \end{bmatrix}$$

The matrices $X$ and $Z$ are similar to those commonly used in single-breed analyses. The matrix $Q$ relates the breed groups to the observations. For purebred animals, each row of the submatrix $Q_{ij}$ $(Q_{2j})$ has a single 1 relating that animal to the appropriate element of the vector of breed groups ($g$). In the case of crossbred animals, each row will have nonzero coefficients corresponding to the fractional breed composition of that animal. If all animals have records, $Z$ is an identity matrix. Parents with no record are represented in $Z$ by a null column (nonparents without records are excluded from the analysis).

Incidence matrices for $S$ and $T$ will be represented symbolically as follows:

$$S = \begin{bmatrix} 0 \\ 0 \\ 0 \\ S_{xp} \\ S_{xn} \end{bmatrix}; \quad T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ T_{1p} & T_{2p} & T_{xp} \\ T_{1n} & T_{2n} & T_{xn} \end{bmatrix}$$

The matrices $S$ and $T$ relate to the fixed and random heterosis effects contained in the phenotypic records of crossbred animals. Under the assumption that reciprocal matings are equal, a given row of $S$ will relate the contribution of $d$ to the record in proportion to the expected fraction of heterozygosity in the individual making the record. This could be expanded further to include reciprocal and maternal effects but will not be done here. Similar to $S$, the nonzero elements of $T$ express the expected fraction of heterozygosity in the individual. The predictions of $\delta$ should therefore not be biased by matings between animals with breed compositions that provide less than the maximum possible heterosis. Rows for crossbred records will have nonzero values in columns corresponding to the parents of the individual. As stated earlier, this is done because heterosis is a genotypic effect due to the contributions of the parents and is not transmissible by the progeny to the next generation. Records of purebred individuals will be represented in $S$ and $T$ by null rows. Parents without crossbred progeny will have a null column in $T$.

The organization of vectors in [2] is given by

$$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}; \quad a = \begin{bmatrix} a_{1p} \\ a_{1n} \\ a_{2p} \\ a_{2n} \\ a_{xp} \\ a_{xn} \end{bmatrix}; \quad d = (d_{11}, d_{12}, d_{2p}, d_{2n}, d_{xp}, d_{xn}); \quad \delta = (\delta_{1p,12}, \delta_{2p,12}, \delta_{xp,12}); \quad e = (e_1, e_2, e_x)$$

The vector $a_{ij}$, $i=\{1,2,x\}$ represents additive genetic deviations within $g$ for individuals of breed 1, 2, or crossbred, respectively. Linear dependencies among the fixed effects in the model make it necessary to impose constraints on the solutions to make the system of equations full rank. With two groups in the model, one type of constraint is simply to eliminate from the system of equations the row and column corresponding to one of the groups.

Vector $d_{ij}$, $(i,j=\{1,2\}, i\neq j)$ is the vector of fixed SB $\times$ DB interaction effects (heterosis) for SB $i$ and DB $j$. The assumption will be made that $d_{12} = d_{21}$ in a nonmaternally influenced trait.

The vector of nonadditive random effects ($\delta_{ip,jk}$, $i=\{1,2,x\}$) represents a random deviation within the fixed heterosis effect represented by $d_{ij}$. In an analysis with more than two breeds, fixed heterosis effects could be modeled for each combination of breeds. In general, $\delta_{ip,jk}$ will contain elements representing a deviation of the parents within $d_{jk}$, where $jk$ stands for the particular mating type subclass. Although the estimation of many separate $\delta$ effects may have theoretical advantages, in practical applications it may be difficult to estimate the necessary nonadditive variance components, or to do so with an acceptable level of accuracy.

In a single-breed, single-trait analysis, both sides of the equations are typically multiplied by
\( \sigma_e^2 \) to remove \( R^{-1} \) from the block matrices. With a heterogeneous residual variance structure, \( R \) is not equal to \( I \cdot \sigma_e^2 \) and this simplification is not as straightforward. One alternative is to divide each record by the appropriate \( \sigma_e^2 \) and adjust the (covariance matrix accordingly. The approach taken here is to retain the residual variance structure for each breed. This is similar to a single-breed, multiple-trait analysis in which each trait has its own genetic and residual (covariances. The residual variance structure for the multibreed model is as follows:

\[
R = \begin{bmatrix}
\sigma_e^2 & 0 & 0 \\
0 & \sigma_e^2 & 0 \\
0 & 0 & \sigma_e^2
\end{bmatrix}
\]

The inverse coefficient of relationship matrix \( (A^{-1}) \) must accurately trace gene flow in the crossbred animals and also must give purebred and crossbred parents credit for the performance of their progeny. There will be a non-null block of \( A^{-1} \) between Breeds 1 and 2 \( (A^{12}) \) corresponding to matings between animals of the two breeds. Additional blocks will account for relationships among crossbred animals \( (A^{xx}) \) and between them and their purebred ancestors or mates \( (A^{1x}) \).

\[
A^{-1} = \begin{bmatrix}
A_{11} & 0 & A_{1X} \\
0 & A_{22} & A_{2X} \\
A_{XX} & A_{XX}
\end{bmatrix}^{-1} = \begin{bmatrix}
A^{11} & A^{12} & A^{1X} \\
A^{22} & A^{2X} \\
A^{XX} & A^{XX}
\end{bmatrix}
\]

\[
Z = \begin{bmatrix}
Z_{1p} & 0 & 0 \\
\frac{1}{2}P_{1n} & 0 & 0 \\
0 & Z_{2p} & 0 \\
0 & \frac{1}{2}P_{2n} & 0 \\
0 & 0 & Z_{xp} \\
\frac{1}{2}P_{1xn} & \frac{1}{2}P_{2xn} & \frac{1}{2}P_{xn}
\end{bmatrix}; 
Q = \begin{bmatrix}
Q_{1p} & 0 & 0 \\
0 & Q_{2p} & 0 \\
Q_{1xp} & Q_{2xp}
\end{bmatrix}; 
\mathbf{a} = \begin{bmatrix}
a_{1p} \\
a_{2p} \\
a_{xp}
\end{bmatrix}; 
\mathbf{e} = \begin{bmatrix}
e_1 \\
e_1 + \Phi_{1n} \\
e_2 \\
e_2 + \Phi_{2n} \\
e_x \\
e_x + \Phi_{xn}
\end{bmatrix}
\]

where \( P_{in} \) is an incidence submatrix relating progeny in \( y_n \) to parental genetic effects. The submatrices \( \frac{1}{2}P_{1xn}, \frac{1}{2}P_{2xn}, \frac{1}{2}P_{xn} \) indicate that crossbred nonparents may be from any combination of either purebred or crossbred parents. The row of \( Z \) for a crossbred nonparent will contain at most two nonzero values in the columns corresponding to the parents. Table 1 presents a sample data set for which the RAM incidence matrices are constructed in Tables 2 and 3. Representations of vectors \( \mathbf{a} \) and \( \delta \) are given in

The inverse relationship matrix must then be combined with the inverse genetic (covariance matrix \( (G_0^{-1}) \) to create \( G^{-1} \). In a typical single-breed model, \( G^{-1} \) may be expressed as the direct product of \( A^{-1} \) with \( G_0^{-1} \). With crossbred animals in the data, this relationship is not as straightforward (Elzo, 1990a). See the Appendix for a description of the procedures for building \( G^{-1} \) with multiple-breed data.

The above matrices are used in (3) to yield the animal model MME for multiple breeds involving heterogeneous variances.

**Reduced Animal Model**

Data processing and computation of solutions for very large systems of equations can be expensive and time-consuming. These problems are compounded in a multiple-breed analysis because of the increase in the size of the data that must be handled. These concerns have previously been addressed (Quaas and Pollak, 1980) for single-breed analyses in the development of the RAM. An analogous approach will be developed here, which will allow equations representing additive genetic values for nonparents to be absorbed by expressing the incidence matrices in terms of the parental values and modifying the error structure. Let:

\[
y_{22} = b_4 + \frac{1}{4}g_1 + \frac{1}{4}g_2 + \frac{1}{4}a_5 + \frac{1}{2}a_{17} + \frac{1}{2}d_{12} + \frac{1}{2}d_{5,12} + \frac{1}{2}d_{17,12} + (e_x + \Phi_{22,x}).
\]

The residual for nonparents, \( \epsilon_j + \Phi_{ij}, j = \{1,2,x\} \), indicates that the Mendelian sampling effect \( (\Phi_{ij}) \) is included in the residual. The error variances under this modified error structure become:

**Table 4:** As an example, the record for Animal 22 is represented in the model as follows:

\[
y_{22} = b_4 + \frac{1}{4}g_1 + \frac{1}{4}g_2 + \frac{1}{4}a_5 + \frac{1}{2}a_{17} + \frac{1}{2}d_{12} + \frac{1}{2}d_{5,12} + \frac{1}{2}d_{17,12} + (e_x + \Phi_{22,x}).
\]
The term \((I_1 \sigma^2_{e_i} + D_1 \sigma^2_{A_i})\) represents the variance of \((e_i + \Phi_i)\) (Quaas and Pollak, 1980). The matrix \(D\) is a diagonal matrix with values of \(\frac{3}{4}\) or \(\frac{1}{2}\), depending on whether one or both parents are known, respectively. If neither parent is known, an animal will not enter the analysis until it produces progeny. The variances for crossbred animals will change depending on the breed composition of the animal and the variances of the parental populations. The variances \(\sigma^2_{e_x}\) and \(\sigma^2_{A_x}\) represent the several variances needed to deal with multiple crossbred groups. These variances will be formed as needed and incorporated into the structure of \(V(e)\). In addition, this model does not account for inbreeding, as has been suggested by Golden et al. (1991).

The multiple-breed model (21), modified for absorption of nonparents, is shown in Table 5. The MME are of the same form as (3) using the incidence matrices and vectors as shown in Table 5. Note that because \(WSd = Sd\) and \(WT6 = T6\) the matrix \(W\) has been omitted from the model representation.

As in the full animal model, \(G^{-1}\) is formed recursively considering the heterogeneity of genetic variances among breeds using the inverse of the numerator relationship matrix \((A^{-1})\) among the parents.

**Discussion**

The models presented in this paper have been intentionally simplified to avoid the problems of space and notation that arise in expansions to more complex models. This basic model may be expanded and modified to incorporate additional effects into the evaluation. Possible extensions include models for maternally influenced traits, multiple-trait models, and the addition of genetic groups within breed to account for heterogeneous base animals. The resulting equations will be correspondingly more complex, although theoretically not fundamentally different from those presented here.

A potentially useful adaptation of this model involves the analysis of multiple breeds of purebred animals. This is essentially a special case of the current model with zero nonadditive effects. The primary advantage of a model of this type would be the evaluation and comparison of additive genetic merit within and across breeds for purebred animals in current national cattle evaluation programs.

Composite breeds such as the Brahman derivative breeds or breeds that allow a grading-up process would benefit from the ability to include animals of intermediate breed type in their analyses. Individuals of similar genetic composition may have a different expression of heterosis due to differences in their parentage. Current analyses of these breeds generally require that records expected to display increased amounts of heterosis be excluded from the analysis due to the possibility of favorable bias toward these animals. Use of a model with a heterosis effect would allow the inclusion of many nonparent records that are currently being eliminated from the analysis. One alternative is to use contemporary groups to segregate the crossbred animals. This would work in some instances but would have the potential disadvantage of eliminating contemporary group ties between animals of the different classes.

This model does not specifically partition the effects into purebred effects and general and specific combining ability as does the more classical least squares analysis of diallel crossbred data (Gregory et al., 1978; Dillard et al., 1980). In the model presented here the breed group effect \((g)\) relates to the overall contribution of a breed to the matings in which it is involved. It will contain elements of both the purebred effect \((pb)\) and the general combining ability \((gca)\) because the estimate of \(g_i\) comes from comparisons among all purebred and crossbred animals containing a fraction of the breed \(i\). Although this is not exactly the same partitioning as the classical approach, \(g_i\)
Table 2. Incidence matrices for reduced animal model example

+ $d_{ij} = p_b_i + g_{cA_i} + m_i + s_{CA_i} + r_{CA_i}$ for a particular breed combination. The fixed heterosis effect ($d$) will contain the specific combining ability ($s_{CA}$) along with any reciprocal ($r_{CA}$) or maternal ($m$) effects. In a maternally influenced trait, the vector for $d$ could be expanded to include estimates for maternal heterosis.

One issue of concern in the application of this model is the possibility of achieving a data structure that allows differences between the various effects to be estimated. Breed groups as defined here provide an estimate of the fixed genetic differences between animals of two different breeds. For this contrast to be estimable, some animals of those breeds must have been raised in common contemporary groups under the same environmental conditions. The definition of comparable environments could take on radically different meaning depending on whether the concern is with absolute equality of environment or equality relative to some measure of total production potential or metabolic requirements. The possibility of genotype × environment interactions becomes a special concern when one compares breeds with greatly different mature sizes and metabolic requirements or specialized environmental adaptations. The models presented here do not attempt to account for any existing genotype × environment interactions. Caution should therefore be exercised when one extends the results of a particular analysis to environments that have not been tested and for which significant genotype × environment interactions may exist.

The nature of assumptions about the random nonadditive genetic component ($\sigma$) will greatly affect the data structure required for the analysis. Inclusion of multiple-locus interactions in multiple-breed data sets may make the dimension of $\sigma$ much larger than that of the vector of additive genetic components. From a practical standpoint, evaluations involving multiple breeds and many crossbred progeny may require supercomputer capability in terms of storage requirements and execution time. This poses potentially severe restrictions on the number of animals that could be analyzed with current computer hardware and computing strategies. The lack of adequate estimates for nonadditive variance components is a serious limitation if random nonadditive effects are important. A concentrated effort will be required to acquire the necessary data.

These models also allow inclusion of random
heterosis for dam × SB interaction effects ($\delta_D$). If adequate estimates of nonrandom variance components were available, it is conceivable that $\delta_D$ could be predicted for individual dams. Realistically, however, this is unlikely for all but a select few dams. It would probably be necessary to estimate just the sire × DB effects ($\delta_S$) and allow $\delta_D$ to be included in the residual. The effect of this would be to increase the prediction error variance. It seems reasonable to think that for most production traits this would be a minor increase. Analysis of traits with strong maternal influences would possibly make prediction of dam contributions to progeny heterosis of more importance, although no less difficult.

**Implications**

Current national genetic evaluation programs generally assume that all animals in the analysis are of uniform genetic composition. For joint analysis of many breeds, this assumption may not be appropriate or may require the culling of many records from the collected data before analysis. In addition, commercial producers often want to compare animals of diverse genetic background. This paper proposes mixed-model equations that account for breed groups, nonadditive effects, and heterogeneity of variances among animals. Animal model and reduced animal model forms of the equations are presented that hold the potential to reduce the loss of data and allow comparisons not possible under single-breed models.

**Literature Cited**


Table 5. Multiple-breed reduced animal model

\[
\begin{align*}
Y_{1p} & = X_{1p} b + e_{1} + f_{1p} \, \text{P}_{1n} + \delta_{1p} \, \text{P}_{2n} + \varepsilon_{1p} \, \text{P}_{2n} \\
Y_{2p} & = X_{2p} b + e_{2} + f_{2p} \, \text{P}_{1n} + \delta_{2p} \, \text{P}_{2n} + \varepsilon_{2p} \, \text{P}_{2n} \\
Y_{3p} & = X_{3p} b + e_{3} + f_{3p} \, \text{P}_{1n} + \delta_{3p} \, \text{P}_{2n} + \varepsilon_{3p} \, \text{P}_{2n} \\
Y_{4p} & = X_{4p} b + e_{4} + f_{4p} \, \text{P}_{1n} + \delta_{4p} \, \text{P}_{2n} + \varepsilon_{4p} \, \text{P}_{2n} \\
\end{align*}
\]

United States beef cattle industry. J. Dairy Sci. 71(Suppl. 2): 35.


This appendix deals with the specific procedures for forming the inverse matrix of (co)variances of the additive genetic effects ($G^{-1}$) for multiple-breed data sets. Although the mixed-model equations in the body of the paper are single-trait models, the procedures for a two-trait model will be presented here to provide a closer comparison with those of Elzo (1990a).

In a two-trait analysis among Traits 1 and 2 for purebred animals we may symbolize the inverse matrix of additive genetic (co)variances by $G^{-1}$:

$$
G_{o}^{-1} = \begin{bmatrix}
  g_{11} & g_{12} \\
  g_{12} & g_{22}
\end{bmatrix}^{-1} = \frac{1}{g_{11}g_{22} - (g_{12})^2}
\begin{bmatrix}
  g_{22} & -g_{12} \\
  -g_{12} & g_{11}
\end{bmatrix}
= \begin{bmatrix}
  g^{11} & g^{12} \\
  g^{12} & g^{22}
\end{bmatrix}
$$

Generally, for a single-breed model, $G^{-1}$ is expressed as the direct product of $A^{-1}$ with $G_{o}^{-1}$. The creation of $G^{-1}$ in the mixed-breed case must recognize that each of the $g_{ij}$ in $G_{o}$ for each individual is assumed to be a composite value depending linearly on the breed group composition of the animal (Elzo, 1990a):

$$
ge_{ij} = \sum_{k=1}^{n} c_{k} g_{kij}
$$

where $c_{k}$ refers to the proportion of genes from the $k^{th}$ breed, $g_{kij}$ is the additive genetic covariance between traits $i,j$ in breed $k$, and summation is over all breeds. This prevents expressing $G^{-1}$ as a direct product for crossbreed data. This discussion will deal with Traits 1 and 2 for Breeds A and B. The elements $g_{ij}$ forming $G_{o}$ for any animal will be given by:

$$
ge_{ij} = c_{A}g_{Aij} + c_{B}g_{Bij}
$$

where $c_{A}$ ($c_{B}$) is the coefficient representing the proportion of breed A (B) in the individual and $c_{A} + c_{B} = 1$.

If we make this substitution in [4] we get:

$$
g^{11} = \frac{(c_{A}g_{A12} + c_{B}g_{B12})}{(c_{A}g_{A11} + c_{B}g_{B11})(c_{A}g_{A22} + c_{B}g_{B22}) - (c_{A}g_{A12} + c_{B}g_{B12})^2};
$$

$$
g^{12} = \frac{- (c_{A}g_{A12} + c_{B}g_{B12})}{(c_{A}g_{A11} + c_{B}g_{B11})(c_{A}g_{A22} + c_{B}g_{B22}) - (c_{A}g_{A12} + c_{B}g_{B12})^2};
$$

$$
g^{22} = \frac{(c_{A}g_{A11} + c_{B}g_{B11})}{(c_{A}g_{A11} + c_{B}g_{B11})(c_{A}g_{A22} + c_{B}g_{B22}) - (c_{A}g_{A12} + c_{B}g_{B12})^2};
$$

The matrix $G^{-1}$ is then created by accumulating coefficients for each animal according to the rules of Henderson (1976). If both parents are known, add:

$$
2 \cdot g^{ij} \text{ to } \{n,n\};
-1 \cdot g^{ij} \text{ to } \{n,s; s,n; n,d; d,n\};
.5 \cdot g^{ij} \text{ to } \{s,s; d,d; s,d; d,s\};
$$

where $g^{ij}$ is the proper element from $G_{o}^{-1}$ for each animal, $\{x,y\}$ is the cell of $G^{-1}$ corresponding to the relationship between individual (n), sire (s), and dam (d).

In the case of only one parent known, the coefficients and cells to augment are:

$$
4/3 \cdot g^{ij} \text{ to } \{n,n\};
$$
Where \( p \) refers to the known parent.

If neither parent is known add:

\[
1 \cdot g^{ij} \to \{n,n\}.
\]

In the purebred case the various \( g^{ij} \) of the mixed-breed analysis reduce to those of a single breed.

The procedure as presented here works when both parents of the animal are known because the variance of the Mendelian sampling effect is equal to one-half the genetic variance of an animal. It also works when neither parent is known because in that case the Mendelian sampling variance equals the genetic variance of the animal.

The method presented here for multibreed data does not work when one parent is completely unknown. The rules of Henderson (1976) work in a single-breed analysis because when one parent is unknown the residual variance is equal to three-quarters of the genetic variance. In the multibreed case the identity of one parent can be unknown as long as the breed group composition of that parent is known. The assumptions regarding the residual and genetic variances will then hold. A similar alternative is to include an imaginary parent in the analysis corresponding to the unknown parent of a particular breed composition.

In either case, under this method the breed composition of all animals must be known so the right genetic (co)variances will be used when building \( G^{-1} \). See Elzo (1990a) for a more general method of constructing \( G^{-1} \).