Methodologies for Population/Quantitative Genetics  
Animal Science 562

Topic: Genetic Groups in Animal Models

“Effects of genetic groups can be included in models used for genetic evaluation to account for prior selection that resulted in base animals for which records are not available to the genetic evaluation (Thompson, 1979; Quaas and Pollak, 1981; Westell, 1984; Robinson, 1986; Quaas, 1988). If base animals are from one birth period and all are selected the same way, then one genetic group is sufficient. A single genetic group is implied when genetic groups are not included in the model used for evaluation. Base animals, by necessity as well as logically, usually are defined as those most recent ancestors that do not have records or collateral (e.g., paternal half-sib) descendants. Nearly all, if not all, data sets for genetic evaluation include records associated with base animals from different time periods and selection paths. If selection has occurred, then ignoring differences in groups has two major consequences in prediction of breeding values. Essentially, the predictions involve regression to the average group effect rather than to the appropriate group effect. Ranking of animals with base ancestors from different time periods and selection paths can be affected. The other consequence arises when new data are added that include base animals from later time periods. Evaluation of older animals with no new information (records or relatives) may change (i.e., float), because the implied assumption that the base for the evaluation has been set by an unchanging base population is not correct. Groups can be assigned arbitrarily but should logically account for different genetic means from different time periods or subpopulations.” Van Vleck (1990)

The use of complete relationship information, as in the case of an animal model, decreases he need for genetic grouping to account for selection (Pollak and Quaas, 1983). Some animals entering the population may have extensive pedigrees, however, the parents of these animals may not contribute ties to the rest of the data or have records themselves to account for selection. Group effects can be used to account for selection not accounted for by records of relatives. In this case animals missing genetic relationships would be assigned to genetic groups. The Westell-Thompson (Westell, 1984; Thompson, 1979) type of grouping allows for genetic differences or similarities of individuals with unknown ancestry to enter the population at different times.

Example 1

Consider the following hypothetical example. Three cows have the same sire, their dams are unknown, and the sire was mated to these dams in different years.

<table>
<thead>
<tr>
<th>Kildee Fencehopper</th>
<th>Kildee Fencehopper</th>
<th>Kildee Fencehopper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suzy</td>
<td>Sweet</td>
<td>Sour</td>
</tr>
</tbody>
</table>
A proxy or “phantom” name could be assigned to each of the unknown dams. Sweet and Sour should be more similar if we assume the dams possess the average genetic merit of the population for the year they were born. Therefore, it is conceptually appealing to assign the unknown dams to one group, estimate the average effects of all dams in a group and then add a weighted average of these estimated group effects to the predicted values of all animals from the same subpopulation. A topic of debate is whether this is better than assuming there is no dam.

**Process for assigning unknown ancestors to a group**

The process for...

…”deciding which animals are represented in a group is typically a recursive process and is somewhat arbitrary (e.g., initially the list of animals in U might include cows with records to be analyzed). From a ‘pedigree’ file, the sires of these cows are looked up and those bulls found are added to the list. Likewise any dams not already on the list might be added. The process is then repeated; the sires and dams of the animals just added are looked up and added to the list. In principle this could be repeated again and again until no more ancestors can be found in the pedigree file, in which case “unidentified” parents are truly unknown. Usually, however, only a few cycles are carried out and perhaps only sires are added to the list. At the stopping point, if an animal’s parent is not on the list, that parent is considered a base parent even though it might be found in the pedigree file.” Quaas (1988)

“In practice, assigning a base animal to a group is usually based on attributes of the base animal’s progeny. It almost always involves the birth date of the progeny to model time trends, but may involve other characteristics as well. For example, Westell (1984) suggested forming separate groups subdivided by time, for sires of sires, sires of dams, dams of sires and dams of dams.” Quaas (1988)

Obviously, there are many ways to assign unknown ancestors to groups. The proper definition will depend on one’s knowledge of the data and the intended purpose of the analysis.

One can also use generation intervals to assign a “year of birth” to phantom parents (e.g., 10 yr sires of sire, 9 yr sire of cow, 7 yr dam of bull and 5 yr dam of cow (Westell and Van Vleck, 1987)). If we add a bull (Big Joe) to our example then Table 1 illustrates some of the possible ways the unknown ancestors could be assigned to groups.

**Computing Algorithm for A⁻¹**

In developing the MME to include groups for unknown ancestors the objective is to use all the information on known relationships to develop A⁻¹. Once the unknown ancestors are assigned
to groups, elements of the group by random and group by group segments of the MME are computed by assigning coefficients that are similar to elements in $A^{-1}$ that are appropriate for the type of unknown ancestor. Although, Westell et al. (1988) give some formal rules for computing these elements, the procedure described by Quaas (1988) seems more straightforward. The procedure described below explains how to compute elements of $A^*$, which is $A^{-1}$ for related animals plus the elements for the animal by group and group by group segments of the mixed model equations.

Table 1. Alternative strategies for assigning unknown ancestors to groups.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Parents</th>
<th>Phantom parent</th>
<th>Group defined by</th>
<th>Year</th>
<th>Four pathways</th>
<th>Sires and dams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suzy</td>
<td>Kildee Fencehopper</td>
<td>a</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>? '78 a</td>
<td>b</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Sweet</td>
<td>Kildee Fencehopper</td>
<td>c</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>? 85 b</td>
<td>d</td>
<td></td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Sour</td>
<td>Kildee Fencehopper</td>
<td>e</td>
<td></td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>? '85 c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Joe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>? '85 d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Let i, j, and k represent the equation numbers of an individual, its sire or its sire’s group (if base sire), and its dam or dam’s group (if base dam) respectively, and $m = 0, 1, 2$ = number of base parents of the individual.

Proceed through the list of animals represented in $U$. For each, calculate $x = 4 / (m+2)$ then add the following contributions to $A^*$:

<table>
<thead>
<tr>
<th>Contribution</th>
<th>to element(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>(i, i)</td>
</tr>
<tr>
<td>$.5x$</td>
<td>(i, j), (j, i), (i, k), (k, i)</td>
</tr>
<tr>
<td>$.25x$</td>
<td>(j, j), (j, k), (k, j), (k, k)</td>
</tr>
</tbody>
</table>
Often, $A^*$ will be formed half-stored, i.e., only contributions on or above the diagonal will be accumulated. In this case it is imperative to check whether $j = k$. This will occur only if both sire and dam are unknown and from the same base group. In the case when $j = k$ and $x = 1$, the rules reduce to adding 1 to $(i, i)$ and $(j, j)$ and $-1$ to $(i, j)$ and $(j, i)$.

**Example 2. Pedigree of animals discussed in text by Quaas (1988)**

Upper case letter are known ancestors and lower case letters are unknown ancestors.

\[
\begin{array}{cccc}
  & a & b & \\
  A & c & d & e \\
  B & & & \\
  C & & & \\
  D & & & \\
\end{array}
\]

Assumptions

Base animals will be assumed to each have a single progeny represented in $u$, and $\text{var } u_b = \sigma_u^2 I$. The latter assumption might hold, e.g., if each base animal is a random selection from one of several large random mating base populations with common additive genetic variance $\sigma_u^2$. Let $g_i$ represent the mean of the $j^{th}$ base population, and take $E(u_b) = Q_b g$ where $Q_b$ is an incidence matrix relating base animals to their respective base population means. Suppose in the example that $a$ and $d$ come from one group while $b$, $c$, and $e$ came from a second, then

\[
Q_b = \begin{bmatrix}
g_1 & g_2 \\
a & 1 & 0 \\
b & 0 & 1 \\
c & 0 & 1 \\
d & 1 & 0 \\
e & 0 & 1 \\
\end{bmatrix}
\]

Application of the rules for computing $A^*$ using the pedigree from Example 2 gives the following values
Sire Models

Application of mixed models that include groups for unknown ancestors is best described and understood in the context of mixed model equations developed by using the Q-P transformation (Quaas and Pollak, 1981; JDS 64:1868). Each sire’s evaluation is a direct solution to the system of modified equations and is different from regular MME’s where the evaluations are linear functions of group plus sire solutions.

The sire model in matrix notation is

\[ Y = Xh + ZQg + Zs + e \]

where \( h \) is a vector of fixed herd-year-season effects, \( g \) is a fixed group effect, \( s \) is the random vector of sire effects, \( X \) is an incidence matrix relating herd-year-seasons to records, \( Z \) is an incidence matrix relating sires to daughter records, \( ZQ \) is an incidence matrix relating group of sire to daughter’s records and \( e \) is a vector of random residuals.

The mixed model equations are

\[
\begin{bmatrix}
\alpha Q A^{-1} Q' & -\alpha Q' A^{-1} \\
-\alpha A^{-1} Q & Z'MZ + \alpha A^{-1}
\end{bmatrix}
\begin{bmatrix}
g \\
u
\end{bmatrix}
= \begin{bmatrix}
0 \\
Z'My
\end{bmatrix}
\]

Where \( M = I - X(X'X)^{-1}X' \) and accounts for the fact that herd-year-seasons have been absorbed. Note that coefficients of the sire equations, \( Z'MZ \) and the right-hand-sides, \( Z'My \), are identically equivalent to the model with no groups.

Animals are often allocated to genetic groups to account for selection that can not be accounted for by known genetic relationships. Consider the case of a sire model that uses the
inverse of relationships among sires and maternal grandsires of all bulls with progeny in the data. Dams are not included and in this case the model assumes that only random portions of sire and MGS effects are inherited, which is not biologically reasonable.

A sire model that includes genetic groups for all unknown ancestors in each bull’s pedigree can be written as

\[ y_{ijkl} = h_{ij} + \sum q_{kr} g_r + s_k + e_{ijkl} \]

where \( y_{ijkl} \) is the observation, \( h_{ij} \) is the herd-year-season effect, \( q_{kr} \) relates sire \( k \) to an unknown ancestor group \( r \) (\( r = 1, \ldots, n \)), \( g_r \) is the effect of the unknown ancestor group \( r \), \( s_k \) is the effect of sire \( k \) and \( e_{ijkl} \) is random residual. For each sire, \( \sum q_{kr} = 1 \) as in Westell et al. (1988) because all ancestors paths eventually trace back to unknown ancestors. Note that this model differs from a model where sires are nested within groups. Here the total genetic group effect for sire \( k \) is

\[ \sum q_{kr} g_r \]

which is a weighted average of several group effects. The weights are determined by the relationship of unknown ancestors to sire \( k \) and the evaluation is

\[ \hat{u} = \sum q_{kr} \hat{g}_r + \hat{s}_k \]

**Computing strategy for \( A^* \) (Van Raden et al., 1990)**

(Sire models using \( A^{-1} \) based on sires and maternal grandsires)

Coefficients for the following 4x4 matrix are added into either \( A^{-1} \) or into appropriate group equations for each bull evaluated:

\[
\begin{bmatrix}
1 & -0.5 & -0.25 & -0.25 \\
-0.5 & 0.25 & 0.125 & 0.125 / p \\
-0.25 & 0.125 & 0.0625 & 0.0625 \\
-0.25 & 0.125 & 0.0625 & 0.0625 \\
\end{bmatrix}
\]

where MGS is maternal grandsire of the bull, MGD refers to a group of unknown maternal granddams, and \( p \) is the proportion of the bull’s genetic variance not attributable to known ancestors. Values of \( p \) are 11/16 if sire and MGS are both known, 3/4 if only the sire is known,
15/16 if only MGS is known, and 1 if neither is known. Coefficients of sire or MGS are assigned to
group equations if sire or MGS are unknown, and MGD coefficients always are assigned to group
equations. Each bull’s pedigree information is then ½ of the sire effect (including fixed and
random portions) plus ¼ of MGS effect (including fixed and random portions) plus ¼ of fixed
MGD group effect.

References

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338-
1345.

64:1868-1872.

Robinson, G. K. 1986. Group effects and computing strategies for models for estimating breeding
values. J. Dairy Sci. 69:3106.


abilities for Holstein type traits. J. Dairy Sci. 73:191-197.

Federation, Nashville, TN May 11-14 (mimeo)

68:3998-4013.

Westell, R. A. 1984. Simultaneous evaluation of sires and cows for a large population. Ph.D.


large population of dairy cattle. J. Dairy Sci. 70:1006.