An equivalence between models of restricted selection and genetic groups

T. R. Famula

Department of Animal Science, University of California, Davis, CA 95616, USA

Summary. An equivalence between a model of restricted selection and a model of genetic groups is presented. This correspondence leads to a realization of how genetic groups account for selection. Specifically, genetic groups act to remove the covariance between predictions of sire merit and functions of the true selection differentials. Further results illustrate a correspondence between models of selection on random effects and models of selection on residuals. Application of the results is useful, not in establishing concrete definitions for the structure of genetic groups, but in the analysis of how groups account for selection.

Introduction

In the genetic evaluation of dairy sires, the concept of genetic groups is recognized as the "weak link" in the sire evaluation process.

Genetic groups have been the focus of several recent research reports (Jensen 1980; Kennedy 1981; Pollak and Quaas 1983; Quaas and Pollak 1981). Genetic groups were introduced by Henderson (1949) in a linear model for predicting the real producing ability of cows from repeated milk records. Later, they were incorporated into sire evaluation models with the idea of accounting for an assumed genetic trend in sire breeding values (Henderson 1973). In general, groups are intended to represent populations of sires for which the mean breeding value may differ from population to population (Famula et al. 1983). Moreover, the results of Quaas and Pollak (1981) and Famula et al. (1983) indicate that the role of genetic groups is intimately linked with artificial selection. Pollak and Quaas (1983) further establish this connection in finding definitions of genetic group effects based on selection

differentials. The conclusion to be drawn from these papers is that genetic groups are a necessary, albeit arbitrarily defined, component of linear models for sire evaluation when the data available are incomplete or have has been subject to non-random sampling (selection).

Questions remain as to how genetic groups account for selection. Famula et al. (1983) have shown that if genetic groups are structured analogous to the selection practiced the predictors are best linear unbiased predictors under Henderson's (1975) selection model. Pollak and Quaas (1983) have shown genetic groups to be accumulations of selection differentials which was developed from an extension of Thompson's (1979) accumulated groups model. These point explain what genetic groups do without really explaining how groups accomplish this task. Consequently, we may still ask how genetic groups account for this selection. Related to this is the concern that predictions of genetic merit computed with and comprised of genetic groups may not be independent of the selection practiced. It may be possible that our current selection decisions are affected by the prior selection practiced within the population. Simply put, we must ask if the true selection differentials, estimated by the genetic groups, are correlated with our estimates of sire breeding value. The potential problem is, even if genetic groups are part of the model and correctly defined, if the true selection differentials (or selection criterion under a selection model) are correlated with our estimates of sire breeding value or deviations of sires within groups, we may be incorrectly ranking sires across or within groups. For example, if the covariance between true selection differentials and estimates of genetic merit were negative, sires from inferior genetic groups may be chosen at the expense of sires with a superior true genetic value. The purpose of this note is to clarify the question as to how genetic groups account for selection. In addition, we can examine whether estimates of breeding value are independent of the group effect or selection differential.

To answer these questions a correspondence between a sire model with genetic groups and a model of restricted best linear unbiased prediction (RBLUP) will be established. Restricted selection, when first and second moments of all distributions are known, was first presented by Kempthorne and Nordskog (1959).

Henderson (1972) appears to be the first to have extended restricted selection to the case of unknown first moments. The results of this work were later presented by Quaas and Henderson (1976) as an abstract. Recall that the goal of restricted selection is to maximize response to selection in traits of interest while "restricting" response in other correlated traits to zero. Accordingly, the objective of this note is first, to establish a correspondence between the genetic groups model and the model of restricted best linear unbiased prediction. This result then can be used to explain how selection on sire proofs (using best linear unbiased prediction) is independent of the selection differentials. This explanation will serve to illustrate how genetic groups account for selection and can allay concerns of incorrect ranking if genetic groups are used.

Statistical preliminaries

The objective of this note is to compare similarities in two linear models used in the genetic evaluation of candidates for selection. Each was developed for a different purpose and, until now, each was thought to be independent of the other. One model is based on the restricted selection index, first introduced by Kempthorne and Nordskog (1959). Whereas this initial work assumed that the first moments of the distribution of observations were known, we shall examine a more general situation, first introduced by Henderson (1972), without this assumption. This is the model of restricted best linear unbiased prediction (RBLUP). The other model is now routinely used, in various forms, for the genetic evaluation of dairy sires. This model includes genetic groups with the original intent to correct for the genetic trend usually found in field collected data.

Let us assume that a vector of observations is available which conforms to the following general mixed linear model:

$$y = Xb + Zu + e \qquad (1)$$

where

y is a vector of observations

$X (Z)$ is an incidence matrix relating fixed (random) effects to observations

$b (u)$ is a vector of unknown fixed (random) effects and

e is a random vector of residuals.

Moreover, we assume in the usual model (i.e., no selection, no restrictions) that

$$E[y] = Xb$$

$$E \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and

$$V \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

$$V[y] = ZGZ' + R.$$

Our objective is to predict $k'b + m'u$ (for some estimable function k' and any m') with a linear function of the observations say $c'y$ so as to minimize the variance of prediction errors. Henderson (1973) presented a solution to this problem by computing $k'\hat{b} + m'\hat{u}$ where $\hat{b}$ and $\hat{u}$ are solutions to the following simultaneous equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \qquad (2)$$

To develop predictors under a model of restricted selection requires some modification of (2), as was presented by Henderson (1972). In keeping with Kempthorne and Nordskog's (1959) original derivation, the phrase "restricted selection" implies that we still wish to rank candidates for selection on the basis of $k'b + m'u$ but with the additional proviso that some other linear function of the random effects, say $T'u$, is uncorrelated with our linear predictor $c'y$. Algebraically, this constraint is equivalent to setting $\text{Cov}(c'y, T'u) = 0$ so that selection on our predictor permits no correlated response in $T'u$. Accordingly we now wish to minimize the variance of prediction errors subject to the additional constraint that $c'ZGT = 0$. To do so we arrive at an extension of (2) which can be written as:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}ZGT \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}ZGT \\ T'GZ'R^{-1}X & T'GZ'R^{-1}Z & T'GZ'R^{-1}ZGT \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \\ \hat{q} \end{bmatrix}$$

$$= \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ T'GZ'R^{-1}y \end{bmatrix} \qquad (3)$$

where q is analogous to a vector of LaGrange multipliers. Before examining equations (3) in depth, at this time we will note only that they correspond to the equations that would be set up under an expansion of model (1) of the form

$$y = Xb + Zu + ZGTq + e \qquad (4)$$

(as was also mentioned by Henderson 1972), and that $\text{Cov}(\hat{u}, T'u) = 0$.

The model of genetic groups is a simple extension of model (1) wherein the assumption that $E(u) = 0$ is dropped and some mean structure is imposed. Specifically, we assume that $u = Qg + s$ for Q a known incidence matrix relating individuals to groups, g an unknown vector of fixed group effects and s an unobservable

random vector of individual's deviations from their group mean, s has null mean and covariance matrix **G** as defined above. Substituting this structure for **u** into model (1) we obtain the genetic groups model

$$y = Xb + ZQg + Zs + e \qquad (5)$$

which we notice has some specific similarities to model (4). When model (5) is used for dairy sire evaluation, predictions of transmitting ability are taken as $\hat{u} = Q\hat{g} + \hat{s}$ where $\hat{g}$ and $\hat{s}$ are computed from

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}ZQ & X'R^{-1}Z \\ Q'Z'R^{-1}X & Q'Z'R^{-1}ZQ & Q'Z'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}ZQ & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{s} \end{bmatrix}$$

$$= \begin{bmatrix} X'R^{-1}y \\ Q'Z'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \qquad (6)$$

It is this model of genetic groups which Quaas and Pollak (1981) found equivalent to Henderson's (1975) selection model and for which they developed the modified mixed model equations to predict **u** directly. Moreover, this model formed the basis of Pollak and Quaas' (1983) definition of group effects,

$$g = (Q'G^{-1}Q)^{-1}Q'G^{-1}s. \qquad (7)$$

This definition can then be explained, depending upon the form of Q and G, as functions of selection differentials.

Relevant to this note, however, is the equivalence between equations (3) and (6) and between models (4) and (5). What we find is the unexpected correspondence between the model of genetic groups and a model of restricted selection. That is, for $T = G^{-1}Q$ the equations yield identical results, which implies that properties derived specific to the restricted selection model also are appropriate for the genetic groups model. The implications of this correspondence warrant investigation.

Discussion

Although discovery of the correspondence between a model of restricted selection and a model of genetic groups is certainly interesting from a pedagogic perspective, however it is the implications of this result that merit discussion. Specifically we must ask what this result tells us about both models that has not been previously observed or is not implicit in their derivation. To do so requires an examination of the principles underlying each model and drawing conclusions from the areas in which they overlap.

To begin, consider the results of the previous section. Under the correspondence of models (4) and (5), the matrix of restrictions (**T**) in the restricted selection model (4) is equivalent to $G^{-1}Q$ of the genetic groups model (5) (i.e. $T = G^{-1}Q$). Thus, a model with genetic groups is also a model of restricted selection with the restriction that Cov $(\hat{s}, Q'G^{-1}s) = 0$. Furthermore, given results derived under the restricted model (and applying them to the genetic groups model) Cov $(k'\hat{b}, Q'G^{-1}s) = 0$ for estimable k'. The equivalent result for the restricted selection model is presented in Henderson (1972). The proof relies on the restriction that $c'ZGT = 0$ and that $k'\hat{b} = c'y - m'\hat{u}$. The proof then follows to determine that Cov $(m'\hat{u}, T'u) = 0$ (a result also indicated in Henderson 1972). The proof is lengthy and tedious but the interested reader is welcome to obtain a copy from the author. Without discovery of this correspondence, the fact that these covariance terms vanish in the groups model has not been previously documented. Rather the assumption was that such covariances were functions of the generalized inverse of the coefficient matrix. We now see that under the correspondence of restricted selection and genetic groups models that these covariance terms must vanish. Recall for the restricted selection model that Henderson (1972) showed that $T'u$ of (3) is null. Extending this to the genetic groups model we derive the familiar result that $Q'G^{-1}s = 0$.

The interpretation of these results and their impact on genetic evaluation is straightforward. Given the null covariances outlined above, it is apparent that predictions of sire deviations from their group mean, $\hat{s}$, are not correlated with the parameter $Q'G^{-1}s$. Depending upon one's perspective $Q'G^{-1}s$ can take one of two forms. According to Pollak and Quaas (1983), $Q'G^{-1}s$ is a function of the true selection differentials and is a component of the true genetic group effects. Alternately, when considering the equivalence between genetic groups model and a model of Henderson's $L'u$ selection, $Q'G^{-1}s$ is the conditional variable upon which selection is assumed to be based. For our purposes the former interpretation permits us to examine the covariance between predictions of sire merit and the true selection differentials. Obviously such a covariance does not exit. Thus the ranking of sires within genetic groups is entirely unaffected by the selection practiced. Thus, considering the genetic groups model while applying results derived in the case of restricted selection shows that predicted deviations from the group mean are not correlated with the true group effects (which are functions of the selection differentials). This independence of predictor and true value does not hold for the complete prediction of genetic merit. The complete prediction of the genetic value of an individual is the group effect plus the

individual deviation, i.e. $\hat{u} = Q\hat{g} + \hat{s}$ from equations (6). Therefore Cov $(\hat{u}, g) \neq 0$ because Cov $(Q\hat{g}, g) \neq 0$. We learn from this result that the covariance between estimates of group effects and the true value is non-null.

The interpretations outlined above obviously make one critical assumption, that is g is a random variable and not a fixed effect. Quaas and Pollak (1981) raised this question previously. Though treated in the model and equations as a fixed effect, g is a function of selection differentials. These, in turn, are functions of random variables. In a theoretical sense it is preferable to regard g as a random variable, particularly for the purpose of examining the effects of selection on models of genetic evaluation. This line of thought grows directly from the correspondence of the genetic groups model with Henderson's (1975) selection model where g is a function of the change in mean of the conditional variable after selection. Consideration of g as a fixed effect obviously leads to null covariances.

Discovery of this correspondence does not change how we define genetic groups. However, the equivalence of models (4) and (5) can aid in the interpretation of how genetic groups account for selection. For example, Famula et al. (1983) suggest that groups be structured analogous to the selection practiced. This suggestion was based on the equivalence of the genetic groups model to Henderson's (1975) selection model, first discovered by Quaas and Pollak (1981). The objective is to choose an appropriate form for Q such that groups are linearly related to the selection practiced so that best linear unbiased predictors of u are obtained. This explained what to do to account for selection without describing how the correction for the action of selection is accomplished. With the results presented here, we see that how groups account for selection is through "breaking" the correlation between what we select on ($\hat{s}$) and the effects of selection — the selection differentials generated by $Q'G^{-1}s$. No longer is the selection model necessary to justify the use of genetic groups. Instead we can justify the necessity of genetic group as a means to correct our predictions of genetic merit so that our selection criterion are not correlated with the changes in mean imposed by selection.

The correspondence of models (4) and (5) also sheds light on the interrelationships among models of genetic groups, restricted selection and Henderson's (1975) selection model. Establishing these equivalences began with Quaas and Pollak (1981) and their discovery of the similarity between models of genetic groups and Henderson's (1975) example of $L'u$ selection. Specifically, their development of modified mixed model equations showed that the groups model was equivalent to $Q'G^{-1}u$ selection (where prior to selection, $E(u) = 0$). Next, Famula (1984) established a link between restricted best linear unbiased prediction (models of restricted selection) and Henderson's (1975) example of $L'e$ selection. In this case, the form of $L'e$ for this correspondence is $T'GZR^{-1}e$. Thus, in a general fashion we conclude that models of genetic groups are equivalent to $L'u$ selection whereas models of restricted selection are equivalent to $L'e$ selection. In this note we have established an equivalence between models of genetic groups and models of restricted selection. The conclusion to be drawn from this syllogistic argument is that each model (i.e., Hen-

derson's selection model, genetic groups and restricted selection) is closely related to the other.

Conclusions

The results presented in this note permit us to draw several general conclusions. First we have seen established the rather unexpected correspondence between the usual model of genetic groups and a general model of restricted selection. This discovery has permitted us to determine that genetic groups account for selection by removing the covariance between predictors of genetic merit and the changes in population mean brought about by selection. As a result, selection decisions made in the present are independent of selection decisions made in the past. We have also found that Henderson's (1975) examples of $L'u$ and $L'e$ selections are equivalent for particular forms of L'.

Finally, we address the practical implications of these results and note that they offer little to change the use of genetic groups. The value of these results is not to be found in application. Instead the results offer further support for the use of genetic groups which are often viewed as the "weak link" of the sire evaluation process.

References

Famula TR (1984) Restricted best linear unbiased prediction and a selection model. Theor Appl Genet 67:337–340

Famula TR, Pollak EJ, Van Vleck LD (1983) Genetic groups in dairy sire evaluation under a selection model. J Dairy Sci 66:927–934

Henderson CR (1949) Estimation of changes in herd environment. J Dairy Sci (Abstract) 32:706

Henderson CR (1972) Restricted selection index. In: Prediction. Mimeo, Dept of Animal Sci, Cornell University, Ithaca, NY

Henderson CR (1973) Sire evaluation and genetic trends. In: Proc Anim Breeding Symp in Honor of Dr JL Lush. Am Soc Anim Sci, Am Dairy Sci Assoc, Blacksburg VA, pp 10–41

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

Jensen EL (1980) Bull groups and relationships among sires in best linear unbiased prediction sire evaluation models. J Dairy Sci 63:2111–2120

Kennedy BW (1981) Bias and mean square error in mixed model sire evaluation when genetic group effects are ignored or included. J Dairy Sci 64:689–697

Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19

Pollak EJ, Quaas RL (1983) Definition of group effects in sire evaluation models. J Dairy Sci 66:1503–1509

Quaas RL, Henderson CR (1976) Selection criteria for altering the growth curve. J Anim Sci (Abstract) 43:221

Quaas RL, Pollak EJ (1981) Modified equations for sire models with groups. J Dairy Sci 64:1868–1872

Thompson R (1979) Sire evaluation. Biometrics 35:339–353