Theor Appl Genet (1984) 67:337–340

Restricted best linear unbiased prediction and a selection model

T. R. Famula

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

Received July 6, 1983
Communicated by D. Van Vleck

Summary. An equivalence between restricted best linear unbiased prediction (and thus restricted selection index) and a particular example of a selection model is presented. Specifically, the equivalence is between restricted selection and a model of selection on the residuals of the general mixed linear model. This result illustrates that restricted selection acts by nonrandomly sampling those genes that act pleiotropically in multiple trait genetic models. An expression for a mixed linear model which includes restrictions is also presented.

Key words: Best linear unbiased prediction – Restricted selection index – Selection models

Introduction

Interest in restricted selection indices began with the original derivation by Kempthorne and Nordskog (1959). Although their work did not receive the same degree of attention as more "traditional" methods of genetic evaluation, recent papers on restricted methods have been presented (Harville and Reeves 1972; Dickerson et al. 1974; Harville 1975; Niebel and Van Vleck 1982). In a parallel manner, the first detailed description of best linear unbiased prediction (and best linear prediction) under a selection model was by Henderson (1975). Interest in further developing Henderson's selection model has only recent begun; principally in the work of Quaas and Pollak (1981) and Famula et al. (1983).

The present note will examine that material common to both restricted selection indices and Henderson's selection model. Specifically, a correspondence between restricted best linear unbiased prediction and a particular example of Henderson's selection model will be presented. The implications of this result will lead toward a more complete understanding of both restricted selection and Henderson's example.

Restricted best linear unbiased prediction

Throughout this note we will assume that a vector of observations is available that adheres to the following general mixed linear model

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{b} + \underset{\sim}{Z}\,\underset{\sim}{u} + \underset{\sim}{e} \tag{1}$$

where

$\underset{\sim}{y}$ is a vector of observations
$\underset{\sim}{X}(\underset{\sim}{Z})$ is an incidence matrix relating fixed (random) effects to observations
$\underset{\sim}{b}(\underset{\sim}{u})$ is a vector of fixed (random) effects and
$\underset{\sim}{e}$ is a random vector of residuals.

In addition it will be assumed, in the no selection model, that

$$E[\underset{\sim}{y}] = \underset{\sim}{X}\,\underset{\sim}{b} \qquad E\begin{bmatrix}\underset{\sim}{u}\\ \underset{\sim}{e}\end{bmatrix} = \begin{bmatrix}\underset{\sim}{0}\\ \underset{\sim}{0}\end{bmatrix}$$

and

$$V\begin{bmatrix}\underset{\sim}{u}\\ \underset{\sim}{e}\end{bmatrix} = \begin{bmatrix}\underset{\sim}{G} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{R}\end{bmatrix} \qquad V[\underset{\sim}{y}] = \underset{\sim}{Z}\,\underset{\sim}{G}\,\underset{\sim}{Z}' + \underset{\sim}{R}\,.$$

Traditionally, the objective of animal geneticists has been to choose among the candidates for selection based on a linear function of the observation vector. The goal is to predict $\underset{\sim}{k}'\underset{\sim}{b} + \underset{\sim}{m}'\underset{\sim}{u}$ (for some estimable function k' and any m') with a linear function of y (say, $c'y$) so as to minimize the variance of prediction error. Henderson (1963) showed that this could be accomplished by the function $k'\hat{b} + m'\hat{u}$ where $\hat{b}$ and $\hat{u}$ are solutions to

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

In addition, $k'\hat{b}$ is the best linear unbiased estimate (BLUE) of $k'b$ and $\hat{u}$ is the best linear unbiased prediction (BLUP) of u.

Extending these principle to the topic of restricted selection index – and restricted best linear unbiased prediction (RBLUP) – requires a slight modification of [2], as was shown by Henderson (1972). The phrase "restricted selection index" implies that we still wish to select candidates on the basis of the value of $k'b + m'u$ but in addition we wish some linear function of the random effects, say $T'u$, to be uncorrelated with our linear predictor $c'y$. This problem was first solved by Kempthorne and Nordskog (1959) for the case where b is known. To my knowledge, no derivation of the form of the restricted linear predictor for the case where b remains unknown (and must be estimated) was made available until Henderson's (1972) notes on prediction methods.

In deriving RBLUP of u, note that $\mathrm{Cov}(c'y, T'u) = c'ZGT$. Accordingly, we minimize $V(c'y-k'b-m'u)$ subject to $c'ZGT = 0$ and $E[c'y] = c'Xb = k'b$. In this way our predictor of u will be uncorrelated with $T'u$ as set out in the derivation of restricted selection indices (Kempthorne and Nordskog 1959). Completing the derivation subject to these restriction, Henderson (1972) arrived at the following set of equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}ZGT \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} & Z'R^{-1}ZGT \\ T'GZ'R^{-1}X & T'GZ'R^{-1}Z & T'GZ'R^{-1}ZGT \end{bmatrix} \cdot \begin{bmatrix} \hat{b} \\ \hat{u} \\ \hat{q} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ T'GZ'R^{-1}y \end{bmatrix} \qquad (3)$$

where $\hat{q}$ is analogous to a vector of Lagrange multipliers. We will return to equations (3) and at this time note only that one can show that $T'\hat{u}$ is null (Henderson 1972).

Briefly, we now consider Henderson's (1975) derivation of BLUE and BLUP under a selection model. Here, Henderson (1975) invokes an additional random vector variable w (which is jointly normally distributed with y, u, and e) upon which selection is practiced. The selection, so it is assumed, alters the means and variances of the variables y, u and e. In the final section of the paper, three possible forms of w are presented of which we will consider only $w = L'e$, for some matrix of full row rank L'. Henderson's intention was that this linear function of the residuals did not conform to the expected first and second moments of such a linear function of the errors. Under $L'e$ selection the following mixed model equations were presented

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'L \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} & Z'L \\ L'X & L'Z & L'RL \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ L'y \end{bmatrix} \qquad (4)$$

Comparing equations (4) with equations (3) one notes that the two sets of equations are equivalent for $L' = T'GZ'R^{-1}$. As a result, a correspondence between a model of restricted best linear unbiased prediction and a model of $L'e$ selection is established. More important than the actual correspondence, however, are the implications of such a result. Examination of this equivalence should shed light on both restricted selection and Henderson's (1975) application of selection on residuals.

Discussion

As we examine the equivalence between a model of restricted selection and an $L'e$ selection model, our first task is to establish what $L'e$ selection implies. Because e is a non-observable random variable, it is not possible to make actual selection decisions based on residuals. Instead we must hypothesize as to what genetic and environmental terms are represented in the residual and postulate possible mechanisms for the nonrandom sampling of these components. For example, in his illustration of $L'e$ selection in a sire model, Henderson (1975) considered differential selection of sire mates. Because the dam contribution in standard sire models is usually partitioned to the residual, suspicion of nonrandom selection of mates of sires can be grounds for the consideration of an $L'e$ selection model. Through the equivalence with restricted selection, we can study other examples of $L'e$ selection. In so doing, however, we must establish what other terms may be represented in the residual under various genetic models.

Quaas and Pollak (1980) presented a gametic model for the simultaneous evaluation of breeding values for weaning weight (direct and maternal) and yearling weight in beef cattle. Their development of appropriate linear models begins with an animal model and eventually develops into a equaivalent gametic model that includes terms for sire effects, dam effects and a component for the Mendelian sampling in the formation of the calf's genotype. As the model is further developed some of the genetic terms (dam or Mendelian sampling) can be partitioned into the residual, illustrating that the "error" term can contain effects which are not strictly environmental. Considering the development of genetic models in this fashion, we can hypothesize as to possible selection mechanisms that lead to a nonrandom sampling of the residuals and thus lead to $\underset{\sim}{L}'\underset{\sim}{e}$ selection.

In particular, as we examine the form of restricted selection (and it's equivalence to $\underset{\sim}{L}'\underset{\sim}{e}$ selection) we can extend the gametic model of Quaas and Pollak (1980) to multiple trait models and the inclusion of pleiotropic gene effects. Thus, a simple gametic genetic model could be developed for milk production and fat test in dairy cows and be represented as

$$\begin{aligned} y_M &= \mu_M + \tfrac{1}{2} u_M^S + \tfrac{1}{2} u_M^D + c_M + e_M \\ y_F &= \mu_F + \tfrac{1}{2} u_F^S + \tfrac{1}{2} u_F^D + c_F + e_F \end{aligned} \tag{5}$$

where y_M (y_F) is the milk (fat test) record, μ_M (μ_F) is a linear function of fixed effects affecting the milk (fat test) record, u_M^S is the breeding value of the sire (dam) and c_M represents the Mendelian sampling of sire and dam genotypes in the formation of the progeny genotype. e_M (e_F) represents all other environmental effects influencing the milk (fat test) record. One can extend this simple representation for the random elements of equation (5) by partitioning each genetic component into some fraction of effect that behaves pleiotropically to the other trait and that part which affects performance in one trait only. The environmental term can be similarly partitioned into some component that affects both traits and some remaining effet that acts only on the specific trait of the model. This partitioning will not influence the expectations or variances of the traits (or models), but does serve to identify specifically that portion of the genetic and environmental effects that are responsible for the existence of genetic and environmental covariances between traits. It is this component, the pleiotropic gene effects and the correlated environmental effects, that are nonrandomly sampled under restricted selection.

Often, in the practice of artificial selection we are interested in the increased genetic potential of one trait only. Yet, due to the genetic correlations between traits there are correlated responses to selection in characters that are not necessarily part of the selection criterion. In a sense, the correlated response is a function of the sampling of pleiotropic effects. If, for example, there is a positive genetic correlation between two traits, selecting for performance in trait 1 will, on the average, pull along positive elements in trait 2 through the pleiotropic gene effects. This can be throught of as a function of random sampling. As positive genes for trait 1 are selected, positive genes for trait 2 will be carried along at random (assuming a positive genetic correlation between traits). Under restricted selection the idea of random samples of pleiotropic gene effects must by abandoned as the equivalence between equations (3) and (4) indicates.

The goal of restricted selection is to "break" the genetic correlation between traits. In effect, we want to choose superior genes for trait 1 that have little or no effect on trait 2 so as to eliminate the correlated response in trait 2. This leads to a nonrandom sampling of pleiotropic genes such that we can statistically remove the genetic correlation. It is this nonrandom sampling of pleiotropic gene effects that leads to the equivalence between restricted selection and Henderson's (1975) $\underset{\sim}{L}'\underset{\sim}{e}$ selection model. Depending upon what terms are included in the residual, there are several places where the nonrandom sampling of pleiotropic genes can take place. Thus, on one level we are interested just in the existence of an equivalence between restricted selection and $\underset{\sim}{L}'\underset{\sim}{e}$ selection. Yet our curiousity is similarly directed to how restrictions change the genetic and statistical models used in the evaluation of candidates for selection, as well as the implications of this correspondence.

One decides to use restricted selection when the correlated response in some performance characters inhibit the overall genetic and economic progress. As equations (3) and (4) illustrate the choice of a restriction ($\underset{\sim}{T}$) defines a specific form of $\underset{\sim}{L}'\underset{\sim}{e}$ selection (i.e., $\underset{\sim}{L}' = \underset{\sim}{T}'\underset{\sim}{G}\underset{\sim}{Z}'\underset{\sim}{R}^{-1}$). The implication is that by choosing a restriction on our predictor of $\underset{\sim}{u}$ we are imposing specific relationships among the residuals where the expected value of this linear function of the residuals is not zero (as under the "usual" prediction model). The linear functions of the residuals involve both genetic and environmental variances and covariances between the observed traits. Note also, that for a two trait selection model, the linear functions involve residuals of both traits not just that trait under selection or restriction. This can be shown by expanding $\underset{\sim}{G}\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{e}$, under an extension provided by the multiple trait model of Henderson and Quaas (1976), and premultiplying this expression by the desired restrictions, $\underset{\sim}{T}'$. Through the equivalence to a selection model, we consider that the imposed linear functions of the residuals imply a nonrandom sampling of some terms contained in the residuals. In effect, by using the restrictions in our selection decisions we are intending to sample specific pleiotropic gene effects that avoid the natural correlation between the observed traits. As another example, if we were to select for increased weaning weight in beef calves, we would also find a correlated response in increased birth weight, as a result of the positive genetic correlation between birth and weaning weights. A desire to restrict the response in birth weight to zero implies that we must non-randomly sample gene effects where the pleiotropic genes are high for weaning weight, but will not contribute substantially to birth weight. This interpretation of restricted selection, though perhaps not new, is made possible through the equivalence of the two models presented.

Some additional observations are also possible under this equivalence. We note that the use of restrictions in the prediction of elements of $\underset{\sim}{u}$ does not bias the predictor of $\underset{\sim}{u}$. This follows from the expectations of $\hat{\underset{\sim}{u}}$ under the illustrations provided by Henderson (1975). However, the restrictions do serve to change the variance of prediction errors. In addition, we can now develop a form for the general mixed linear model, after (1), when restrictions are included. Derivation of this model follows from a result of Thompson (1979) and the equivalence described above. Specifically, we note that the linear model which generates equation (3) is

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{Z}\,\underset{\sim}{u} + \underset{\sim}{C}\,[\underset{\sim}{C}'\,R^{-1}\,\underset{\sim}{C}]^{-1}\,\underset{\sim}{C}'\,\underset{\sim}{e} + \underset{\sim}{e} \tag{6}$$

where $\underset{\sim}{C} = \underset{\sim}{Z}\underset{\sim}{G}\underset{\sim}{T}$ (Thompson 1979). It is important to note the $E[\underset{\sim}{C}'\,\underset{\sim}{e}] \neq \underset{\sim}{0}$, as a result of the restrictions. The arguments of the preceding paragraphs explain why.

References

Dickerson GE, Kunzi N, Cundiff LV, Koch RM, Arthaud VH, Gregory KE (1974) Selection criteria for efficient beef production. J Anim Sci 39:659–673

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

Harville DA (1975) Index selection with proportionality constraints. Biometrics 31:223–225

Harville DA, Reeves TE (1972) Optimal linear indexes for some selection problems. In: Tech Rep No 72-0123, Aerospace Research Laboratories, Wright-Patterson AFB, Ohio

Henderson CR (1963) Selection index and expected genetic advance. NAS-NRC 982, pp 144–163

Henderson CR (1972) Restricted selection index. In: Prediction. Mimeo, Dept of Animal Science, Cornell University, Ithaca NY, p 43

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

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

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

Niebel E, Van Vleck LD (1982) Selection with restriction in cattle. J Anim Sci 55:439–457

Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing programs. J Anim Sci 51:1277–1287

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

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