

2-3-1987

# ASPECTS OF SELECTION FOR PERFORMANCE IN SEVERAL ENVIRONMENTS WITH HETEROGENEOUS VARIANCES

D. J. Garrick

*Cornell University*, [dorian@iastate.edu](mailto:dorian@iastate.edu)

L. Dale Van Vleck

*University of Nebraska-Lincoln*, [dvan-vleck1@unl.edu](mailto:dvan-vleck1@unl.edu)

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

---

Garrick, D. J. and Van Vleck, L. Dale, "ASPECTS OF SELECTION FOR PERFORMANCE IN SEVERAL ENVIRONMENTS WITH HETEROGENEOUS VARIANCES" (1987). *Faculty Papers and Publications in Animal Science*. 344.

<https://digitalcommons.unl.edu/animalscifacpub/344>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# ASPECTS OF SELECTION FOR PERFORMANCE IN SEVERAL ENVIRONMENTS WITH HETEROGENEOUS VARIANCES

D. J. Garrick and L. D. Van Vleck<sup>1</sup>

Cornell University  
Ithaca, New York 14853

## ABSTRACT

Dairy cattle evaluation schemes routinely assume homogeneous variance with respect to environment. Increasing evidence suggests the presence of systematic changes in variance components associated with mean level of performance. Best linear unbiased prediction procedures that account for heterogeneity are reviewed. The consequences of incorrectly assuming homogeneity for evaluation are demonstrated for a progeny test and an artificial breeding program that screens dams of sires from heterogeneous populations. Selection assuming homogeneity can be very efficient when heritability, and therefore accuracy of selection, is greatest in the more variable environment. Conversely, appreciable reduction in response results when heritability is greater in the less variable environment.

(Key Words: Heterogeneity, Best Linear Unbiased Predictor, Genotype Environment Interaction, Selection, Simulation, Genetic Gain.)

## Introduction

The possibility of encountering variance heterogeneity with respect to environment is not a new concept for animal breeders. Lush (1945) recommended that animals be kept in the environments in which they will be used so that desirable genes have a chance to express their effects. Hammond (1947) concluded that selection should be practiced in the most favorable environment to improve accuracy of selection due to greater expression of genes of interest. Falconer (1952) introduced the concept of a genetic correlation between performance in different environments and used the ratio of indirect to direct response to selection to determine the optimum environment for selection. Robertson et al. (1960) identified the need to recognize whether heritability differed between the environments and whether the ranking of dairy bulls was affected.

Variance component estimation using dairy cattle data has included investigations of homogeneity of variance among herd-year-seasons. Although the number of records in a herd-year-season is generally insufficient for reliable within-herd estimation of variance

components, similar herds have been grouped together and components obtained separately for each grouping (e.g., Hill et al., 1983; Mirande and Van Vleck, 1985). These and other studies provide considerable evidence for heterogeneous variances among environments. However, these studies did not find the genetic correlation between performance of dairy cattle in different environments to be significantly different from unity.

Most dairy cattle evaluation schemes assume constant sire and residual variance applicable to all herds regardless of level of production or estimated within-herd variance. In some cases (e.g., Everett and Keown, 1984) this has involved a logarithmic transformation prior to analysis.

The objective of this paper is to review breeding value estimation with mixed models for various situations involving genotype-environment interactions. In fact, this amounts to evaluation involving heterogeneous variance components, possibly with a singular genetic variance-covariance matrix.

Properties of solutions obtained using simpler evaluations (ignoring heterogeneity) are outlined. These are used via simulation to determine the relative loss of efficiency resulting from ignoring interactions. In practice, variance components are frequently estimated ignoring genotype-environment interactions, and then routinely used for evaluation purposes. It is shown that depending on the population

---

<sup>1</sup> Dept. of Anim. Sci.  
Received August 18, 1986.  
Accepted February 3, 1987.

parameters and the intensity of selection, this may result in a substantial reduction in response.

#### Possible Manifestations of Genotype-Environment Interactions

A given model equation could be associated with any of the following situations with respect to variance components.

1) Unit genetic correlation between merit in each environment. This situation includes genotype-environment interactions that result in varying absolute differences between evaluations of candidates in different environments. The following possibilities were given by Henderson (1984):

(i) Equal additive genetic and residual variances in all environments.

(ii) Equal additive genetic variances but residual variances with magnitudes dependent on the environment. Consequently, heritability will vary between environments.

(iii) Additive genetic variances differing according to the environment and residual variances constant. Heritability will vary with environments.

(iv) Additive genetic and residual variances changing proportionally such that heritability remains constant across environments.

(v) Additive genetic and residual variances changing such that heritability is variable.

2) Genetic correlation of less than one between performance in different environments.

#### Evaluating Genetic Merit using Best Linear Unbiased Prediction

Consider the application of Best Linear Unbiased Prediction (BLUP) for breeding values. It will be assumed that the appropriate model equation is

$$Y = X\beta + Zu + e, \quad [1]$$

where,

$Y$  is the vector of performance records;

$\beta$  is a vector of unknown fixed effects;

$u$  is a vector of unknown random additive genetic effects;

$X$  is a known design matrix corresponding to the fixed effects;

$Z$  is a known incidence matrix correspond-

ing to the additive genetic effects and

$e$  is a vector of random residual effects, uncorrelated with other effects in the model.

Situation 1(i) is the simplest single trait situation, and breeding values can be estimated by setting up and solving the well-known mixed model equations (Henderson, 1963) shown below.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}. \quad [2]$$

The additive genetic variance-covariance matrix for  $u$  is  $G$ . In the single trait setting  $G$  is the matrix of additive genetic relationships (denoted  $A$ ) in product with the scalar additive genetic variance. The diagonal matrix consisting of constant residual variance is  $R$ ; therefore multiplying throughout by the scalar residual variance results in the equivalent equations:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\alpha A^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \quad [3]$$

where  $\alpha$  is the ratio of residual to additive genetic variance, and  $A^{-1}$  is the inverse of the numerator relationship matrix.

Situation 1(ii) involves equations [2], although  $R$  is no longer the product of an identity matrix and a scalar. Computationally, the least-squares partitions are those used for weighted regression with weights equal to the reciprocals of the residual variances. This is identical to using equations [3] if  $X$ ,  $Z$  and  $Y$  are transformed by dividing each row of these matrices by their relevant residual standard deviations and  $\alpha$  is redefined as the reciprocal of the additive genetic variance.

Situations 1(iii), (iv) and (v) involve multiple trait equations with genetic merit in each environment as a different trait. These equations are characterized by singular genetic variance-covariance matrices because breeding values for performance in each environment are linearly related, with the multiplier being the ratio of the additive genetic standard deviations in the two environments. The prediction error variances will not be related in this way. The matrix  $G$  is now defined as the direct product

of the singular variance-covariance matrix between the traits with the matrix of additive genetic relationships. This will be singular, so  $G^{-1}$  does not exist. Situation 1(iii) may appear simpler than the other cases of singular  $G$  matrices. However, when a sire model is fitted the residual component will include 75% of the additive genetic variance, so that both sire and residual variances will be heterogeneous. Accordingly, no further distinction need be made between these three situations. Harville (1976) presented equivalent mixed model equations (MME) that can be used for singular  $G$ . These equations [4] have as many singularities as exist among the fixed effects. A possible disadvantage of these equations is their lack of symmetry.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ GZ'R^{-1}X & GZ'R^{-1}Z+I \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ GZ'R^{-1}Y \end{bmatrix} \quad [4]$$

An equivalent symmetric set of equations is:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}ZG \\ GZ'R^{-1}X & GZ'R^{-1}ZG+G \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u}^+ \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ GZ'R^{-1}Y \end{bmatrix}, \quad [5]$$

where  $\hat{u} = G\hat{u}^+$ . The symmetric matrix  $G$  can always be decomposed into its canonical form under orthogonal similarity (Searle, 1982) such that  $U'GU = D$ , where  $U$  is an orthogonal matrix and  $D$  is a diagonal matrix of eigenvalues. Because [5] can be written as:

$$\begin{bmatrix} I & O \\ O & U' \end{bmatrix} \begin{bmatrix} X'R^{-1}X & X'R^{-1}ZG \\ GZ'R^{-1}X & GZ'R^{-1}ZG+G \end{bmatrix} \begin{bmatrix} I & O \\ O & U \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u}^+ \end{bmatrix} = \begin{bmatrix} I & O \\ O & U' \end{bmatrix} \begin{bmatrix} X'R^{-1}Y \\ GZ'R^{-1}Y \end{bmatrix} \quad [6]$$

then,

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}ZUD \\ DU'Z'R^{-1}X & DU'Z'R^{-1}ZUD+D \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{u}^\# \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ DU'Z'R^{-1}Y \end{bmatrix}, \quad [7]$$

where  $\hat{u} = U\hat{u}^\#$ . There is always a  $D$  such that only the first  $n$  diagonal elements are nonzero when there are  $n$  animals in the relationship matrix and  $t$  different environments. (The relationship matrix is full rank unless it includes identical twins.) The order of  $D$  is  $n \times t$  by  $n \times t$ , and the order of  $\beta$  is  $q$ . Consequently all elements of equations [7] will be null except the leading  $q + n$  submatrix of the coefficient matrix and right-hand side. These  $q + n$  equations have the advantage of symmetry and order equal to the equations that would be formed if the model from situation 1(i) or (ii) was assumed. The non-zero elements of  $\hat{u}^\#$  will be scalar multiples of the elements of  $\hat{u}$  from one of the environments. An example of equations equivalent to [7] incorporating  $A^{-1}$ , scaled to provide solutions to  $\hat{u}$  in the first

environment, can be found in Henderson (1984).

Situation 2 is an example of a multiple trait problem. Equations [2] can be used with suitable redefinition of  $X, Z, Y, G$  and  $R$  (Henderson and Quaas, 1976);  $G$  becomes the non-singular direct product of the numerator relationship matrix and the additive genetic variance-covariance matrix. If each individual has a performance record in only one environment the  $R$  matrix will have a diagonal structure. Computational difficulties arise as the number of traits increases. In special circumstances canonical transformations (brought to the attention of animal breeders by Thompson, 1979) can be used to reduce the computations to those for a set of single trait equations. In a progeny test situation the circumstance is

unlikely to occur unless all sires have progeny in all environments (herds).

The application of BLUP is straightforward when the appropriate model equation and  $R$  and  $G$  matrices are known, a priori. In practice, variance components are never known without error, and the best estimates are used as if these represent the true values. In large populations with homogeneous variances this approach is likely to be satisfactory. If variances differ by environment the difficulty arises in determining which components to use for each environment. One approach to this problem would be to use a Bayesian procedure to combine prior information such as the population estimates of components with within-herd estimates of components (Henderson, 1984; Hill, 1984; Gianola, 1986). An alternative, if biologically reasonable, would be to model the variances in different environments, perhaps as a function including herd size and level of production. Weller et al. (1985) have suggested this approach with reference to dairy breeding evaluations. Everett and Keown (1984) presented correlations between herd characteristics and within-herd residual variance.

It would be useful to know the relative efficiency of selection based on a simplified model, such as 1(i) when in reality 1(v) or 2 reflect the underlying model. Henderson (1975) has presented methods for determining the consequences of certain violations of the model, notably the inclusion of unnecessary or exclusion of required fixed effects and the consequence of incorrect  $G^{-1}$ . Hill (1984) has investigated the effect of varying residual variances when heritability is constant across

environments. Hill's results suggest that the loss of efficiency may be relatively minor, and scaling with the estimated standard deviations would result in close to optimal rates of improvement.

#### Consequences of Using Estimates of Variance Components for Predicting Breeding Values Using BLUP

The results of Henderson (1975) are extended here to examine the consequences of simultaneously using incorrect  $G$  and  $R$  matrices with  $G$  not necessarily full rank. It is assumed that model equation [1] is correct, i.e., the underlying model is linear and all relevant fixed and random effects are included. The true residual and genetic variance-covariance matrices will be denoted by  $R$  and  $G$  respectively, and  $\tilde{R}$  and  $\tilde{G}$  will denote the estimated variance-covariance matrices.

From [4] a solution is obtained as

$$\begin{bmatrix} \beta^0 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ GZ'R^{-1}X & GZ'R^{-1}Z+I \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}Y \\ GZ'R^{-1}Y \end{bmatrix}. \quad [8]$$

Alternatively, using estimated (co-)variance components

$$\begin{aligned} \begin{bmatrix} \beta^{0*} \\ \hat{u}^* \end{bmatrix} &= \begin{bmatrix} X'\tilde{R}^{-1}X & X'\tilde{R}^{-1}Z \\ \tilde{G}Z'\tilde{R}^{-1}X & \tilde{G}Z'\tilde{R}^{-1}Z+I \end{bmatrix}^{-1} \begin{bmatrix} X'\tilde{R}^{-1}Y \\ \tilde{G}Z'\tilde{R}^{-1}Y \end{bmatrix} \\ &= \begin{bmatrix} T^{11} & T^{12} \\ T^{21} & T^{22} \end{bmatrix} \begin{bmatrix} X'\tilde{R}^{-1}Y \\ \tilde{G}Z'\tilde{R}^{-1}Y \end{bmatrix}. \quad [9] \end{aligned}$$

Then,

$$\begin{aligned} V(\hat{u}^*) &= \begin{bmatrix} T^{21} & T^{22} \end{bmatrix} \begin{bmatrix} X'\tilde{R}^{-1}Z \\ \tilde{G}Z'\tilde{R}^{-1}Z \end{bmatrix} G \begin{bmatrix} Z'\tilde{R}^{-1}X & Z'\tilde{R}^{-1}Z\tilde{G} \end{bmatrix} \begin{bmatrix} T^{21'} \\ T^{22'} \end{bmatrix} \\ &+ \begin{bmatrix} T^{21} & T^{22} \end{bmatrix} \begin{bmatrix} X' \\ \tilde{G}Z' \end{bmatrix} \tilde{R}^{-1}R\tilde{R}^{-1} \begin{bmatrix} X & Z\tilde{G} \end{bmatrix} \begin{bmatrix} T^{21'} \\ T^{22'} \end{bmatrix}. \end{aligned}$$

$$\text{Cov}(\hat{u}^*, u') = \begin{bmatrix} T^{21} & T^{22} \end{bmatrix} \begin{bmatrix} X'\tilde{R}^{-1}Z \\ \tilde{G}Z'\tilde{R}^{-1}Z \end{bmatrix} G.$$

$$V(u) = G.$$

Note that  $V(\hat{u}^*)$  does not equal  $Cov(\hat{u}^*, u')$  as with [8]. Accordingly,  $b_{u, \hat{u}^*}$ , the regression of  $u$  on  $\hat{u}^*$  is not unity. Prediction error variance can be obtained as:

$$V(\hat{u}^* - u) = V(\hat{u}^*) + V(u) - 2 \text{cov}(\hat{u}^*, u')$$

and the correlation between true and predicted breeding value as:

$$r_{u, \hat{u}^*} = \text{Cov}(\hat{u}^*, u') / (V(\hat{u}^*)V(u))^{.5}$$

For a given situation with known incidence matrices and true variance components,  $V(\hat{u}^*)$

and  $b_{u, \hat{u}^*}$  can be computed for differing values of estimated variance components. These enable computation of the effect of using estimated variance structures for breeding value estimation when more complex structures exist in practice. The following section uses this method to determine the loss of selection efficiency from assuming homogeneity of variance when heterogeneity is present.

**Numerical Examples**

Suppose progeny produce a first-lactation record in one of three environments with the following true (co-)variance structure.

Parameter set I:

Environment	Sire variances			Residual variances		
	1	2	3	1	2	3
1	1.894			57.891		
2	2.337	3.056			64.450	
3	2.615	3.362	4.880			72.390
Heritabilities (on diagonal) and genetic correlations						
1	.13					
2	.97	.19				
3	.86	.87	.27			

These parameters were scaled from estimates calculated using dairy records from the north-eastern United States (De Veer, 1986) with the environments categorized on the basis of mean production. Environment 3 represents the high production herds, which also are characterized by increased residual variance and heritability.

Parameter set II:

Environment	Sire variances			Residual variances		
	1	2	3	1	2	3
1	3.056			34.930		
2	2.965	3.056			48.903	
3	2.629	2.659	3.056			81.504
Heritabilities (on diagonal) and genetic correlations						
1	.35					
2	.97	.25				
3	.86	.87	.15			

This set of parameters was contrived to provide the same correlation structure as parameter set I, with sire variance constant across the differing environments. The third environment still corresponds to that with the greatest residual variance, but now has the smallest heritability. These parameters are similar after scaling to results from De Veer (1986) for a logarithmic transformation. (Environment 1 would now correspond to De Veer's category with the highest mean production level.)

*Application 1, Determining the Accuracy of Progeny Testing Bulls*

Sires were evaluated on an index involving the sum of their estimated transmitting abilities (ETA) in each of the three environments. An ideal situation would involve each sire having progeny equally distributed in all three environments. The selection differential, which could be obtained from selection on the evaluations using the correct variance components, determines a ceiling response. In practice, progeny of some sires may be poorly distributed across different environments. This is likely to result in selection bias if homogeneous variances are assumed, and some sires have all of their progeny in the more variable environments. An extreme situation was generated in which each sire had a total of 15 progeny with one of the following six distributions of daughters:

- (i) 5 daughters in each of the three environments;
- (ii) 7 and 8 daughters in environments 1 and 2, respectively;
- (iii) 7 and 8 daughters in environments 2 and 3;
- (iv) 15 daughters in environment 1;
- (v) 15 daughters in environment 2 and
- (vi) 15 daughters in environment 3.

A large group of unrelated progeny-test bulls was arbitrarily assumed to be equally represented with each of the six daughter distributions. Furthermore, the genetic and residual effects were distributed normally and the means of the six daughter distribution types were identical. This results in half the available sires having progeny confounded with respect to variance.

Three evaluations (a, b and c) were obtained for parameter set I and two evaluations (a and c) were derived for parameter set II:

- (a) true  $R$  and  $G$  matrices used,

- (b) true variances for  $R$  and  $G$  used but unit genetic correlations among the environments were assumed (thus  $G$  is a singular matrix),

- (c) homogeneous genetic and residual variances were assumed using those for environment 2 as if they applied to all environments.

The variances of the index from each evaluation and the corresponding covariances with true index merit are shown in table 1 for the situation with no fixed effects in the model. Analyses were also obtained, given daughter records of the sampling bulls contributing to the estimates of the herd-year-season fixed effects. The results were relatively similar so are not presented here.

To investigate the effects of daughter distribution and of using the incorrect evaluation (b and c), the expected superiorities from selection of varying proportions of progeny-tested bulls chosen on the basis of indices a, b or c were compared. These genetic selection differentials were calculated in a three-step procedure. For each evaluation, Newton's method was used to obtain the truncation point at which the sum of the integrals of the six normal density functions with variance  $V(\hat{u}^*)$  was equal to the required proportion to be retained (described by Ducrocq, 1984). The selection differential in terms of  $\hat{u}^*$  was then obtained by calculating the mean of these truncated normal distributions. The corresponding genetic selection differentials (table 2) were given by multiplying the  $\hat{u}^*$  selection differentials with the regression coefficient  $b_{u,\hat{u}^*}$ , weighted by the number of bulls chosen from each daughter distribution class.

Given parameter set I, greater inefficiency results from failure to distribute daughters of bulls across different environments than occurs from using incorrect parameters. There is virtually no detrimental effect from assuming the absence of genotype-environment interactions that alter rankings. Furthermore, there is little reduction in efficiency from ignoring the heterogeneity present. This can be rationalized in that the use of homogeneous parameters favors the selection of sires with progeny in environment 3 where there was the greatest residual variance, but this is the environment with the greatest heritability and therefore a preferred environment for selecting animals because the accuracy of evaluation is greater.

Parameter set II is more sensitive to evaluation using incorrect variance components. For selection intensities from 20 to 12%, failure to

TABLE 1. VARIANCES AND COVARIANCES OF PREDICTED AND ACTUAL TRANSMITTING ABILITIES ( $u = u_1 + u_2 + u_3$ ) FOR SIX BULL-DAUGHTER DISTRIBUTIONS

Parameters <sup>a</sup>		Daughter distribution types					
True	In MME	(i)	(ii)	(iii)	(iv)	(v)	(vi)
$V(\hat{u}^*)$							
I	a	10.759	9.488	11.749	8.145	10.426	12.145
I	b	11.715	10.705	12.868	9.363	11.824	14.302
I	c	11.300	10.166	12.916	8.946	11.434	15.093
$Cov(\hat{u}^*, u')$							
I	a	10.759	9.488	11.749	8.145	10.426	12.145
I	b	11.226	10.978	12.293	8.733	11.103	13.179
I	c	10.998	9.807	12.316	8.536	10.918	13.539
$V(\hat{u}^*)$							
II	a	12.012	13.034	10.417	13.893	11.929	8.200
II	c	13.753	12.298	15.336	11.347	13.310	17.890
$Cov(\hat{u}^*, u')$							
II	a	12.012	13.034	10.417	13.893	11.929	8.200
II	c	12.423	12.578	12.340	12.556	12.600	12.112

<sup>a</sup> $V(u) = 26.458$  for parameter set I and  $V(u) = 25.674$  for set II. a, true R and G matrices used; b, true variances and unit genetic correlations; c, homogeneous variances using the components applicable to environment 2.

distribute progeny results in a reduction of gain comparable to assuming homogeneity. Given very intense selection, assuming homogeneity can severely reduce the efficiency of selection. In this case the most variable environment (which will be over-represented when homogeneous parameters are used) is the least accu-

rate for evaluations. The selection differential achieved from bulls with all six distributions of progeny ranges from 98 to 92% of that achieved using the correct parameters, with greater reduction when selection is more intense. For a dairy bull progeny test, typically close to 10% would be retained.

TABLE 2. EXPECTED SUPERIORITY OF SELECTED BULLS IN STANDARD DEVIATIONS OF TRANSMITTING ABILITY FOR VARYING VARIANCE COMPONENTS AND PARAMETER SETS

Parameters <sup>a</sup>		Bull daughter distrib. <sup>b</sup>	Proportion of tested bulls retained						
True	In MME		20%	15%	12%	10%	8%	5%	1%
I	a	(i)	.893	.991	1.063	1.119	1.185	1.315	1.700
I	a	All	.879	.977	1.049	1.105	1.171	1.301	1.690
I	b	All	.879	.977	1.048	1.104	1.170	1.301	1.689
I	c	All	.878	.976	1.047	1.103	1.169	1.299	1.685
II	a	(i)	.957	1.063	1.140	1.200	1.271	1.411	1.823
II	a	All	.939	1.044	1.121	1.181	1.252	1.393	1.814
II	c	All	.921	1.020	1.090	1.145	1.209	1.333	1.677

<sup>a</sup>a, true R and G matrices used; b, true variances and unit genetic correlations; c, homogeneous variances using the components applicable to environment 2.

<sup>b</sup>(i) = daughters equally represented in each environment; all = daughters belong to one of six distributions as described in text.



*Application 2, Asymptotic Genetic Gains in a Dairy Breeding Scheme*

It is apparent from the application to a progeny test that only moderate reduction in selection differential results from using simplified parameters for data analysis. Intuitively, it would seem that the bias may be greater in the selection of bull dams from heterogeneous environments because cows only produce a record within one of the environments, whereas sires may have progeny in all environments. A

deterministic model (figure 1, tables 3 and 4) with the following assumptions was developed to describe a large-scale dairy breeding program.

*Cow Population.* A population of 750,000 cows equally represented in three environments was artificially bred. Accordingly, an index of combined merit from the three environments was the selection objective. Population size determined the semen requirements and thus the selection intensity for sires of cows. Cows were bred to calve from 2 to 9 yr of age in the

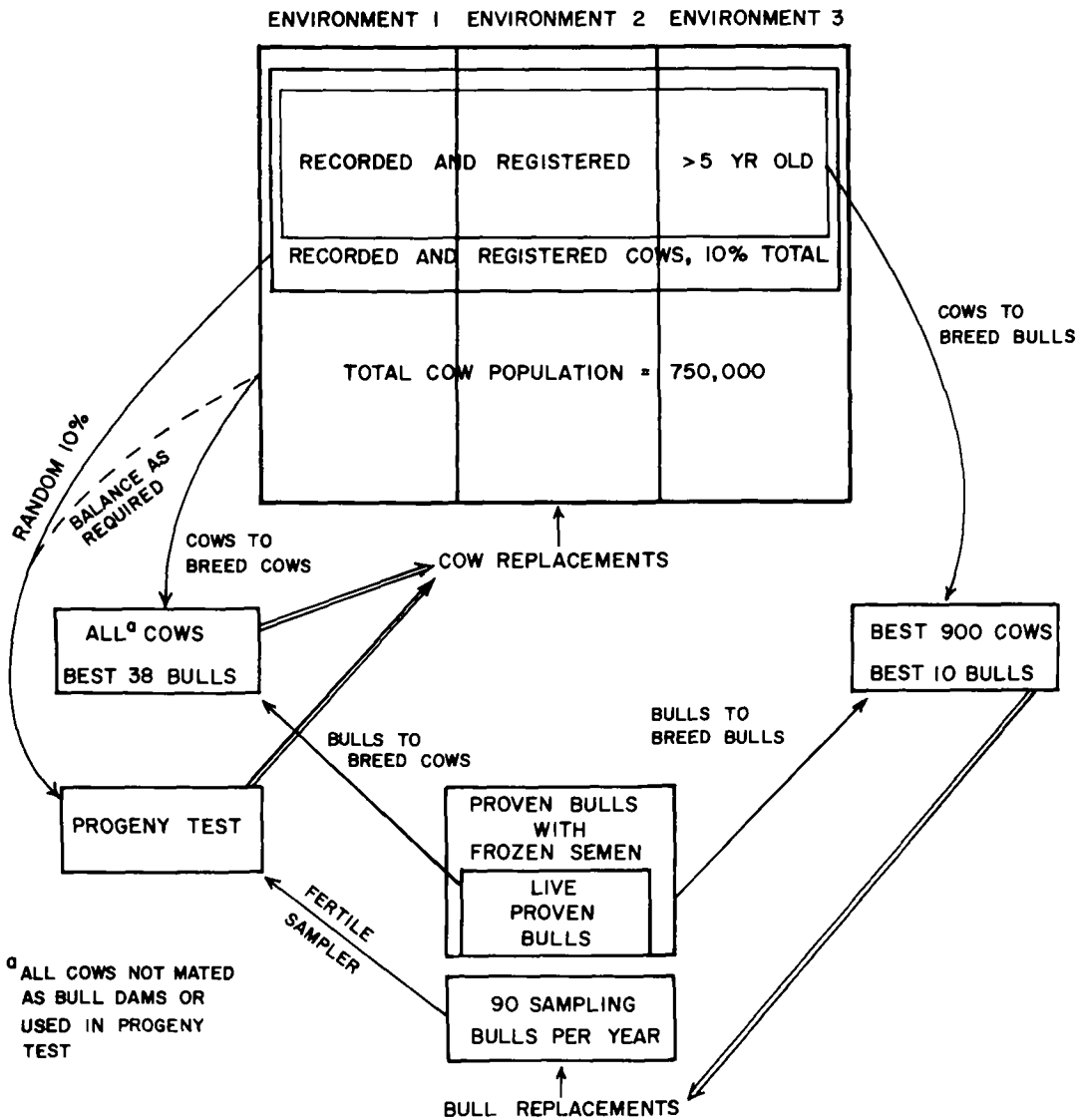


Figure 1. Diagrammatic representation of dairy breeding scheme.

TABLE 3. STRUCTURE OF THE COW POPULATION IN EACH OF THE THREE ENVIRONMENTS

Age of cow, yr	Numbers for bull dams	Numbers for cow dams	Mean cow breeding value by birth year <sup>a</sup>
9	800	8,000	$\mu_c$
8	1,025	10,250	$\mu_c + 1 \times \Delta G$
7	1,600	16,000	$\mu_c + 2 \times \Delta G$
6	2,125	21,250	$\mu_c + 3 \times \Delta G$
5	3,350	33,500	$\mu_c + 4 \times \Delta G$
4	4,375	43,750	$\mu_c + 5 \times \Delta G$
3	NA <sup>b</sup>	52,750	$\mu_c + 6 \times \Delta G$
2	NA	64,500	$\mu_c + 7 \times \Delta G$

<sup>a</sup> $\mu_c$  = genetic mean of cow population 9 yr previously.  $\Delta G$  = asymptotic genetic gain per year.

<sup>b</sup>NA = not applicable.

following proportions; 25.8, 21.1, 17.5, 13.4, 8.5, 6.4, 4.1 and 3.2% (Everett et al., 1976).

Recorded and registered cows (10% of the population) were candidates for selection as bull dams and were equally represented in the three environments. It was assumed that 10% of these cows were bred to young sires for progeny testing. The remainder of the progeny-test mates of the young sires was chosen from the other 90% of the cow population. Young bulls were randomly mated to cows of all age groups in proportion to their numbers. Bull dams were restricted to cows of 5 yr or older because cow selection is done on first lactation results. The breeding values for the cows were not available until after their third mating (early in their second lactation). Hence the first mating to produce bull replacements is the fourth and the bull calves are born when the cows are 5 yr old.

*Bull Population.* Each year, 90 fertile sires were born. (Culling for injury and infertility

had been accounted for in choosing the number of bull dams.) The bulls had a 1% annual mortality rate until their evaluations were obtained at 6 yr. Their second-crop daughters were born when the sires were 7 yr or more. The bulls had a 5% annual mortality rate from the time their evaluations were known. Sires to breed cows were obtained from the live bulls of 6 yr or greater. Sires to breed sires were obtained from any bull that survived until evaluation because sufficient frozen semen was assumed to be available to breed young bulls. Each bull annually produced sufficient semen for 40,000 inseminations. On average, two inseminations were needed per conception. Thus a total of 38 bulls was required for breeding cows to produce cows. The best 10 bulls were used as sires to breed sires each year. The top 900 cows were retained from the bull dam population for mating to these sires. Progeny tests on young bulls were based on an effective number of 45 daugh-

TABLE 4. STRUCTURE OF THE BULL POPULATION

Age of bull, yr	Numbers for bull sires	Numbers for cow sires	Mean bull breeding value by birth year <sup>a</sup>
12	84.7	65.6	$\mu_b$
11	84.7	69.0	$\mu_b + 1 \times \Delta G$
10	84.7	72.6	$\mu_b + 2 \times \Delta G$
9	84.7	76.5	$\mu_b + 3 \times \Delta G$
8	84.7	80.5	$\mu_b + 4 \times \Delta G$
7	84.7	84.7	$\mu_b + 5 \times \Delta G$

<sup>a</sup> $\mu_b$  = genetic mean of bull population 12 yr previously.  $\Delta G$  = asymptotic genetic gain per year.

ters equally represented in the three environments. Evaluation of cows to select as bull dams was based on their own and their dam's first lactation record from the same variance environment and the progeny test of their sire. Second-crop daughters were not used to improve the accuracy of evaluations of aged bulls.

To account for overlapping generations and therefore the effect of genetic trend increasing the mean of each subsequent crop of young bulls or cows, the simulation included a separate distribution for each age group of cows in each of the three environments and for each age group of bulls (Bichard et al., 1973). The sizes of these populations were determined from the assumptions described earlier. A routine using Newton's method to obtain the truncation points was applied as previously described. The variance of each of these populations was based on the variance of the estimated breeding value for total merit. The mean of each age group was the genetic mean of that sex born in that particular year and environment. This was determined by obtaining the expected value of  $u$  (given  $\hat{u}$ ) for cows and bulls used in each path, and then averaging these to obtain expected progeny averages. Simulation was continued iteratively until the annual change in mean  $u$  for the bulls and cows were identi-

cal. This asymptotic gain can be obtained using the formula [10] presented by Rendel and Robertson (1950). The selection differential on the choice of cows to breed cows was assumed to be zero.

$$\text{Asymptotic genetic gain} = \frac{I_{SS} + I_{SC} + I_{CS} + I_{CC}}{L_{SS} + L_{SC} + L_{CS} + L_{CC}}, \quad [10]$$

where

$I$  refers to selection differential,

$L$  refers to generation interval and the subscripts refer to the paths of selection, i.e., SS, sires to breed sires; SC, sires to breed cows; CS, cows to breed sires; CC, cows to breed cows.

Parameter sets I and II and the three sets of estimated variance components described for the progeny test were used for calculating variances of index values and covariances between index and true merit (table 5). In this case,  $u = u_1 + u_2 + u_3$  is defined as the breeding value rather than transmitting ability. The calculated selection differentials and generation intervals for each selection path and each evaluation method are presented in table 6 (parameter set I) and table 7 (parameter set II).

TABLE 5. VARIANCES AND COVARIANCES OF PREDICTED AND ACTUAL BREEDING VALUES ( $u = u_1 + u_2 + u_3$ )

Parameters <sup>a</sup>		Environment			
True	In MME	1 Cows	2 Cows	3 Cows	All Bulls' PT
$V(\hat{u}^*)$					
I	a	29.175	34.031	39.137	71.188
I	b	31.547	36.867	43.584	74.571
I	c	32.545	35.939	40.213	73.148
$Cov(\hat{u}^*, u')$					
I	a	29.175	34.031	39.137	71.188
I	b	30.330	35.413	41.269	72.856
I	c	30.666	34.940	39.641	72.064
$V(\hat{u}^*)$					
II	a	45.080	38.517	30.263	74.289
II	c	36.296	41.948	54.858	79.882
$Cov(\hat{u}^*, u')$					
II	a	45.080	38.517	30.263	74.289
II	c	39.919	40.006	39.051	75.759

<sup>a</sup> $V(u) = 105.832$  for set I and  $V(u) = 102.696$  for set II. a, true R and G matrices used; b, true variances and unit genetic correlations; c, homogeneous variances using the components applicable to environment 2.

TABLE 6. ASYMPTOTIC RESULTS FROM SIMULATION OF DAIRY BREEDING SCHEME USING PARAMETER SET I

Path <sup>b</sup>	Parameters <sup>a</sup> in MME					
	a		b		c	
	Generation interval, yr	Genetic selection differential, SD	Generation interval, yr	Genetic selection differential, SD	Generation interval, yr	Genetic selection differential, SD
SS	8.1934	1.8847	8.2162	1.8886	8.2082	1.8848
DS	5.5052	1.2828	5.5310	1.2865	5.5172	1.2819
SD	7.7055	1.2775	7.7228	1.2804	7.7168	1.2777
DD	4.0930	.0000	4.0930	.0000	4.0930	.0000
Mean	6.3743	1.1113	6.3908	1.1139	6.3838	1.1111
Annual gain, SD/yr		.1743		.1743		.1741
Environment					No. of bull dams from each environment	
1		210		203		236
2		297		290		294
3		393		407		370

<sup>a</sup>a, true R and G matrices used; b, true variances and unit genetic correlations; c, homogeneous variances using the components applicable to environment 2.

<sup>b</sup>Paths are defined as follows: SS, sires to breed cows; SC, sires to breed sires; CC, cows to breed cows.

TABLE 7. ASYMPTOTIC RESULTS FROM SIMULATION OF DAIRY BREEDING SCHEME USING PARAMETER SET II

Path <sup>b</sup>	Parameters <sup>a</sup> in MME			
	a		c	
	Generation interval, yr	Genetic selection differential, SD	Generation interval, yr	Genetic selection differential, SD
SS	8.1832	1.9527	8.2511	1.9322
DS	5.5240	1.3821	5.5869	1.2927
SD	7.6977	1.3234	7.7492	1.3104
DD	4.0930	.0000	4.0930	.0000
Mean	6.3745	1.1646	6.4200	1.1338
Annual gain, SD/yr	.1827		.1766	
Environment	No. of bull dams from each environment			
1	418		185	
2	306		262	
3	176		453	

<sup>a</sup>a, true R and G matrices used; c, homogeneous variances using the components applicable to environment 2.

<sup>b</sup>Paths are defined as follows: SS, sires to breed sires; SC, sires to breed cows; CS, cows to breed sires; CC, cows to breed cows.

In agreement with the progeny test results, the effect of incorrect variance components in the mixed model equations is greatest when the most variable environment has the lowest heritability (table 7). When heritability increases with the residual variance, there is little loss in efficiency of selection from assuming homogeneity (table 6). Furthermore, there is a negligible effect from assuming a unit genetic correlation between performance in different environments when in reality a small interaction is present. This demonstrates the robustness of BLUP to certain violations in the assumptions. The results from parameter set I are in agreement with Powell et al. (1983), who suggested that the effects of bias caused by larger variances in high producing herds would be offset by higher heritabilities.

Examination of the numerical contributions of bull dams from each environmental class (tables 6 and 7) demonstrates the sampling bias resulting from assuming homogeneity. The bias is most severe with parameter set II where almost half of the bull dams should be chosen from environment 1, yet the simplified parameters result in half the bull dams being identified from the high variance herds (environment 3), resulting in reduced genetic gain.

### Discussion

In practical situations simplifying assumptions are often essential to reduce computational requirements for breeding value estimation. It is useful to know the efficiency of these simpler methods relative to theoretically appropriate procedures. The sire model is a commonly used simplification based on the assumption that dams are chosen randomly. Including the maternal grandsire is slightly more complex but can partially account for selection on dams. In selection index schemes it is common to ignore many of the genetic relationships that exist between animals to be evaluated. In many respects the assumption of homogeneity of variances can be viewed as a simplification in much the same way.

In due course a method comparable to BLUP may be developed to account for intra-herd variances coming from a sample of some population of variances. In addition, techniques for pooling within-herd and population estimates may be used more widely. In the meantime, it is worthwhile to investigate the reduction in efficiency of selection from assuming heterogeneity for individual breeding applications.

In situations involving greater heritability in more variable populations, there is likely to be

little reduction in progress from assuming homogeneity. However, some caution should be used when considering the application of logarithmic transformations. If the transformation is to achieve normality (due to multiplicative effects), then it should be used. If the transformation is used in an attempt to remove heterogeneity for situations in which the more variable environments have a higher heritability, then this will result in a decrease in the efficiency of selection if the transformed records have higher heritability in the less variable environments.

Bull studs should also ensure that sampling bulls are represented in a variety of different herd environments. Failure to do so will lead to less accurate evaluations for bulls with progeny in low heritability herds, and correspondingly reduced efficiency of selection. Although the use of correct variance components for evaluation would ensure that bulls with all progeny in particular variance herds would be assessed fairly, simplified parameters will result in these bulls being under- or overevaluated.

#### Literature Cited

- Bichard, M., A.H.R. Pease, P. H. Swales and K. Özkütük. 1973. Selection in a population with overlapping generations. *Anim. Prod.* 17:215.
- De Veer, J. C. 1986. Genetic parameters for first lactation milk yields at three levels of production. Ph.D. Dissertation. Cornell Univ., Ithaca, NY.
- Ducrocq, V. 1984. Effects of selection for secondary traits on genetic gain for milk production in cattle. *Genet. Sel. Evol.* 16:467.
- Everett, R. W. and J. F. Keown. 1984. Mixed model sire evaluation with dairy cattle—Experience and genetic gain. *J. Anim. Sci.* 59:529.
- Everett, R. W., J. F. Keown and E. E. Clapp. 1976. Production and stayability trends in dairy cattle. *J. Dairy Sci.* 59:1532.
- Falconer, D. S. 1952. The problem of environment and selection. *Amer. Nat.* 86:293.
- Gianola, D. 1986. On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor. Appl. Genet.* 72:671.
- Hammond, J. 1947. Animal breeding in relation to nutrition and environmental conditions. *Biol. Rev.* 22:195.
- Harville, D. A. 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Math. Statist.* 4:384.
- Henderson, C. R. 1963. Selection index and expected genetic advance. In W. D. Hanson and H. F. Robinson (Ed.) *Statistical Genetics and Plant Breeding*. Pub. 982. National Academy of Sciences-National Research Council, Washington, DC.
- Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. *J. Anim. Sci.* 41:760.
- Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. of Guelph, Guelph, Ontario, Canada.
- Henderson, C. R. and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43:1188.
- Hill, W. G. 1984. On selection among groups with heterogeneous variance. *Anim. Prod.* 39:473.
- Hill, W. G., M. R. Edwards and M.-K.A. Ahmed. 1983. Heritability of milk yield and composition at different levels and variability of production. *Anim. Prod.* 36:59.
- Lush, J. L. 1945. *Animal Breeding Plans*. Iowa State College Press, Ames.
- Mirande, S. L. and L. D. Van Vleck. 1985. Trends in genetic and phenotypic variances for milk production. *J. Dairy Sci.* 68:2278.
- Powell, R. L., H. D. Norman and B. T. Weinland. 1983. Cow evaluation at different milk yields of herds. *J. Dairy Sci.* 66:148.
- Rendel, J. M. and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J. Genet.* 50:1.
- Robertson, A., L. K. O'Connor and J. Edwards. 1960. Progeny testing dairy bulls at different management levels. *Anim. Prod.* 2:141.
- Searle, S. R. 1982. *Matrix Algebra Useful for Statistics*. p 291. J. Wiley and Sons, Inc., New York.
- Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339.
- Weller, J. I., M. Ron and R. Bar-Anan. 1985. Accounting for environmentally dependent variance components in BLUP sire evaluation. *J. Dairy Sci.* 68(Suppl. 1):212.