

January 1980

Sire Evaluation by Best Linear Unbiased Prediction for Categorically Scored Type Traits

L. Dale Van Vleck

University of Nebraska-Lincoln, dvan-vleck1@unl.edu

P.J. Karner

Cornell University

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



Part of the [Animal Sciences Commons](#)

Van Vleck, L. Dale and Karner, P. J., "Sire Evaluation by Best Linear Unbiased Prediction for Categorically Scored Type Traits" (1980).
Faculty Papers and Publications in Animal Science. 378.
<http://digitalcommons.unl.edu/animalscifacpub/378>

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.

Sire Evaluation by Best Linear Unbiased Prediction for Categorically Scored Type Traits

L. D. VAN VLECK and P. J. KARNER

Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

Best linear unbiased prediction to predict category frequencies of future progeny of a sire for type traits scored in mutually exclusive categories is described. The method accounts for automatic covariances among categories and is comparable to prediction for multiple traits. The method does not require linearity of measurements and also allows nonlinear economic values to be assigned to each category after frequencies are predicted. Evaluations were for 12 descriptive type traits for 712 Brown Swiss bulls having daughters in more than one herd. Problems in obtaining solutions to the mixed-model equations for multiple traits are discussed.

INTRODUCTION

Genetic evaluations for type characteristics have changed rapidly in the last few years after many years of little change. Workers at Virginia Tech (23, 24) in cooperation with the Holstein Association developed herdmate procedures for Holstein data. Even more recently the United States Department of Agriculture (USDA) (4,14) has adapted best linear unbiased prediction (BLUP) procedures (5, 6, 7, 8, 18, 19, 20) for Jersey, Guernsey, and Holstein records. A BLUP system was begun in Canada for Holsteins (16,17). These evaluations have been for final score and score card traits that have been recorded on a linear scale.

Analysis of descriptive traits has been complicated by their categorical and nonlinear scoring. The National Association of Animal Breeders (11) has been instrumental in attempting to standardize a linear method of scoring descriptive traits. Another method of analysis

of categorically scored traits is to treat each category as a separate subtrait and utilize the covariance structure of the subtraits to predict frequencies of future progeny for each category (1, 2, 3, 15). The covariance structure can be utilized and, perhaps more important, nonlinear economic values can be assigned to predicted frequencies in the categories.

The purpose of this paper is to describe a best linear unbiased prediction procedure which was used for categorically scored type traits recorded by the Brown Swiss Cattle Breeders' Association.

METHODS

The model treats categories of a trait as separate traits so that the model is the same as for multiple trait evaluation (10,12) as modified in (15). The modification essentially is to delete one category so that the variance-covariance matrices will be nonsingular.

Let $y' = (y'_1 \ y'_2 \ \dots \ y'_n)$, where y'_l is the vector of observations for animal l on c categories for a trait with $c + 1$ categories. All elements of y_l are zero unless the corresponding category is scored; then the element will be a 1. Thus, y is the vector of observations for the model

$$y = X\beta + Zu + e$$

where

$\beta' = (\beta_1 \ \beta_2 \ \dots \ \beta_c)'$ with β_i the fixed effect for category i which is a constant corresponding to frequency in the category;

$u' = (h'_1 \ h'_2 \ \dots \ h'_H \ s'_1 \ s'_2 \ \dots \ s'_S)$ with h_i the vector of herd-year effects of the c categories for herd-year i and s_j the vector of sire effects for the c categories for sire j ;

e = the vector of random residual effects;

X = the incidence matrix describing which fixed effect is associated with each

Received December 14, 1979.

record and is a column of $n_c \times c$ where identity matrices; and

$Z' = (Z'_h Z'_s)$ = the incidence matrix describing which herd-year and sire effect is included in each record. Each set of c columns of the Z_h matrix will contain a $c \times c$ identity matrix for each animal associated with that herd-year and $c \times c$ blocks of zeros when the records are not associated with that herd-year. A similar pattern holds for Z_s . The mean vector is

$$E(y) = X\beta,$$

and the variances are

$$V \begin{bmatrix} h \\ s \\ e \end{bmatrix} = \begin{bmatrix} H & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & R \end{bmatrix}$$

H is a block diagonal matrix with number of blocks equal to the number of herd-years and the identical blocks (H_1) are the variance-covariance matrices of herd-year effects for the remaining c categories of the trait;

S is a similar block diagonal matrix corresponding to the number of sires and with blocks (S_1) consisting of the variance and covariance matrix of sire effects (if relationships among the sires are considered, then $S = A^*S_1$, the direct product of the numerator relationship among the sires and S_1); and

R is a block diagonal matrix of n blocks with elements corresponding to variances and covariances among residual effects (R_1).

The usual BLUP equations are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_h & X'R^{-1}Z_s \\ Z'_hR^{-1}X & Z'_hR^{-1}Z_h + H^{-1} & Z'_hR^{-1}Z_s \\ Z'_sR^{-1}X & Z'_sR^{-1}Z_h & Z'_sR^{-1}Z_s + S^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{h} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'_hR^{-1}y \\ Z'_sR^{-1}y \end{bmatrix}$$

Due to the nature of X , Z_h , and Z_s , which are made up of zeros and identity matrices,

the equations can be rewritten in a form similar to (12) as

$$\begin{bmatrix} n_{.1}R_1^{-1} & n_{1.}R_1^{-1} & \dots & n_{H.}R_1^{-1} & n_{.1}R_1^{-1} & \dots & n_{.S}R_1^{-1} \\ n_{1.}R_1^{-1} & n_{1.}R_1^{-1} + H_1^{-1} & \dots & 0 & n_{11}R_1^{-1} & \dots & n_{1S}R_1^{-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{H.}R_1^{-1} & 0 & \dots & n_{H.}R_1^{-1} + H_1^{-1} & n_{H1}R_1^{-1} & \dots & n_{HS}R_1^{-1} \\ n_{.1}R_1^{-1} & n_{11}R_1^{-1} & \dots & n_{H1}R_1^{-1} & n_{.1}R_1^{-1} + S^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{.S}R_1^{-1} & n_{1S}R_1^{-1} & \dots & n_{HS}R_1^{-1} & 0 & \dots & n_{.S}R_1^{-1} + S^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{h}_1 \\ \vdots \\ \hat{h}_H \\ \hat{s}_1 \\ \vdots \\ \hat{s}_S \end{bmatrix} = \begin{bmatrix} R_1^{-1}y_{.1} \dots \\ R_1^{-1}y_{1.} \dots \\ \vdots \\ R_1^{-1}y_{H.} \dots \\ R_1^{-1}y_{.1} \dots \\ \vdots \\ R_1^{-1}y_{.S} \end{bmatrix}$$

where n_{ij} is the number of animals with records in the i th herd-year sired by the j th sire, and y_{ijk} is the vector of c observations on the k th

animal in the i th herd-year sired by the j th sire and corresponds to the y_l defined earlier.

The computing strategy was to absorb the

herd-year equations into the $\hat{\beta}$ and \hat{s} equations before obtaining solutions by modified Gauss-Seidel iteration with number of iterations being 25 or less than 25 if the sum of absolute changes from the previous round was less than .0001 N, where N is the number of equations after absorption ($[\text{number of sires} + 1] \times c$).

The variance-covariance matrices H_1 , S_1 , and R_1 were those from the same data (21,22).

Estimates and predictions for the deleted category were obtained by difference as described by (15) so that

$$\hat{\beta}_{c+1} = 1 - \sum_{m=1}^c \hat{\beta}_m \quad \text{and}$$

$$\hat{s}_{j\ c+1} = -\sum_{m=1}^c \hat{s}_{jm} \quad ,$$

where \hat{s}_{jm} is the solution for category m for the *j*th sire. The evaluation for a sire is then the sum of the $\hat{\beta}$ and \hat{s}_j vectors.

The data consisted of official classification records of 12,838 Brown Swiss cows classified by 47 mo of age between 1971 and 1976. The data were restricted so that each of the 712 sires had to have daughters in more than one herd. There were 824 herds represented and 2295 herd-years. The descriptive traits were not adjusted, but final score, final classification, and the eight score card traits were adjusted as described by Moreno (13) to the basis of a lactating cow, 47 mo of age classified in the fall. Predictions for these adjusted traits were as single traits but are not reported here since the procedure is similar to that described by Norman et al. (14). Relationships among the sires were not considered although a slight modification easily would account for numerator relationships (9).

DISCUSSION

Any discussion of sire evaluation for categorical traits must include a listing of any difficulties in computing and especially of potential computing time. The computing sequence is reasonably standard for situations when the entire coefficient matrix cannot be stored in memory. 1) The data are sorted sire within herd-year. 2) The herd-year equations are absorbed, and the resulting β and sire coefficients together with their right-hand side terms are written in half-stored form on tape for each herd-year. 3) The coefficients and right-hand

sides are sorted together. 4) The coefficients and right-hand sides are summed, expanded to full-stored form, and written on tape. 5) The full-stored coefficients and right-hand sides are sorted into equation order. 6) Solutions are obtained by modified Gauss-Seidel iteration (successive over relaxation). 7) Identification from step 2) is merged with the solutions to provide the final listing illustrated in Table 2.

Although the procedure requires generalized inverses of the variance-covariance matrices, which has the effect of deleting one category for each trait, there were some cases where the variance-covariance matrices of the remaining categories were singular or nearly singular. These cases generally occurred when the ignored category had a low frequency. Table 1 presents the raw frequencies as well as the solutions for the frequencies in categories and indicates the four traits that required deletion of two categories, the only solution to the problem attempted. The result was that the two deleted categories effectively were lumped together as a single category.

A major concern before computing was the possible difficulty in obtaining convergence of the solutions in a reasonable number of rounds of iteration because of experience with similar equations for evaluation of calving difficulty (2,3). Therefore, solutions for a small example were examined with slight modifications of the basic iteration procedure. The example included three categories (one was ignored), 21 animals in 10 herd-years by 14 sires. Thus, there were 30 equations after absorption.

The solutions for $\hat{\beta}$ seemed to dominate the equations because of their absolute size relative to sire solutions. Thus, the first attempt was to use the raw frequencies as first or guessed solutions for the solutions for $\hat{\beta}$ and zeros for the sire solutions. A better starting point appeared to be to solve the equations for $\hat{\beta}$ after absorption but ignoring the sire equations and using those solutions as a starting point. Use of a relaxer, however, and the other equations caused the solutions for $\hat{\beta}$ to change considerably in the first few rounds. Therefore, the solutions for $\hat{\beta}$ were forced to be the same for 10 rounds of iteration. The average absolute change of all solutions from the previous round, however, took an upward jump in the 11th round. When the solutions for $\hat{\beta}$ were forced to be the same for the first three rounds, this upward jump did

TABLE 1. Descriptive type traits and their frequencies in the data and as estimated from the analyses.

Trait, category	Frequencies	
	Average	BLUP solutions
Stature		
Upstanding	.408	.385
Intermediate	.543	.560
Low set ^a	.049	.055
Head		
Superior	.254	.254
Acceptable	.646	.652
Plain, coarse ^b	.094	.094
Weak ^b	.006	...
Front end		
Smooth, strong	.322	.315
Desirable	.564	.568
Coarse	.054	.054
Narrow, weak	.036	.041
Low ^a	.024	.022
Back and loin		
Straight, full, wide	.401	.401
Medium strength	.481	.476
Low chine	.017	.019
Weak loin, back ^a	.101	.104
Rump		
Long, wide	.096	.093
Medium	.421	.416
High, coarse	.341	.348
Narrow	.051	.051
Sloping ^a	.091	.092
Hind legs		
Strong, clear, square	.172	.171
Acceptable	.529	.532
Sickled, close	.268	.267
Too light, refined ^b	.021	.030
Coarse, blemished ^b	.008	...
Feet		
Strong	.228	.222
Acceptable	.563	.567
Front toe out	.026	.027
Shallow heel	.118	.119
Weak pasterns ^a	.065	.065
Fore udder		
Moderate, strong	.139	.135
Acceptable	.502	.500
Short	.218	.227
Bulgy, loose ^b	.127	.138
Broken, very loose ^b	.014	...

Trait, category	Frequencies	
	Average	BLUP solutions
Rear udder		
Firm, high	.169	.160
Intermediate	.515	.516
Low	.194	.196
Narrow	.091	.097
Loose, broken ^a	.030	.031
Udder support		
Strong	.398	.377
Acceptable	.465	.479
Floor low	.015	.017
Tilted	.103	.107
Broken ^a	.019	.020
Udder quality		
Soft, pliable	.567	.562
Intermediate	.391	.397
Not determinable	.026	.025
Meaty ^b	.015	.016
Persistent edema ^b	.001	...
Teat size, placement		
Plumb, desirable	.403	.384
Acceptable	.379	.387
Wide front	.129	.138
Other undesirable	.050	.051
Objectionable ^a	.039	.040

^aIndicates one category deleted in analysis.

^bIndicates two categories deleted in analysis.

not occur and did not seem as likely to cause problems if the solutions otherwise might converge by close to 10 rounds. Several relaxers also were tried including 1.0 (usual Gauss-Seidel) and ranging otherwise from 1.5 to 1.8. The smallest number of rounds of iteration was for a relaxer of 1.7. The combination of good initial estimates of the solutions for $\hat{\beta}$ and the relaxer of 1.7 gave a smaller sum of absolute changes in 14 rounds than 50 rounds of usual Gauss-Seidel iteration. The standard modification for the complete data set for all traits was to hold the initial solutions for $\hat{\beta}$ for the three rounds and use a relaxer of 1.7.

There was at least one case when the sum of absolute changes from round to round began to become larger after several rounds of iteration which indicated that convergence would not occur. Other experience (Roger Cady, personal communication) indicated that changing the

TABLE 2. Best linear unbiased prediction of future daughter frequencies for stature.

Sire reg. #	No. daughters	Unadjusted daughter frequencies, category			Predicted future frequencies, category		
		1	2	3	1	2	3
Population	12,838	.408	.543	.049	.385	.560	.055
102954	39	.205	.692	.103	.261	.652	.087
107915	35	.629	.371	0	.505	.471	.024
109152	32	.563	.438	0	.519	.461	.020
110327	36	.389	.528	.083	.394	.548	.058
114123	80	.425	.550	.025	.409	.549	.042
116497	38	.447	.500	.053	.410	.540	.050
117219	13	.231	.692	.077	.337	.598	.065
117640	11	.364	.636	0	.383	.567	.050
118619	76	.342	.579	.079	.334	.592	.074
121816	206	.340	.617	.044	.342	.606	.052
156561	11	.455	.545	0	.413	.543	.044
156805	12	.333	.667	0	.384	.569	.047
157024	17	.529	.412	.059	.427	.526	.047
157399	13	.231	.769	0	.370	.578	.052
158105	12	.500	.500	0	.432	.527	.041
158262	13	.538	.462	0	.478	.493	.029
158853	10	.400	.600	0	.435	.526	.039
159857	43	.326	.605	.070	.342	.593	.065
159978	10	0	.300	.700	.164	.690	.146
161653	10	.200	.800	0	.346	.600	.054

relaxer to 1.00 (equivalent of usual Gauss-Seidel) would solve that problem. Thus, programming was incorporated to change the relaxer to 1.00 if the sum of absolute changes increased. More efficient iteration may result by incorporating a larger relaxer after solutions begin to converge again, but this was not attempted.

The time to obtain solutions for a trait was naturally dependent on the number of categories as well as the number of rounds of iteration. Timing was dependent on other jobs the computer was doing, but generally the time for steps 2 through 5 was proportional to the number of categories. For three categories (two analyzed) the time required was about 22 min, for five categories (four analyzed) the time needed was about 36 min. Timing for iteration was nearly proportional to the square of the number of equations. For 1426 equations (two analyzed categories) with 78,908 nonzero coefficients time required was about 21 min; for 2139 equations, time was about 49 min; and for 2852 equations with 315,632 nonzero coefficients the time needed was about 81 min for 25 rounds of iteration. Faster methods of

iteration would reduce costs.

Analysis of final score took about 10 min for steps 2 through 5 and 5 min for iteration, which also would be the time required if the categories were considered as measurements on a linear trait.

An excerpt of the evaluation for stature is in Table 2 by registration number for the first and last 10 sires that had 10 or more daughters in the evaluation. Changes from the unadjusted to BLUP predictions of category frequencies generally show the typical pattern of regression toward mean frequencies. Stature, however, is a high heritability trait as compared with some other traits that generally show a more pronounced regression towards mean frequencies.

In a few cases predicted frequencies of progeny were slightly negative for categories with a small mean frequency. There is nothing in the procedure that forces positive evaluations for each category, although category frequencies must sum to 100%. Some arbitrary rule probably should be imposed before publication to set the frequency to zero or to a small positive number. A more troublesome problem occurs when two

categories must be combined and the resulting evaluation predicts the frequency of the combined categories. Again, one of the categories is likely to have a small frequency and possibly could be ignored.

Sampling errors of predictions were not obtained but could, perhaps, be approximated by selection index methods.

CONCLUSIONS

Best linear unbiased prediction of frequencies in categories for descriptive type traits of future daughters of sires is feasible. The effectiveness of this system should be compared with predictions from linearized systems of scoring such traits. Assignment of economic values, usefulness in predicting herd survival of daughters, standard errors of prediction, and effect of including effects (e.g., herd-year effects) as fixed to account for association of sires and herds are all factors which should be considered. How to adjust categorical data for age is another problem that was not investigated.

ACKNOWLEDGMENT

The support of the Brown Swiss Cattle Breeders' Association in providing the data made this research possible.

REFERENCES

- 1 Cady, R. A., R. D. Anderson, and L. D. Van Vleck. 1978. Evaluation of sires for calving difficulty. Page 96 in *Genetics research, 1977-78 Rep. Eastern AI Coop.*
- 2 Cady, R. A., and L. D. Van Vleck. 1979. Evaluation of bulls for calving difficulty. Page 74 in *Genetics research, 1978-79 Rep. Eastern AI Coop.*
- 3 Cady, R. A., and L. D. Van Vleck. 1979. Evaluation of Holstein bulls for dystocia using a categorical analysis. *J. Dairy Sci.* 62(Suppl. 1):145. (Abstr.)
- 4 Dickinson, F. N., H. D. Norman, B. G. Cassell, E. P. Barton, and M. L. Dawdy. 1979. Relationships between the proposed uniform type system and the official Guernsey and Jersey classification programs. *J. Dairy Sci.* 62(Suppl. 1):143. (Abstr.)
- 5 Henderson, C. R. 1963. Selection index and expected genetic advance. In *Statistical genetics and plant breeding*. Publ. 982. Nat. Acad. Sci., Nat. Res. Council, Washington, DC.
- 6 Henderson, C. R. 1973. Sire evaluation and genetic trends. Page 10 in *Proc. Anim. Breeding Genet. Symp. in Honor of Dr. Jay L. Lush*. Am. Soc. Anim. Sci., Champaign, IL.
- 7 Henderson, C. R. 1974. General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.* 57:963.
- 8 Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
- 9 Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
- 10 Henderson, C. R., and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43:1188.
- 11 Hoyt, R. S. 1978. Is better information needed to enhance type improvement? *Holstein-Friesian World* 75:41.
- 12 Lee, A. J. 1979. Mixed model, multiple trait evaluation of related sires when all traits are recorded. *J. Anim. Sci.* 48:1079.
- 13 Moreno, A. M. 1978. Heritability estimates for type traits of Brown Swiss cattle. M. S. thesis, Cornell University, Ithaca, NY.
- 14 Norman, H. D., B. G. Cassell, and R. L. Powell. 1979. New procedures for estimating type transmitting ability. *J. Dairy Sci.* 62(Suppl. 1):129. (Abstr.)
- 15 Quass, R. L., and L. D. Van Vleck. 1980. Categorical trait sire evaluation by best linear unbiased prediction of future progeny category frequencies. *Biometrics* 36:117.
- 16 Schaeffer, L. R., M. Hunt, and E. B. Burnside. 1978. Evaluation of Holstein-Friesian dairy sires for conformation. *Can. J. Anim. Sci.* 58:409.
- 17 Schaeffer, L. R., M. S. Hunt, G. C. Smith, and E. B. Burnside. 1977. Sire evaluation for conformation of Canadian Holstein-Friesians. *J. Dairy Sci.* 60(Suppl. 1):148. (Abstr.)
- 18 Ufford, G. R., C. R. Henderson, J. F. Keown, and L. D. Van Vleck. 1979. Accuracy of first lactation versus all lactation sire evaluation by best linear unbiased prediction. *J. Dairy Sci.* 62:603.
- 19 Ufford, G. R., C. R. Henderson, and L. D. Van Vleck. 1978. Derivation of computing algorithms for sire evaluation using all lactation records and natural service sires. *Anim. Sci. Mimeo. Ser. No. 39*, Cornell University, Ithaca, NY.
- 20 Ufford, G. R., C. R. Henderson, and L. D. Van Vleck. 1979. Computing algorithms for sire evaluation with all lactation records and natural service sires. *J. Dairy Sci.* 62:511.
- 21 Van Vleck, L. D., P. J. Karner, and G. R. Wiggans. 1979. Supplemental tables of variances and covariances of type traits of Brown Swiss cattle. Dept. Anim. Sci. Mimeo., Cornell University, Ithaca, NY.
- 22 Van Vleck, L. D., P. J. Karner, and G. R. Wiggans. 1980. Relationships among type traits and milk yield of Brown Swiss cattle. *J. Dairy Sci.* 63:120.
- 23 Vinson, W. E., J. M. White, and R. H. Kliever. 1976. Overall classification as a selection criterion for improving categorical scored components of type in Holsteins. *J. Dairy Sci.* 59:2104.
- 24 White, J. M., W. E. Vinson, and R. H. Kliever. 1976. Procedures in sire evaluations for type of registered Holstein sires. *Holstein-Friesian Assoc. Am., Brattleboro, VT.*