HAL
open science

Linear versus nonlinear methods of sire evaluation for categorical traits : a simulation study

A. Meijering, D. Gianola

To cite this version:

A. Meijering, D. Gianola. Linear versus nonlinear methods of sire evaluation for categorical traits : a simulation study. Génétique sélection évolution, 1985, 17 (1), pp.115-132. hal-00893632

HAL Id: hal-00893632

https://hal.science/hal-00893632v1

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Génét. Sél. Evol., 1985, **17** (1), 115-132

Linear versus nonlinear methods of sire evaluation for categorical traits : a simulation study

A. MEIJERING *, ** and D. GIANOLA *

* *Department of Animal Science, University of Illinois, Urbana, Illinois 61801, U.S.A.*
** *On leave from : Research Institute for Animal Production « Schoonoord » 3700 AM Zeist, The Netherlands*

Summary

Linear (BLUP) and nonlinear (GFCAT) methods of sire evaluation for categorical data were compared using Monte Carlo techniques. Binary and ordered tetrachotomous responses were generated from an underlying normal distribution via fixed thresholds, so as to model incidences in the population as a whole. Sires were sampled from a normal distribution and family structure consisted of half-sib groups of equal or unequal size ; simulations were done at several levels of heritability (h^2_y). When a one-way model was tenable or when responses were tetrachotomous, the differences between the 2 methods were negligible. However, when responses were binary, the layout was highly unbalanced and a mixed model was appropriate to describe the underlying variate, GFCAT elicited significantly larger responses to truncation selection than BLUP at h^2_y = .20 or .50 and when the incidence in the population was below 25 p. 100. The largest observed difference in selection efficiency between the 2 methods was 12 p. 100.

Key words : Categorical data, sire evaluation, threshold traits, nonlinear models, simulation.

Résumé

Méthodes linéaires et non linéaires d'évaluation des pères sur des caractères discrets : étude par simulation

Des méthodes linéaires (BLUP) et non linéaires (GFCAT) d'évaluation des pères sur données discrètes ont été comparées à l'aide des techniques de Monte Carlo. On a simulé des réponses selon 2 ou 4 catégories à partir d'une distribution normale sous-jacente munie de seuils fixés. Les pères ont été échantillonnés dans une distribution normale. La structure famille comportait des groupes de demi-germains de taille égale ou inégale. Les simulations ont été effectuées pour plusieurs niveaux d'héritabilité (h^2_y). Les différences entre les 2 méthodes d'évaluation sont négligeables avec un modèle à une voie ou des réponses en 4 classes. Toutefois, en présence de réponses binaires, d'un dispositif fortement déséquilibré et d'une sous-jacente décrite en modèle mixte, la procédure GFCAT procure des réponses après sélection par troncature significativement supérieures à celles obtenues avec le BLUP pour h^2_y = 0,20 et 0,50 et une incidence du caractère dans la population inférieure à 25 p. 100. La différence maximum d'efficacité de sélection observée entre ces deux méthodes s'est située à 12 p. 100.

Mots clés : Données discrètes, évaluation des pères, caractères à seuils, modèle non linéaire, simulation.

I. Introduction

Prediction of genetic merit of individuals from observations on relatives is of basic importance in animal breeding. If the records and the genetic values to be predicted follow a joint normal distribution, best linear unbiased prediction (BLUP) is the method of choice, because it yields the maximum likelihood estimator of the best predictor, it maximizes the probability of correct pairwise ranking (HENDERSON, 1973) and more relevantly, it maximizes genetic progress among translation invariant rules when selecting a fixed number of candidates (GOFFINET, 1983 ; FERNANDO, 1983). However, a number of traits of importance in animal production (e.g., calving ease, livability, disease susceptibility, type scores) are measured as a response in a small number of mutually exclusive, exhaustive and usually ordered categories. These variates are not normally distributed and, in this case, linear predictors may behave poorly for ranking purposes (PORTNOY, 1982). GIANOLA (1980, 1982) discussed additional potential drawbacks of linear predictors for sire evaluation with categorical data, arguing from the viewpoint of « threshold » models for meristic traits (DEMPSTER & LERNER, 1950 ; FALCONER, 1981).

SCHAEFFER & WILTON (1976) examined a modified version of a (fixed) linear model for analysis of categorical data developed by GRIZZLE *et al.* (1969). They suggested that the use of BLUP methodology in sire evaluation for categorical responses might be justified given certain sampling conditions which unfortunately are inconsistent with the assumptions required by their model. This work gave impetus for widespread use of BLUP in evaluation of sires for categorical variates (e.g., BERGER & FREEMAN, 1978 ; VAN VLECK & KARNER, 1979 ; CADY & BURNSIDE, 1982 ; WESTELL *et al.*, 1982).

GIANOLA & FOULLEY (1983a) developed a Bayesian nonlinear method of sire evaluation for categorical variates based on the « threshold » concept. In this approach (GFCAT = *G*ianola-*F*oulley-*Cat*egorical), the probability of response in a given category is assumed to follow a normal integral with an argument dependent on fixed thresholds and on a location parameter in a conceptual underlying distribution. The location parameter is modeled as a linear combination of fixed effects and random variables. Prior information on the distribution of the parameters of the model is combined with the likelihood of the data to yield a posterior density function, the mode of which is then taken as an approximation to the posterior mean or optimum ranking rule in the sense of COCHRAN (1951), BULMER (1980), FERNANDO (1983) & GOFFINET (1983). Solution of the resulting equations requires an iterative implementation. A conceptually similar method has been developed by HARVILLE & MEE (1982). Although these procedures are theoretically appealing, computations are more complicated than those arising in linear methodology.

Although BLUP has become a standard method of sire evaluation in many countries, its robustness to departures from linearity has not been examined. Non linearity arises with categorical data and, therefore, a comparison between BLUP and the procedure developed by GIANOLA & FOULLEY (1983 a) is of interest. The objective of this paper is to present results of a Monte Carlo comparison of the ability of the above 2 methods to rank sires correctly when applied to simulated categorical data.

II. Methodology

A. *Experimental design and simulation of data*

Three experimental settings were considered to compare the 2 methods of evaluation :

1) a one-way sire model with equal progeny group size within a data set ;
2) a one-way sire model with unequal progeny group size within a data set ; and
3) a mixed model with unequal group size within a data set.

In the 1st setting 36 independent data sets were generated per replicate. These data sets represented all combinations of 3 progeny group sizes (10, 50 or 250 progeny records for each of 50 sires), 3 levels of heritability in a conceptual underlying scale (h^2_y = 0.05, 0.20 or 0.50), and 4 types of categorization which will be described later. Phenotypic values in the underlying scale were generated (RÖNNINGEN, 1974 ; OLAUSSON & RÖNNINGEN, 1975) as :

$$y_{ij} = \frac{1}{2} h_y a_i + a_{ij} \sqrt{\left(1 - \frac{1}{4} h^2_y\right)} \qquad [1]$$

where :

y_{ij} : phenotype of individual j in progeny group i, with $y_{ij} \sim N(0,1)$;
h^2_y : heritability in the underlying scale ;
a_i : standard normal random variate common to all individuals in progeny group i with $a_i \sim N(0,1)$, and
a_{ij} : standard normal random variate for individual j in progeny group i, with $a_{ij} \sim N(0,1)$.

The phenotypes y_{ij} were categorized using fixed thresholds in the standard normal distribution function. The first 3 categorizations reflected either a 1 p. 100 ($y_{ij} > 2.33$), 5 p. 100 ($y_{ij} > 1.65$) or 25 p. 100 ($y_{ij} > 0.68$) incidence of a binary trait in the population as a whole. The 4th type of categorization created a tetrachotomous trait reflecting incidences of 40 p. 100-40 p. 100-15 p. 100-5 p. 100 in the population as a whole ; this was made using 3 thresholds ($y_{ij} \leq -.25$; $-.25 < y_{ij} \leq .84$; $.84 < y_{ij} \leq 1.65$; $y_{ij} > 1.65$). Binary responses were coded as 0-1, and tetrachotomies were coded using the integer values 1 to 4. The difference in heritability in a categorical scale resulting from using integer verus « optimal » scores is negligible (GIANOLA & NORTON, 1981).

In the 2nd setting 12 independent data sets were generated per replicate, representing all combinations of the above levels of heritability and categorization. However, the 50 progeny groups represented in each data set varied between 5 and 250 in steps of 5. Data were simulated as outlined for Setting 1.

In Setting 3, 15 independent data sets were generated per replicate. Combinations of the 3 heritability levels with a 10 p. 100 incidence level ($y_{ij} > 1.28$) of a binary trait were added to those used in Setting 2. Data were generated as before. Prior to categorization, the effects of 2 fixed classifications, factor A (2 levels) and factor B (10 levels), were superimposed, as indicated in table 1. Each progeny group was almost equally represented in the levels of factor A, but only in 2 levels of factor B (20 p. 100 in

level B_ℓ and 80 p. 100 in level $B_{\ell+1}$; ℓ = 1, 3, 5, 7 or 9). Consequently, 80 p. 100 of the A × B × sire cells had no observations so as to approximate the situation in field data sets. The disconnectedness of data subsets with respect to factor B and sires does not hamper the comparison of predictors of genetic merit, as these are uniquely defined and obtainable regardless of connectedness if the sires are a random sample from one population (FERNANDO *et al.*, 1983). The phenotypic values in the underlying scale modified by the effects of the levels of the A and B factors, were categorized as follows. With $y_{ij} \sim N(0,1)$ as in [1], let :

$$w_{ijk\ell} = y_{ij} + A_k + B_\ell \qquad [2]$$

TABLE 1

Effects of levels of fixed factors A and B, in units of standard deviation, and allocation of progeny to sire groups.

Valeurs des niveaux des effets fixés A et B en unité d'écart-type et répartition des descendances paternelles.

Factor A	Factor B	Progeny group									Total
		1	2	3	4	5	6	7	. . .	50	
1.0	1.0	1	× (a)	×	×	×	3	×	. . .	×	120
	0.8	2	×	×	×	×	12	×	. . .	×	470
	0.6	×	1	×	×	×	×	4	. . .	×	125
	0.4	×	4	×	×	×	×	14	. . .	×	490
	0.2	×	×	2	×	×	×	×	. . .	×	130
	0.0	×	×	6	×	×	×	×	. . .	×	510
	− 0.2	×	×	×	2	×	×	×	. . .	×	135
	− 0.4	×	×	×	8	×	×	×	. . .	×	530
	− 0.6	×	×	×	×	3	×	×	. . .	25	140
	− 0.8	×	×	×	×	10	×	×	. . .	100	550
− 1.0	1.0	0	×	×	×	×	3	×	. . .	×	115
	0.8	2	×	×	×	×	12	×	. . .	×	470
	0.6	×	1	×	×	×	×	3	. . .	×	120
	0.4	×	4	×	×	×	×	14	. . .	×	490
	0.2	×	×	1	×	×	×	×	. . .	×	125
	0.0	×	×	6	×	×	×	×	. . .	×	510
	− 0.2	×	×	×	2	×	×	×	. . .	×	130
	− 0.4	×	×	×	8	×	×	×	. . .	×	530
	− 0.6	×	×	×	×	2	×	×	. . .	25	135
	− 0.8	×	×	×	×	10	×	×	. . .	100	550
		5	10	15	20	25	30	35	. . .	250	6 375

(a) × indicates empty cells.

Clearly, $w_{ijk\ell} \sim N(A_k + B_\ell, 1)$ represents phenotypic values in 20 « sub-populations » corresponding to the filled cells in Table 1. The categories were then formed as :

Incidence and thresholds in N (0,1)	Mapping function	Expected incidence in whole population of table 1
1 p. 100 ; t = 2.33	$w_{ijk\ell} > t$	6.5 p. 100
5 p. 100 ; t = 1.65	$w_{ijk\ell} > t$	15.1 p. 100
10 p. 100 ; t = 1.28	$w_{ijk\ell} > t$	21.6 p. 100
25 p. 100 ; t = .68	$w_{ijk\ell} > t$	34.3 p. 100
40 p. 100 - 40 p. 100 - 15 p. 100 - 5 p. 100 ; $t_1 = .25$, $t_2 = .84$, $t_3 = 1.65$	$w_{ijk\ell} \leq t_1$; $t_1 < w_{ijk\ell} \leq t_2$; $t_2 < w_{ijk\ell} \leq t_3$; $t_3 < w_{ijk\ell}$	43.7 p. 100 - 25.5 p. 100 - 15.7 p. 100 - 15.1 p. 100

In order to limit computing costs, each data set in each setting was replicated 10 times. Further replication depended on the Monte Carlo estimates of the difference between methods of evaluation and of its sampling variance based on the first 10 replicates.

B. *Methods of sire evaluation and computing procedures*

1) In sire evaluations with linear models (BLUP ; HENDERSON, 1973),

$$\mathbf{x} = \mathbf{1}\mu + \mathbf{Zu} + \mathbf{e} \quad \text{(Settings 1 and 2)} \qquad [3]$$

$$\mathbf{x} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e} \quad \text{(Setting 3)} \qquad [4]$$

where :

- $\mathbf{x}$: vector of categorical responses,
- $\mathbf{1}$: vector of ones,
- μ : fixed effect common to all observations,
- $\mathbf{X}, \mathbf{Z}$: known incidence matrices,
- $\boldsymbol{\beta}$: vector of unknown fixed effects,
- $\mathbf{u}$: vector of unknown sire effects,
- $\mathbf{e}$: vector of residuals.

Assumptions in [3] and [4] were :

$$E(\mathbf{x}) = \mathbf{1}\mu \quad \text{(Settings 1 and 2)} \qquad [5]$$

and :

$$E(\mathbf{x}) = \mathbf{X}\boldsymbol{\beta} \quad \text{(Setting 3)} \qquad [6]$$

Further, in the 3 settings :

$$\mathrm{Var}\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{ZZ}'\,\sigma_u^2 + \mathbf{I}_e\,\sigma_e^2 & \mathbf{Z}\,\sigma_u^2 & \mathbf{I}_e\sigma_e^2\\ & \mathbf{I}_u\sigma_u^2 & \mathbf{0}\\ \text{symm.} & & \mathbf{I}_e\sigma_e^2\end{bmatrix} \qquad [7]$$

where σ_u^2 and σ_e^2 are the sire and residual variances, respectively, and $\mathbf{I}_u$ and $\mathbf{I}_e$ are identity matrices of appropriate order. With progeny consisting of halb-sib groups :

$$\frac{\sigma_e^2}{\sigma_u^2} = \frac{4 - h_c^2}{h_c^2} \qquad [8]$$

where h_c^2 is « heritability in the categorical scale ». The latter was calculated from the « true » underlying heritability (h_y^2) and from the expected incidences for each of the settings using the formula (VINSON *et al.*, 1976 ; GIANOLA, 1979).

$$h_c^2 = h_y^2 \frac{[\sum_{i=1}^{m-1} z_i (w_{i+1} - w_i)]^2}{[\sum_{i=1}^{m} w_i^2 p_i - (\sum_{i=1}^{m} w_i p_i)^2]} \qquad [9]$$

where m is the number of response categories (2 or 4), p_i is the expected incidence in the ith category, $\{z_i\}$ are ordinates of the standard normal density function evaluated at the abscissae corresponding to $\{p_i\}$, and $\{w_i\}$ are the scores assigned to the categories (0-1 or 1-4). Mixed model equations corresponding to the models [3] and [4] were formed using variance ratios as in [8] pertaining to the appropriate levels of heritability used in the simulation. Sire solutions to the mixed model equations were taken as predictors of the transmitting abilities of the 50 sires.

2) In the non linear method (GFCAT ; GIANOLA & FOULLEY, 1983a) the thresholds and the unknown effects which affect location in the conceptual underlying distribution are estimated jointly. The location parameters ($\boldsymbol{\eta}$) were modeled as :

$$\boldsymbol{\eta} = \mathbf{Z}\mathbf{u}^* \qquad \text{(Settings 1 and 2)} \qquad [10]$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{u}^* \qquad \text{(Setting 3)} \qquad [11]$$

where : $\mathbf{X},\mathbf{Z}$: known incidence matrices,
$\boldsymbol{\beta}^*$: vector of unknown fixed effects,
$\mathbf{u}^*$: vector of unknown sire effects.

The parameter vector to estimate was either :

$$\boldsymbol{\theta}' = [\mathbf{t}', \mathbf{u}^{*\prime}] \qquad \text{(Settings 1 and 2)} \qquad [12]$$

$$\boldsymbol{\theta}' = [\mathbf{t}', \boldsymbol{\beta}^{*\prime}, \mathbf{u}^{*\prime}] \qquad \text{(Setting 3)} \qquad [13]$$

In [12] and [13], $\mathbf{t}$ is a vector of unknown fixed thresholds ; t is a scalar when response variables are dichotomous, or a vector of order 3×1 when there are 4 categories of response. Prior information about $\mathbf{t}$ and $\boldsymbol{\beta}^*$ was assumed to be vague, and $\mathbf{u}^* \sim \mathbf{N}(\mathbf{0}, \mathbf{I}h_y^2/4)$. The log-posterior density to maximize is :

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{n} \sum_{k=1}^{m} \delta_{jk} \ell n (P_{jk}) - \frac{1}{2} \mathbf{u}^{*\prime} \mathbf{G}^{-1} \mathbf{u}^* + \text{const.} \qquad [14]$$

where :

n : number of observations,

m : number of categories,

δ_{jk} : Kronecker delta, taking the value 1 if observation j is in category k, and 0 otherwise,

P_{jk} : $\Phi(t_k - \eta_j) - \Phi(t_{k-1} - \eta_j)$, is the probability of response in category k given the location parameter η_j, and $\Phi(.)$ denotes the standard normal distribution function ($t_o = -\infty$, $t_m = \infty$), and

$\mathbf{G}$: Diag $\{h_y^2/4\}$.

The parameters ($\boldsymbol{\theta}$) were estimated iteratively using the modification of the Newton-Raphson algorithm suggested by GIANOLA & FOULLEY (1983a). Starting values used for t were 0 in the case of binary responses, or the threshold values used for categorization into 4 classes when the data were generated. Starting values for $\boldsymbol{\beta}^*$ and $\mathbf{u}^*$ were always zero. In random models, iteration continued until $\boldsymbol{\Delta}' \boldsymbol{\Delta}/p < 10^{-10}$, where $\boldsymbol{\Delta} = \hat{\boldsymbol{\theta}}^{[i]} - \hat{\boldsymbol{\theta}}^{[i-1]}$ is a vector of corrections at the ith iterate, and p is the order of $\boldsymbol{\theta}$. In the mixed model [11] the system does not converge if all responses in a subclass of a fixed effect are in the same extreme category, a problem recognized by HARVILLE & MEE (1982). These authors suggested ignoring the data from such subclasses or to impose upper and lower bounds on the parameter values. In the present study the main interest was in the sire solutions. Because these converge more rapidly than the solutions for $\mathbf{t}$ and $\boldsymbol{\beta}^*$, convergence was monitored by restricting attention to the sire part of the parameter vector. The criterion used was :

$$\{\hat{\mathbf{u}}^{*[i]} - \hat{\mathbf{u}}^{*[i-1]}\}' \{\hat{\mathbf{u}}^{*[i]} - \hat{\mathbf{u}}^{*[i-1]}\}/50 < 10^{-10} \qquad [15]$$

The above test, while suitable for the purpose of this study, cannot be recommended for more general puposes, e.g., field data sets with large numbers of sparsely filled subclasses from combinations of levels of fixed effects.

As the residual standard deviation is the unit of measurement implicit in the method developed by GIANOLA & FOULLEY (1983a), all solutions were multiplied by $\sqrt{1 - h_y^2/4}$ to express them in the scale of the simulation. This, of course, does not affect sire rankings.

C. *Comparison of methods*

The analysis of each data set generated yielded 2 vectors of estimated transmitting abilities (BLUP : $\hat{\mathbf{u}}$; GFCAT : $\hat{\mathbf{u}}^*$) ; the vector of true transmitting abilities ($\mathbf{a}$) was stored during simulation. Sires were ranked using $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}^*$, and the corresponding average true transmitting abilities for the 10 lowest ranking sires were computed ; let these values be $\bar{a}$ and $\bar{a}^*$ for rankings based on $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}^*$, respectively. As the categories of response were scored in ascending order, this is tantamount to selection against a « rare » categorical trait or « lower tail selection ». Because of symmetry, only « lower tail selection » needs to be considered. Further, because $E(a_i) = 0$, $\bar{a}$ and $\bar{a}^*$ can be viewed as expressing « effectiveness » of lower tail selection based on $\hat{\mathbf{u}}$ or $\hat{\mathbf{u}}^*$, or as a realized genetic response. The method of evaluation which on average (over replicates) yields the lowest values ($\bar{a}$ or $\bar{a}^*$) would be preferred.

Differences between $\bar{a}$ and $\bar{a}^*$ were examined using paired t-tests within each of the treatment combinations (i.e., progeny group size × heritability × level of categorization). The statistic used is :

$$t_{r-1} = \frac{\bar{d}}{s_{\bar{d}}} \qquad [16]$$

where :

$$\bar{d} = \frac{\sum_{i=1}^{r} (\bar{a}_i^* - \bar{a}_i)}{r} \qquad [17]$$

$$s_{\bar{d}} = \frac{[\hat{V}ar\,(\bar{a}_i^* - \bar{a}_i)]^{1/2}}{\sqrt{r}} \qquad [18]$$

and r is the number of Monte-Carlo replicates.

Efficiency of selection, i.e., realized genetic progress as a percentage of maximum genetic progress, was also assessed. Maximum genetic progress was defined as the genetic selection differential occurring if the true transmitting abilities were observable. For example, in the case of selection using BLUP evaluations, efficiency of selection was calculated as :

$$\bar{e} = \frac{100}{r} \sum_{i=1}^{r} \left(\frac{\bar{a}_i}{\bar{a}_i^T}\right) \qquad [19]$$

where $\bar{a}_i^T$ is the average transmitting ability of the sires with the lowest 10 true values.

III. Results

A. *Setting 1*

After 2 replications, it became apparent that the 2 procedures, linear and non linear, gave exactly the same ranking of sires when progeny group size was constant and responses were dichotomous. The log-posterior density in GFCAT (GIANOLA & FOULLEY, 1983a) is equal to :

$$L = \sum_{i=1}^{s} \left\{(n - n_i)\, \ell n\, [\Phi\, (t - u_i^*)] + n_i\, \ell n\, [1 - \Phi\, (t - u_i^*)] - \frac{2u_i^{*2}}{h_y^2}\right\} + \text{const.} \qquad [20]$$

where :

n : constant progeny group size,
n_i : number or responses for sire i,
t : unknown threshold, and
s : number of sires.

Substituting $v_i = u_i^* - t$ in [20], v_i and t are solved from :

$$\frac{\partial L}{\partial v_i} = \left[\frac{n_i}{\Phi\, (\hat{v}_i)} + \frac{n_i - n}{1 - \Phi\, (\hat{v}_i)}\right] \phi\, (\hat{v}_i) - \frac{4\, (\hat{v}_i + t)}{h_y^2} = 0 \qquad [21a]$$

and

$$\frac{\partial L}{\partial t} = \sum_{i=1}^{s} - \frac{4\, (\hat{v}_i + \hat{t})}{h_y^2} = 0 \quad \Rightarrow \quad \hat{t} = -\frac{1}{s} \sum_i \hat{v}_i = -\bar{v} \qquad [21b]$$

where : $\phi\, (.)$: normal probability density function.

Observe that $\bar{u}^* = \frac{1}{s} \sum_i u_i^* = \bar{v} + t = 0$

It is informative to express n_i in [21a] as a function of v_i, using [21b] :

$$n_i = \frac{4\, (\hat{v}_i - \bar{v})\, \Phi\, (\hat{v}_i)\, [1 - \Phi\, (\hat{v}_i)]}{h_y^2\, \phi\, (\hat{v}_i)} + n\, \Phi\, (\hat{v}_i) \qquad [22]$$

It can be shown (proof available on request) than n_i is a monotonically increasing function of $\hat{v}_i$, and hence of $\hat{u}_i^*$. It is easy to see that this is the case by replacing $\Phi(v_i)$ by its logistic approximation (GIANOLA & FOULLEY, 1983a) so :

$$n_i = \frac{4(\hat{v}_i - \bar{w})}{h_y^2} + n\,[1 + e - \hat{v}_i]^{-1} \qquad [23]$$

which is clearly a monotonically increasing function of $\hat{v}_i$ and thus of $\hat{u}_i^*$. Because of the monotonicity, as n_i increases, so does $\hat{u}_i^*$. Similarly, in BLUP, when $\mu = 0$, the transmitting ability of the sire is calculated from :

$$\left[n + \frac{4 - h_c^2}{h_c^2}\right]\hat{u}_i = n_i \qquad [24]$$

so $\hat{u}_i$ is a linear and, therefore, monotonically increasing function of n_i. We conclude that for a one-way random model, binary responses and constant progeny group size :

$$n_i > n_j \quad \Rightarrow \quad \hat{u}_i^* > \hat{u}_j^* \quad \Rightarrow \quad \hat{u}_i > \hat{u}_j$$

so GFCAT and BLUP yield exactly the same ranking of sires.

With 4 categories of response and constant progeny group size, BLUP and GFCAT gave, in general, similar sire rankings (table 2). The average difference (eq. [17]) between methods was generally not significant and lower than 2 p. 100, except for

TABLE 2

Effects of lower tail selection (10 sires out of 50) using BLUP and GFCAT [(a)] : one way random model, constant progeny group size and tetrachotomous responses (Average of 10 replicates).

Effets d'une sélection sur la partie inférieure de la distribution (10 pères retenus parmi 50) en utilisant le BLUP ou GFCAT : modèle aléatoire à 1 facteur, nombre constant de descendants, réponse en 4 catégories (moyenne de 10 réplications).

h_y^2	n [(d)]	h_c^2	Mean true T.A. of sires selected by :		SD [(b)]	Significance [(c)]	Efficiency (%)	
			BLUP	GFCAT			BLUP	GFCAT
0.05	10	0.041	− 0.324	− 0.310	0.162	NS	22.9	21.9
	50		− 0.892	− 0.910	0.072	NS	61.6	63.2
	250		− 1.207	− 1.217	0.040	NS	82.1	82.9
0.20	10	0.163	− 0.826	− 0.832	0.085	NS	61.4	61.7
	50		− 1.167	− 1.181	0.043	NS	80.0	81.0
	250		− 1.363	− 1.366	0.009	NS	95.6	95.8
0.50	10	0.407	− 1.038	− 0.977	0.066	*	76.8	72.4
	50		− 1.393	− 1.392	0.021	NS	92.8	92.7
	250		− 1.283	− 1.283	0.015	NS	97.9	97.9

(a) Symbols defined in text.
(b) Standard deviation of difference.
(c) NS : not significant ; * : $P < .05$.
(d) Progeny group size.

$h_y^2 = .50$ and $n = 10$. In this case, BLUP was « better » in 7 of the 10 replicates, and equal to GFCAT in the remaining 3 ; for this combination of h_y^2 and n BLUP was 4.4 p. 100 better than GFCAT, ($p < .05$). However, in view of the overall pattern of results in Table 2, it is doubtful whether this « significance » should be taken seriously. As expected, the efficiency of selection as defined in this paper increased with h_y^2 and, particularly, with n. The results indicate a « consistency » property of the 2 methods : as n increases, BLUP and GFCAT converge in probability to the true transmitting ability of a sire, and more rapidly so at a higher level of heritability.

B. *Setting 2*

When the data were described by a one-way random model and progeny group size was variable (5 to 250 progeny per sire), BLUP and GFCAT did not always yield the same sire rankings (Table 3). However, on the basis of 10 replications, the 2 methods gave virtually similar results, as indicated by the almost null variance of their difference. As in the previous case, the efficiency of selection increased with heritability and incidence, and also with the extent of polychotomization (binary vs. tetrachotomous variables).

TABLE 3

Effects of lower tail selection (10 sires out of 50) using BLUP and GFCAT [(a)] : one way random model with unequal progeny group sizes (Average of 10 replications).

Effets d'une sélection sur la partie inférieure de la distribution (10 pères retenus sur 50) en utilisant le BLUP ou GFCAT : modèle aléatoire à 1 facteur avec un nombre variable de descendants par père (moyenne de 10 réplications).

h_y^2	Incidence (%)	h_c^2	Mean true T.A. of sires selected by :		SD [(b)]	Efficiency (%)	
			BLUP	GFCAT		BLUP	GFCAT
0.05	1	0.004	− 0.397	− 0.397	0	27.5	27.5
	5	0.011	− 0.591	− 0.579	0.036	40.7	39.8
	25	0.027	− 0.990	− 0.990	0	70.3	70.3
	IV [(c)]	0.041	− 1.112	− 1.111	0.075	74.0	74.0
0.20	1	0.014	− 0.670	− 0.674	0.036	49.6	49.8
	5	0.045	− 0.941	− 0.941	0	71.4	71.4
	25	0.108	− 1.115	− 1.115	0	85.9	85.9
	IV [(c)]	0.163	− 1.234	− 1.236	0.005	86.9	87.1
0.50	1	0.035	− 0.879	− 0.879	0	60.8	60.8
	5	0.112	− 1.143	− 1.124	0.044	81.3	79.9
	25	0.269	− 1.257	− 1.261	0.013	89.4	89.7
	IV [(c)]	0.407	− 1.381	− 1.390	0.031	95.3	95.8

(a) Symbols defined in text.
(b) Standard deviation of difference.
(c) Tetrachotomous response (40 p. 100 - 40 p. 100 - 15 p. 100 - 5 p. 100).

C. *Setting 3*

Under the more realistic assumptions of this setting, GFCAT performed significantly better than BLUP when responses were binary, heritability in the underlying scale was moderate (h_y^2 = .20) or high (h_y^2 = .50), and when low incidences (1 p. 100, 5 p. 100) were used to categorize the underlying variate (Table 4). GFCAT was also better when h_y^2 = .50 and incidence was 10 p. 100. In these instances, the increase in efficiency ranged between 3.9 p. 100 (h_y^2 = .50 and 5 p. 100 incidence) to 12.2 p. 100 (h_y^2 = .20 and 1 p. 100 incidence). The 2 methods did not differ significantly at h_y^2 = .05, or when the incidence of a binary trait was 25 p. 100, or when the response was tetrachotomous.

TABLE 4

Effects of lower tail selection (10 sires out of 50) using BLUP and GFCAT [a] : mixed model with unequal progeny group sizes (Average of 20 replicates).

Effets d'une sélection sur la partie inférieure de la distribution (10 pères retenus parmi 50) en utilisant le BLUP ou GFCAT : modèle mixte avec un nombre variable de descendants par père (moyenne de 20 réplications).

h_y^2	Incidence (%)	h_c^2	Mean true T.A. of sires selected by :		SD [b]	Significance [c]	Efficiency (%)	
			BLUP	GFCAT			BLUP	GFCAT
0.05	1	0.013	− 0.594	− 0.611	0.159	NS	42.7	44.3
	5	0.021	− 0.675	− 0.682	0.093	NS	47.1	47.6
	10	0.026	− 0.757	− 0.732	0.108	NS	53.4	51.6
	25	0.034	− 0.875	− 0.860	0.081	NS	59.4	58.3
	IV [d]	0.040	− 0.791	− 0.782	0.093	NS	60.1	59.3
0.20	1	0.053	− 0.646	− 0.771	0.237	*	47.0	55.8
	5	0.085	− 0.868	− 0.942	0.134	*	63.6	69.0
	10	0.102	− 0.955	− 0.978	0.094	NS	69.9	71.5
	25	0.136	− 1.108	− 1.116	0.093	NS	77.0	77.5
	IV	0.161	− 1.080	− 1.081	0.083	NS	77.5	77.5
0.50	1	0.134	− 0.775	− 0.950	0.210	**	55.6	67.8
	5	0.215	− 1.007	− 1.063	0.110	*	73.5	77.4
	10	0.255	− 1.080	− 1.203	0.208	*	74.4	82.3
	25	0.339	− 1.129	− 1.124	0.066	NS	86.3	86.2
	IV	0.402	− 1.264	− 1.276	0.072	NS	89.2	89.7

(a) Symbols defined in text.
(b) Standard deviation of difference.
(c) * $p < .05$; ** $p < .01$; NS : not significant.
(d) Tetrachotomous response (see text).

As pointed out before, the intended incidence levels in the mixed model setting do not correspond to the realized incidence levels ; the reason for this is that each combination of fixed effects represents a distinct statistical population.

IV. Discussion

This study addressed ranking properties of linear (BLUP) and non-linear (GFCAT) methods of sire evaluation for dichotomous or ordered categorical responses. The end-point measured was the Monte Carlo realized response to truncation selection upon predicted sire values. The impetus for the study was provided by shortcomings expected in theory when linear predictors are used with categorical responses (GIANOLA, 1980, 1982) ; these shortcomings are addressed by GFCAT. As BLUP has become in many countries the standard procedure for sire evaluation, a change in methodology for certain traits could be justified only if the alternative method, in this case GFCAT, leads to improved selection decisions. This was the rationale for the choice of end-point measured.

Under normality, BLUP is the maximum likelihood estimator of E ($\mathbf{u} \mid \mathbf{y}$) or best predictor (HENDERSON, 1973). The best predictor maximizes the correlation between true and predicted values, or accuracy of selection (HENDERSON, 1973 ; BULMER, 1980). In order to illustrate, consider a one-way sire model with known mean. If the sires are unrelated, the squared accuracy of selection for the ith sire, using the best linear predictor as a ranking rule, is :

$$\rho_i^2 = n_i \Big/ \left[n_i + \frac{4 - h_c^2}{h_c^2} \right] \qquad [25]$$

However, under the threshold model and with binary responses (DEMPSTER & LERNER, 1950) :

$$h_c^2 = \frac{\phi^2 (t)}{\Phi (t) [1 - \Phi (t)]} h_y^2 = w\, h_y^2 \qquad [26]$$

where $t = \Phi^{-1} (\alpha)$ is the inverse probability transformation corresponding to an overall incidence α in the population. Using [26] in [25], it is clear that ρ_i^2 increases with w at a given h_y^2. However w is maximum when $t = 0$ (α = 50 p. 100), and symmetric about this value. Hence, ρ_i^2 is frequency dependent, and the accuracy of selection of a linear predictor declines as α departs from 50 p. 100, irrespective of the direction. Although ρ_i is only an approximate measure of efficiency of selection when E ($\mathbf{u} \mid \mathbf{y}$) is not linear in $\mathbf{y}$ (BULMER, 1980), the above argument illustrates the impact of the incidence of a binary trait on efficiency of selection (see, for example, table 3). In GFCAT, the posterior density is well approximated by a multivariate normal distribution as the margins of the contingency table (GIANOLA & FOULLEY, 1983a) become large. In a one way-sire model, the squared accuracy of selection with GFCAT is approximately :

$$\rho_i^{*2} = n_i w_i^* \Big/ \left[n_i w_i^* + \frac{4 - h_y^2}{h_y^2} \right] \qquad [27]$$

where :

$$w_i^* = \frac{\phi^2 (t - u_i)}{\Phi (t - u_i) [1 - \Phi (t - u_i)]} \qquad [28]$$

and u_i is the transmitting ability of the ith sire in the underlying scale. Note that the accuracy of selection depends not only on n_i and h_y^2 but on the distance between the true transmitting ability of the ith sire and the threshold. This is automatically estimated

in GFCAT and not taken into account in BLUP. Nevertheless, [27] is maximum when $t = u_i$, and decreases as the proportion of the progeny of the sire exhibiting a response deviates from 50 p. 100. This is also borne out by the results in table 3. All in all, the results in tables 2 and 3 clearly suggest that BLUP, as measured by the criterion considered in this study, is a very satisfactory method of prediction of breeding value for categorical responses when the one-way sire evaluation model is tenable. In view of the lower computational requirements of BLUP relative to GFCAT, the adoption of non linear methodology is difficult to justify in this type of sampling scheme.

In one-way layouts, many assumptions violated by linear models when applied to binary responses are not strained (GIANOLA, 1980, 1982). For example, the phenotypic variance, Φ (t) [1 − Φ (t)], is homogeneous. This is not true in the mixed model situation where, in the usual notation (e.g., GIANOLA & FOULLEY, 1983b), the residual variance is Φ ($\mathbf{X\beta} + \mathbf{Zu}$) [1 − Φ ($\mathbf{X\beta} + \mathbf{Zu}$)]. When a mixed model was applied to generate and to analyze the data, GFCAT was significantly better than BLUP in a number of heritability-incidence combinations for binary responses (table 4). This occurred at $h_y^2 = .20$ and .50, and when incidence was low. Note that at these levels of h_y^2, the heritability in the « observed » scale for the significant comparisons varied between .05 and .26, depending on the incidence. The range of incidences encompassed by the significant comparisons was 1 p. 100 (6.5 p. 100 of « effective incidence » ; see previous sections) to 10 p. 100 (21.6 p. 100 of « effective incidence »). It is not immediately obvious, at least when responses were binary, why « significance » occurred for some treatment combinations but not for others. Because a plot of the standard normal distribution function against its argument is particularly non linear in the tails, we conjecture that a linear approximation is fairly robust at intermediate frequencies, say 20 to 80 p. 100, but breaks down otherwise. The levels of incidence (1-10 p. 100, or effectively 6.1 p. 100-21 p. 100) and the « observed » heritabilities (.05-.26) at which « significances » occurred, suggest that GFCAT should be considered for application to genetic evaluation of binary traits related to reproduction and fitness, e.g., calf survival, conception rate, or abortion rate under tropical or sub-tropical conditions (A. MENENDEZ, Cuba ; personal communication). When responses were tetrachotomous the 2 methods did not differ significantly for any of the treatment combinations considered. This suggests that the linear combination $\mathbf{w'v}$ ($\mathbf{w}$: vector of scores ; $\mathbf{v}$: 4 × 1 vector containing the observations in the 4 categories for a particular subclass) tends to normality rapidly so that a linear approximation does not result in any appreciable loss in response to selection.

A conceptual difficulty encountered when implementing the linear analysis in the simulation under the assumptions of a mixed model, was arriving at a meaningful value of h_c^2. In a single population problem, h_c^2 can be readily calculated from h_y^2 and from the incidences in the population (ROBERTSON, 1950 ; VINSON *et al.*, 1976 ; GIANOLA, 1979) ; simulation studies conducted by VAN VLECK (1972) and OLAUSSON & RÖNNINGEN (1975) suggest that this approximation is fairly accurate, at least for binary responses. However, under a mixed model, there are as many h_c^2's as there are combinations of levels of fixed effects or sub-populations (GIANOLA, 1980, 1982). This implies that the variance ratio used in BLUP would need to vary from sub-population to sub-population. However, because a sire leaves progeny in many sub-populations, this poses the problem of which variance ratio applies to which sire. The approach taken in this paper, e.g., for binary responses, was to approximate h_c^2 as :

$$h_c^2 = \frac{h_y^2 \, \bar{\phi}^2}{\bar{\Phi}\,(1 - \bar{\Phi})} \qquad [29]$$

where :

$\bar{\Phi}$: $\sum_i p_i \Phi_i$, with Φ_i being the incidence in the sub-population i ;

p_i : proportion of observations in the data set in the ith sub-population ;

$\bar{\phi}$: ordinate of the standard normal density function appropriate to $\bar{\Phi}$.

TABLE 5

Efficiency (p. 100) of lower tail selection (10 sires out of 50) using BLUP, as a function of « operational » heritability : mixed model setting and dichotomous responses (Average of 10 replications).

Efficacité (en p. 100) d'une sélection par le BLUP sur la partie inférieure de la distribution (10 pères retenus parmi 50) en fonction de la valeur opérationnelle de l'héritabilité : modèle mixte et réponses tout-ou-rien (moyenne de 10 réplications).

Incidence	« Operational » heritability	Underlying heritability		
		.05	.20	.50
1 p. 100	$\left(\frac{1}{2}\right) h_c^2$	40.6	58.8	54.4
	$\left(\frac{3}{4}\right) h_c^2$	40.8	57.2	53.2
	h_c^2	40.9	56.3	52.9
	$\left(\frac{5}{4}\right) h_c^2$	41.4	55.2	51.6
	$\left(\frac{3}{2}\right) h_c^2$	40.2	54.0	50.8
5 p. 100	$\left(\frac{1}{2}\right) h_c^2$	37.2	70.0	75.7
	$\left(\frac{3}{4}\right) h_c^2$	38.0	71.7	77.1
	h_c^2	39.5	70.3	77.1
	$\left(\frac{5}{4}\right) h_c^2$	39.3	68.9	77.1
	$\left(\frac{3}{2}\right) h_c^2$	39.2	69.3	77.2

While this is a heuristic solution, to which alternatives exist, the difficulty of modeling correctly the threshold concept with linear models is well illustrated. It is possible to speculate that use of an « incorrect » heritability might have affected negatively the effectiveness of selection using BLUP as a sire ranking criterion. Mixed linear

model predictors of breeding value are believed to be insensitive to changes in heritability (FREEMAN, 1979). We examined this problem for the case of binary responses via an additional simulation. Six data sets were generated under the mixed model assumptions of Setting 3 ; the data sets corresponded to 2 levels of incidence in the population (1 p. 100 or 5 p. 100), and 3 levels of heritability (h_y^2 = .05, .20 or .50). In each data set, the value of h_c^2 used in the mixed model computations was varied from 50 p. 100 to 150 p. 100 of the « true » h_c^2. As shown in table 5, the efficiency of selection averaged over 10 replications was virtually insensitive to the value of h_c^2 used in the calculations.

GFCAT is based on the concept of an underlying continuous distribution of genetic merits and environmental influences. The data simulation procedure applied implicitly assumes this concept to be reality. In many cases the evidence to support this concept may not be sufficient however. A trait may be categorical down to the genetic level (e.g. halothane sensitivity in pigs). On theoretical grounds both methods compared would be hard to justify then.

V. Conclusions

The results of this study indicate that a non linear method of sire evaluation for categorical responses, GFCAT, does not always outperform BLUP. In the one-way sire evaluation models, differences between methods, as measured by the ability to elicit larger response to selection, were negligible. However, this type of layout is seldom realistic in practice. Similarly, when responses were polychotomous, i.e., more than 2 meaningful categories of response, there was little difference between methods, irrespective of the model used to generate and to analyze the data. In the above cases, given the additional computational requirements of GFCAT and the apparent robustness of BLUP, it seems doubtful that non linear methodology could be justified from a practical point of view.

When a mixed model was required to describe variation of binary responses, GFCAT performed significantly better than BLUP when heritability in the conceptual underlying scale was moderate to high, and when the expected incidence was below 25 p. 100 ; for some combination of parameters the gain in efficiency of selection amounted to 12 p. 100. Hence, it appears that for the type of sampling situations that arise in animal breeding practice, non linear methods should be given serious consideration for the analysis of binary responses. At least in theory, it is expected that the superiority of GFCAT over BLUP would be proportional to the number of fixed effects required in the model and to the extent of heterogeneity in incidence across statistical sub-populations.

The cost of data processing is usually small relative to the other outlays associated with a large scale breeding program, e.g., field personnel, testing facilities, and overhead costs. Small increases in accuracy of selection stemming from improved evaluations of candidates are usually cost effective because the total cost of the evaluation is increased only to a limited extent. Further, the improved evaluations have multiplicative effects as potentially increased returns from improved stock are spread industry-wide (albeit unequally among tiers) and over generations. In the data sets considered in the present study between 5 to 10 rounds of iteration were required to attain convergence using GFCAT. This may give an indication of the additional computational requirements of the non linear methodology.

In the present study, the underlying variance-covariance structure was known. This might have favored GFCAT somewhat over BLUP because the prior distribution for the former method could be specified without « error ». As pointed out previously, when applying linear methodology to categorical responses a « heritability » value needs to be contrived ; in the general case it is not obvious how to do this. However, the results displayed in table 5 suggest that this issue should not have been an important source of difference between the 2 methods. On the other hand, it is possible that the linear methodology was « helped » by using a contrived value of heritability. This is so because of the inability of linear methods to account for the relationship between mean and variance arising in categorical responses. Methods for estimating variance components in mixed linear models (at least 10 such methods have been described in the literature !) may give severely biased estimates of the underlying variance-covariance structure. This is an area for further work. Clearly, a single simulation study cannot address all possible combinations of parameters, data structures, models and methods.

A question of considerable interest is the ability of BLUP versus GFCAT to account for selection bias. While under normality BLUP is unbiased by selections based on translation invariant functions of the records (HENDERSON, 1973 ; FERNANDO, 1983 ; GOFFINET, 1983), this property does not hold for other distributions, e.g. multinomial. FERNANDO (1983) has shown that when selecting a fixed number of candidates, genetic progress is maximized by ranking individuals with conditional means, calculated as if selection had not occurred, irrespective of the number of stages or of generations involved in the selection program. With categorical responses, GFCAT can be thought of as an approximation to the posterior mean or conditional expectation of the predictands given the data. We conjecture that GFCAT should be less prone to bias than BLUP for categorical data in a population undergoing selection.

Acknowledgements

The contributions that ROHAN L. FERNANDO & STEPHEN D. KACHMAN made to this study are sincerely appreciated. A. MEIJERING wishes to express his gratitude to the Department of Animal Science, University of Illinois, for support and hospitality during his stay in Urbana. The study was partially supported by the Holstein Association, Brattleboro, Vermont, U.S.A.

Received May 16, 1984.

Accepted August 22, 1984.

References

BERGER P.J., FREEMAN A.E., 1978. Prediction of sire merit for calving difficulty. *J. Dairy Sci.*, **61**, 1146-1150.

BULMER M.G., 1980. *The mathematical theory of quantitative genetics*, 254 pp., Clarendon Press, Oxford.

CADY R.A., BURNSIDE E.B., 1982. Evaluation of dairy bulls in Ontario for calving ease of offspring. *J. Dairy Sci.*, **65**, 2150-2156.

COCHRAN W.G., 1951. Improvement by means of selection. *Proceedings of the second Berkeley Symposium on mathematical statistics and probability*. Neyman J. (Ed.), 449-470, University of California Press, Berkeley CA.

DEMPSTER E.R., LERNER I.M., 1950. Heritability of threshold characters. *Genetics*, **35**, 212-236.

FALCONER D.S., 1981. *Introduction to quantitative genetics*, 340 pp., Longman, New York.

FERNANDO R.L., 1983. *Optimal rules for selection*. Mimeo. Department of Animal Science, University of Illinois, Urbana.

FERNANDO R.L., GIANOLA D., GROSSMAN M., 1983. Identifying all connected subsets in a two-way classification without interaction. *J. Dairy Sci.*, **66**, 1399-1402.

FREEMAN A.E., 1979. Components of variance : Their history, use and problems in animal breeding. *Proceedings of a conference in honor of C.R. HENDERSON, Ithaca NY, July 16-17, 1979*, 43-57. Cornell University, Ithaca, NY.

GIANOLA D., 1979. Heritability of polychotomous characters. *Genetics*, **93**, 1051-1055.

GIANOLA D., 1980. A method of sire evaluation for dichotomies. *J. Anim. Sci.*, **51**, 1266-1271.

GIANOLA D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.*, **54**, 1079-1096.

GIANOLA D., FOULLEY J.L., 1983a. Sire evaluation for ordered categorical data with a threshold model. *Génét. Sél. Evol.*, **15**, 201-224.

GIANOLA D., FOULLEY J.L., 1983b. New techniques of prediction of breeding value for discontinuous traits. *Paper presented at the 32nd Annual National Breeders Roundtable, May 6, 1983*, Saint-Louis, Missouri.

GIANOLA D., NORTON H.W., 1981. Scaling threshold characters. *Genetics*, **99**, 357-364.

GOFFINET B., 1983. Selection on selected records. *Génét. Sél. Evol.*, **15**, 91-97.

GRIZZLE J.E., STARMER C.F., KOCH G.C., 1969. Analysis of categorical data by linear models. *Biometrics*, **25**, 489-504.

HARVILLE D.A., MEE R.W., 1982. *A mixed model procedure for analyzing ordered categorical data*. Mimeo, Department of Statistics, Iowa State University, Ames.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. *Proceedings of the animal breeding and genetics symposium in honor of Dr. JAY L. LUSH, Blacksburg, Virginia, July 29, 1972*. American Society of Animal Science and American Dairy Science Association, 10-41, Champaign, IL.

HENDERSON C.R., 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-449.

OLAUSSON A., RÖNNINGEN K., 1975. Estimation of genetic parameters for threshold characters. *Acta Agric. Scand.*, **25**, 201-208.

PORTNOY S., 1982. Maximizing the probability of correctly ordering random variables using linear predictors. *J. Mult. Anal.*, **12**, 256-269.

ROBERTSON A., 1950. Appendix of « Heritability of threshold characters ». *Genetics*, **35**, 212-236.

RÖNNINGEN K., 1974. Monte Carlo simulation of statistical-biological models which are of interest in animal breeding. *Acta Agric. Scand.*, **24**, 135-142.

SCHAEFFER L.R., WILTON J.W., 1976. Methods of sire evaluation for calving ease. *J. Dairy Sci.*, **59**, 544-551.

VAN VLECK L.D., 1972. Estimation of heritability of threshold characters. *J. Dairy Sci.*, **55**, 218-225.

VAN VLECK L.D., KARNER P.J., 1980. Sire evaluation by best linear unbiased prediction for categorically scored type traits. *J. Dairy Sci.*, **63**, 1328-1333.

VINSON W.E., WHITE J.M., KLIEWER R.H., 1976. Overall classification as a selection criterion for improving categorically scored components of type in Holsteins. *J. Dairy Sci.*, **59**, 2104-2114.

WESTELL R.A., BURNSIDE E.B., SCHAEFFER L.R., 1982. Evaluation of Canadian Holstein Friesian sires on disposal reasons of their daughters. *J. Dairy Sci.*, **65**, 2366-2372.