

Heteroskedastic threshold models with applications to the analysis of calving difficulties

JL Foulley.

INRA Quantitative Genetics, 78352 Jouy-en-Josas Cedex, F

Abstract

In the standard threshold model, differences among statistical subpopulations in the distribution of ordered polychotomous responses are modeled via differences in location parameters of an underlying continuous scale. A new model is proposed whereby subpopulations can also differ in dispersion parameters. Heterogeneity in such parameters is described using a structural linear model and a loglink function involving continuous or discrete covariates. Inference (estimation, testing procedures, goodness-of-fit) about parameters in fixed effects models is based on likelihood procedures. Bayesian techniques are also described to deal with mixed effects model structures. Numerical applications to the analysis of calving difficulties in cattle are shown and discussed.

1. Introduction

An appealing model for the analysis of ordered categorical data is the so-called threshold model. The threshold model postulates an underlying (liability) continuous distribution rendered discrete via threshold values. The probability of response in a given category can be expressed as the difference between continuous cumulative distribution functions having as arguments the upper and lower thresholds minus the mean liability for subpopulation divided by the corresponding standard deviation. Usually the standard deviations are assumed to be known and equal, or proportional to known quantities. The purpose of this paper is to extend the standard procedure (S-TM) to a heteroskedastic threshold model (H-TM) with a structural approach on the unknown dispersion parameters. For simplicity, the theory will be presented using a fixed effects model and likelihood procedures for inference. Mixed model extensions based on Bayesian techniques will also be outlined.

2. Theory

2.1. Statistical model

The overall population is assumed to be stratified into several subpopulations (*eg*: subclasses of sex, parity, age, genotypes, etc) indexed by $i = 1, 2, \dots, I$ representing potential sources of variation. Let J be the number of ordered response categories indexed by j , and $\mathbf{y}_{i+} = \{y_{ij+}\}$ be the $(J \times 1)$ vector whose

element y_{ij+} is the total number of responses in category j for subpopulation i . The vector \mathbf{y}_{i+} can be written as a sum

$$\mathbf{y}_{i+} = \sum_{r=1}^{n_i} \mathbf{y}_{ir}, \quad [1]$$

(over the n_i observations made in subpopulation i) of indicator vectors $\mathbf{y}_{ir} = (y_{i1r}, y_{i2r}, \dots, y_{ijr}, \dots, y_{iJr})$ such that $y_{ijr} = 1$ or 0 depending on whether a response for observation (r) in population (i) is in category (j) or not.

Given n_i independent repetitions of \mathbf{y}_{ir} , the sum \mathbf{y}_{i+} is multinomially distributed

$$\mathbf{y}_{i+} \sim M(n_i, \mathbf{\Pi}_i), \quad [2]$$

with parameters $n_i = \sum_{j=1}^J y_{ij+}$ and probability vector $\mathbf{\Pi}_i = \{\Pi_{ij}\}$.

In the threshold model, the probabilities Π_{ij} are connected to the underlying continuous variables X_{ir} with threshold values τ_j via the statement

$$\Pi_{ij} = \Pr(\tau_{j-1} < X_{ir} \leq \tau_j), \quad [3]$$

with $\tau_0 = -\infty$ and $\tau_J = +\infty$, so that there are $J - 1$ finite thresholds.

Assuming $X_{ir} \sim id(\eta_i, \sigma_i^2)$, this becomes:

$$\Pi_{ij} = F\left(\frac{\tau_j - \eta_i}{\sigma_i}\right) - F\left(\frac{\tau_{j-1} - \eta_i}{\sigma_i}\right), \quad [4]$$

where $F(x)$ is the CDF of some continuous distribution. Classical choices are: i) $F(x) = \Phi(x)$ for the normal and ii) $F(x) = [1 + \exp(-x)]^{-1}$ for the logistic.

The mean liability (η_i) for the i^{th} subpopulation is modelled as in Gianola and Foulley (1983) and Harville and Mee (1984), and as in generalized linear models (McCullagh and Nelder, 1989) in terms of the linear predictor

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta}. \quad [5]$$

Here, the vector of unknowns ($\boldsymbol{\beta}$) involves fixed effects only and \mathbf{x}_i is the corresponding ($p \times 1$) vector of qualitative or quantitative covariates.

In the H-TM, a structure is imposed on the scaling parameters. As in Foulley and Quaas (1995), the natural logarithm of σ_i is written as a linear combination of some unknown ($r \times 1$) real-valued vector of parameters ($\boldsymbol{\delta}$),

$$\ln(\sigma_i) = \mathbf{p}_i' \boldsymbol{\delta}, \quad [6]$$

\mathbf{p}_i being the corresponding row incidence vector of qualitative or continuous covariates.

2.2. Estimation

Let $\boldsymbol{\tau} = \{\tau_j\}$ for $j=1,2,\dots, J-1$ and $\boldsymbol{\alpha} = (\boldsymbol{\tau}, \boldsymbol{\beta}', \boldsymbol{\delta}')$. In fixed effects models with multinomial data, inferences about $\boldsymbol{\alpha}$ can be based on likelihood procedures. Here, the loglikelihood $L(\boldsymbol{\alpha}; \mathbf{y})$ can be expressed, apart from an additive constant, as:

$$L(\boldsymbol{\alpha}; \mathbf{y}) = \sum_{i=1}^I \sum_{j=1}^J y_{ij+} \ln(\Pi_{ij}), \quad [7]$$

The maximum likelihood (ML) estimator of $\boldsymbol{\alpha}$ can be computed using a second order algorithm. A convenient choice for multinomial data is the scoring algorithm, because the Fisher information measure is simple here. The system of equations to solve iteratively can be written as:

$$\mathbf{J}(\boldsymbol{\alpha})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{[k]}} \Delta \boldsymbol{\alpha}^{[k+1]} = \mathbf{U}(\boldsymbol{\alpha}; \mathbf{y})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{[k]}}, \quad [8]$$

where $\mathbf{U}(\boldsymbol{\alpha}; \mathbf{y}) = \partial L(\boldsymbol{\alpha}; \mathbf{y}) / \partial \boldsymbol{\alpha}$ and $\mathbf{J}(\boldsymbol{\alpha}) = -E[\partial^2 L(\boldsymbol{\alpha}; \mathbf{y}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}']$ are the score function and the Fisher information matrix, respectively; k is an iterate number. Analytical expressions for the elements of $\mathbf{U}(\boldsymbol{\alpha}; \mathbf{y})$ and $\mathbf{J}(\boldsymbol{\alpha})$

are given in the appendix. for the probit and logit links. These are generalizations of formulae given by Gianola and Foulley (1983) and Harville and Mee (1984).

2.3. Goodness-of-fit and Hypothesis testing

The two usual statistics, the Pearson X^2 and the (scaled) deviance D^* can be used to check the overall adequacy of a model. These are

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J (y_{ij+} - n_i \hat{\Pi}_{ij})^2 / n_i \hat{\Pi}_{ij}, \quad [9]$$

where $\hat{\Pi}_{ij} = \Pi_{ij}(\hat{\boldsymbol{\alpha}})$ is the ML estimate of Π_{ij} , and

$$D^*(y, \hat{\boldsymbol{\alpha}}) = 2[L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\alpha}}; \mathbf{y})]. \quad [10]$$

Above, D^* is based on the likelihood ratio statistic for fitting the proposed model against a saturated model having as many parameters as there are algebraically independent variables in the data vector, *ie* $I(J-1)$ here. Data should be grouped as much as possible for the asymptotic chi-square distribution to hold in [9] and [10] (Mc Cullagh and Nelder, 1989; Fahrmeir and Tutz, 1994). The degrees of freedom to consider here are $I(J-1)$ (saturated model) minus $[(J-1) + \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{P})]$ (model under study), where \mathbf{X} and \mathbf{P} are the incidence matrices for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ respectively. Formulae [9] and [10] are special cases of the power divergent statistics introduced by Read and Cressie (1984).

Hypothesis testing about $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \boldsymbol{\delta}')$ can be carried out via either the Wald test or the likelihood ratio (or deviance) test.

2.4. Numerical application

2.4.1. Material

The data set analyzed was a contingency table of calving difficulty scores (from 1 to 4) recorded on purebred Blonde d'Aquitaine cows (Menissier and Sapa, 1995 personal communication) distributed according to sex of calf (males, females) and calving parity (0=unknown; 1,2 and 3 for first, second, third calvings and above respectively). Scores 3 and 4 were pooled on account of the low frequency of score 4. The distribution of the 12183 records by sex-parity combinations is displayed in table 1. The raw data revealed the usual pattern of highest calving difficulty in male calves out of younger dams.

Method/ Data were analyzed with standard (S-TM) and heteroskedastic (H-TM) threshold models using either the normal or the logistic as the distribution of liability. Location and dispersion parameters were described using fixed models involving sex (S) and

parity (P) as factors of variation. In both cases, inference was based on maximum likelihood procedures. A log-link function was used for standard deviations.

2.4.2. Results

The S-TM does not fit the data whatever the distribution assumed or the model considered (full and additive). The deviance (D) of the full model (S+P+SP) was equal to 22.2 and 52.9 for 7 degrees of freedom with the normal and logistic respectively and the SP interaction effects were not significant (D=1.31, P-value=0.72 and D=1.58, P value=0.66 with 3 degrees of freedom for the normal and logistic respectively). A closer look at the observed and expected numbers indicates that the S-TM leads to an underestimation of difficult calvings in males and an overestimation in females. Fitting a H-TM decreases the chi-square and deviance. A model with even fewer parameters and having a better fit than the full S-TM can be adjusted to the data. This model includes sex and parity as additive covariates for location parameters and sex as the only covariate for dispersion parameters. Liability turns out to be more concentrated around the mean in females than in males, the estimation of the standard deviation of females (vs males set to 1) being smaller under the logistic (0.806) than under the normal (0.885).

3. Mixed models

Correlations can be accounted for conveniently via a mixed model structure on the η_i 's, written now as

$$\eta_i = \mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \mathbf{u}, \quad [11]$$

where the fixed component $\mathbf{x}_i \boldsymbol{\beta}$ is as before, and \mathbf{u} is a ($qx1$) vector of Gaussian random effects with corresponding incidence row vector \mathbf{z}_i .

For simplicity, we will consider a one-way random model, *ie* $\mathbf{u} \sim N(0, \mathbf{A} \sigma_u^2)$ (\mathbf{A} is a positive definite matrix of known elements such as kinship coefficients), but the extension to several u-components is straightforward. The random part of the location is rewritten as in Foulley and Quaas (1995) as $\mathbf{z}_i \sigma_{u_i} \mathbf{u}^*$ where \mathbf{u}^* is a vector of standard normal deviates, and σ_{u_i} is the square root of the u-component of variance, the value of which may be specific to subpopulation i. For instance, the sire variance may vary according to the environment in which the progeny of the sires is raised. Furthermore, it will be assumed that the ratio σ_{u_i} / σ_i , where σ_i is now the residual variance, is constant across subpopulations. In a sire by environment layout, this is tantamount to assuming

homogeneous intraclass correlations (or heritability) across environments, which seems to be a reasonable assumption in practice. Thus, the argument h_{ij} of the CDF in [4] and [7] becomes

$$h_{ij} = \frac{\tau_j - \mathbf{x}_i \boldsymbol{\beta} - \mathbf{z}_i \sigma_{u_i} \mathbf{u}^*}{\sigma_i} = \frac{\tau_j - \mathbf{x}_i \boldsymbol{\beta}}{\exp(\mathbf{p}_i \boldsymbol{\delta})} - \mathbf{z}_i \boldsymbol{\rho} \mathbf{u}^*, \quad [12]$$

where $\boldsymbol{\rho} = \sigma_{u_i} / \sigma_i$.

In the fixed model, parameters $\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\delta}$ were estimated by maximum likelihood. Given $\boldsymbol{\rho}$, a natural extension would be to estimate these and \mathbf{u}^* by the mode of their joint posterior distribution (MAP). To mimic a mixed model structure, one can take flat priors on $\boldsymbol{\tau}, \boldsymbol{\beta}$ and $\boldsymbol{\delta}$. The only informative prior is then on \mathbf{u}^* , *ie*, $\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A})$. Thus

$$\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\mathbf{u}}^* = \text{Argmax}[\mathbf{p}(\mathbf{y} | \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{u}^*, \boldsymbol{\rho}) \mathbf{p}(\mathbf{u}^*)]. \quad [13]$$

MAP solutions can be computed with minor modifications from [8]. The only changes to implement are to replace: i) $\boldsymbol{\beta}$ by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{u}^{\#'})'$ with $\mathbf{u}^{\#} = \boldsymbol{\rho} \mathbf{u}^*$, ii) $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_I)'$ by $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_i, \dots, \mathbf{s}_I)'$ with $\mathbf{s}_i = (\mathbf{x}_i, \sigma_i \mathbf{z}_i)'$; iii) add $\boldsymbol{\rho}^{-2} \mathbf{A}^{-1}$ to the coefficients of the $\mathbf{u} \mathbf{x} \mathbf{u}$ block pertaining to the random effects on the left hand side and $-\boldsymbol{\rho}^{-2} \mathbf{A}^{-1} \mathbf{u}^{[k]}$ to the \mathbf{u} -part of the right hand side. The system to solve can be written as:

$$\begin{bmatrix} \mathbf{T} & \mathbf{L}_\theta \mathbf{S} & \mathbf{L}_\delta \mathbf{P} \\ \mathbf{S}' \mathbf{L}_\theta & \mathbf{S}' \mathbf{W}_{\theta\theta} \mathbf{S} + \boldsymbol{\Sigma}^- & \mathbf{S}' \mathbf{W}_{\theta\delta} \mathbf{P} \\ \mathbf{P}' \mathbf{L}_\delta & \mathbf{P}' \mathbf{W}_{\delta\theta} \mathbf{S} & \mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \end{bmatrix}^{[k]} \begin{bmatrix} \Delta \boldsymbol{\tau} \\ \Delta \boldsymbol{\theta} \\ \Delta \boldsymbol{\delta} \end{bmatrix}^{[k+1]} = \begin{bmatrix} \boldsymbol{\xi} \\ \mathbf{S}' \mathbf{v}_\theta - \boldsymbol{\Sigma}^- \boldsymbol{\theta} \\ \mathbf{P}' \mathbf{v}_\delta \end{bmatrix}^{[k]}, \quad [14]$$

where $\boldsymbol{\Sigma}^- = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\rho}^{-2} \mathbf{A}^{-1} \end{bmatrix}$ and $\mathbf{T}, \mathbf{v}_\theta = \mathbf{v}_\beta$,

$\mathbf{L}_\theta = \mathbf{L}_\beta$, $\mathbf{W}_{\theta\theta} = \mathbf{W}_{\beta\beta}$, $\mathbf{W}_{\theta\delta} = \mathbf{W}_{\beta\delta}$ as expressed in the appendix.

A further step would be to estimate $\boldsymbol{\rho}$ using an EM marginal maximum likelihood procedure based on

$$\boldsymbol{\rho}^{2[k+1]} = \mathbf{E}(\mathbf{u}^{\#'} \mathbf{A}^{-1} \mathbf{u}^{\#} | \mathbf{y}, \boldsymbol{\rho}^{2[k]}) / q, \quad [15]$$

This may involve either an approximate calculation of the conditional expectation of the quadratic in $\mathbf{u}^{\#}$ as in Harville and Mee (1984) or a Monte Carlo calculation of this conditional expectation via, for example, the so-called Stochastic-EM algorithm

(Celeux and Diebolt, 1985) or Monte-Carlo EM (Natarajan, 1995). Alternative procedures for estimating ρ might be also envisioned, such as the iterated re-weighted REML of Engel *et al* (1995). A test example based on simulated calving scores with sire and maternal grandsire effects is shown in table 3 and 4. Notice that genetic evaluation and selection can be directly carried out on MAP values of \mathbf{u}^* (σ_{u_i} units) or $\mathbf{u}^\# = \rho \mathbf{u}^*$ (σ_{e_i} units).

4. Conclusion

The threshold model offers great flexibility for fitting ordinal categorical data such as those arising in animal breeding (eg calving scores). In particular a further step can be achieved by implementing a structural heteroskedastic model for the argument of the CDF describing the cumulative probability of response. Relationships between the H-TM and models with variable thresholds have been already discussed by Foulley and Gianola (1996), the main feature of H-TM being its lower dimensionality and its simpler interpretation.

From a genetic point of view, the potential increase in selection response by selecting on H-TM versus S-TM based EBV's is likely to be not very large but within the usual range of improvement already observed after the implementation of new statistical methods. Preliminary simulation work carried out on a sire - maternal grand sire design indicates an increase of the coefficient of determination of about 4 to 7% depending on the sample characteristics and parameter values. Further research is also needed to assess the question of the number of categories considered in relation to the corresponding misclassification issues or other aspects of subjective appraisal of performance . In the analysis of calving difficulty, it might also be interesting to develop a joint analysis of continuous (birth weight, gestation length) and ordered polychotomous traits using an heteroskedastic approach.

References

- Celeux G, Diebolt J (1985) The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Statist. Quart* 2, 73-82.
- Engel B, Buist W, Visscher A (1995) Inference for the threshold models with variance components from the generalized linear mixed model perspective. *Genet Sel Evol* 27, 15-22
- Fahrmeir L, Tutz G (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag, NY, Berlin

- Foulley JL, Quaas RL (1995) Heterogeneous variances in Gaussian linear mixed models. *Genet Sel Evol* 27, 211-228
- Foulley JL, Gianola D (1996) Statistical analysis of ordered categorical data via a structural threshold model. *Genet Sel Evol* 28, 249-273
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Génét Sél Evol* 15, 201-224
- Harville DA, Mee RW (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408
- Mc Cullagh P, Nelder J (1989) *Generalized linear models*. 2nd ed. Chapman and Hall, London
- Natarajan R(1995) *Estimation in models for multinomial response data: Bayesian and frequentist approaches*. PhD thesis, Cornell University, Ithaca
- Read I, Cressie N (1988) *Goodness-of-fit statistics for discrete multivariate data*. Springer-Verlag, Berlin

This presentation is widely drawn from an article published by Foulley and Gianola in *GSE* 28, 249-273 (1996). The author is grateful to the beef cattle group at INRA-Jouy (F Ménissier, J Sapa, D Laloe) for providing the data analyzed here and to Mrs E Thompson for the English revision of the manuscript

Table 1. Goodness-of-fit of threshold models applied to calving scores in the Blonde d'Aquitaine breed. (regular TM; probit)

SP #	Observed & Expected Numbers*						Statistics of fit [@]			
	1		2		3		1	2	3	X2
10	488	(484.3)	423	(430.9)	118	(113.8)	-0.17	+0.38	-0.39	0.33
11	55	(51.4)	65	(72.3)	33	(29.3)	-0.50	+0.86	-0.69	1.48
12	563	(558.2)	535	(545.2)	162	(156.7)	-0.21	+0.44	-0.43	0.41
13	1990	(1958.3)	1525	(1592.5)	424	(388.2)	-0.72	+1.69	-1.82	6.67
20	300	(305.5)	178	(165.9)	22	(28.6)	+0.31	-0.94	+1.23	2.50
21	69	(69.2)	51	(50.6)	11	(11.2)	+0.02	-0.06	+0.07	0.01
22	761	(769.3)	433	(414.7)	61	(71.0)	+0.30	-0.90	+1.19	2.31
23	2559	(2583.5)	1216	(1161.1)	141	(171.3)	+0.48	-1.61	+2.32	8.20

S: 1=male, 2=female; P: 0=unknown, 1=parity 1, 2=parity 2, 3=parity 3 and above;

* Observed and (Expected) numbers; [@](E-O)/ \sqrt{E} , and Pearson's X^2 in the last column

Table 2. ML estimates and their asymptotic standard errors (SE) for parameters of threshold models applied to calving difficulty scores in the Blonde d'Aquitaine breed

Parameters	Standard TM				Hetero TM	
		Normal	Logistic	Normal	Logistic	
Threshold	2-3	1.297 \pm 0.017	*1.331 \pm 0.019	1.239 \pm 0.021	*1.237 \pm 0.023	
	μ 1x1	0.423 \pm 0.091	*0.424 \pm 0.090	0.350 \pm 0.065	*0.321 \pm 0.061	
	S 2	-0.494 \pm 0.136	*-0.488 \pm 0.134	-0.341 \pm 0.026	*-0.296 \pm 0.023	
	P 0	-0.350 \pm 0.097	*-0.350 \pm 0.097	-0.274 \pm 0.070	*-0.247 \pm 0.066	
	P 2	-0.280 \pm 0.096	*-0.281 \pm 0.095	-0.236 \pm 0.067	*-0.215 \pm 0.063	
Location	3	-0.416 \pm 0.092	*-0.420 \pm 0.092	-0.359 \pm 0.065	*-0.329 \pm 0.061	
	0	0.139 \pm 0.151	*0.157 \pm 0.149			
	S x P 2	0.064 \pm 0.144	*0.078 \pm 0.142			
	3	0.075 \pm 0.139	*0.097 \pm 0.137			
Dispersion	S 2			-0.123 \pm 0.027	-0.215 \pm 0.031	
	X2 (D)	21.90 (22.33)	50.71 (52.88)	3.65 (3.63)	5.63 (5.54)	
Test	DF	7	7	9	9	
	P-val	0.0026	1E-8	0.93	0.77	

Sample size: N=12183

Logistic: *original entries for threshold values and position parameters are divided by $15\pi/16\sqrt{3}$

S: Sex effect stands for $s_2 - s_1$ (females - males) in additive models and $\mu_{21} - \mu_{11}$

(females - males in parity 1) in models with interaction.

P: Parity effects stands for $p_j - p_1$ (j^{th} parity - 1st parity) for $j = 0, 2, 3$ in additive models

and $\mu_{1j} - \mu_{11}$ (j^{th} parity - first parity among males) in models with interaction.

SP: Interaction effects (SP) stand for $sp_j = (\mu_{2j} - \mu_{1j}) - (\mu_{21} - \mu_{11})$ (differences between female - male deviations in parity j and parity 1) for $j = 0, 2, 3$.

Table 3. A test example for mixed threshold model analyses

No	Environmental factors		Genetic factors		Categories		
	a	b	s	t	[1]	[2]	[3]
1	1	1	1	4	23	27	13
2	1	2	1	4	14	21	22
3	1	1	1	7	20	16	6
4	1	2	1	7	13	5	3
5	1	2	1	8	9	5	4
6	1	1	2	6	6	18	12
7	1	1	2	7	13	5	3
8	1	2	2	8	39	10	5
9	2	1	2	8	51	26	4
10	2	2	3	5	8	8	5
11	2	1	3	5	16	24	17
12	2	2	3	5	14	13	3
13	2	3	3	2	60	37	14
14	2	2	3	2	22	13	4
15	2	3	4	7	27	9	3
16	2	3	4	8	9	5	4
17	2	2	4	9	30	19	8
18	2	2	4	5	23	11	2

a,b,s,t: stand for factors a, b (fixed) and sire and maternal grand sire (random) respectively

Non zero elements (i,j)=(j,i) of the numerator relationship matrix are:

(1,2)=(8,9)=1/4; (1,5)=(2,5)=(3,7)=(4,6)=(8,10)=(9,10)=1/2 (i,i)=1, for any i=1,2,...,10

Table 4. MAP estimates of parameters and their precision (SE)

	Parameter	Estimate	SE	Parameter	Estimate	SE
Fixed ¹	$\tau_2 - \tau_1$	1.0096	0.0980	η_R^*	0.0000	0.0000
	$\eta_R - \tau_1$	0.2802	0.3056	$a_2^* - a_1^*$	-0.2072	0.1159
	$a_2 - a_1$	-0.1392	0.2038	$b_2^* - b_1^*$	0.1080	0.1145
	$b_2 - b_1$	-0.1920	0.1190	$b_3^* - b_1^*$	0.2425	0.1595
	$b_3 - b_1$	-0.1520	0.2115			
Random ²	s_1	0.0214	0.2240	t_1	0.0120	0.3492
	s_2	-0.2973	0.2166	t_2	-0.3888	0.2475
	s_3	0.3031	0.2157	t_3	-0.0257	0.3451
	s_4	-0.0906	0.2078	t_4	0.2466	0.2224
	s_5	-0.0764	0.2838	t_5	-0.0288	0.2383
	s_6	0.0896	0.2747	t_6	0.6076	0.2514
	s_7	0.0752	0.2782	t_7	-0.2869	0.2057
	s_8	-0.0486	0.2923	t_8	-0.1746	0.2069
	s_9	0.0252	0.2943	t_9	0.0904	0.2418
	s_{10}	-0.0094	0.3015	t_{10}	-0.0337	0.3363

1: τ_1, τ_2 : threshold values; $\eta_R = \eta + a_1 + b_1$; $\eta_R^* = \eta^* + a_1^* + b_1^*$

$a_1, a_2; (a_1^*, a_2^*)$: location and log-scaling effects of levels 1 and 2 of factor a

$b_1, b_2, b_3; (b_1^*, b_2^*, b_3^*)$: location and log-scaling effects of levels 1,2 and 3 of factor b

2: estimates of components of variance and covariance for sire (s) and maternal grand sire effects (t) are: **sire / residual variance=0.0937; mgs/residual variance=0.1492** and **sire-mgs covariance/residual variance= 0.0416** (correlation =0.3514)

All values are expressed in residual standard deviation units, the link function being the probit.

Appendix

Expressions for the score function \mathbf{U} and the information matrix \mathbf{J}

\mathbf{U} is defined as: $\mathbf{U} = [(\partial L / \partial \boldsymbol{\tau})', (\partial L / \partial \boldsymbol{\beta})', (\partial L / \partial \boldsymbol{\delta})']$.

Letting $h_{ij} = \frac{\tau_j - \eta_i}{\sigma_i}$, a general expression for the probability Π_{ij} is $\Pi_{ij} = F(h_{i,j}) - F(h_{i,j-1})$,

where $F(x) = \Phi(x)$ and $F(x) = [1 + \exp(-x)]^{-1}$ for the probit and logit links respectively.

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\tau}} &= \boldsymbol{\xi} \\ \boldsymbol{\xi} = \{\xi_j\} &= \left\{ \frac{\partial L}{\partial \tau_j} \right\} = \left\{ \sum_{i=1}^I \left(\frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) f(h_{ij}) \sigma_i^{-1} \right\}; j = 1, 2, \dots, J-1. \end{aligned} \quad [\text{A.1}]$$

where y_{ij} stands for $y_{ij+} = \sum_{r=1}^{n_i} y_{ijr}$.

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^I v_{\boldsymbol{\beta},i} \mathbf{X}_i = \mathbf{X}' \mathbf{v}_{\boldsymbol{\beta}}, \\ \mathbf{v}_{\boldsymbol{\beta}} = \{v_{\boldsymbol{\beta},i}\} &= \left\{ \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} [f(h_{i,j-1}) - f(h_{ij})] \boldsymbol{\beta}_i^{-1} \right\}; i = 1, 2, \dots, I. \end{aligned} \quad [\text{A.2}]$$

Similarly, letting $\psi(x) = xf(x)$:

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\delta}} &= \sum_{i=1}^I v_{\boldsymbol{\delta},i} \mathbf{P}_i = \mathbf{P}' \mathbf{v}_{\boldsymbol{\delta}}, \\ \mathbf{v}_{\boldsymbol{\delta}} = \{v_{\boldsymbol{\delta},i}\} &= \left\{ \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} [\psi(h_{i,j-1}) - \psi(h_{ij})] \right\}; i = 1, 2, \dots, I. \end{aligned} \quad [\text{A.3}]$$

Notice the remarkable symmetry in expressions [A.2] and [A.3].

Finally, \mathbf{U} can be expressed as:

$$\mathbf{U}(\boldsymbol{\alpha}; \mathbf{y}) = \begin{bmatrix} \boldsymbol{\xi} \\ \mathbf{X}' \mathbf{v}_{\boldsymbol{\beta}} \\ \mathbf{P}' \mathbf{v}_{\boldsymbol{\delta}} \end{bmatrix},$$

with expressions for $\boldsymbol{\xi}$, $\mathbf{v}_{\boldsymbol{\beta}}$ and $\mathbf{v}_{\boldsymbol{\delta}}$ given in [A.1], [A.2] and [A.3]. Elements of the information matrix $\mathbf{J}(\boldsymbol{\alpha})$ include the expectations of minus the second derivatives. The following derivatives will be considered: threshold-threshold; $\boldsymbol{\beta}$ -threshold; $\boldsymbol{\delta}$ -threshold; $\boldsymbol{\beta}$ - $\boldsymbol{\beta}$; $\boldsymbol{\beta}$ - $\boldsymbol{\delta}$, and $\boldsymbol{\delta}$ - $\boldsymbol{\delta}$.

a) threshold-threshold derivatives

$$\mathbb{E} \left(-\frac{\partial^2 L}{\partial \tau_j^2} \right) = \sum_{i=1}^I \frac{n_i}{\sigma_i^2} [f(h_{ij})]^2 \left(\frac{1}{\Pi_{ij}} + \frac{1}{\Pi_{i,j+1}} \right). \quad [\text{A.4}]$$

$$\mathbb{E} \left(-\frac{\partial^2 L}{\partial \tau_j \partial \tau_{j+1}} \right) = -\sum_{i=1}^I \sigma_i^{-2} \frac{n_i}{\Pi_{i,j+1}} f(h_{ij}) f(h_{i,j+1}), \quad [\text{A.5}]$$

and

$$\frac{\partial^2 L}{\partial \tau_j \partial \tau_\ell} = 0, \text{ for } |j - \ell| > 1.$$

b) β -threshold derivatives

$$E\left(-\frac{\partial^2 L}{\partial \tau_j \partial \beta}\right) = \sum_{i=1}^I \ell_{\beta,ij} \mathbf{x}_i = \mathbf{X}' \mathbf{L}_{\beta,j}, \quad [\text{A.6a}]$$

$$\mathbf{L}_{\beta,j} = \{\ell_{\beta,ij}\} = \left\{ \frac{n_i}{\sigma_i^2} f(h_{ij}) \left[\frac{f(h_{i,j-1}) - f(h_{ij})}{\Pi_{ij}} - \frac{f(h_{ij}) - f(h_{i,j+1})}{\Pi_{i,j+1}} \right] \right\}; i = 1, 2, \dots, I,$$

$$E\left(-\frac{\partial^2 L}{\partial \beta \partial \tau}\right) = \mathbf{X}' \mathbf{L}_{\beta}, \quad [\text{A.6b}]$$

$$\mathbf{L}_{\beta_{[I \times (J-1)]}} = (\mathbf{L}_{\beta,1}, \mathbf{L}_{\beta,2}, \dots, \mathbf{L}_{\beta,j}, \dots, \mathbf{L}_{\beta,J-1}).$$

c) δ -threshold derivatives

$$E\left(-\frac{\partial^2 L}{\partial \tau_j \partial \delta}\right) = \sum_{i=1}^I \ell_{\delta,ij} \mathbf{p}_i = \mathbf{P}' \mathbf{L}_{\delta,j}, \quad [\text{A.7a}]$$

$$\mathbf{L}_{\delta,j} = \{\ell_{\delta,ij}\} = \left\{ \frac{n_i}{\sigma_i} f(h_{ij}) \left[\frac{\Psi(h_{i,j-1}) - \Psi(h_{ij})}{\Pi_{ij}} - \frac{\Psi(h_{ij}) - \Psi(h_{i,j+1})}{\Pi_{i,j+1}} \right] \right\}; i = 1, 2, \dots, I$$

$$E\left(-\frac{\partial^2 L}{\partial \delta \partial \tau}\right) = \mathbf{P}' \mathbf{L}_{\delta} \quad [\text{A.7b}]$$

$$\mathbf{L}_{\delta_{[I \times (J-1)]}} = (\mathbf{L}_{\delta,1}, \mathbf{L}_{\delta,2}, \dots, \mathbf{L}_{\delta,j}, \dots, \mathbf{L}_{\delta,J-1}).$$

The remaining elements can be easily obtained using Fisher's information measure for a multinomial distribution,

ie

$$E\left(-\frac{\partial^2 L}{\partial \varphi_k \partial \varphi_\ell}\right) = \sum_{i=1}^I n_i \sum_{j=1}^J \frac{1}{\Pi_{ij}} \frac{\partial \Pi_{ij}}{\partial \varphi_k} \frac{\partial \Pi_{ij}}{\partial \varphi_\ell}. \quad [\text{A.8}]$$

Applying [A.8] here leads to:

d) $\beta - \beta$ derivatives

$$E\left(-\frac{\partial^2 L}{\partial \beta \partial \beta'}\right) = \sum_{i=1}^I w_{\beta\beta,ii} \mathbf{x}_i \mathbf{x}_i' = \mathbf{X}' \mathbf{W}_{\beta\beta} \mathbf{X}, \quad [\text{A.9}]$$

$$\mathbf{W}_{\beta\beta} = \text{Diag}\{w_{\beta\beta,ii}\} = \text{Diag}\left\{ \frac{n_i}{\sigma_i^2} \sum_{j=1}^J \frac{[f(h_{i,j-1}) - f(h_{ij})]^2}{\Pi_{ij}} \right\}, i = 1, 2, \dots, I.$$

e) $\beta - \delta$ derivatives

$$E\left(-\frac{\partial^2 L}{\partial \beta \partial \delta'}\right) = \sum_{i=1}^I w_{\beta\delta,ii} \mathbf{x}_i \mathbf{p}_i' = \mathbf{X}' \mathbf{W}_{\beta\delta} \mathbf{P}, \quad [\text{A.10}]$$

$$\mathbf{W}_{\beta\delta} = \text{Diag}\{w_{\beta\delta,ii}\} = \text{Diag}\left\{ \frac{n_i}{\sigma_i} \sum_{j=1}^J \frac{[f(h_{i,j-1}) - f(h_{ij})][\Psi(h_{i,j-1}) - \Psi(h_{ij})]}{\Pi_{ij}} \right\}, i = 1, 2, \dots, I.$$

$$\begin{aligned} \mathbf{E}\left(-\frac{\partial^2 L}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}}\right) &= \sum_{i=1}^I w_{\boldsymbol{\delta}\boldsymbol{\delta},ii} \mathbf{P}_i \mathbf{P}_i' = \mathbf{P}' \mathbf{W}_{\boldsymbol{\delta}\boldsymbol{\delta}} \mathbf{P}, & [\text{A.11}] \\ \mathbf{W}_{\boldsymbol{\delta}\boldsymbol{\delta}} &= \text{Diag}\{w_{\boldsymbol{\delta}\boldsymbol{\delta},ii}\} = \text{Diag}\left\{n_i \sum_{j=1}^J \frac{[\Psi(h_{i,j-1}) - \Psi(h_{ij})]^2}{\Pi_{ij}}\right\}, i=1,2,\dots,I. \end{aligned}$$

Finally $\mathbf{J}(\boldsymbol{\alpha})$ can be written in condensed form as:

$$\mathbf{J}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{T} & \mathbf{L}'_{\beta} \mathbf{X} & \mathbf{L}'_{\delta} \mathbf{P} \\ \mathbf{X}' \mathbf{L}_{\beta} & \mathbf{X}' \mathbf{W}_{\beta\beta} \mathbf{X} & \mathbf{X}' \mathbf{W}_{\beta\delta} \mathbf{P} \\ \mathbf{P}' \mathbf{L}_{\delta} & \mathbf{P}' \mathbf{W}_{\beta\delta} \mathbf{X} & \mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \end{bmatrix}, \quad [\text{A.12}]$$

where $\mathbf{T} = \{t_{jk}\}$ is a $(J-1) \times (J-1)$ symmetric band matrix having as elements: $t_{jj} = \mathbf{E}(-\partial^2 L / \partial \tau_j^2)$, and $t_{j,j+1} = \mathbf{E}(-\partial^2 L / \partial \tau_j \partial \tau_{j+1})$, given in [A.4] and [A.5].

These expressions can be extended to obtain the MAP of parameters in a mixed model structure by replacing i) $\boldsymbol{\beta}$ by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{u}\#')'$ with $\mathbf{u}\# \sim N(\mathbf{0}, \rho^2 \mathbf{A})$ ($\rho^2 = \sigma_{u_i}^2 / \sigma_i^2 = \text{constant}$); ii) \mathbf{X} by $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_i, \dots, \mathbf{s}_I)'$ with $\mathbf{s}_i = (\mathbf{x}_i', \sigma_i \mathbf{z}_i')$;

and making the appropriate adjustments for prior information as shown below

$$\begin{bmatrix} \mathbf{T} & \mathbf{L}'_{\theta} \mathbf{S} & \mathbf{L}'_{\delta} \mathbf{P} \\ \mathbf{S}' \mathbf{L}_{\theta} & \mathbf{S}' \mathbf{W}_{\theta\theta} \mathbf{S} + \boldsymbol{\Sigma}^- & \mathbf{S}' \mathbf{W}_{\theta\delta} \mathbf{P} \\ \mathbf{P}' \mathbf{L}_{\delta} & \mathbf{P}' \mathbf{W}_{\theta\delta} \mathbf{S} & \mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \end{bmatrix}^{[k]} \begin{bmatrix} \Delta \boldsymbol{\tau} \\ \Delta \boldsymbol{\theta} \\ \Delta \boldsymbol{\delta} \end{bmatrix}^{[k+1]} = \begin{bmatrix} \boldsymbol{\xi} \\ \mathbf{S}' \mathbf{v}_{\theta} - \boldsymbol{\Sigma}^- \boldsymbol{\theta} \\ \mathbf{P}' \mathbf{v}_{\delta} \end{bmatrix}^{[k]}, \quad [\text{A.13}]$$

where $\boldsymbol{\Sigma}^- = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \rho^{-2} \mathbf{A}^{-1} \end{bmatrix}$ and $\mathbf{v}_{\theta} = \mathbf{v}_{\beta}$, $\mathbf{L}_{\theta} = \mathbf{L}_{\beta}$, $\mathbf{W}_{\theta\theta} = \mathbf{W}_{\beta\beta}$, $\mathbf{W}_{\theta\delta} = \mathbf{W}_{\beta\delta}$ as expressed in [A.2], [A.6ab], [A.9] and [A.10] respectively.