



**HAL**  
open science

## Statistical analysis of ordered categorical data via a structural heteroskedastic threshold model

Jl Foulley, D Gianola

► **To cite this version:**

Jl Foulley, D Gianola. Statistical analysis of ordered categorical data via a structural heteroskedastic threshold model. *Genetics Selection Evolution*, 1996, 28 (3), pp.249-273. hal-00894134

**HAL Id: hal-00894134**

**<https://hal.science/hal-00894134>**

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.

# Statistical analysis of ordered categorical data via a structural heteroskedastic threshold model

JL Foulley<sup>1</sup>, D Gianola<sup>2</sup>

<sup>1</sup> *Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, centre de recherches de Jouy, 78352 Jouy-en-Josas cedex, France;*

<sup>2</sup> *Department of Meat and Animal Science, University of Wisconsin-Madison, Madison, WI 53706, USA*

(Received 2 October 1995; accepted 25 March 1996)

**Summary** – 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 normal scale. A new model is proposed whereby subpopulations can also differ in dispersion (scaling) 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. An application to calving ease scores in the US Simmental breed is presented; the heteroskedastic threshold model had a better goodness of fit than the standard one.

**threshold character / heteroskedasticity / maximum likelihood / mixed linear model / calving difficulty**

**Résumé** – **Analyse statistique de variables discrètes ordonnées par un modèle à seuils hétéroscédastique.** Dans le modèle à seuils classique, les différences de réponses entre sous-populations selon des catégories discrètes ordonnées sont modélisées par des différences entre paramètres de position mesurés sur une variable normale sous-jacente. L'approche présentée ici suppose que ces sous-populations diffèrent aussi par leurs paramètres de dispersion (ou paramètres d'échelle). L'hétérogénéité de ces paramètres est décrite par un modèle linéaire structurel et une fonction de lien logarithmique impliquant des covariables discrètes ou continues. L'inférence (estimation, qualité d'ajustement, test d'hypothèse) sur les paramètres dans les modèles à effets fixes est basée sur les méthodes du maximum de vraisemblance. Des techniques bayésiennes sont également proposées pour le traitement des modèles linéaires mixtes. Une application aux notes de difficultés de vêlage en race Simmental américaine est présentée. Le modèle à seuils hétéroscédastique améliore dans ce cas la qualité de l'ajustement des données par rapport au modèle standard.

**caractères à seuils / hétéroscédasticité / maximum de vraisemblance / modèle linéaire mixte / difficultés de vêlage**

## INTRODUCTION

An appealing model for the analysis of ordered categorical data is the so-called threshold model. Although introduced in population and quantitative genetics by Wright (1934a,b) and discussed later by Dempster and Lerner (1950) and Robertson (1950), it dates back to Pearson (1900), Galton (1889) and Fechner (1860). This model has received attention in various areas such as human genetics and susceptibility to disease (Falconer, 1965; Curnow and Smith, 1975), population biology (Bulmer and Bull, 1982; Roff, 1994), neurophysiology (Brillinger, 1985), animal breeding (Gianola, 1982), survey analysis (Grosbas, 1987), psychological and social sciences (Edwards and Thurstone, 1952; McKelvey and Zavoina, 1975), and econometrics (Kaplan and Urwitz, 1979; Levy, 1980; Bryant and Gerner, 1982; Maddala, 1983).

The threshold model postulates an underlying (liability) normal distribution rendered discrete via threshold values. The probability of response in a given category can be expressed as the difference between normal 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 location parameter ( $\eta_i$ ) for a subpopulation is expressed as a linear function  $\eta_i = \mathbf{t}'_i \boldsymbol{\theta}$  of some explanatory variables (row incidence vector  $\mathbf{t}'_i$ ) (see theory of generalized linear models, McCullagh and Nelder, 1989; Fahrmeir and Tutz, 1994). The vector of unknowns ( $\boldsymbol{\theta}$ ) may include both fixed and random effects and statistical procedures have been developed to make inferences about such a mixed-model structure (Gianola and Foulley, 1983; Harville and Mee, 1984; Gilmour et al, 1987). In all these studies, the standard deviations (also called the scaling parameters) are assumed to be known and equal, or proportional to known quantities (Foulley, 1987; Misztal et al, 1988).

The purpose of this paper is to extend the standard threshold model (S-TM) to a model allowing for heteroskedasticity (H-TM) with modeling of the unknown scaling 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. The theory will be illustrated with an example on calving difficulty scores in Simmental cattle from the USA.

## THEORY

### *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

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

and (3)  $I - 1$  contrasts among log-scaling parameters (eg,  $\ln(\sigma_i) - \ln(\sigma_1)$  for  $i = 2, \dots, I$ ) or, equivalently,  $I - 1$  standard deviation ratios (eg,  $\sigma_1/\sigma_1$ ), with one of these arbitrarily set to a fixed value (eg,  $\sigma_1 = 1$ ). This makes a total of  $2I + J - 3$  identifiable parameters, so that the full H-TM reduces to the saturated model for  $J = 3$  categories, see examples in Falconer (1960), chapter 18.

More parsimonious models can also be envisioned. For instance, in a two-way crossclassified layout with  $A$  rows and  $B$  columns ( $I = AB$ ), there are 16 additive models that can be used to describe the location ( $\eta_i$ ) and the scaling ( $\sigma_i$ ) parameters. The simplest one would have a common mean and standard deviation for the  $I = AB$  populations. The most complete one would include the main effects of  $A$  and  $B$  factors for both the location and dispersion parameters. Here there are  $2(A+B)+J-5$  estimable parameters, ie,  $J-1$  thresholds plus twice  $(A-1)+(B-1)$ . Under an additive model for the location parameters  $\eta_i$ , it is possible to fit the H-TM to binary data. For the crossclassified layout with  $A$  rows and  $B$  columns, there are  $AB - 2(A + B) + 3$  degrees of freedom available which means that we need  $A$  (or  $B$ )  $\geq 4$  to fit an additive model using all the levels of  $A$  and  $B$  at both the location and dispersion levels. Finally, it must be noted that in this particular case, dispersion parameters act as substitutes of interaction effects for location parameters.

**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 log likelihood  $L(\boldsymbol{\alpha}; \mathbf{y})$  can be expressed, apart from an additive constant, as:

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

with, given [4], [5] and [6],

$$\Pi_{ij} = \Phi \left[ \frac{\tau_j - \eta_i}{\exp(\mathbf{p}'_i \boldsymbol{\delta})} \right] - \Phi \left[ \frac{\tau_{j-1} - \eta_i}{\exp(\mathbf{p}'_i \boldsymbol{\delta})} \right]$$

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 Fisher's information measure is simple here. The system of equations to solve iteratively can be written as:

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

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 Fisher's information matrix respectively;  $k$  is iterate number. Analytical expressions for the elements of  $\mathbf{U}(\boldsymbol{\alpha}; \mathbf{y})$  and  $\mathbf{J}(\boldsymbol{\alpha})$  are given in *Appendix 1*.

These are generalizations of formulae given by Gianola and Foulley (1983) and Misztal et al (1988). In some instances, one may consider a backtracking procedure (Denis and Schnabel, 1983) to reach convergence, ie, at the beginning of the iterative process, compute  $\boldsymbol{\alpha}^{[k+1]}$  as  $\boldsymbol{\alpha}^{[k+1]} = \boldsymbol{\alpha}^{[k]} + \varpi^{[k+1]} \boldsymbol{\Delta} \boldsymbol{\alpha}^{[k+1]}$  with  $0 < \varpi^{[k+1]} \leq 1$ . A constant value of  $\varpi = 0.8$  has been satisfactory in all the examples run so far with the H-TM.

(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

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

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

In the threshold model, the probabilities  $\Pi_{ij}$  are connected to the underlying normal variables  $X_{ir}$  with threshold values  $\tau_j$  via the statement

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

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

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

$$\Pi_{ij} = \Phi\left(\frac{\tau_j - \eta_i}{\sigma_i}\right) - \Phi\left(\frac{\tau_{j-1} - \eta_i}{\sigma_i}\right) \tag{4}$$

where  $\Phi(\cdot)$  is the CDF of the standardized normal distribution.

The mean liability ( $\eta_i$ ) for the  $i$ th subpopulation is modeled 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

$$\eta_i = \mathbf{x}'_i \boldsymbol{\beta} \tag{5}$$

Here, the vector ( $p \times 1$ ) 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 et al (1990, 1992) and Foulley and Quaas (1995), the natural logarithm of  $\sigma_i$  is written as a linear combination of some unknown ( $r \times 1$ ) real-valued vector of parameters ( $\boldsymbol{\delta}$ ),

$$\ln(\sigma_i) = \mathbf{p}'_i \boldsymbol{\delta} \tag{6}$$

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

### Identifiability of parameters

In the case of  $I$  subpopulations and  $J$  categories, there is a maximum of  $I(J - 1)$  identifiable (or estimable) parameters if the margins  $n_i$  are fixed by sampling. These are the parameters of the so-called saturated model.

What is the most complete H-TM (or 'full' model) that can be fitted to the data using the approach described here? One can estimate: (1)  $J - 1$  finite threshold values or, equivalently,  $J - 2$  differences among these (eg,  $\tau_j - \tau_1$  for  $j = 2, \dots, J - 1$ ) plus a baseline population effect (eg,  $\eta_1 - \tau_1$ ); (2)  $I - 1$  contrasts among  $\eta_i$  values;

**Goodness of fit**

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

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \left( y_{ij+} - n_i \widehat{\Pi}_{ij} \right)^2 / n_i \widehat{\Pi}_{ij} \tag{9}$$

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

$$D^*(y, \widehat{\alpha}) = 2[L(y; y) - L(\widehat{\alpha}; y)] \tag{10}$$

Above,  $D^*$  is based on the likelihood ratio statistic for fitting the entertained 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] (McCullagh 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  $\beta$  and  $\delta$  respectively. It should be noted that [9] and [10] can be computed as particular cases of the power divergent statistics introduced by Read and Cressie (1984).

**Hypothesis testing**

Tests of hypotheses about  $\gamma = (\beta', \delta')'$  can be carried out via either Wald's test or the likelihood ratio (or deviance) test. The first procedure relies on the properties of consistency and asymptotic normality of the ML estimator.

For linear hypotheses of the form  $H_0 : \mathbf{K}'\gamma = \mathbf{m}$  against its alternative  $H_1 = \overline{H_0}$ , the Wald statistic is:

$$w = (\mathbf{K}'\widehat{\gamma} - \mathbf{m})' [\mathbf{K}'\Gamma(\widehat{\gamma})\mathbf{K}]^{-1} (\mathbf{K}'\widehat{\gamma} - \mathbf{m})' \tag{11}$$

which under  $H_0$  has an asymptotic chi-square distribution with  $\text{rank}(\mathbf{K})$  degrees of freedom. Above,  $\Gamma(\widehat{\gamma})$  is an appropriate block of the inverse of Fisher's information matrix evaluated at  $\gamma = \widehat{\gamma}$ , where  $\widehat{\gamma}$  is the ML estimator.

The likelihood ratio statistic (LRS) allows testing nested hypotheses of the form  $H_0 : \gamma \in \Omega_0$  against  $H_1 : \gamma \in \Omega - \Omega_0$  where  $\Omega_0$  and  $\Omega$  are the restricted and unrestricted parameter spaces respectively, pertaining to  $H_0$  and  $H_0 \cup H_1$ . The LRS is:

$$\lambda^\# = 2[L(\widehat{\gamma}) - L(\widetilde{\gamma})] \tag{12}$$

where  $\widetilde{\gamma}$  and  $\widehat{\gamma}$  are the ML estimators of  $\gamma$  under the restricted and unrestricted models respectively. The criterion  $\lambda^\#$  also can be computed as the difference in (scaled) deviances of the restricted and unrestricted models

$$\lambda^\# = D^*(y; \widetilde{\gamma}) - D^*(y; \widehat{\gamma}) \tag{13}$$

This is equivalent to what is usually done in ANOVA except that residual sums of squares are replaced by deviances.

Under  $H_0$ ,  $\lambda^\#$  has an asymptotic chi-square distribution with  $r = \dim(\Omega) - \dim(\Omega_0)$  degrees of freedom. Under the same null hypothesis, the Wald and LR statistics have the same asymptotic distribution. However, Wald's statistic is based on a quadratic approximation of the loglikelihood around its maximum.

### Including random effects

In many applications, the  $\mathbf{y}_{ir}$ 's cannot be assumed to be independent repetitions due to some cluster structure in the data. This is the case in quantitative genetics and animal breeding with genetically related animals, common environmental effects and repeated measurements on the same individuals.

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

$$\eta_i = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u} \quad [14]$$

where the fixed component  $\mathbf{x}'_i \boldsymbol{\beta}$  is as before, and  $\mathbf{u}$  is a  $(q \times 1)$  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 deviation, and  $\sigma_{u_i}$  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  $\sigma_{u_i}/\sigma_i$ , where  $\sigma_i$  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 (Visscher, 1992). Thus, the argument  $h_{ij}$  of the normal CDF in [4] and [7] becomes

$$h_{ij} = \frac{\tau_j - \mathbf{x}'_i \boldsymbol{\beta} - \mathbf{z}'_i \sigma_{u_i} \mathbf{u}^*}{\sigma_i} = \frac{\tau_j - \mathbf{x}'_i \boldsymbol{\beta}}{\exp(\mathbf{p}'_i \boldsymbol{\delta})} - \mathbf{z}'_i \rho \mathbf{u}^* \quad [15]$$

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

In the fixed model, parameters  $\boldsymbol{\tau}$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  were estimated by maximum likelihood. Given  $\rho$ , 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

$$\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\mathbf{u}}^* = \text{Argmax}[\mathbf{p}(\mathbf{y}|\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{u}^*, \rho)\mathbf{p}(\mathbf{u}^*)] \quad [16]$$

MAP solutions can be computed with minor modifications from [8]. The only changes to implement are to replace: (i)  $\boldsymbol{\alpha} = (\boldsymbol{\tau}', \boldsymbol{\beta}', \boldsymbol{\delta}')$  by  $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\beta}', \boldsymbol{\delta}', \mathbf{u}^{\#})'$  with  $\mathbf{u}^{\#} = \rho \mathbf{u}^*$ ; (ii)  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_I)'$  by  $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_i, \dots, \mathbf{s}_I)'$  with  $\mathbf{s}'_i = (\mathbf{x}'_i, \sigma_i \mathbf{z}'_i)$ ; (iii) add  $\rho^{-2} \mathbf{A}^{-1}$  to the coefficients of the  $\mathbf{u}^{\#} \times \mathbf{u}^{\#}$  block pertaining

to the random effects on the left hand side and  $-\rho^{-2}\mathbf{A}^{-1}\mathbf{u}^{[k]}$  to the  $\mathbf{u}^\#$ -part of the right hand side (see *Appendix 1*). A test example is shown in *Appendix 2*.

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

$$\rho^{2[k+1]} = \mathbf{E} \left( \mathbf{u}^\# \mathbf{A}^{-1} \mathbf{u}^\# | \mathbf{y}, \rho^{2[k]} \right) / q \quad [17]$$

This may involve either an approximate calculation of the conditional expectation of the quadratic in  $\mathbf{u}^\#$  as in Harville and Mee (1984), Hoeschele et al (1987) and Foulley et al (1990), or a Monte-Carlo calculation of this conditional expectation using, for example, a Gibbs sampling scheme (Natarajan, 1995). Alternative procedures for estimating  $\rho$  might be also envisioned, such as the iterated re-weighted REML of Engel et al (1995).

## NUMERICAL APPLICATION

### *Material*

The data set analyzed was a contingency table of calving difficulty scores (from 1 to 4) recorded on purebred US-Simmental cows distributed according to sex of calf (males, females) and age of dam at calving in years. Scores 3 and 4 were pooled on account of the low frequency of score 4. Nine levels were considered for age of dam: < 2 years, 2.0–2.5, 2.5–3.0, 3.0–3.5, 3.5–4.0, 4.0–4.5, 4.5–5.0, 5.0–8.0, and > 8.0 years. In the analysis of the scaling parameters, six levels were considered for this factor: < 2 years, 2.0–2.5, 2.5–3.0, 3.0–4.0, 4.0–8.0, and > 8.0 years. The distribution of the 363 859 records by sex-age of dam combinations is displayed in table I, as well as the frequencies of the three categories of calving scores. The raw data reveal the usual pattern of highest calving difficulty in male calves out of younger dams. However, more can be said about the phenomenon.

### *Method*

Data were analyzed with standard (S-TM) and heteroskedastic (H-TM) threshold models. Location and scaling parameters were described using fixed models involving sex (S) and age of dam (A) as factors of variation. In both cases, inference was based on maximum likelihood procedures. A log-link function was used for scaling parameters.

With  $J = 3$  categories, the most highly parameterized S-TM that can be fitted for the location structure includes  $J - 1 = 2$  threshold values (or, equivalently, the difference between thresholds ( $\tau_2 - \tau_1$ ) and a baseline population effect  $\mu$ ), plus sex (one contrast), age of dam (eight contrasts) and their interaction (eight contrasts) as elements of  $\boldsymbol{\beta}$ ; this gives  $r(\mathbf{X}) = 17$  which yields 19 as the total number of parameters to be estimated. There were  $I = 18$  sex  $\times$  age subpopulations so that the maximum number of parameters which can be estimated (in the saturated model) is  $(3 - 1) \times 18 = 36$ . The degrees of freedom ( $df$ ) were thus  $36 - 19 = 17$ .



**Table I.** Distribution of calving difficulty scores by sex  $\times$  age of dam subclasses in Simmental cattle: absolute and relative (%) frequencies.

<i>Sex Dam age (year)</i>		<i>Score</i>						
		<i>1</i>		<i>2</i>		<i>3+4</i>		<i>Total</i>
M	< 2.0	7 836	(61.56)	3 139	(24.66)	1 754	(13.78)	12 729
	2.0-2.5	16 656	(70.82)	4 532	(19.27)	2 330	(9.91)	23 518
	2.5-3.0	9 542	(86.16)	1 092	(9.86)	441	(3.98)	11 075
	3.0-3.5	17 942	(91.48)	1 223	(6.24)	449	(2.29)	19 614
	3.5-4.0	9 556	(93.96)	458	(4.50)	156	(1.53)	10 170
	4.0-4.5	14 453	(94.62)	592	(3.88)	229	(1.50)	15 274
	4.5-5.0	8 159	(95.65)	272	(3.19)	99	(1.16)	8 530
	5.0-8.0	40 484	(96.04)	1 205	(2.86)	464	(1.10)	42 153
	> 8.0	15 991	(96.37)	433	(2.61)	170	(1.02)	16 594
	All	140 619	(88.07)	12 946	(8.11)	6 092	(3.82)	159 657
F	< 2.0	12 846	(79.14)	2 455	(15.13)	930	(5.73)	16 231
	2.0-2.5	26 023	(85.42)	3 355	(11.01)	1 087	(3.57)	30 465
	2.5-3.0	14 207	(93.37)	809	(5.32)	199	(1.31)	15 215
	3.0-3.5	23 486	(96.31)	711	(2.92)	190	(0.78)	24 387
	3.5-4.0	13 202	(97.19)	315	(2.32)	67	(0.49)	13 584
	4.0-4.5	18 100	(97.63)	336	(1.81)	104	(0.56)	18 540
	4.5-5.0	10 833	(97.75)	201	(1.81)	48	(0.43)	11 082
	5.0-8.0	52 397	(98.26)	763	(1.43)	167	(0.31)	53 327
	> 8.0	21 943	(98.47)	252	(1.18)	76	(0.36)	21 371
	All	192 137	(94.09)	9 197	(4.50)	2 868	(1.41)	204 202

M (F): male (female) calves; 1 = easy; 2 = assisted; 3 + 4 = difficult calving.

The H-TM to start with was as in the S-TM for location parameters  $\beta$ . With respect to dispersion parameters  $\delta$ , the model was an additive one, with sex ( $S^*$ : one contrast) and age of dam ( $A^*$ : five contrasts) so that  $r(\mathbf{P}) = 6$  ( $\sigma = 1$  in male calves and < 2.0 year old dams). Thus, the total number of parameters was  $19 + 6 = 25$  and, the  $df$  were equal to  $36 - 25 = 11$ .

## RESULTS

All factors considered in the S-TM were significant ( $P < 0.01$ ), especially the sex by age of dam interactions (except the first one, as shown in table II). Hence, the model cannot be simplified further. This means that differences between sexes were not constant across age of dam subclasses, contrary to results of a previous study in Simmental also obtained with a fixed S-TM (Quaas et al, 1988). Differences in liability between male and female calves decreased with age of dam. However the S-TM did not fit well to the data, as the Pearson statistic (or deviance) was  $X^2 = 419$  on 17 degrees of freedom, resulting in a nil  $P$ -value. An examination of the Pearson residuals indicated that the S-TM leads to an underestimation of the probability of difficult calving (scores 3 + 4) in cows older than 3 years of age, and to an overestimation in younger cows.

**Table II.** Maximum likelihood estimates of parameters and their asymptotic standard errors (SE) for threshold models applied to calving difficulty scores in Simmental cattle (H-TM: S + A + SA for location; S + A for scale).

<i>Parameter</i>		<i>Standard TM</i>		<i>Hetero TM</i>	
		<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>
Threshold	2 – 3	0.695	0.005	0.777	0.010
Location					
$\mu$	1 × 1	-0.324	0.011	-0.300	0.011
S	2	-0.501	0.015	-0.493	0.020
A	2	-0.234	0.014	-0.279	0.019
	3	-0.758	0.018	-0.938	0.040
	4	-1.038	0.017	-1.404	0.046
	5	-1.218	0.022	-1.627	0.053
	6	-1.271	0.020	-1.928	0.056
	7	-1.374	0.026	-2.067	0.063
	8	-1.416	0.015	-2.127	0.057
	9	-1.455	0.021	-2.373	0.114
	S × A	2	-0.003	0.020	-0.021
3		0.077	0.026	0.042	0.032
4		0.081	0.025	0.015	0.034
5		0.136	0.033	0.086	0.044
6		0.124	0.030	0.028	0.046
7		0.199	0.038	0.131	0.056
8		0.136	0.023	0.046	0.039
9		0.129	0.032	0.005	0.056
Scale					
S*	2			-0.018	0.013
A*	2			0.053	0.016
	3			0.132	0.025
	4*			0.218	0.022
	5*			0.325	0.025
	6*			0.398	0.040
	$\chi^2(D)$		419.11 (398.22)		18.17 (18.34)
Test					
df		17		11	
P-value		0		0.078	

S: Sex effect stands for  $s_2 - s_1$  (females – males) in additive models and  $\mu_{21} - \mu_{11}$  (females – males in age of dam 1) in models with interaction. A: age of dam effects stand for  $a_j - a_1$  ( $j$ th parity – 1st parity) for  $j = 2, 3, \dots, 9$  in additive models ( $4^* = 4 + 5$ ,  $5^* = 6 + 7 + 8$ ,  $6^* = 9$ ) and  $\mu_{1j} - \mu_{11}$  ( $j$ th age of dam – 1st age of dam within males) in models with interaction. SA: interaction effects (SA) stand for  $sa_j = (\mu_{2j} - \mu_{1j}) - (\mu_{21} - \mu_{11})$  (differences between female – male deviations in age of dam  $j$  and age of dam 1) for  $j = 2, 3, \dots, 9$ .  $X^2(D)$ : Pearson's and deviance statistics respectively.

As shown in table II, fitting the H-TM decreased the  $X^2$  and deviance by a factor of 20 and led to a satisfactory fit. The significance of many interactions vanished, and this was reflected in the LRS ( $P < 0.088$ ) for the hypothesis of no  $S \times A$  interaction in the most parameterized model. Several models were tried and tested as shown in table III. The scaling parameters depended on the age of the dam, the effect of sex being not significant ( $P < 0.163$ ). Relative to the baseline population, the standard deviation increased by a factor of about 1.05, 1.15, 1.25, 1.40 and 1.50 for cows of 2.0–2.5, 2.5–3.0, 3.0–4.0, 4.0–8.0, and  $> 8.0$  years of age at calving respectively (table IV).

**Table III.** Wald and deviance (LR) statistics for testing effects of factors in threshold models applied to calving difficulty scores in Simmental cattle.

	Model §					
	$S + A + SA; S^* + A^*$			$S + A + SA; A^*$		
	df	$X^2$	<i>P-value</i>	df	$X^2$	<i>P-value</i>
Model for $\beta$	17	2 757.01	0	17	3 937.66	0
S <sup>a</sup>	1	630.12	0	1	1 131.51	0
A <sup>b</sup>	8	2 124.46	0	8	2 141.10	0
$S \times A^c$	8	14.23 (13.74) <sup>d</sup>	0.075 0.088	8	11.87 (11.82) <sup>e</sup>	0.157 0.159
Model for $\sigma$	6	366.54	0	5	365.42	0
S <sup>*a</sup>	1	1.94 (1.94) <sup>f</sup>	0.163 0.163			0
A <sup>*b</sup>	5	360.84 (372.58) <sup>g</sup>	0 0	5	365.42	0
Overall fit	11	18.17	0.078	12	20.01	0.067

§ Model for location; model for scale\*; <sup>a</sup> sex effect (S) stands for  $s_2 - s_1$  (females - males) in additive models and  $\mu_{21} - \mu_{11}$  (females - males in age of dam 1) in models with interaction; <sup>b</sup> age of dam effects (A) stand for  $a_j - a_1$  ( $j$ th age of dam - 1st age of dam) for  $j = 2, 3, \dots, 9$  ( $4^* = 4 + 5; 5^* = 6 + 7 + 8; 6^* = 9$ ) in additive models and  $\mu_{1j} - \mu_{11}$  ( $j$ th age of dam - 1st age of dam within males) in models with interaction; <sup>c</sup> interaction effects (SA) stand for  $sa_j = (\mu_{2j} - \mu_{1j}) - (\mu_{21} - \mu_{11})$  (differences between female - male deviations in age of dam  $j$  and age 1) for  $j = 2, 3, \dots, 9$ ; <sup>d</sup> LR statistic contrasting  $S + A + SA; S^* + A^*$  ( $D = 18.34$ ) against  $S + A; S^* + A^*$  ( $D = 32.08$ ); <sup>e</sup> LR statistic contrasting  $S + A + SA; A^*$  ( $D = 20.28$ ) against  $S + A; A^*$  ( $D = 32.10$ ); <sup>f</sup> LR statistic contrasting  $S + A + SA; S^* + A^*$  ( $D = 18.34$ ) against  $S + A + SA; A^*$  ( $D = 20.28$ ); <sup>g</sup> LR statistic contrasting  $S + A + SA; S^* + A^*$  ( $D = 18.34$ ) against  $S + A + SA; S^*$  ( $D = 390.92$ ).

The H-TM made differences between sex liabilities across ages of dam practically constant as the interaction effects were negligible relative to the main effects. The difference between male and female calves was about 0.5. Eventually, a model including sex plus age of dam (without interaction) for the location structure and only age of dam for the scaling part seemed to account well for the variation in the

**Table V.** Maximum likelihood estimates of parameters and their asymptotic standard errors (SE) for threshold models applied to calving difficulty scores in Simmental cattle (H-TM: S + A or S + A + SA for location; A\* for scale).

<i>Parameter</i>		<i>S + A + SA; A*</i>		<i>S + A; A*</i>	
		<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>
Threshold	2 – 3	0.783	0.010	0.783	0.009
Location					
$\mu$	1 × 1	-0.298	0.011	-0.296	0.010
S	2	-0.510	0.015	-0.513	0.009
A	2	-0.279	0.019	-0.285	0.018
	3	-0.940	0.040	-0.942	0.040
	4	-1.408	0.046	-1.421	0.045
	5	-1.631	0.054	-1.615	0.051
	6	-1.933	0.056	-1.953	0.054
	7	-2.074	0.063	-2.049	0.058
	8	-2.132	0.057	-2.146	0.056
	9	-2.378	0.114	-2.359	0.110
	S × A	2	-0.026	0.021	
3		0.027	0.031		
4		-0.008	0.030		
5		0.059	0.041		
6		-0.006	0.040		
7		0.097	0.052		
8		0.010	0.030		
9		-0.036	0.049		
Scale		A*			
	2	0.053	0.016	0.047	0.015
	3	0.132	0.025	0.142	0.024
	4*	0.219	0.022	0.224	0.021
	5*	0.326	0.021	0.334	0.019
	6*	0.400	0.040	0.388	0.036
$\chi^2(D)$		20.01 (20.28)		32.18 (32.10)	
Test					
<i>df</i>		12		20	
<i>P</i> -value		0.067		0.041	

S: Sex effect stands for  $s_2 - s_1$  (females – males) in additive models and  $\mu_{21} - \mu_{11}$  (females – males in age of dam 1) in models with interaction. A: age of dam effects stand for  $a_j - a_1$  ( $j$ th parity – 1st parity) for  $j = 2, 3, \dots, 9$  in additive models ( $4^* = 4 + 5$ ,  $5^* = 6 + 7 + 8$ ,  $6^* = 9$ ) and  $\mu_{1j} - \mu_{11}$  ( $j$ th age of dam – 1st age of dam within males) in models with interaction. S × A: interaction effects (SA) stand for  $sa_j = (\mu_{2j} - \mu_{1j}) - (\mu_{21} - \mu_{11})$  (differences between female – male deviations in age of dam  $j$  and age of dam 1) for  $j = 2, 3, \dots, 9$ .  $X^2(D)$ : Pearson's and deviance statistics respectively.

Logistic heteroskedastic models have been considered by McCullagh (1980) and Derquenne (1995). Formulae are given in *Appendix 1* to deal with this distribution.

When the Simmental data are analyzed with the logistic (table VI), the homoskedastic model is also rejected although the fit is not as poor as with the

**Table IV.** Maximum likelihood estimates of the scaling parameter by sex  $\times$  parity subclasses of calving difficulty in the Simmental breed.

Model <sup>a</sup>	Sex	Age of dam (year)					
		< 2.0	2.0-2.5	2.5-3.0	3.0-4.0	4.0-8.0	> 8.0
S + A + SA; S* + A*	1	1.000	1.054	1.141	1.243	1.384	1.489
S + A + SA; S* + A*	2	0.982	1.035	1.120	1.221	1.360	1.462
S + A + SA; A*	1 = 2	1.000	1.054	1.142	1.245	1.387	1.492
S + A; A*	1 = 2	1.000	1.048	1.152	1.250	1.396	1.473

<sup>a</sup> Model for location; \* model for scale; S = sex; A = age of dam.

data. Wald's and deviance statistics were in very good agreement in that respect, with  $P$  values of 0.08 and 0.16 respectively, for the SA interaction. It should be observed that this heteroskedastic model has even fewer parameters (16) than the two-way S-TM considered initially (19 parameters). In spite of this, the Pearson's chi-square (and also the deviance) was reduced from about 419 (table II) to 32 (table V) with a  $P$ -value of 0.04. This fit is remarkable for this large data set ( $N = 363859$ ), where one would expect many models to be rejected.

Although the H-TM may have captured some extra hidden variation due to ignoring random effects, it is unlikely that the poor fit of the S-TM can be attributed solely to the overdispersion phenomenon resulting from ignoring genetic and other clustering effects. The large value of the ratio of the observed  $X^2$  to its expected value ( $419/17 = 25$ ) suggests that the dependency of the probabilities  $\Pi_{ij}$  with respect to sex of calf and age of dam is not described properly by a model with constant variance. Whether the poor fit of the S-TM is the result of ignoring random effects, heterogeneous variance, or both, require further study, perhaps simulation.

These results suggest that in beef cattle breeding the goodness of fit of a constant variance threshold model for calving ease can be improved by incorporating scale effects for age of dam either as discrete classes, as in this study, or alternatively  $\sigma_i$  as a polynomial regression of  $\log \sigma_i$  on age.

## DISCUSSION

### Other distributional assumptions

The underlying distribution was supposed to be normal which is a standard assumption of threshold models in a genetic context (Gianola, 1982). However, other distributions might have been considered for modeling  $\Pi_{ij}$  in [3]. A classical choice, especially in epidemiology, would be the logistic distribution with mean  $\eta_i$  and variance  $\pi^2\sigma_i^2/3$  (Collett, 1991, page 93), where

$$\Pr(X_{ir} \leq \tau_j) = \frac{\exp[(\tau_j - \eta_i)/\sigma_i]}{1 + \exp[(\tau_j - \eta_i)/\sigma_i]} \quad [18]$$

probit. Interestingly, there is not much difference between the complete (S+A+SA) and the additive model (S + A), the interaction (SA) being non-significant ( $P = 0.30$ ). Taking into account the variation in variance in addition to that explained by the additive model on location parameters does not improve the fit greatly. In that respect, the main source of variation turned out to be sex rather than age of dam.

**Table VI.** Deviance statistics for logistic models applied to calving difficulty scores in Simmental cattle.

No	Model		df	Deviance (D)	Covariate tested	Models contrasted	Statistic ( $\Delta D$ )	df	P- value
	Location <sup>a</sup>	Dispersion <sup>b</sup>							
(1)	S + A + SA	None	17	132.99					
(2)	S + A	None	25	142.53	SA	(2) - (1)	9.54	8	0.30
(3)	S + A	S* + A*	19	103.85	S* + A*	(2) - (3)	38.68	6	$10^{-6}$
(4)	S + A	S*	24	109.75	A*	(4) - (3)	5.90	5	0.32

<sup>a</sup> Location covariates: S, A, SA stand for sex, age of dam and sex by age of dam interaction;

<sup>b</sup> dispersion covariates: same symbols with a star (\*) and age defined as in table II with six classes.

Other options include the  $t$ -distribution (Albert and Chib, 1993), the Edgeworth series distribution (Prentice, 1976) and other non-normal classes of distribution functions (Singh, 1987). In fact, the  $t$  distribution  $t_v(\eta_i, s^2)$  with spread parameter  $s^2$  and  $v$  degrees of freedom is the marginal distribution of a mixture of a normal distribution  $N(\eta_i, \sigma_i^2)$ , with  $\sigma_i^2$  randomly varying according to a scaled inverted  $\chi^2(v, s^2)$  distribution (Zellner, 1976). Therefore, a threshold model based on such a  $t$  distribution is embedded in our procedure by taking a one-way random model for  $\ln \sigma_i^2$ , ie,  $\ln \sigma_i^2 = \ln s^2 + a_i$  with the density function  $p(a_i|v)$  of the random variable  $a_i$  as presented in Foulley and Quaas (1995, formulae 21 and 22); see also Gianola and Sorensen (1996) for a specific study of the threshold model based on the  $t$ -distribution in animal breeding.

### Relationships with variable thresholds

Conceptually, heterogeneity of the  $\sigma_i^2$ 's is viewed here in the same way as in Gaussian linear models since it applies to an underlying random variable that is normally distributed. However, the underlying variables are not observable, and the corresponding real line includes cutoff points, the thresholds, that make the outcomes discrete. It is of interest to address the question of how changes in dispersion can be interpreted with respect to the threshold concept.

Let us illustrate this by a simple example involving  $J = 3$  categories, and a one-way classification model ( $i = 1, 2, \dots, I$ ) as, for example, sex of calf in the Simmental breed. We will assume that the origin is at the first threshold, and that the unit of measurement is the standard deviation within males ( $\sigma_M = 1$ ). The difference between the first and second threshold values in males is expressible as:

$$\frac{\tau_2 - \tau_1}{\sigma_M} = \Phi^{-1}(\Pi_{M1} + \Pi_{M2}) - \Phi^{-1}(\Pi_{M1})$$

where  $\Pi_{M1}$ ,  $\Pi_{M2}$  are the probabilities of response in the first and second categories, respectively, for male calves. A similar expression is obtained for female calves (F), so

$$\frac{\sigma_F}{\sigma_M} = \frac{\Phi^{-1}(\Pi_{M1} + \Pi_{M2}) - \Phi^{-1}(\Pi_{M1})}{\Phi^{-1}(\Pi_{F1} + \Pi_{F2}) - \Phi^{-1}(\Pi_{F1})} \quad [19]$$

This is precisely the ratio of the difference between thresholds 1 and 2 that would be obtained when evaluated separately in each sex. Thus

$$\frac{\sigma_F}{\sigma_M} = \frac{(\tau_2 - \tau_1)_M}{(\tau_2 - \tau_1)_F} \quad [20]$$

Formulae [19] and [20] are analogous to expressions given by Wright (1934b) (the reciprocal of the distance between the thresholds on this scale gives the standard deviation on the postulated scale on which the thresholds are separated by a unit distance, p 545) and Falconer (1989, formula 18.5, p 307) except that these authors set to unity the difference between thresholds in the baseline population, rather than the standard deviation, which we find more appealing conceptually.

In the case of the Simmental data shown in table I, applying formulae [19] and [20] using observed frequencies of responses gives:

$$\frac{\sigma_F}{\sigma_M} = \frac{\Phi^{-1}(0.8807) - \Phi^{-1}(0.9618)}{\Phi^{-1}(0.9860) - \Phi^{-1}(0.9409)} = \frac{1.7725 - 1.1788}{2.1960 - 1.5625} = 0.937$$

If more than three categories are observed, this formula also holds for the differences  $\tau_3 - \tau_2$ ,  $\tau_4 - \tau_3, \dots, \tau_{J-1} - \tau_{J-2}$ , so that the ratio between standard deviations in subpopulation ( $i$ ) and a reference population ( $R$ ) can be expressed as:

$$\frac{\sigma_i}{\sigma_R} = \frac{(\tau_j - \tau_1)_R}{(\tau_j - \tau_1)_i}, \forall j = 2, \dots, J - 1 \quad [21]$$

which involves  $(J - 2)$  algebraically independent equalities.

In the case of three categories and a single classification, the saturated model has  $2I$  parameters  $(\tau_2 - \tau_1, \mu_1, \mu_2, \dots, \mu_I, \sigma_2/\sigma_1, \dots, \sigma_i/\sigma_1, \dots, \sigma_I/\sigma_1)$ . Numerical values of  $\sigma_i/\sigma_1$  computed from [20] are also ML estimates (eg,  $\hat{\sigma}_2/\hat{\sigma}_1 = 0.973$ ).

Formula [21] indicates that there is a link between H-TM and models with variable thresholds (Terza, 1985). As compared to these, the main features of the H-TM are:

- i) a multiplicative model on ratios of standard deviations or differences between thresholds, rather than a linear model on such differences;
- ii) a lower dimensional parameterization due to the proportionality assumption made in [18] rather than a category-specific parameterization, ie:

$$(\tau_j - \tau_1)_i = \mathbf{p}'_i \boldsymbol{\delta}_j$$

where  $\boldsymbol{\delta}_j$  is the vector of unknowns pertaining to the difference  $(\tau_j - \tau_1)$ .

For  $J = 3$ , the two models generate the same number of parameters but they are still different vis-a-vis to (i).

### Further extensions

The H-TM opens new perspectives for the analysis of ordinal responses. Interesting extensions may include:

- i) implementing other inference approaches for mixed models such as Gilmour's procedure based on quasi-likelihood, or a fully Bayesian analysis of parameters using Monte-Carlo Markov-Chain methods along the lines of Sorensen et al (1995);
- ii) assessing the potential increase in response to selection by selecting on estimated breeding values calculated from an H-TM versus an S-TM;
- iii) incorporating a mixed linear model on log-variances as described in San Cristobal et al (1993) and Foulley and Quaas (1995) for Gaussian observations;
- iv) carrying out a joint analysis of continuous and ordered polychotomous traits as already proposed for the S-TM by Foulley et al (1983), Janss and Foulley (1993) and Hoeschele et al (1995).

Further research is also needed at the theoretical level to look at the sampling properties of estimators based on mis-specified models. For instance, one may be interested in the asymptotic properties of the ML estimator of  $(\boldsymbol{\tau}', \boldsymbol{\beta}', \boldsymbol{\delta}')$  derived under the assumption of independence of the  $\mathbf{y}_{ir}$ 's when this hypothesis does not hold. This problem has been discussed in general by White (1982), and it may be conjectured from the results of Liang et al (1992) that the ML estimators of these parameters remain consistent. It might also be worthwhile to assess the effect of departures from independence on testing procedures. The generalized chi-square procedure for goodness of fit derived by McLaren et al (1994) might be useful in that respect for analyses based on large samples.

### ACKNOWLEDGMENTS

Part of this research was conducted while JL Foulley was a Visiting Professor at the University of Wisconsin-Madison. He greatly acknowledges the support of this institution and of the Direction des productions animales and Direction des relations internationales, INRA. The manuscript is largely drawn from invited lectures given by the senior author at the Midwestern Animal Science Meeting (Des Moines, Iowa, April 12 1995), the University of Wisconsin-Madison (12 April 1995), Cornell University (15 May 1995) and the University of Illinois (28 August 1995). Special thanks are expressed to all those who contributed substantially to the discussions generated thereby, and also to B Klei and B Cunningham from the US Simmental Association for providing the data set on calving ease scores analyzed in this study. The H-TM has been applied to calving performance of other cattle breeds (US Holstein, French Blonde d'Aquitaine, Charolais, Limousin, Maine Anjou). It has also been applied to prolificacy records in sheep. We are grateful to J Berger (Iowa State University, Ames), F Ménéssier, J Sapa (INRA Genetics, Jouy) and JP Poivey (INRA Genetics, Toulouse) for providing the corresponding data sets.

### REFERENCES

- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88, 669-679
- Brillinger DR (1985) *What do seismology and neurophysiology have in common? - statistics*. Technical report 50, University of California, Berkeley



- Bryant WK, Gerner (1982) The demand for service contracts. *J Business* 55, 345-366
- Bulmer MG, Bull JJ (1982) Models of polygenic sex determination and sex ratio control. *Evolution* 36, 13-26
- Collet (1991) *Modelling Binary Data*. Chapman and Hall, London
- Curnow R, Smith C (1975) Multifactorial models for familial diseases in man. *J R Stat Soc A* 138, 131-169
- Dempster ER, Lerner IM (1950) Heritability of threshold characters. *Genetics* 35, 212-236
- Denis JE Jr, Schnabel RB (1983) *Numerical Methods for Unconstrained optimization and Non-Linear Equations*. Prentice-Hall Inc, Englewood Cliffs
- Derquenne (1995) Heteroskedastic logit models. 50th Session of the International Statistical Institute (ISI), Beijing, China, 21-29 August
- Edwards AL, Thurstone LL (1952) An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika* 17, 169-180
- 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
- Falconer DS (1960) *An Introduction to Quantitative Genetics*. 1st edition, Oliver and Boyd, London
- Falconer DS (1965) The inheritance of liability to certain diseases estimated from the incidence among relatives. *Ann Hum Genet* 29, 51-76
- Falconer DS (1989) *An Introduction to Quantitative Genetics*. 3rd edition, Longman Scientific and Technical, Harlow
- Fahrmeir L, Tutz G (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York
- Fechner GT (1860) *Elemente der Psychophysik*, 2 vols: Vol 1 translated (1966). In: *Elements of Psychophysics* (DH Howes, EG Boring, eds), Holt, Rinehart and Winston, New York
- Foulley JL (1987) Méthodes d'évaluation des reproducteurs pour des caractères discrets à déterminisme polygénique en sélection animale. PhD thesis, University of Paris-Sud-Orsay
- Foulley JL, Gianola D, Thompson R (1983) Prediction of genetic merit from data on categorical and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet Sel Evol* 15, 401-424
- Foulley JL, Gianola D, San Cristobal M, Im S (1990) A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci* 73, 1612-1624
- Foulley JL, San Cristobal M, Gianola D, Im S (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Comput Stat Data Anal* 13, 291-305
- Foulley JL, Quaas RL (1995) Heterogeneous variances in Gaussian linear mixed models. *Genet Sel Evol* 27, 211-228
- Galton F (1889) *Natural Inheritance*. Macmillan, London
- Gianola D (1982) Theory and analysis of threshold characters. *J Anim Sci* 54, 1079-1096
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-224
- Gianola D, Sorensen D (1996) Abstract to the EAAP, Lillehammer, Norway, 25-29 August
- Gilmour A, Anderson RD, Rae A (1987) Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear model. *J Anim Breed Genet* 104, 149-155
- Grosbras JM (1987) Les données manquantes. In: *Les sondages* (JJ Dreesbeke, B Fichet, P Tassi, eds), Economica, Paris, 173-195

- Harville DA, Mee RW (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408
- Höschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-349
- Höschele I, Tier B, Graser HU (1995) Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. *J Anim Sci* 73, 1609-1627
- Janss L, Foulley JL (1993) Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty. *Livest Prod Sci* 33, 183-198
- Kaplan RS, Urwitz G (1979) Statistical models of bond ratings: a methodological inquiry. *J Business* 52, 231-261
- Levy F (1980) Changes in employment prospects for black males. *Brooking Papers* 2, 513-538
- Liang KY, Zeger SL, Qaqish B (1992) Multivariate regression analysis for categorical data. *J R Stat Soc B* 54, 3-40
- Maddala GS (1983) *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York
- McCullagh P (1980) Regression models for ordinal data. *J R Stat Soc B* 42, 109-142
- McCullagh P, Nelder J (1989) *Generalized Linear Models*. 2nd ed, Chapman and Hall, London
- McKelvey R, Zavoina W (1975) A statistical model for the analysis of ordinal level dependent variables. *J Math Soc* 4, 103-120
- McLaren CE, Lecler JM, Brittenham GM (1994) The generalized chi square goodness-of-fit test. *Statistician* 43, 247-258
- Misztal I, Gianola D, Höschele I (1988) Threshold model with heterogeneous residual variance due to missing information. *Genet Sel Evol* 20, 511-516
- Natarajan R (1995) Estimation in Models for Multinomial Response Data: Bayesian and Frequentist Approaches. PhD thesis, Cornell University, Ithaca
- Pearson K (1900) Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. *Phil Trans Roy Soc London A* 195,79
- Prentice RL (1976) Generalisation of probit and logit models. *Biometrika* 32, 761-768
- Quaas RL, Zhao Y, Pollak EJ (1988) Describing interactions in dystocia scores with a threshold model. *J Anim Sci* 66, 396-399
- Read I, Cressie N (1988) *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, New York
- Robertson A (1950) Proof that the additive heritability on the  $p$  scale is given by the expression  $\bar{z}^2 h_x^2 / \bar{p}\bar{q}$ . *Genetics* 35, 234-236
- Roff DA (1994) The evolution of dimorphic traits: predicting the genetic correlation between environments. *Genetics* 136, 395-401
- San Cristobal M, Foulley JL, Manfredi E (1993) Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet Sel Evol* 25, 3-30
- Singh M (1987) A non-normal class of distribution for dose binary response curve. *J Appl Stat* 14, 91-97
- Sorensen DA, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* 27, 229-249
- Terza JV (1985) Ordinal probit: a generalization. *Commun Stat Theor Meth* 14, 1-11
- Visscher PM (1992) Power of likelihood ratio tests for heterogeneity of intraclass correlation and variance in balanced half-sib designs. *J Dairy Sci* 75, 1320-1330

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25  
 Wright S (1934a) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19, 506-536  
 Wright S (1934b) The results of crosses between inbred strains of guinea pigs differing in number of digits. *Genetics* 19, 537-551  
 Zellner A (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *J Am Stat Assoc* 71, 400-405

**APPENDIX 1**

**Expressions for the score function *U* and the information matrix *J***

**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  $\Pi_{ij}$  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. Then,

$$\frac{\partial L}{\partial \tau_j} = \sum_{i=1}^I \left( y_{ij} \frac{\partial \ln \Pi_{ij} \partial \Pi_{ij}}{\partial \Pi_{ij}} \frac{\partial \Pi_{ij}}{\partial \tau_j} + y_{i,j+1} \frac{\partial \ln \Pi_{i,j+1}}{\partial \Pi_{i,j+1}} \frac{\partial \Pi_{i,j+1}}{\partial \tau_j} \right)$$

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

Now,

$$\frac{\partial \Pi_{ij}}{\partial \tau_j} = \sigma_i^{-1} f(h_{ij}); \quad \frac{\partial \Pi_{i,j+1}}{\partial \tau_j} = -\sigma_i^{-1} f(h_{ij}) \tag{A.0}$$

where  $f(x) = dF(x)/dx$ , ie,  $f(x) = \phi(x)$ , and  $f(x) = F(x)[1 - F(x)]$  for the probit and logit links respectively.

Thus:

$$\frac{\partial L}{\partial \boldsymbol{\tau}} = \boldsymbol{\xi}$$

with

$$\boldsymbol{\xi} = \{\xi_j\} = \left\{ \frac{\partial L}{\partial \tau_j} \right\} = \left\{ \sum_{i=1}^I \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) f(h_{ij}) \sigma_i^{-1} \right\} j = 1, 2, \dots, J - 1 \tag{A.1}$$

Concerning derivatives with respect to the vector  $\boldsymbol{\beta}$  of location parameters, one has:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^I \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} \frac{\partial \Pi_{ij}}{\partial \eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$

Now,

$$\begin{aligned} \frac{\partial \Pi_{ij}}{\partial \eta_i} &= \sigma_i^{-1} [f(h_{i,j-1}) - f(h_{ij})]; \\ \frac{\partial \eta_i}{\partial \beta} &= \mathbf{x}_i \end{aligned}$$

so that:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^I \nu_{\beta,i} \mathbf{x}_i = \mathbf{X}' \mathbf{v}_{\beta} \tag{A.2}$$

with

$$\mathbf{v}_{\beta} = \{\nu_{\beta,i}\} = \left\{ \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} [f(h_{i,j-1}) - f(h_{ij})] \sigma_i^{-1} \right\}; \quad i = 1, 2, \dots, I$$

Similarly,

$$\frac{\partial L}{\partial \delta} = \sum_{i=1}^I \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} \frac{\partial \Pi_{ij}}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \ln \sigma_i} \frac{\partial \ln \sigma_i}{\partial \delta}$$

Now,

$$\begin{aligned} \frac{\partial \Pi_{ij}}{\partial \sigma_i} &= f(h_{ij}) \frac{\partial h_{ij}}{\partial \sigma_i} - f(h_{i,j-1}) \frac{\partial h_{i,j-1}}{\partial \sigma_i} \\ \frac{\partial h_{ij}}{\partial \sigma_i} &= -\frac{(\tau_j - \eta_i)}{\sigma_i^2} = -\frac{h_{ij}}{\sigma_i} \\ \frac{\partial \Pi_{ij}}{\partial \sigma_i} &= \sigma_i^{-1} [h_{i,j-1} f(h_{i,j-1}) - h_{ij} f(h_{ij})] \\ \frac{\partial \sigma_i}{\partial \ln \sigma_i} &= \sigma_i \\ \frac{\partial \ln \sigma_i}{\partial \delta} &= \mathbf{p}_i \end{aligned}$$

Thus, letting  $\Psi(x) = xf(x)$ :

$$\frac{\partial L}{\partial \delta} = \sum_{i=1}^I \nu_{\delta,i} \mathbf{p}_i = \mathbf{P}' \mathbf{v}_{\delta} \tag{A.3}$$

with

$$\mathbf{v}_{\delta} = \{\nu_{\delta,i}\} = \left\{ \sum_{j=1}^J \frac{y_{ij}}{\Pi_{ij}} [\Psi(h_{i,j-1}) - \Psi(h_{ij})] \right\}; \quad i = 1, 2, \dots, I$$

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

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

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

with expressions for  $\xi$ ,  $\mathbf{v}_\beta$  and  $\mathbf{v}_\delta$  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;  $\beta$ -threshold;  $\delta$ -threshold;  $\beta - \beta$ ;  $\beta - \delta$ ; and  $\delta - \delta$ .

**Threshold-threshold derivatives**

$$\begin{aligned} \frac{\partial^2 L}{\partial \tau_j^2} &= \sum_{i=1}^I \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) \frac{\partial f(h_{ij})}{\partial \tau_j} \sigma_i^{-1} \\ &\quad + \sum_{i=1}^I \left( \frac{y_{i,j+1}}{\Pi_{i,j+1}^2} \frac{\partial \Pi_{i,j+1}}{\partial \tau_j} - \frac{y_{ij}}{\Pi_{ij}^2} \frac{\partial \Pi_{ij}}{\partial \tau_j} \right) f(h_{ij}) \sigma_i^{-1} \end{aligned}$$

Now,  $E(y_{ij}) = n_i \Pi_{ij}$  so that the expectation of the first term vanishes.  $\frac{\partial \Pi_{ij}}{\partial \tau_j}$  and  $\frac{\partial \Pi_{i,j+1}}{\partial \tau_j}$  were already derived in [A.0].

Thus:

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

Similarly

$$\frac{\partial^2 L}{\partial \tau_j \partial \tau_{j+1}} = - \sum_{i=1}^I \frac{y_{i,j+1}}{\Pi_{i,j+1}^2} \frac{\partial \Pi_{i,j+1}}{\partial \tau_j} f(h_{i,j+1}) \sigma_i^{-1}$$

As seen previously

$$\frac{\partial \Pi_{i,j+1}}{\partial \tau_j} = -\sigma_i^{-1} f(h_{ij})$$

Then

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

and

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

**β-Threshold derivatives**

$$\frac{\partial^2 L}{\partial \tau_j \partial \beta} = \frac{\partial}{\partial \beta} \left( \frac{\partial L}{\partial \tau_j} \right)$$

and using [A.1], one has:

$$\begin{aligned} \frac{\partial^2 L}{\partial \tau_j \partial \beta} &= \sum_{i=1}^I \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) \frac{\partial f(h_{ij})}{\partial \beta} \sigma_i^{-1} \\ &\quad + \sum_{i=1}^I f(h_{ij}) \sigma_i^{-1} \left[ \frac{\partial}{\partial \eta_i} \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) \right] \mathbf{x}_i \end{aligned}$$

The expectation of the first term vanishes because  $E(y_{ij}) = n_i \Pi_{ij}$ .  
 Moreover,

$$\frac{\partial \Pi_{ij}^{-1}}{\partial \eta_i} = -\frac{\partial \Pi_{ij} / \partial \eta_i}{\Pi_{ij}^2} = -\frac{1}{\sigma_i \Pi_{ij}^2} [f(h_{i,j-1}) - f(h_{ij})]$$

Thus:

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

with

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

$i = 1, 2, \dots, I$

and

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

with

$$\mathbf{L}_{\beta_{\{Jx(J-1)\}}} = (\mathbf{L}_{\beta,1}, \mathbf{L}_{\beta,2}, \dots, \mathbf{L}_{\beta,j}, \dots, \mathbf{L}_{\beta,J-1})$$

**δ-Threshold derivatives**

$$\begin{aligned} \frac{\partial^2 L}{\partial \tau_j \partial \delta} &= \sum_{i=1}^I \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) \frac{\partial [f(h_{ij}) \sigma_i^{-1}] }{\partial \delta} \\ &\quad + \sum_{i=1}^I f(h_{ij}) \sigma_i^{-1} \left[ \frac{\partial}{\partial \sigma_i} \left( \frac{y_{ij}}{\Pi_{ij}} - \frac{y_{i,j+1}}{\Pi_{i,j+1}} \right) \right] \frac{\partial \sigma_i}{\partial \ln \sigma_i} \mathbf{p}_i \end{aligned}$$

Again, the expectation of the first term is equal to zero.

$$\frac{\partial \Pi_{ij}^{-1}}{\partial \sigma_i} = -\frac{\partial \Pi_{ij} / \partial \sigma_i}{\Pi_{ij}^2} = -\frac{1}{\sigma_i \Pi_{ij}^2} [\Psi(h_{i,j-1}) - \Psi(h_{ij})]$$

so that:

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

with

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

$i = 1, 2, \dots, I$

and

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

with

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

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

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

### $\beta - \beta$ Derivatives

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

with

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

$i = 1, 2, \dots, I$

### $\beta - \delta$ Derivatives

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

with

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

$i = 1, 2, \dots, I$

**$\delta$ - $\delta$  Derivatives**

$$\mathbb{E} \left( -\frac{\partial^2 L}{\partial \delta \partial \delta'} \right) = \sum_{i=1}^I w_{\delta\delta,ii} \mathbf{P}_i \mathbf{P}_i' = \mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \tag{A.11}$$

with

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

$i = 1, 2, \dots, I$

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

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

where  $\mathbf{T} = \{t_{jk}\}$  is a  $(J - 1) \times (J - 1)$  symmetric band matrix having as elements:  $t_{ij} = \mathbb{E}(-\partial^2 L / \partial \tau_j^2)$ , and  $t_{j,j+1} = \mathbb{E}(-\partial^2 L / \partial \tau_j \partial \tau_{j+1})$ , 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:

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

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.



APPENDIX 2

*A numerical example for the mixed-model approach of the H-TM*

**Table A2.I.** Numerical example: distribution of records.

No	a	b	s	t	Category			
					[1]	[2]	[3]	[4]
1	1	1	1	4	6	11	1	3
2	1	2	1	4	4	14	1	0
3	1	1	1	7	2	8	2	2
4	1	2	1	7	3	4	0	0
5	1	2	1	8	5	1	0	0
6	1	1	2	6	1	3	6	2
7	1	1	2	7	5	2	0	0
8	1	2	2	8	9	8	1	0
9	2	1	2	8	17	4	0	6
10	2	2	3	5	1	2	1	3
11	2	1	3	5	2	5	2	10
12	2	2	3	5	0	3	2	5
13	2	3	3	2	17	4	3	13
14	2	2	3	2	2	4	3	4
15	2	3	4	7	2	2	1	8
16	2	3	4	8	2	1	0	3
17	2	2	4	9	0	9	3	7
18	2	2	4	5	0	2	3	7

a, b, s and 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  $i = 1, 2, \dots, 10$ .

**Table A2.II.** MAP estimates of parameters and their estimated posterior precision.

	<i>Parameter</i>	<i>Estimate</i>	<i>Precision</i> <sup>-0.5</sup>	<i>Parameter</i>	<i>Estimate</i>	<i>Precision</i> <sup>-0.5</sup>
Fixed <sup>a</sup>	$\tau_2 - \tau_1$	1.077	0.148	$\eta_R^*$	0.000	0.000
	$\tau_3 - \tau_1$	1.540	0.205	$a_2^* - a_1^*$	0.545	0.175
	$\eta_R - \tau_1$	0.568	0.361	$b_2^* - b_1^*$	-0.717	0.170
	$a_2 - a_1$	0.874	0.292	$b_3^* - b_1^*$	0.496	0.312
	$b_2 - b_1$	-0.237	0.248			
	$b_3 - b_1$	-0.303	0.787			
Random <sup>b</sup>	$s_1$	-0.202	0.282	$t_1$	0.066	0.331
	$s_2$	-0.296	0.266	$t_2$	-0.145	0.278
	$s_3$	-0.037	0.260	$t_3$	-0.015	0.325
	$s_4$	0.365	0.268	$t_4$	0.192	0.250
	$s_5$	-0.142	0.337	$t_5$	0.198	0.259
	$s_6$	0.278	0.330	$t_6$	0.479	0.279
	$s_7$	-0.021	0.330	$t_7$	-0.016	0.229
	$s_8$	-0.126	0.347	$t_8$	-0.503	0.227
	$s_9$	-0.054	0.349	$t_9$	-0.214	0.270
	$s_{10}$	-0.072	0.351	$t_{10}$	-0.287	0.317

<sup>a</sup>  $\tau_1, \tau_2, \tau_3$ : 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 factor  $b$ ; Wald statistics for the  $a$  and  $b$  factors on the log-scaling parameters are: 9.74 ( $df = 1$ ) and 30.35 ( $df = 2$ ) respectively with corresponding  $P$ -values of 0.0018 and  $2.5 \times 10^{-7}$ ; <sup>b</sup> components of variance and covariance used for sire ( $s$ ) and maternal grand sire effects ( $t$ ) are:  $\sigma_s^2 = \sigma_t^2 = 0.125$ ;  $\sigma_{st} = 0.03125$ , assuming that the residual component  $\sigma_e^2 = 1$ .