

# Cumulative *t*-link threshold models for the genetic analysis of calving ease scores

Kadir KIZILKAYA<sup>a</sup>, Paolo CARNIER<sup>b</sup>,  
Andrea ALBERA<sup>c</sup>, Giovanni BITTANTE<sup>b</sup>, Robert J. TEMPELMAN<sup>a\*</sup>

<sup>a</sup> Department of Animal Science, Michigan State University,  
East Lansing 48824, USA

<sup>b</sup> Department of Animal Science, University of Padova,  
Agripolis, 35020 Legnaro, Italy

<sup>c</sup> Associazione Nazionale Allevatori Bovini di Razza Piemontese,  
Strada Trinità 32a, 12061 Carrù, Italy

(Received 24 June 2002; accepted 10 March 2003)

**Abstract** – In this study, a hierarchical threshold mixed model based on a cumulative *t*-link specification for the analysis of ordinal data or more, specifically, calving ease scores, was developed. The validation of this model and the Markov chain Monte Carlo (MCMC) algorithm was carried out on simulated data from normally and  $t_4$  (*i.e.* a *t*-distribution with four degrees of freedom) distributed populations using the deviance information criterion (DIC) and a pseudo Bayes factor (PBF) measure to validate recently proposed model choice criteria. The simulation study indicated that although inference on the degrees of freedom parameter is possible, MCMC mixing was problematic. Nevertheless, the DIC and PBF were validated to be satisfactory measures of model fit to data. A sire and maternal grandsire cumulative *t*-link model was applied to a calving ease dataset from 8847 Italian Piemontese first parity dams. The cumulative  $t_4$ -link model was shown to lead to posterior means of direct and maternal heritabilities ( $0.40 \pm 0.06$ ,  $0.11 \pm 0.04$ ) and a direct maternal genetic correlation ( $-0.58 \pm 0.15$ ) that were not different from the corresponding posterior means of the heritabilities ( $0.42 \pm 0.07$ ,  $0.14 \pm 0.04$ ) and the genetic correlation ( $-0.55 \pm 0.14$ ) inferred under the conventional cumulative probit link threshold model. Furthermore, the correlation ( $> 0.99$ ) between posterior means of sire progeny merit from the two models suggested no meaningful rerankings. Nevertheless, the cumulative *t*-link model was decisively chosen as the better fitting model for this calving ease data using DIC and PBF.

**threshold model / *t*-distribution / Bayesian inference / calving ease**

---

\* Correspondence and reprints  
E-mail: tempelma@msu.edu

## 1. INTRODUCTION

Data quality is an increasingly important issue for the genetic evaluation of livestock, both from a national and international perspective [13]. Breed associations and government agencies typically invoke arbitrary data quality control edits on continuously recorded production characters in order to minimize the impact of recording error, preferential treatment and/or injury/disease on predicted breeding values [5]. These edits are used in the belief that the data residuals should be normally distributed.

It has been recently demonstrated that the specification of residual distributions in linear mixed models that are heavier-tailed than normal densities may effectively mute the impact of residual outliers, particularly in situations where preferential treatment of some breedstock may be anticipated [41]. Based on the work of Lange *et al.* [24] and others, Strandén and Gianola [42] developed the corresponding hierarchical Bayesian models for animal breeding, using Markov chain Monte Carlo (MCMC) methods for inference. In their models, residuals are specified as either having independent (univariate)  $t$ -distributions or multivariate  $t$ -distributions within herd clusters. Outside of possibly longitudinal studies, the multivariate specification is of dubious merit [36,41,42] such that all of our subsequent discussion pertains to the univariate  $t$ -error specification only.

Auxiliary traits such as calving ease or milking speed are often subjectively scored on an ordinal scale. It might then be anticipated that data quality, including the presence of outliers, would be an issue of greater concern in these traits than more objectively measured production characters, particularly since record keeping is generally unsupervised, being the responsibility of the attending herdsman. As one example of preferential treatment, a herdsman may more quickly decide to assist or even surgically remove a calf from a highly valued dam. Luo *et al.* [25] has furthermore suggested that a decline in the diligence of data recording was partially responsible for their lower heritability estimates of calving ease relative to earlier estimates from the same Canadian Holstein population.

The cumulative probit link (CP) generalized linear mixed model, otherwise called the threshold model, is currently the most commonly used genetic evaluation model for calving ease [4,49]. MCMC methods are particularly well suited to this model since the augmentation of the joint posterior density with normally distributed underlying or latent liability variables facilitate implementations very similar to those developed for linear mixed effects models [2, 39]. A cumulative  $t$ -link (CT) model has been proposed by Albert and Chib [2] for the analysis of ordinal categorical data, thereby providing greater modeling flexibility relative to the CP model. The CT model can be created by simply augmenting the joint posterior density with  $t$ -distributed rather than normally

distributed underlying liability variables [18]. Since outliers on the observed categorical scale also correspond to outliers on the underlying liability scale [1], the CT model might be anticipated to be more robust to residual outliers relative to the CP model.

The objectives of this study were to validate MCMC inference of the CT generalized linear mixed (sire) model *via* a simulation study and to compare the fit of this model with the CP model for the quantitative genetic analysis of calving ease scores in Italian Piemontese cattle. In section 2, the CT model is constructed hierarchically. We then present a discussion of two model choice criteria that we believe are appropriate for the comparisons of the CP with the CT model in section 3. In section 4, we describe a simulation study that is used to validate posterior inference and model choice criteria for the CP and CT models, presenting the results of this study along with an application to Italian Piemontese calving ease data in section 5. We conclude with a discussion of these results in section 6.

## 2. MODEL CONSTRUCTION

Suppose that elements of the  $n \times 1$  data vector  $\mathbf{Y} = \{Y_i\}_{i=1}^n$  can take values in any one of  $C$  mutually exclusive ordered categories. The classical CP model for ordinal data [17] can be written as follows:

$$\text{Prob}(Y_i = j | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}) = \Phi\left(\frac{\tau_j - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\sigma_e}\right) - \Phi\left(\frac{\tau_{j-1} - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\sigma_e}\right), \tag{1}$$

where  $j = 1, 2, \dots, C$  denotes the index for categories. Also,  $\Phi(\cdot)$  denotes the standard normal cumulative distribution function,  $\boldsymbol{\beta}$  and  $\mathbf{u}$  are the vectors of unknown fixed and random effects, and  $\boldsymbol{\tau}' = [\tau_0 \ \tau_1 \ \dots \ \tau_C]$  is a vector of unknown threshold parameters satisfying  $\tau_1 < \tau_2 < \dots < \tau_C$  with  $\tau_0 = -\infty$  and  $\tau_C = +\infty$ . Furthermore,  $\mathbf{x}'_i$  and  $\mathbf{z}'_i$  are known incidence row vectors. Latent liability variables ( $\mathbf{L} = \{L_i\}_{i=1}^n$ ) can be introduced to alternatively define the same specification as in equation (1) but in two hierarchical stages:

$$\text{Prob}(Y_i = j | L_i, \boldsymbol{\tau}) = \sum_{j=1}^C 1(\tau_{j-1} < L_i < \tau_j) 1(Y_i = j), \tag{2a}$$

$$L_i | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2 \sim N(\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u}, \sigma_e^2), \tag{2b}$$

for  $i = 1, 2, \dots, n$ . Here  $1(\cdot)$  denotes an indicator function, which is equal to 1 when the expression in the function is true and is equal to 0 otherwise. As shown by Albert and Chib [2], and in an animal breeding context by Sorensen *et al.* [39], this model augmentation using  $\mathbf{L}$  facilitates a tractable MCMC implementation.

The CT model is a simple generalization of (1), that is,

$$\text{Prob}(Y_i = j | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}, \nu, \sigma_e^2) = F_\nu \left( \frac{\tau_j - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\sigma_e} \right) - F_\nu \left( \frac{\tau_{j-1} - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\sigma_e} \right), \quad (3)$$

for  $j = 1, 2, \dots, C$  where  $F_\nu$  represents the cumulative density function of a standard Student  $t$ -distribution with degrees of freedom  $\nu$ . Note that as  $\nu \rightarrow \infty$ , (3)  $\rightarrow$  (1) such that the standard CP model is simply a special case of the CT model. Like the CP model, the CT model can also be represented as a two-stage specification, with the first stage as in equation (2a) but the second stage specified as:

$$p(L_i | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right) (\nu\sigma_e^2)^{\frac{1}{2}}} \left(1 + \frac{(L_i - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u}))^2}{\nu\sigma_e^2}\right)^{-\frac{1}{2}(\nu+1)}, \quad (4)$$

*i.e.*,  $L_i$  is Student  $t$ -distributed with location parameter  $\mu_i = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u}$ , scale parameter  $\sigma_e^2 > 0$  and degrees of freedom  $\nu > 2$  for  $i = 1, 2, \dots, n$ . In turn, equation (4) can be represented by a two-stage scale mixture of normals:

$$L_i | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \lambda_i \sim N\left(\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u}, \frac{\sigma_e^2}{\lambda_i}\right), \quad (5a)$$

$$p(\lambda_i | \nu) = \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_i^{\frac{\nu}{2}-1} \exp\left\{-\frac{\lambda_i}{2}\nu\right\}. \quad (5b)$$

Note that (5b) specifies a Gamma density with parameters  $\nu/2$  and  $\nu/2$ , thereby having an expectation of 1. The remaining stages of our hierarchical model are characteristic of animal breeding models. We write

$$\boldsymbol{\beta} \sim p(\boldsymbol{\beta}), \quad (6)$$

where  $p(\boldsymbol{\beta})$  is a subjective prior, typically specified to be flat or vaguely informative. Furthermore, the random effects are typically characterized by a structural multivariate prior specification:

$$\mathbf{u} | \boldsymbol{\varphi} \sim p(\mathbf{u} | \boldsymbol{\varphi}) \sim N(\mathbf{0}, \mathbf{G}(\boldsymbol{\varphi})). \quad (7)$$

Here  $\mathbf{G}(\boldsymbol{\varphi})$  is a variance-covariance matrix that is a function of several unknown variance components or variance-covariance matrices in  $\boldsymbol{\varphi}$ , depending on

whether or not there are multiple sets of random effects and/or specified covariances between these sets; an example of the latter is the covariance between additive and maternal genetic effects. Furthermore, flat priors, inverted Gamma densities, inverted Wishart densities or products thereof may be specified for the prior density  $p(\boldsymbol{\varphi})$  on  $\boldsymbol{\varphi}$ , depending, again, on the number of sets of random effects and whether there are any covariances thereof [21].

Finally, a prior is required for the degrees of freedom parameter  $\nu$  to ensure a proper joint posterior density. We use the prior:

$$p(\nu) \propto \frac{1}{(1 + \nu)^2}; \tag{8}$$

which is consistent with a vaguely informative Uniform(0,1) prior on  $1/(1 + \nu)$ .

As with the CP models, there are identifiability issues involving elements of  $\boldsymbol{\tau}$  with  $\sigma_e^2$  such that constraints are necessary. The origin and scale are arbitrary so that, as done by others (e.g. [17]),  $\tau_1$  is set here to zero and  $\sigma_e^2$  to 1. We chose this parameterization such that inference on  $\sigma_e^2$  is not subsequently considered in this paper.

Presuming that the elements of  $\mathbf{Y}$  are conditionally independent given  $\boldsymbol{\beta}$  and  $\mathbf{u}$ , we can write the joint posterior density of all unknown parameters and latent variables ( $\mathbf{L}$ ) as follows:

$$p(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}, \boldsymbol{\varphi}, \nu, \mathbf{L}, \boldsymbol{\lambda} | \mathbf{y}) \propto \left( \prod_{i=1}^n \text{Prob}(y = y_i | L_i, \boldsymbol{\tau}) p(L_i | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \lambda_i) p(\lambda_i | \nu) \right) \times p(\boldsymbol{\beta}) p(\mathbf{u} | \boldsymbol{\varphi}) p(\boldsymbol{\varphi}) p(\nu), \tag{9}$$

where  $\boldsymbol{\lambda} = \{\lambda_i\}_{i=1}^n$ .

An MCMC inference strategy involves determining and generating random variables from the full conditional densities (FCD) of each parameter or blocks thereof. Many of the FCD can be directly derived using results from Sorensen *et al.* [39] jointly with the results from Strandén and Gianola [42]. Let  $\boldsymbol{\theta} = [\boldsymbol{\beta}' \ \mathbf{u}']'$ . It can be readily shown that the FCD of  $\boldsymbol{\theta}$  is multivariate normal:

$$\boldsymbol{\theta} | \mathbf{y}, \nu, \boldsymbol{\lambda}, \mathbf{L} \sim N(\hat{\boldsymbol{\theta}}, \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \boldsymbol{\Sigma}^-), \tag{10}$$

where  $\mathbf{R}^{-1} = \text{diag}\{\lambda_i\}_{i=1}^n$  is a diagonal matrix,  $\boldsymbol{\Sigma}^- = \begin{bmatrix} \mathbf{0}_{p \times p} & \mathbf{0}_{p \times q} \\ \mathbf{0}_{q \times p} & (\mathbf{G}(\boldsymbol{\varphi}))^{-1} \end{bmatrix}$ ,  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}]$  for  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]'$ ,  $\mathbf{Z} = [\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_n]'$ , and

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = [\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \boldsymbol{\Sigma}^-]^{-1} \mathbf{W}'\mathbf{R}^{-1}\mathbf{L}. \tag{11}$$

The generation of individual elements  $\theta_j$ ,  $j = 1, 2, \dots, p + q$  or blocks thereof of  $\boldsymbol{\theta}$  from their respective FCD is straightforward using the strategy presented by Wang *et al.* [48].

The FCD of individual elements of  $\mathbf{L}$  and  $\boldsymbol{\tau}$  are straightforward to generate from, using results from Sorensen *et al.* [39]. We, however, preferred the Metropolis-Hastings and method of composition joint update of  $\mathbf{L}$  and  $\boldsymbol{\tau}$  presented by Cowles [11]. She demonstrated and we have further noted in our previous applications [23] that the resulting MCMC mixing properties using this joint update are vastly superior to using separate Gibbs updates on individual elements of  $\mathbf{L}$  and  $\boldsymbol{\tau}$  as outlined by Sorensen *et al.* [39]. A lucid exposition on Cowles' update is also provided by Johnson and Albert [22].

If some partitions of  $\boldsymbol{\varphi}$  form a variance-covariance matrix, then their respective FCD can be readily shown to be inverted-Wishart [21] whereas if other partitions of  $\boldsymbol{\varphi}$  involve scalar variance but no covariance components, then the FCD of each component can be shown to be inverted-gamma.

The FCD of  $\lambda_i$  can be shown to be:

$$p(\lambda_i | \boldsymbol{\lambda}_{-i}, \mathbf{L}, \boldsymbol{\theta}, \nu) \propto \lambda_i^{\left(\frac{\nu+1}{2}\right)-1} \exp\left(-\frac{\lambda_i}{2}((L_i - \mathbf{x}'\boldsymbol{\beta} - \mathbf{z}'_i\mathbf{u})^2 + \nu)\right), \quad (12)$$

that is, the kernel in (12) specifies that the distribution to be

$$\text{Gamma}\left(\frac{\nu + 1}{2}, \frac{1}{2}(v + (L_i - \mathbf{x}'\boldsymbol{\beta} - \mathbf{z}'_i\mathbf{u})^2)\right).$$

Here  $\boldsymbol{\lambda}_{-i}$  denotes all elements of  $\boldsymbol{\lambda} = \{\lambda_i\}$  except for  $\lambda_i$ ,  $i = 1, 2, \dots, n$ .

Finally, the FCD of  $\nu$  can be shown to be:

$$p(\nu | \boldsymbol{\beta}, \mathbf{u}, \mathbf{L}, \boldsymbol{\lambda}) \propto \left(\frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma(\nu/2)}\right)^n \left(\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_i\right)\right) \frac{1}{(1 + \nu)^2}, \quad (13)$$

given the specification for  $p(\nu)$  in (8). Equation (13) is not a recognizable density such that a Metropolis-Hastings update is required. We utilized a random walk implementation [10] of Metropolis-Hastings sampling; specifically, a normal density with expectation equal to the parameter value from the previous MCMC cycle was used as the proposal density for drawing from the FCD of  $\kappa = \log(\nu)$ , using equation (13) and the necessary Jacobian for this transformation. The Metropolis-Hastings acceptance ratio was tuned to intermediate rates (40–50%) during the MCMC burn-in period to optimize MCMC mixing [10], adapting the tuning strategy of Müller [32]. Since the variance of a  $t$ -density is not defined for  $\nu \leq 2$ , we truncate the sample from (13) such that  $\nu > 2$ , or equivalently  $\kappa > \log(2)$ , consistent with work by previous investigators ([42,47]).

### 3. MODEL COMPARISON

Model choice is an important issue that has not received considerable attention in animal breeding until only very recently (*e.g.* [20,35]). Likelihood ratio tests have been used to compare differences in fit between various models and their reduced subsets; however, these tests do not facilitate more general model comparisons. The Bayes factor has a strong theoretical justification as a general model choice criterion; however algorithms for Bayes factor computations are either computationally intensive (*e.g.* [9]) or numerically unstable [33]. Furthermore, as Gelfand and Ghosh [15] indicate, Bayes factors lack clear interpretation in the case of improper priors which are particularly frequent specifications in animal breeding hierarchical models. The Akaike information criterion or Schwarz Bayesian criterion are analytical measures that provide an asymptotic representation of Bayes factors and reflect a compromise between goodness of fit and number of parameters. Since the total number of parameters and latent variables often exceeds the number of observations in animal breeding (*e.g.* animal model) analysis, the effective number of parameters in hierarchical models is not always so obvious. The MCMC sample average of the posterior log likelihoods, or data sampling log densities, may be used as a means for comparing different models [12]; however, as Spiegelhalter *et al.* [40] indicate, it is not always so obvious how to proceed when these densities are similar but the number of parameters and/or the numbers of hierarchical stages of the candidate models vary. Spiegelhalter *et al.* [40] proposed the deviance information criterion (DIC) for comparing alternative constructions of hierarchical models. The DIC is based on the posterior distribution of the deviance statistic, which is  $-2$  times the sampling distribution of the data as specified in the first stage of a hierarchical model. However, it may not be obvious how to specify the data sampling stage in a hierarchical model. For example, the data sampling stage for the CT model may be specified in one way as:

$$\begin{aligned} &\text{Prob}(Y_i = j | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}, \lambda_i, \sigma_e^2) \\ &= \Phi \left( \frac{\tau_j - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\frac{\sigma_e}{\sqrt{\lambda_i}}} \right) - \Phi \left( \frac{\tau_{j-1} - (\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u})}{\frac{\sigma_e}{\sqrt{\lambda_i}}} \right), \quad (14) \end{aligned}$$

given the specifications of (2a) and (5a) or it may be specified more marginally using (3). We prefer a more marginalized or heavier-tailed first stage specification such as (3) for CT and (1) for CP, potentially leading to a more stable implementation with justification provided by Satagopan *et al.* [38] but with their context being the stabilization of the Bayes factor estimator of Newton and Raftery [33].

The DIC is computed as the sum of average Bayesian deviance ( $\bar{D}$ ) plus the “effective number of parameters” ( $p_D$ ) with respect to a model, such that smaller DIC values indicate better fit to the data. Let  $G$  denote the number of cycles after convergence in an MCMC chain. Furthermore, we represent all unknown parameters in the marginalized first stage specification by  $\vartheta = (\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}, \nu)$  with  $\vartheta$  excluding  $\nu = \infty$  in the CP model. Then, for the CT model, the average Bayesian deviance can be estimated using (3) by

$$\bar{D} = -2 \left( \sum_{g=1}^G \sum_{i=1}^n \log (\text{Prob}(Y = y_i | \boldsymbol{\beta}^{[g]}, \mathbf{u}^{[g]}, \boldsymbol{\tau}^{[g]}, \nu^{[g]})) \right),$$

where the superscript  $^{[g]}$  denotes the MCMC cycle  $g$ ,  $g = 1, 2, \dots, G$  for the sampled value of the corresponding parameter. Furthermore,  $p_D$  can be estimated as  $p_D = \bar{D} - D(\bar{\vartheta})$  where

$$D(\bar{\vartheta}) = -2 \left( \sum_{i=1}^n \log \text{Prob}(Y = y_i | \bar{\boldsymbol{\beta}}, \bar{\mathbf{u}}, \bar{\boldsymbol{\tau}}, \bar{\nu}) \right).$$

Here the bar notation (*e.g.*  $\bar{\vartheta}$ ) denotes the corresponding posterior mean vector.

We alternatively considered the conditional predictive ordinate (CPO) as the basis for model choice [14]. Defined for observation  $i$ , we write the CPO as:

$$\hat{\pi}(y_i | \mathbf{y}_{-i}, M_1) \approx \left[ \frac{1}{G} \sum_{g=1}^G (\text{Prob}(Y = y_i | \boldsymbol{\beta}^{[g]}, \mathbf{u}^{[g]}, \boldsymbol{\tau}^{[g]}))^{-1} \right]^{-1},$$

using (1) for the CP model (Model  $M_1$ ) and

$$\hat{\pi}(y_i | \mathbf{y}_{-i}, M_2) \approx \left[ \frac{1}{G} \sum_{g=1}^G (\text{Prob}(Y = y_i | \boldsymbol{\beta}^{[g]}, \mathbf{u}^{[g]}, \boldsymbol{\tau}^{[g]}, \nu^{[g]}))^{-1} \right]^{-1},$$

using (3) for the CT model (Model  $M_2$ ). Here  $\mathbf{y}_{-i}$  denotes all observations other than  $y_i$ . The log marginal likelihood (LML) of the data for a certain model, say  $M_k$ , can then be estimated as:

$$\text{LML}_k = \sum_{i=1}^n \log (\hat{\pi}(y_i | \mathbf{y}_{-i}, M_k)).$$

A pseudo Bayes factor (PBF) between two models, say Model  $M_1$  and Model  $M_2$ , can be determined by computing the antilog of their LML difference, that is,

$$\text{PBF}_{1,2} = \exp(\text{LML}_1 - \text{LML}_2). \tag{15}$$

Under the assumption of equal prior model probabilities,  $\text{PBF}_{1,2}$  can be interpreted as a surrogate Bayes factor measure [14] and hence the approximate posterior odds of Model 1 relative to Model 2.



## 4. DATA

### 4.1. Simulation study

A simulation study was used to validate the CT model and the utility of the DIC and the PBF for model choice between CP and CT. Three replicated datasets were generated from each of two different populations as characterized by the distribution of the liability residuals. Population I had a residual density that was standard Student- $t$  distributed with scale parameter  $\sigma_e^2 = 1$  and degrees of freedom  $\nu = 4$  whereas Population II had a residual density that was standard normal. All datasets were generated based on a simple random effects (sire) model with a null mean. Liability data for 50 progeny from each of 50 unrelated sires was generated by summing independently drawn sire effects from  $N(0, \sigma_s^2 = 0.10)$  with independently drawn residuals from  $N(0, \sigma_e^2 = 1.00)$  for a total of 2500 records. These underlying liabilities were mapped to ordinal data with four categories based on the threshold parameter values of  $\tau_1 = -0.50$ ,  $\tau_2 = 1.00$ , and  $\tau_3 = 2.00$  for all populations. Ordinal data from each replicated dataset was analyzed using both CP and CT sire models. For the purpose of parameter identifiability, we invoked the restrictions  $\sigma_e^2 = 1$  and  $\tau_1 = -0.50$ . As a positive control, the underlying liability data for each replicate was analyzed using both normal and  $t$  distributed error mixed linear models. For the  $t$ -distributed error model, the MCMC procedure adapted was similar to that presented in Stranden and Gianola [42], except that the degrees of freedom parameter ( $\nu > 2$ ) was inferred as a continuous (rather than discrete) parameter, using the Metropolis-Hastings update as presented earlier. Graphical inspection of the chains based on preliminary analyses was used to determine a common length of burn-in period. For each replicated data set within each population, a burn-in period of 20 000 cycles was seen to be sufficiently large upon which random draws from each of an additional 100 000 MCMC cycles were subsequently saved. Furthermore, DIC and LML values were computed for each model on each replicated dataset to validate those measures as model choice criteria. For the direct mixed linear model analysis of liability data, DIC and LML measures were based on normal and  $t$ -error data sampling densities for their respective models, similar to that implemented for the robust regression example in Spiegelhalter *et al.* [40]. In all cases, flat unbounded priors were invoked on the variance components and on the fixed effects and the vaguely informative prior in (8) was used for  $\nu$ . Furthermore, the effective number of independent samples (ESS) for each parameter was determined using the initial positive sequence estimator of Geyer [16] as adapted by Sorensen *et al.* [39].

### 4.2. Italian Piemontese calving ease data

First parity calving ease scores recorded on Italian Piemontese cattle from January, 1989 to July, 1998 by ANABORAPI (Associazione nazionale

allevatori bovini di razza Piemontese, Strada Trinità 32a, 12061 Carrù, Italy) were used for this study. In order to limit computing demands, only the 66 herds that were represented by at least 100 records over that nine-year period were considered for the demonstration of the proposed methods in this paper, leaving a total of 8847 records. Calving ease was coded into five categories by breeders and subsequently recorded by technicians who visited the breeders monthly. The five ordered categories are: (1) unassisted delivery; (2) assisted easy calving (3) assisted difficult calving (4) caesarean section and (5) foetotomy. Since the incidence of foetotomy was less than 0.5%, the last two ordinal categories were combined, leaving a total of four mutually exclusive categories. The general frequencies of first parity calving ease scores in the data set were 951 (10.75%) for unassisted delivery; 5514 (62.32%) for assisted easy calving; 1316 (14.88%) for assisted difficult calving; and 1066 (12.05%) for caesarean section and foetotomy.

The effects of dam age, sex of the calf, and their interaction were considered by combining eight different age groups (20 to 23, 23 to 25, 25 to 27, 27 to 29, 29 to 31, 31 to 33, 33 to 35, and 35 to 38 months) with the sex of the calf for a total of 16 nominal subclasses. A total of 1212 herd-year-season (HYS) contemporary subclasses were created from combinations of herd, year, and two different seasons (from November to April and from May to October) as in Carnier *et al.* [7] who also analyzed calving ease data from this same population. The sire pedigree file was further pruned by striking out identifications of sires having no daughters with calving ease records and appearing only once as either a sire or a maternal grandsire of a sire having daughters with records in the data file. Pruning results in no loss of pedigree information on parameter estimation yet is effective in reducing the number of random effects and hence computing demands. The number of sires remaining in the pedigree file after pruning was 1929.

As in Kizilkaya *et al.* [23], the CP and CT models used for the analysis of calving ease data included the fixed effects of age of dam classifications, sex of calf and their interaction in  $\beta$ , the random effects of independent herd-year-season effects in  $\mathbf{h}$ , random sire effects in  $\mathbf{s}$  and random maternal grandsire effects in  $\mathbf{m}$ . We assume:

$$\begin{pmatrix} \mathbf{s} \\ \mathbf{m} \end{pmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A} \right),$$

and

$$\mathbf{h} \sim N(\mathbf{0}, \mathbf{I}\sigma_h^2),$$

where  $\mathbf{G}_0 = \begin{pmatrix} \sigma_s^2 & \sigma_{sm} \\ \sigma_{sm} & \sigma_m^2 \end{pmatrix}$ , with  $\sigma_s^2$  denoting the sire variance,  $\sigma_m^2$  denoting the maternal grandsire variance,  $\sigma_{sm}$  denoting the sire-maternal grandsire covariance, and  $\sigma_h^2$  denoting the HYS variance. Furthermore,  $\otimes$  denotes the

Kronecker (direct) product and  $\mathbf{A}$  is the numerator additive relationship matrix between sires due to identified male ancestors [19]. Also,  $\mathbf{h}$  is assumed to be independent of  $\mathbf{s}$  and  $\mathbf{m}$ . Flat unbounded priors were placed on all fixed effects and variance components. Based on the poor mixing results from the simulation study and the increased relative computing demands for this larger data set,  $\nu$  was not inferred upon. That is, since the simulation study was much simpler in design compared to the calving ease dataset, any attempt to infer upon  $\nu$  would prove even more difficult. To provide a stark contrast to the CP model,  $\nu$  was then held constant to 4 in the CT model.

MCMC inference was based on the execution of three different chains for each model. For each chain in the CP model, a total of 5000 cycles of the burn-in period followed by saving samples from each of 100 000 additional cycles was executed based on the experiences of Kizilkaya *et al.* [23]. Because of initially anticipated slower mixing, the corresponding burn-in period for each chain in the CT model was 10 000 cycles followed by saving each of 200 000 additional cycles. To facilitate diagnosis of sufficient MCMC convergence, the starting values on variance components for each chain within a model were widely discrepant, with one chain starting at the posterior mean of all (co)variance components based on the analysis of Kizilkaya *et al.* [23], another chain starting at the posterior mean minus 3 posterior standard deviations for each (co)variance component and the final chain starting at the posterior mean plus 3 posterior standard deviations for each (co) variance component.

As with the simulation study, the ESS for each inferred parameter was determined. Furthermore, key genetic parameters, specifically direct heritability ( $h_d^2$ ), maternal heritability ( $h_m^2$ ) and the direct-maternal genetic correlation ( $r_{dm}$ ) were inferred upon in the calving ease data using the functions of  $\mathbf{G}_o$  as presented by Kizilkaya *et al.* [23] and Luo *et al.* [25], for example. The only difference in the computation of heritabilities between the CP and the CT model was that the marginal residual variance for the underlying liabilities was not  $\sigma_e^2$  in CT, as it is in CP, but is equal to  $\frac{\nu}{\nu - 2}\sigma_e^2$  [42]. Posterior means and the standard deviation of elements of  $\mathbf{s}$  were also compared between the CP and the CT model.

## 5. RESULTS

### 5.1. Simulation study

Table I summarizes inferences on  $\nu$  based on the replicated datasets from the two populations, comparing the CP *versus* CT models for the analysis of ordinal categorical data and comparing the Gaussian linear mixed model *versus* the *t*-error linear mixed model for the analysis of the matched latent or underlying normal liabilities, as if they were directly observed. Inference on  $\nu$

**Table I.** Posterior inferences on the residual degrees of freedom parameter ( $\nu$ ) in a simulation study using the cumulative  $t$ -link model.

Population-Dataset <sup>a</sup>	Liability data			Ordinal data		
	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>
I-1	3.70 $\pm$ 0.30	3.15–4.31	1849	5.58 $\pm$ 1.79	3.58–10.36	74
I-2	4.22 $\pm$ 0.35	3.59–4.97	1927	7.28 $\pm$ 3.02	4.01–15.20	86
I-3	4.26 $\pm$ 0.36	3.61–5.03	1781	9.57 $\pm$ 15.85	3.84–35.38	38
II-1	276 $\pm$ 373.21	42.53–1328.10	21	20.13 $\pm$ 38.93	4.08–126.20	20
II-2	144 $\pm$ 176.50	21.99–653.56	20	121.05 $\pm$ 209.46	6.04–862.12	17
II-3	324 $\pm$ 444.38	30.69–1576.98	13	39.75 $\pm$ 97.82	4.20–411.63	16

<sup>a</sup> Population I specified with  $t$  residual distribution with 4 degrees of freedom; Population II specified with Gaussian residual distribution. Each population replicated for 3 datasets each.

<sup>b</sup> posterior mean  $\pm$  standard deviation.

<sup>c</sup> 95% equal-tailed posterior probability interval based on 2.5th and 97.5th percentiles.

<sup>d</sup> the effective number of independent samples using the initial positive sequence estimator of Geyer [16].

was surprisingly sharp and seemingly unbiased for the  $t$ -error mixed model analysis of liability data from Population I ( $\nu = 4$ ), with 95% equal-tailed posterior probability intervals (PPI) not exceeding 1.5 in width; furthermore, the corresponding ESS were relatively large indicating stable MCMC inference. We define a 95% PPI to be the range of the posterior density falling between the 2.5th and 97.5th percentile. Conversely, inference on  $\nu$  based on the  $t$ -error mixed model analysis of liability data from Population II ( $\nu = \infty$ ) indicated extremely wide 95% PPI and posterior means exceeding 100, indicating stronger evidence of Gaussian distributed *versus*  $t$ -distributed residuals for data from that population. Furthermore, ESS were generally very small ( $\sim 20$ ) indicating that inference on  $\nu$  was rather unreliable for data from Population II, at least given the specified MCMC sampling scheme. That is, five times as many MCMC samples would be needed to attain a minimum of 100 as advocated by previous investigators [6,44].

Inference on  $\nu$  in ordinal data under the CT model was also interesting. In Population I, the 95% PPI correctly concentrated on low values for  $\nu$  although the PPI were understandably wider than for the corresponding analyses of liability data under  $t$ -error linear mixed models. Also, the ESS on  $\nu$  were considerably smaller ( $< 100$ ) for the CT model analysis of ordinal data than for the corresponding matched linear model analyses of liability data, such that acceptably accurate inference on  $\nu$  would require substantially more sampling than what we considered in this study. In replicated ordinal data from Population II, the 95% PPI on  $\nu$  were wide and concentrated on high values of  $\nu$ , consistent with what was expected. Furthermore, as with the  $t$ -error mixed model analysis of liability data, MCMC mixing on  $\nu$  using the CT model on ordinal data was seen to be particularly problematic in Population II as manifested by the small ESS.

Table II summarizes inference on  $\sigma_s^2$  based on the replicated datasets from the two populations, comparing the CP *versus* CT models for the analysis of ordinal categorical data and comparing the Gaussian linear mixed model *versus* the  $t$ -error linear mixed model for the analysis of the underlying liabilities. In the analyses of liability data from replicated datasets from both populations, the 95% PPI were in good agreement with  $\sigma_s^2 = 0.10$ ; furthermore, very large ESS indicated very good MCMC mixing. MCMC mixing on  $\sigma_s^2$  was understandably slower in the analysis of ordinal categorical data, particularly in replicated data from Population II using CT, due to the generally high posterior sampling correlation between  $\sigma_s^2$  and  $\nu$ .

Because of the problem of MCMC mixing of  $\nu$  in ordinal data, the MCMC chains were rerun with  $\nu$  held constant ( $\nu = 4$ ) for more reliable DIC and LML comparisons between models. Both DIC and LML are reported in Table III for each replicated dataset within each population for the CP *versus* the CT model analyses of ordinal data and for the Gaussian linear mixed model *versus* the

**Table II.** Posterior inference on sire variance ( $\sigma_s^2$ ) in a simulation study using the cumulative  $t$ -link model.

Population-dataset <sup>a</sup>	Liability data			Ordinal data		
	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>
I-1	0.13 $\pm$ 0.03	0.08–0.21	39 722	0.11 $\pm$ 0.03	0.06–0.19	2998
I-2	0.14 $\pm$ 0.04	0.08–0.22	41 817	0.13 $\pm$ 0.04	0.08–0.22	1775
I-3	0.10 $\pm$ 0.03	0.05–0.16	31 969	0.08 $\pm$ 0.03	0.04–0.14	966
II-1	0.10 $\pm$ 0.03	0.06–0.15	46 915	0.11 $\pm$ 0.03	0.06–0.19	185
II-2	0.12 $\pm$ 0.03	0.07–0.18	47 911	0.13 $\pm$ 0.04	0.08–0.22	130
II-3	0.08 $\pm$ 0.02	0.05–0.13	42 514	0.12 $\pm$ 0.03	0.07–0.19	187

<sup>a</sup> Population I specified with  $t$  residual distribution with 4 degrees of freedom; Population II specified with Gaussian residual distribution. Each population replicated for 3 datasets. <sup>b</sup> posterior mean  $\pm$  standard deviation. <sup>c</sup> 95% equal-tailed posterior probability interval based on 2.5th and 97.5th percentiles. <sup>d</sup> the effective number of independent samples using the initial positive sequence estimator of Geyer [16].

**Table III.** Deviance information criteria (DIC) and log marginal likelihood (LML) comparisons between models in a simulation study.

Population-dataset <sup>a</sup>	Analysis of liability data				Analysis of categorical data			
	Gaussian error model		Student $t_4$ -error model		Cumulative probit link model		Cumulative $t_4$ -link model	
	DIC	LML	DIC	LML	DIC	LML	DIC	LML
I-1	8782	−4392	8426	−4213	5596	−2798	5587	−2794
I-2	8711	−4356	8344	−4172	5734	−2867	5735	−2868
I-3	8816	−4409	8492	−4246	5727	−2864	5726	−2863
II-1	7131	−3566	7246	−3625	5322	−2661	5325	−2663
II-2	7180	−3590	7287	−3646	5172	−2586	5184	−2592
II-3	7237	−3618	7358	−3682	5443	−2722	5447	−2723

<sup>a</sup> Population I specified with  $t$  residual distribution with 4 degrees of freedom; Population II specified with Gaussian residual distribution. Each population replicated for 3 datasets.

$t$ -error linear mixed model analyses of liability data. Spiegelhalter *et al.* [40] suggested that a DIC difference exceeding 7 was a substantial indication of an important difference in model fit. Raftery [34] suggested Bayes factors exceeding 12 or, equivalently, LML differences exceeding 2.5, as being important.

Given that, the model choices based on DIC and PBF for the linear mixed model analyses of liability data were always resoundingly in favor of the correct model. However, in the comparison between CP and CT models for ordinal data analysis, the correct (CT) model was decisively chosen in only one of the replicates of Population I ( $\nu = 4$ ) and, similarly, the correct (CP) model was decisively chosen in only one of the three replicates of Population II ( $\nu = \infty$ ), with all other comparisons being indecisive (*i.e.* DIC differences  $< 7$  and LML differences  $< 2.5$ ). These results should not be too surprising since the much lower information content of ordinal data relative to underlying continuous liability data would make it that much more difficult to clearly resolve model choice. It is interesting to note, however, that LML and DIC measures led to identical conclusions with respect to decisive or indecisive model choice. In particular,  $-2\text{LML}$  and DIC values were virtually identical to each other.

## 5.2. Application to calving ease scores in Italian Piemontese cattle

### 5.2.1. Genetic parameter inference

Sire and maternal grandsire CP and CT models were used for the analysis of calving ease scores in Italian Piemontese cattle. Because of the MCMC mixing problems encountered in inferring upon  $\nu$ , this parameter was held constant to  $\nu = 4$ . Posterior inferences on key genetic parameters are summarized in Table IV and are based on the combined results from each of the three separate chains. The posterior mean, median and modal estimates (not shown) of the two heritabilities, and the genetic correlation using the MCMC algorithms were similar to each other across models, implying that the posterior densities were symmetric and unimodal. This property was further manifested by the fact that the 95% PPI are closely matched by the posterior mean  $\pm 2$  standard deviations. In this study, the total ESS for dispersion parameters across the three chains ranged from 1420 to 9305, indicating sufficient MCMC mixing under both models. Table IV shows that the ESS from the CT model were found to be almost double those of the CP model, attributable to the twice as large post-burn-in period for the CT model. Considering  $\nu$  as known improved mixing of, particularly, genetic parameters, in the CT model relative to joint inference with  $\nu$  (results not shown). Although the  $n \times 1$  auxiliary variable vector  $\lambda$  is additionally included in the CT model, this augmentation does not appear to adversely impact ESS and hence mixing of key genetic parameters relative to the CP model.

In this study, the CT model produced posterior means of genetic variance components that were nearly twice as large as those estimated using the CP model. Furthermore, the marginal residual variance is  $\frac{\nu}{\nu - 2}\sigma_e^2$  in the CT model such that seemingly twice as much residual variance is inferred under a CT model (with  $\nu = 4$ ) than under a CP model (with  $\nu = \infty$ ). Although these

**Table IV.** Posterior inference on genetic parameters of calving ease scores in Italian Piemontese cattle.

Parameter <sup>a</sup>	Cumulative probit link model			Cumulative <i>t</i> -link model		
	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>	PM $\pm$ SD <sup>b</sup>	95% PPI <sup>c</sup>	ESS <sup>d</sup>
$\sigma_s^2$	0.146 $\pm$ 0.026	[0.101; 0.200]	3281 [925–1261]	0.274 $\pm$ 0.047	[0.189; 0.373]	5517 [1745–1984]
$\sigma_m^2$	0.039 $\pm$ 0.011	[0.019; 0.063]	1346 [336–569]	0.060 $\pm$ 0.019	[0.027; 0.100]	2022 [573–845]
$\sigma_{sm}$	0.027 $\pm$ 0.013	[0.002; 0.053]	1791 [557–655]	0.054 $\pm$ 0.023	[0.010; 0.099]	2871 [897–1068]
$\sigma_h^2$	0.153 $\pm$ 0.019	[0.118; 0.192]	9305 [2899–3232]	0.308 $\pm$ 0.036	[0.241; 0.382]	12 264 [3666–4503]
$h_d^2$	0.42 $\pm$ 0.07	[0.30; 0.56]	2873 [833–1086]	0.40 $\pm$ 0.06	[0.28; 0.53]	4827 [1486–1786]
$h_m^2$	0.14 $\pm$ 0.04	[0.07; 0.23]	1461 [461–506]	0.11 $\pm$ 0.04	[0.05; 0.19]	2309 [748–801]
$r_{dm}$	–0.55 $\pm$ 0.14	[–0.78; –0.23]	1420 [456–498]	–0.58 $\pm$ 0.15	[–0.82; –0.24]	2135 [620–877]

<sup>a</sup> see text for description.<sup>b</sup> posterior mean  $\pm$  standard deviation.<sup>c</sup> 95% equal-tailed posterior probability interval based on 2.5th and 97.5th percentiles.<sup>d</sup> the total effective number of independent samples across the three MCMC chains using the initial positive sequence estimator of Geyer [16]; interval values [ ] refer to a range in ESS across the three chains.



**Table V.** MCMC chain-specific deviance information criterion (DIC) and log marginal likelihood (LML) values for the analysis of calving ease scores in Italian Piemontese cattle.

	$\bar{D}^a$	$D(\bar{\vartheta})^b$	$P_D^c$	DIC	LML
<u>Cumulative Probit</u>					
Chain 1	16 564	15 782	782	17 346	−8705
Chain 2	16 562	15 780	782	17 344	−8705
Chain 3	16 562	15 779	783	17 345	−8705
<u>Cumulative <i>t</i>-link</u>					
Chain 1	16 348	15 525	823	17 172	−8634
Chain 2	16 346	15 521	825	17 171	−8634
Chain 3	16 348	15 524	824	17 172	−8634

<sup>a</sup> the mean Bayesian deviance (see text). <sup>b</sup> bayesian deviance based on posterior mean vector (see text). <sup>c</sup> the effective number of parameters (see text).

results might at first sight imply that greater genetic and residual variation is directly captured by the CT model, it should be realized that the variance of underlying variables **L** is only defined proportionately to the marginal residual variance. This is further apparent in that the 95% PPI of heritabilities were only very slightly concentrated towards lower values in the CT model relative to the CP model such that the corresponding 95% PPI for both  $h_d^2$  and  $h_m^2$  overlapped substantially between the two models. Furthermore, the posterior density of  $r_{dm}$  was very similar between CT and CP, with most of the density being between −0.2 and −0.8.

In order to compare the CP and CT models for fit to the calving ease data, LML and DIC values, broken down into its components  $\bar{D}$  and  $P_D$ , are reported in Table V. Since it is difficult to quantify the degree of Monte Carlo error on DIC [50], we report DIC and LML values separately for each of the three chains under each model. It can be seen that there are relatively inconsequential differences in the measures from one chain to the next within each model relative to chains between models, thereby indicating minutely small Monte Carlo errors on the DIC difference between the two models. Both model choice criteria were overwhelmingly in favor of the CT model with  $\nu = 4$ . As anticipated, the model complexity, as measured by  $P_D$ , is higher for the CT model; however, the complexity penalty was strongly counteracted by a smaller mean deviance  $\bar{D}$ , thereby resulting in a smaller DIC favoring choice of the CT model.

**5.2.2. Inferences on sire effects**

Posterior means of elements of **s** were determined to be corresponding point estimates of progeny differences (EPD) under both CT and CP models.

The relationships between these estimates are shown to be strongly linear with no hint of substantial reranking as indicated by a Pearson correlation of 0.99. The CT model had greater spread in EPD's compared to the CP model, as further manifested by a least-squares estimated slope of 1.38 of CT on CP EPD's. This is not too surprising since a larger additive genetic (sire) variance was inferred in the CT model such that posterior means of elements of  $\mathbf{s}$  should be more dispersed in the CT model relative to the CP model. However, as discussed later, this is not practically important since the variance of  $\mathbf{L}$  is defined only proportionately to the marginal residual variability which is also larger in the CT model.

Posterior standard deviations of elements of  $\mathbf{s}$  are analogous to standard errors of prediction (SEP) in mixed effects model analysis and can be used to derive approximate reliabilities of EPD's [49]. That is, the SEP's were simply determined as the standard deviation of the MCMC samples of elements of  $\mathbf{s}$ . In this case, the correlation between the estimated standard errors was near unity. The estimated slope of CT on CP SEP's was nearly 1.41 indicating that the posterior standard errors were on average 41% larger under the CT model than under the CP model; nevertheless, for the purpose of reporting reliabilities, these standard errors need to be considered relative to  $\sigma_s^2$  which was concentrated on higher values under the CT model.

## 6. DISCUSSION

Given the recent momentum in using heavy-tailed residual specifications for the analysis of production data in animal breeding [36, 41, 42, 47] a hierarchical threshold (CT) mixed model based on a cumulative  $t$ -link specification was developed, validated by simulation and applied to a small calving ease dataset from Italian Piemontese cattle. The simulation study indicated that inference on  $\nu$  is possible in a CT model; however, it appears that either a more suitable MCMC strategy is needed or many more samples are required compared to that considered in our study to ensure a more reliable inference on  $\nu$ . Until this issue is satisfactorily resolved, we advocate fixing  $\nu$  at some arbitrarily low value in a CT model analysis. We chose  $\nu = 4$  since this value minimally assures defined first, second, and third moments while providing a liability variable distribution that is maximally heavy-tailed. One can then use model choice criteria such as DIC to assess whether or not the CT model is a better data fit than the CP model. We further note that for the case where  $\nu$  is fixed, that MCMC mixing was not negatively affected using the CT model, even though our data augmentation (of  $\lambda$ ) implementation might be of concern to those who might prefer Metropolis-Hastings sampling on all parameters [47] instead of introducing augmented variables. More recently, it has been demonstrated that data augmentation can be strategically used to enhance MCMC mixing; the strategies discussed by

van Dyk and Meng [45] may facilitate more favorable mixing on  $\nu$  and hence deserve further consideration in CT model applications to animal breeding. One reviewer pointed out that several Metropolis-Hastings updates on  $\nu$  per each MCMC cycle, as utilized by Bink *et al.* [6], and Uirmari *et al.* [44] may facilitate improvements in mixing. Of particular note (results not reported) was that due to the implementation of the algorithm of Cowles [11], MCMC mixing of  $\tau$  was not seen to have been the most limiting in terms of ESS in contrast to previous animal breeding implementations [39,46].

Our point estimates for heritabilities are substantially higher than corresponding threshold model estimates for calving ease reported by Manfredi *et al.* [27,28], McGuirk *et al.* [30,31], Varona *et al.* [46], Luo *et al.* [25], and Bennett and Gregory [3]. Nevertheless, our inference on a strongly negative direct-maternal genetic correlation is in agreement with previous work on calving ease using threshold models [3,25,46] and linear mixed models using data from the same source [7]. Hence, it appears from our results, in agreement with other studies, that selection of sires for calving ease of their progeny as calves should result in antagonistic effects in the ability of their daughters to calve easily as dams in successive generations. What was the most surprising is that the 95% PPI for  $h_d^2$  are greater and do not overlap with corresponding PPI in Kizilkaya *et al.* [23] who used a larger data set on first parity records from herds with greater than 50 rather than 100 records over the nine year period from the same data source. This result may be indicative of heterogeneity in genetic and residual variance due to the size of the herd or other confounding factors (*e.g.* region); this is an area for further research that our group has started with respect to residual variability, extending the work of San Cristobal-Gaudy *et al.* [37].

Two Bayesian model choice criteria, DIC and LML, were used to choose between the CP and CT models. In a simulation study, it was demonstrated that both DIC and LML were able to decisively choose the correct model in some cases whereas, in the remaining cases, these measures were too similar between the two models to allow a definitive choice. In the analysis of calving ease scores in Italian Piemontese cattle, the CT model was overwhelmingly chosen as the best fitting model by both model choice criteria. Nevertheless, in the examination of EPD's there were no real tangible differences between the two models in terms of sire genetic rankings.

Our study involved sire models, where calf records are connected directly to sires with genetic relationships identified only through known paternal ancestry. Sire models are different from animal models where each record is connected directly to a calf identification with all known genetic (maternal and paternal) relationships explicitly modeled. For animals (*e.g.* dams), with effectively less data information, we would anticipate greater differences in predicted genetic merit between CT and CP animal models than those given in our study.

However, a sire model has been seen to be more stable than an animal model in CP implementations [29], particularly when paternal half-sib relationships are dominant as in our data. Furthermore, the fact that a sire and maternal grandsire model only accounts directly for a portion of the additive genetic variance and of the maternal genetic variance implies that a *t*-error assumption is essentially placed on a composite source of error, *i.e.* the sum of the residual variance and remaining genetic variation, attributable to unknown dams and to Mendelian sampling. From the perspective of using heavy tailed densities to mute residual outliers, this is significant since deviant dam and/or Mendelian sampling effects may be muted as well in a CT sire model specification.

Presently, it does not appear to be feasible to apply MCMC methods to the very large datasets used for routine genetic evaluation of livestock by breed associations and national recording organizations. Kizilkaya *et al.* [23] recently demonstrated, however, very little differences in predicted genetic merit and standard error of prediction in a CP sire model between inference provided by MCMC and by approximate empirical Bayes procedures currently utilized by the industry (*e.g.* [4]). Empirical Bayes procedures are based on using the joint posterior modes of the sire effects as the EPD's, conditional on estimated variance components as if they were known with certainty. Based on results from Gianola and Foulley [17], the CT model can be readily implemented using empirical Bayes methods since the probability density function and the cumulative density function of a standard Student *t*-distribution could be substituted for the corresponding Gaussian functions needed to derive the necessary scoring equations used to compute the joint posterior modes. Furthermore, the degrees of freedom parameter,  $\nu$ , could be jointly estimated with the variance components using the Laplace method and tested for statistical significance using marginal likelihood ratio tests [43]. Empirical Bayes implementation of the CT sire model and the comparison of results with MCMC inference under the CT sire model deserve further consideration. Unfortunately, however, these comparisons may not necessarily apply to CP or CT animal models since joint modal estimates of EPD's in the CP animal model can be badly biased [29].

In this study, we have considered only two cumulative link models for the analysis of calving ease; conceptually, there are many others including those proposed by Chen and Dey [8]. In fact, Albert and Chib [2] demonstrate that the cumulative logistic link model is roughly equivalent to the CT model with  $\nu = 8$ . Other models can be contrived by considering alternative heavy-tailed distributions for the underlying variables, such as those considered by Rosa [36]. Some of the resulting models may be shown to have better fit to calving ease data than demonstrated with the CT model in our paper. Specifications based on skewness [47] may also have merit.

The substantial residual and genetic correlations between birth weight and calving difficulty imply that genetic evaluations of calving ease would

substantially benefit from a bivariate threshold/linear multiple trait analysis with birth weight [26,46]. Further work on jointly providing modeling flexibility with *t*-distributed specifications on both traits is needed given our results and those already presented by Strandén and Gianola [41,42] and Rosa [36].

## ACKNOWLEDGEMENTS

This research was supported by the Michigan Agricultural Experiment Station (Project MICL 01822). ANABORAPI (Associazione nazionale allevatori bovini di razza Piemontese, Strada Trinità 32a, 12061 Carrù, Italy) is gratefully acknowledged for providing the data for this study. Dr. Ignacy Misztal is profusely acknowledged for his provision of FSPAK90 Fortran subroutines used in this study. We also thank Dr. Guilherme Rosa for his comments on the manuscript.

## REFERENCES

- [1] Albert J.H., Chib S., Bayesian analysis of binary and polychotomous response data, *J. Am. Stat. Assoc.* 88 (1993) 669–679.
- [2] Albert J., Chib S., Bayesian residual analysis for binary response regression models, *Biometrika* 82 (1995) 747–759.
- [3] Bennett G.L., Gregory K.E., Genetic (co)variances for calving difficulty score in composite and parental populations of beef cattle. I. Calving difficulty score, birth weight, weaning weight, and postweaning gain, *J. Anim. Sci.* 79 (2001) 45–51.
- [4] Berger P.J., Genetic prediction for calving ease in the United States: Data, models, and use by the dairy industry, *J. Dairy Sci.* 77 (1994) 1146–1153.
- [5] Bertrand J.K., Wiggans G.R., Validation of data and review of results from genetic evaluation systems for US beef and dairy cattle, in: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, 11–16 January, University of New England, Armidale, NSW, Australia, 1998, Vol. 27, pp. 327–330.
- [6] Bink M.C.A.M., Quaas R.L., van Arendonk J.A.M., Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects, *Genet. Sel. Evol.* 30 (1998) 103–125.
- [7] Carnier P., Albera A., Dal Zotto R., Groen A.F., Bona M., Bittante G., Genetic parameters for direct and maternal calving performance over parities in Piemontese cattle, *J. Anim. Sci.* 78 (2000) 2532–2539.
- [8] Chen M-H., Dey D.K., Bayesian analysis for correlated ordinal data models, in: Dey D.K., Ghosh S.K., Mallick B.K. (Eds.), *Generalized linear models: A Bayesian perspective*, Marcel Dekker, New York, 2000, pp. 133–158.
- [9] Chib S., Marginal likelihood from the Gibbs output, *J. Am. Stat. Assoc.* 90 (1995) 773–795.

- [10] Chib S., Greenberg E., Understanding the Metropolis Hastings algorithm, *Am. Stat.* 49 (1995) 327–335.
- [11] Cowles M.K., Accelerating Monte Carlo Markov Chain convergence for cumulative link generalized linear models, *Stat. Comp.* 6 (1996) 101–111.
- [12] Dempster A.P., The direct use of likelihood for significance testing, *Stat. Comp.* 7 (1997) 247–252.
- [13] Emanuelson U., Fikse F., Banos G., Impact of national genetic evaluation models on international comparisons, in: *Computational Cattle Breeding '99*, 18–20 March 1999, Tuusula, Finland (<http://www.csc.fi/ttn/ccb99/>).
- [14] Gelfand A.E., Model determination using sampling-based methods, in: Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), *Markov Chain Monte Carlo in practice*, Chapman & Hall, New York, 1996, pp. 145–162.
- [15] Gelfand A.E., Ghosh S.K., Model choice: A minimum posterior predictive loss approach, *Biometrika* 85 (1998) 1–11.
- [16] Geyer C.J., Practical Markov chain Monte-Carlo (with discussion), *Stat. Sci.* 7 (1992) 467–511.
- [17] Gianola D., Foulley J.L., Sire evaluation for ordered categorical data with a threshold model, *Génét. Sélect. Évol.* 15 (1983) 201–224.
- [18] Gianola D., Sorensen D.A., A mixed effects threshold model with *t*-distributions. Book of Abstracts of the 47th Annual Meeting of the European Association for Animal Production, 1996, Wageningen Press, the Netherlands, p. 47.
- [19] Henderson C.R., Inverse of a matrix of relationships due to sires and maternal grandsires in an inbred population, *J. Dairy Sci.* 59 (1976) 1585–1588.
- [20] Heringstad B., Rekaya R., Gianola D., Klemetsdal G., Weigel K.A., Bayesian analysis of liability of clinical mastitis in Norwegian cattle with a threshold model: Effects of data sampling method and model specification, *J. Dairy Sci.* 84 (2001) 2337–2346.
- [21] Jensen J., Wang C.S., Sorensen D.A., Gianola D., Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler, *Acta Agric. Scand. A-An.* 44 (1994) 193–201.
- [22] Johnson V.E., Albert J.H., *Ordinal data modeling*, Springer, New York, 1999.
- [23] Kizilkaya K., Banks B.D., Carnier P., Albera A., Bittante G., Tempelman R.J., Bayesian inference strategies for the prediction of genetic merit using threshold models with an application to calving ease scores in Italian Piemontese cattle, *J. Anim. Breed. Genet.* 119 (2002) 209–220.
- [24] Lange K.L., Little R.J.A., Taylor J.M.G., Robust statistical modeling using the *t*-distribution, *J. Am. Stat. Assoc.* 84 (1989) 881–896.
- [25] Luo M.F., Boettcher P.J., Dekkers J.C.M., Schaeffer L.R., Bayesian analysis for estimation of genetic parameters of calving ease and stillbirth for Canadian Holsteins, *J. Dairy Sci.* 82 (1999) 1848.
- [26] Luo M.F., Boettcher P.J., Schaeffer L.R., Dekkers J.C.M., Bayesian inference for categorical traits with an application to variance components estimation, *J. Dairy Sci.* 84 (2001) 694–704.
- [27] Manfredi E.J., San Cristobal M., Foulley J.L., Some factor affecting the estimation of genetic parameters for cattle dystocia under a threshold model, *Anim. Prod.* 53 (1991) 151–156.

- [28] Manfredi E.J., Ducrocq V., Foulley J.L., Genetic analysis of dystocia in dairy cattle, *J. Dairy Sci.* 74 (1991) 1715–1723.
- [29] Mayer M., Inequality of maximum *a posteriori* estimators with equivalent sire and animal models for threshold traits, *Genet. Sel. Evol.* 27 (1995) 423–435.
- [30] McGuirk B.J., Going I., Gilmour A.R., The genetic evaluation of beef sires used for crossing with dairy cows in the UK. 2. Genetic parameters and sire merit predictions for calving survey traits, *Anim. Sci.* 66 (1998) 47–54.
- [31] McGuirk B.J., Going I., Gilmour A.R., The genetic evaluation of UK Holstein Friesian sires for calving ease and related traits, *Anim. Sci.* 68 (1999) 413–422.
- [32] Müller P., A generic approach to posterior integration and Gibbs sampling, Technical Report, Purdue University, 1993.
- [33] Newton M.A., Raftery A.E., Approximate Bayesian inference with the weighted likelihood bootstrap, *J. R. Stat. Soc. Ser. B* 56 (1994) 3–48.
- [34] Raftery A.E., Hypothesis testing and model selection, in: Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), *Markov Chain Monte Carlo in practice*, Chapman & Hall, New York, 1996, pp. 163–187.
- [35] Rekaya R., Weigel K.A., Gianola D., Application of a structural model for genetic covariances in international dairy sire evaluations, *J. Dairy Sci.* 84 (2001) 1525–1530.
- [36] Rosa G.J.M., Robust mixed linear models in quantitative genetics: Bayesian analysis *via* Gibbs sampling, in: *Proceedings of International Symposium on Animal Breeding and Genetics*, 21–24 September 1999, Brazil, pp. 133–159.
- [37] San Cristobal-Gaudy M., Bodin L., Elsen J.-M., Chevalet C., Genetic components of litter size variability in sheep, *Genet. Sel. Evol.* 33 (2001) 249–271.
- [38] Satagopan J.M., Newton M.A., Raftery A.E., Easy estimation of normalizing constants and Bayes factors from posterior simulation: stabilizing the harmonic mean estimator, Technical Report 1028, Department of Statistics, 2001. <http://www.stat.wisc.edu/newton/papers/abstracts/tr1028a.html>.
- [39] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, *Genet. Sel. Evol.* 27 (1995) 229–249.
- [40] Spiegelhalter D.J., Best N.G., Carlin B.P., van der Linde A., Bayesian measures of model complexity and fit (with discussion), *J. R. Statist. Soc. B* 64 (2002) 583–639.
- [41] Stranden I., Gianola D., Attenuating effects of preferential treatment with Student-*t* mixed linear models: a simulation study, *Genet. Sel. Evol.* 30 (1998) 565–583.
- [42] Stranden I., Gianola D., Mixed effects linear models with *t*-distributions for quantitative genetic analysis: a Bayesian approach, *Genet. Sel. Evol.* 31 (1999) 25–42.
- [43] Tempelman R.J., Gianola D., A mixed effects model for overdispersed count data in animal breeding, *Biometrics* 52 (1996) 265–279.
- [44] Uimari P., Thaller G., Hoeschele I., The use of multiple markers in a Bayesian method for mapping quantitative trait loci, *Genetics* 143 (1996) 1831–1842.
- [45] van Dyk D.A., Meng X.-L., The art of data augmentation (with discussion), *J. Comp. Graph. Stat.* 10 (2001) 1–50.
- [46] Varona L., Misztal I., Bertrand J.K., Threshold-linear *versus* linear-linear analysis of birth weight and calving ease using an animal model. I. Variance component estimation, *J. Anim. Sci.* 77 (1999) 1994–2002.

- [47] von Rohr P., Hoeschele I., Bayesian QTL mapping using skewed Student-*t*-distributions, *Genet. Sel. Evol.* 34 (2002) 1–21.
- [48] Wang C.S., Rutledge J.J., Gianola D., Bayesian analysis of mixed linear models *via* Gibbs sampling with an application to litter size in Iberian pigs, *Genet. Sel. Evol.* 26 (1994) 91–115
- [49] Wang C.S., Quaas R.L., Pollak E.J., Bayesian analysis of calving ease scores and birth weights, *Genet. Sel. Evol.* 29 (1997) 117–143.
- [50] Zhu L., Carlin B.P., Comparing hierarchical models for spatio-temporally mis-aligned data using the deviance information criterion, *Stat. Med.* 19 (2000) 2265–2278.