Genet. Sel. Evol. 38 (2006) 565–581

DOI: 10.1051/gse:2006022

Original article

A bivariate quantitative genetic model for a threshold trait and a survival trait

Lars Holm DAMGAARD[a,b*], Inge Riis KORSGAARD[b]

[a] Department of Animal Science and Animal Health, Royal Veterinary and Agricultural University, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark
[b] Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, P.O. Box 50, 8830 Tjele, Denmark

(Received 16 November 2005; accepted 2 June 2006)

Abstract – Many of the functional traits considered in animal breeding can be analyzed as threshold traits or survival traits with examples including disease traits, conformation scores, calving difficulty and longevity. In this paper we derive and implement a bivariate quantitative genetic model for a threshold character and a survival trait that are genetically and environmentally correlated. For the survival trait, we considered the Weibull log-normal animal frailty model. A Bayesian approach using Gibbs sampling was adopted in which model parameters were augmented with unobserved liabilities associated with the threshold trait. The fully conditional posterior distributions associated with parameters of the threshold trait reduced to well known distributions. For the survival trait the two baseline Weibull parameters were updated jointly by a Metropolis-Hastings step. The remaining model parameters with non-normalized fully conditional distributions were updated univariately using adaptive rejection sampling. The Gibbs sampler was tested in a simulation study and illustrated in a joint analysis of calving difficulty and longevity of dairy cattle. The simulation study showed that the estimated marginal posterior distributions covered well and placed high density to the true values used in the simulation of data. The data analysis of calving difficulty and longevity showed that genetic variation exists for both traits. The additive genetic correlation was moderately favorable with marginal posterior mean equal to 0.37 and 95% central posterior credibility interval ranging between 0.11 and 0.61. Therefore, this study suggests that selection for improving one of the two traits will be beneficial for the other trait as well.

bivariate genetic model / survival trait / ordered categorical trait / Bayesian analysis

1. INTRODUCTION

Because of their economic and ethical importance, functional traits have been given increasing priority in breeding programs for livestock during the last decade. Functional traits can generally be regarded as those traits, which

* Corresponding author: lars.damgaard@agrsci.dk

increase net income by reducing the cost of input rather than increasing the output of saleable products. Numerous functional traits are considered in dairy cattle breeding including longevity, conformation scores, calving difficulty, and resistance to diseases (*e.g.* [10]). In pig breeding focus has mainly been on leg characteristics [22, 30] and resistance to diseases [19].

As with many other traits, it is assumed that the genotypic value affecting a functional trait results from the sum of a very large number of independent contributions from independently segregating loci, each with a small effect. The Central Limit Theorem leads to the result that the additive genetic value is approximately normally distributed [4, 14]. However, phenotypically these traits often have non-normal distributions, and many of the functional traits considered in animal breeding can be analyzed as threshold traits or survival traits with examples including disease resistance, conformation scores, calving difficulty and longevity.

Analysis of threshold characters often relies on the threshold liability concept first proposed by Wright [41]. Application of this model in animal breeding dates back to Robertson and Lerner [34]. During the last decade, survival analysis based on the proportional hazards model has become the method of choice for inferring longevity [11]. Survival analysis was first proposed in animal breeding by Smith and Quaas [35] for studying longevity of dairy cows. Since then survival models have also been used to infer environmental and genetic aspects of resistance to diseases in beef bulls [23], in fish [20] and in pigs [19].

Knowledge of genetic parameters such as heritabilities and genetic correlations are required to predict response to selection, to select among various breeding programs based on *e.g.* their economic revenue, and to estimate breeding values of selection candidates.

Multivariate quantitative genetic models for inferring an arbitrary number of threshold characters, survival traits and linear Gaussian traits only exist for censored linear Gaussian survival traits [25]. A recent methodology contribution includes a bivariate quantitative genetic model for a linear Gaussian trait and a Weibull survival trait [8].

The objective of this study was to extend the methodology of Damgaard and Korsgaard [8] to a bivariate quantitative genetic model of a threshold character and a survival trait that are genetically and environmentally correlated. Firstly, the Bayesian model is presented and the fully conditional distributions needed for implementing the Gibbs sampler are described. Secondly, the Gibbs sampler is tested by simulation and a joint analysis of longevity and calving difficulty of dairy cattle is presented for illustration of the model.

2. MATERIALS AND METHODS

Let Y_{1i} be a random variable of the ordered categorical trait of animal i for $i = 1, \ldots, n$, where n is the total number of animals with records. Y_{1i} can take values in one out of K for $K \geq 2$ mutually exclusive ordered categories. The outcome of Y_{1i} is equal to k if $\tau_{k-1} < L_{1i} \leq \tau_k$ for $k = 1, \ldots, K$, where L_{1i} is a continuous unobserved random variable often denoted the liability and $\tau = (\tau_0, \tau_1, \ldots, \tau_K)$ is a vector with $K + 1$ thresholds defined on the liability scale with $\tau_0 = -\infty$ and $\tau_K = \infty$. For the survival trait let T_i and C_i be random variables representing a survival time and a censoring time. In what follows, we assume that all animals have records of both traits such that data on animal i is given by $(y_{1i}, y_{2i}, \delta_{2i})$, where y_{1i} is an observed value of Y_{1i}, y_{2i} is an observed value of $Y_{2i} = \min(T_i, C_i)$, and δ_{2i} is the outcome of a censoring indicator variable equal to 1 if $T_i \leq C_i$ and 0 otherwise. Later we consider the case where data on one of the two traits are missing at random.

In this paper we augment the joint posterior distribution with the vector of unobserved liabilities. By doing so the model specification is very similar to the one already given for a bivariate model of a survival trait and a Gaussian trait [8]. In this paper we define parameters and give the prior distribution and the augmented posterior distribution. Regarding the fully conditional posterior distributions we will only explicitly give them for liabilities and thresholds. For the remaining parameters they are identical to the ones already given for a bivariate model of a Gaussian trait and survival trait [8] if the sampled liabilities are considered as data from a Gaussian process.

The sampling distribution for the bivariate model will be represented by the conditional hazard function of T_i and the joint distribution of $\mathbf{L}_1$ and $\mathbf{e}_2$

$$\begin{aligned} \lambda_i(t|\boldsymbol{\theta}, \mathbf{e}_2) &= \rho t^{(\rho-1)} \exp\left\{\mathbf{x}'_{2i}(t)\boldsymbol{\beta}_2 + \mathbf{z}'_{2i}\mathbf{a}_2 + e_{2i}\right\} \\ \begin{matrix}\mathbf{L}_1 \\ \mathbf{e}_2\end{matrix} \Bigg| \boldsymbol{\theta} &\sim N\left(\begin{pmatrix}\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{a}_1 \\ \mathbf{0}\end{pmatrix}, \mathbf{R}_e \otimes \mathbf{I}_n\right) \end{aligned} \tag{1}$$

where $\mathbf{e}_2$ with elements $(e_{2i})_{i=1,\ldots,n}$ is a vector of residual effects of the survival traits on the log-frailty scale, which accounts for variation in log-frailty not otherwise accounted for by the specification of the model with covariates and random effects. Here $\lambda_i(t|\boldsymbol{\theta}, \mathbf{e}_2)$ is the hazard function of T_i conditional on model parameters $(\boldsymbol{\theta}, \mathbf{e}_2)$, where $\theta = (\rho, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{G}, \mathbf{R}_e)$. The Weibull baseline hazard function is generally given as $\lambda^\rho \rho t^{(\rho-1)}$ with parameters ρ and λ. Here the term λ^ρ is included on the log-frailty scale as $\rho \log(\lambda)$ and is the first element of the vector $\boldsymbol{\beta}_2$. The p_1 dimensional vector $\boldsymbol{\beta}_1$ and the p_2 dimensional vector $\boldsymbol{\beta}_2$ represent systematic effects of the threshold trait and

the survival trait. $\mathbf{a}_1$ and $\mathbf{a}_2$ of dimension q are the vectors of additive genetic effects, where q is the total number of animals in the pedigree, and the vectors $\mathbf{x}'_{1i}$, $\mathbf{x}'_{2i}$, $\mathbf{z}'_{1i}$, $\mathbf{z}'_{2i}$ are incidence arrays relating parameter effects to observations. Finally $\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ is the genetic covariance matrix, and $\mathbf{R}_e = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$ is the residual covariance matrix.

The time-dependent covariates of animal i are assumed to be left-continuous and piecewise constant on the intervals $h_{i(m-1)} < t \leq h_{i(m)}$ for $m = 1, \ldots, M_i$, where $h_{i(0)} = 0$ and $h_{i(M_i)} = y_{2i}$, and $h_{i(m)}$ for $m = 1, \ldots, (M_i - 1)$ are the ordered time points at which one or more of the time-dependent covariates of animal i changes. Finally $M_i - 1$ is the number of different time points with changes in one or more of the time-dependent covariates associated with animal i.

Prior specification

A priori model parameters $(\beta_{1b})_{b=1,\ldots,p_1}$, $(\beta_{2b})_{b=1,\ldots,p_2}$, $\boldsymbol{\tau}$, ρ, $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{G})$ and $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{R}_e)$ are assumed to be mutually independent. For $K \geq 3$ we assume that all elements of the residual covariance matrix $\mathbf{R}_e$ are stochastic, implying that only $K - 3$ thresholds can be identified [36]. We set $\tau_1 = 0$ and $\tau_{K-1} = 1$, and assume *a priori* that the remaining unknown thresholds are distributed as order statistics from a *uniform* (0, 1) distribution according to $p(\tau_2, \ldots, \tau_{K-2}) = (K - 3)! I(\tau \in \Upsilon)$, where $\Upsilon = \{(\tau_2, \ldots, \tau_{K-2}) | 0 \leq \tau_2 \leq \ldots \leq \tau_{K-2} \leq 1\}$. For the binary response ($K = 2$) it is necessary for reasons of identifiability to constrain both the threshold and the residual variance component: we set $\tau_1 = 0$ and $R_{e22} = 1$. Improper uniform priors are assigned to $(\boldsymbol{\beta}_{1b})_{b=1,\ldots,p_1}$, $(\boldsymbol{\beta}_{2b})_{b=1,\ldots,p_2}$ and ρ over their range of positive support. The prior distribution of additive genetic effects is by assumption of the additive genetic infinitesimal model [4] assumed to be multivariate normally distributed $\left(\mathbf{a}'_1, \mathbf{a}'_2\right)' | \mathbf{G} \sim N\left(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}_q\right)$, where $\mathbf{A}_q$ is the additive genetic relationship matrix. Finally the covariance matrices $\mathbf{G}$ and $\mathbf{R}_e$ are *a priori* assumed to be inverse Wishart distributed according to $\mathbf{G} \sim IW(\mathbf{F}_g, f_g)$ and $\mathbf{R}_e \sim IW(\mathbf{F}_e, f_e)$.

2.1. Augmented posterior distribution

The augmented posterior distribution of $(\boldsymbol{\theta}, \mathbf{e}_2, \mathbf{L}_1)$ is obtained using Bayes' theorem. Here we augment the parameter vector with the vector of unobserved

liabilities $\mathbf{L}_1 = (L_{11}, \ldots, L_{1n})'$ and posterior distribution is given by

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{e}_2, \mathbf{l}_1 | \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_2) &\propto p(\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_2 | \mathbf{l}_1, \boldsymbol{\theta}, \mathbf{e}_2) p(\boldsymbol{\theta}, \mathbf{e}_2, \mathbf{l}_1) \\ &= p(\mathbf{y}_2, \boldsymbol{\delta}_2 | \mathbf{e}_2, \boldsymbol{\theta}) p(\mathbf{y}_1, \mathbf{l}_1, \mathbf{e}_2 | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &= p(\mathbf{y}_2, \boldsymbol{\delta}_2 | \mathbf{e}_2, \boldsymbol{\theta}) p(\mathbf{y}_1 | \mathbf{l}_1, \mathbf{e}_2, \boldsymbol{\theta}) p(\mathbf{l}_1, \mathbf{e}_2 | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \end{aligned} \tag{2}$$

where in the second step we used that $(\mathbf{Y}_2, \boldsymbol{\delta}_2)$ and $(\mathbf{Y}_1, \mathbf{L}_1)$ are assumed to be conditional independent given $\boldsymbol{\theta}$ and $\mathbf{e}_2$. Further, conditional on $(\boldsymbol{\theta}, \mathbf{e}_2)$, censoring is assumed to be independent and non-informative [3] implying that $p(\mathbf{y}_2, \boldsymbol{\delta}_2 | \boldsymbol{\theta}, \mathbf{e}_2) \propto \prod_i S_i(y_{2i} | \boldsymbol{\theta}, \mathbf{e}_2) \left[\lambda_i(y_{2i} | \boldsymbol{\theta}, \mathbf{e}_2)\right]^{\delta_{2i}}$, where $S_i(y_{2i} | \boldsymbol{\theta}, \mathbf{e}_2) = \exp\left\{ -\int_0^{y_{2i}} \lambda_i(s | \boldsymbol{\theta}, \mathbf{e}_2) \mathrm{d}s \right\}$ is the conditional survival function. Under the assumption of a proportional Weibull log-normal animal frailty model, $p(\mathbf{y}_2, \boldsymbol{\delta}_2 | \mathbf{e}_2, \boldsymbol{\theta})$ is up to proportionality given by

$$\begin{aligned} &\rho^{\sum_{i=1}^{n} \delta_{2i}} \left(\prod_{i=1}^{n} y_{2i}^{\delta_{2i}} \right)^{(\rho-1)} \exp\left\{ \sum_{i=1}^{n} \delta_{2i} (\mathbf{x}'_{2i}(y_{2i}) \boldsymbol{\beta}_2 + \mathbf{z}'_{2i} \mathbf{a}_2 + e_{2i}) \right\} \\ &\times \exp\left\{ -\sum_{i=1}^{n} \sum_{m=1}^{M_i} \exp(\mathbf{x}'_{2i}(h_{i(m)}) \boldsymbol{\beta}_2 + \mathbf{z}'_{2i} \mathbf{a}_2 + e_{2i}) (h^{\rho}_{i(m)} - h^{\rho}_{i(m-1)}) \right\}. \end{aligned} \tag{3}$$

It follows that the distribution $p(\mathbf{y}_1 | \mathbf{l}_1, \mathbf{e}_2, \boldsymbol{\theta})$ is degenerate because the probability that a categorical record falls in a given category conditional on liabilities and thresholds is completely specified. $p(\mathbf{y}_1 | \mathbf{l}_1, \mathbf{e}_2, \boldsymbol{\theta})$ is written as

$$\prod_{i=1}^{n} \left[\sum_{k=1}^{K} I\left[\tau_{k-1} \leq l_{1i} < \tau_k\right] I\left[y_{1i} = k\right] \right].$$

Finally, the conditional distribution of $\mathbf{L}_1, \mathbf{e}_2 | \boldsymbol{\theta}$ is multivariate normally distributed according to (1).

2.2. Fully conditional distributions

The fully conditional distributions of each or groups of model parameters were obtained up to proportionality by retaining from the posterior distribution (2) the terms depending on the parameter of interest. By regarding the sampled values of the liabilities as observations from a linear Gaussian trait, the model parameters, with exception of thresholds, have fully conditional posterior distribution which are identical to those given for a bivariate model of a linear Gaussian trait and a survival trait [8]. For the liabilities assume that

$Y_{1i} = k$ for $k \in \{1, \ldots, K\}$, then the fully conditional distribution of a liability L_{1i} for $i = 1, \ldots, n$ follows a truncated normal distribution on the interval $[\tau_{k-1} < l_{1i} \leq \tau_k)$ with mean $\mu_{l_{1i}}$ and variance $V_{l_{1i}}$ before truncation given by

$$\mu_{l_{1i}} = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{z}'_{1i}\mathbf{a}_1 + R_{e_{12}}R^{-1}_{e_{22}}e_{2i}$$
$$V_{l_{1i}} = R_{e_{11}} - (R_{e_{12}})^2 R^{-1}_{e_{22}}.$$

The fully conditional distribution of a threshold τ_k for $k = 2, \ldots, K-2$ is uniformly distributed on the interval

$$\left[\max\left\{\max\left\{l_{1i}|y_{1i} = k\right\}, \tau_{k-1}\right\}, \min\left\{\min\left\{l_{1i}|y_{1i} = k+1\right\}, \tau_{k+1}\right\}\right].$$

Note that this notation accommodates for the possibility of missing observations in one or more categories [1].

2.3. Model with missing observations

Missing observations for one of the two traits are often the case in field data. If the observations are missing at random [27], then a common approach in Bayesian analysis is to augment the joint posterior distribution with the residual effects associated with missing records (*e.g.* [36]). The augmented residuals are treated as unknown parameters, so at each iteration of the Gibbs sampler the augmented residual effects are sampled from their fully conditional posterior distributions. This approach was described in details for the bivariate model of a linear Gaussian trait and a survival trait [8]. Again if we regard the sampled values of liabilities as observations from a linear Gaussian trait the fully conditional distributions of augmented residuals are normally distributed with the same mean and variance as those given for the bivariate model of a linear Gaussian trait and a survival trait [8].

2.4. Implementation

A Gibbs sampler for the bivariate animal model (1) with no random environmental effects was implemented in Fortran 90 for data without missing observations of the two traits. The implementation is an extension of the one described already for a bivariate model of a linear Gaussian trait and a survival trait [8] and will therefore only be described briefly in this study. Inferences of model parameters were based on a single Gibbs chain. Updating of the model parameters $\left(\boldsymbol{\beta}_1, (\tau)_{i=2,\ldots,K-2}, (a_{1i})_{i=1,\ldots,n}, \mathbf{G}, \mathbf{R}_e\right)$ and of the liabilities $(L_{1i})_{i=1,\ldots,n}$

for which the fully conditional distributions can be recognized in closed form, was performed using standard methods. Note that the elements of the vector $\boldsymbol{\beta}_1$ and the elements of the matrices $\mathbf{G}$ and $\mathbf{R}_e$ were jointly updated. The parameters $((\beta_{2i})_{i=2,\ldots,p_2}, (a_{2i})_{i=1,\ldots,n}, (e_{2i})_{i=1,\ldots,n})$ for which the fully conditional distributions could not be recognized in closed form, were updated univariately using adaptive rejection sampling (ARS) [16]. Finally, the two Weibull baseline parameters (ρ, β_{21}) were updated jointly by a Metropolis-Hastings step [18,32] using a large sample bivariate normal distribution as proposal distribution [8].

2.5. Simulation study

The proposed bivariate model was illustrated in a simulation study in which the same model was used to generate and analyze data. Thus, focus was on the estimation of parameters under conditions where all model assumptions were satisfied.

Records of both traits were generated for 6000 animals after 100 unrelated sires each having 60 offspring (balanced half-sib design) using the bivariate model (1). The number of categories for the threshold character was three with observed frequencies; 1: (32%), 2: (45%) and 3: (23%). Lifetimes higher than 1500 were right censored resulting in a data set with approximately 15% censored records. The model of the threshold character included a mean effect, a sire effect and a residual effect. The model for the survival trait included the two baseline Weibull parameters, a systematic effect with two levels, a sire effect and a residual effect [8]. We adopted improper uniform priors for the genetic sire covariance matrix ($\mathbf{G}_s$) and the residual covariance matrix ($\mathbf{R}_{\tilde{e}}$).

The model parameters used to simulate data and the results from the Bayesian analysis are given in Table I. Starting values of the Gibbs sampler of the sire and residual variances and of the two Weibull parameters were set equal to the true values used in the simulation of data. All of the remaining model parameters were initially set to zero. A Gibbs chain of length, 600 000 iterations, was run. The first 10 000 iterations were considered as burnin and therefore discarded from the post Gibbs analysis. The interval between saved sampled values was 100, so that the total number of iterations kept was 4900. Effective number of samples (N_e) of each parameter was calculated by the method of batching based on 30 batches (*e.g.* [36]).

Table I. Posterior summary statistics for the simulation study: marginal posterior mode, mean, 2.5% and 97.5% percentiles, and effective sample size (N_e) of Weibull parameters (ρ, β_{21}), of time-independent systematic effects (β_{22}, $\beta_{23} = 0$), of sire variance components (G_{s11}, G_{s22}) and sire correlation (ρ_{G_s}), of residual variance components ($R_{\bar{e}11}$, $R_{\bar{e}22}$) and residual correlation ($\rho_{R_{\bar{e}}}$).

Parameter	True	Mode	Mean	2.5%	97.5%	N_e
ρ	2.2	2.16	2.14	2.02	2.28	557
β_{21}	−14.8	−14.20	−14.44	−15.33	−13.64	568
β_{22}	−0.4	−0.35	−0.36	−0.45	−0.29	1945
G_{s11}	0.05	0.053	0.055	0.038	0.077	2538
G_{s22}	0.1	0.096	0.10	0.065	0.15	1798
ρ_{G_s}	0.5	0.56	0.50	0.25	0.70	4293
$R_{\bar{e}11}$	0.65	0.66	0.66	0.62	0.70	3051
$R_{\bar{e}22}$	1.0	0.91	0.90	0.64	1.21	606
$\rho_{R_{\bar{e}}}$	−0.20	−0.21	−0.20	−0.25	−0.16	3967

2.6. Example, calving difficulty and longevity

2.6.1. Data

Since 1985, Danish dairy farmers have recorded calving difficulty in one out of five ordered categories; 1: easy without assistance, 2: easy with assistance, 3: difficult without veterinary assistance, 4: difficult with veterinary assistance, and 5: caesarian delivery. Because of few observations, categories 3, 4, 5 were grouped together so that calving difficulty in first lactation was defined by three categories. Note that here we only analyzed calving difficulty in first lactation with observed frequencies; 1 : 53%, 2: 38% and (3, 4, 5): 9%. Longevity was defined as time from first calving until culling.

Data was extracted from the Danish national database for cattle [5] and consisted of records of both traits from 16 345 Danish Holstein cows originating from 33 herds. Only cows having their first calving in the period from January 1990 to June 2002 were included in the analysis. The herds were selected so that the number of first calvings in 1990 was higher than 100 and in the following years was within plus and minus 15% of the level in 1990. This strategy was chosen in order to avoid herds with substantial changes in herd size during the study period. Lifetimes of cows sold or still alive at the time of last registered milk recording date in the extracted data were right censored, corresponding to 26% censored records. Pedigree of the cows with records was traced back only to their sires, implying that sires were assumed unrelated (half-sib design). The cows were daughters after 590 sires, and the daughter-group size ranged between 5 and 939 with an average of 28.

2.6.2. Model and data analysis

The data was analyzed using a sire model equivalent to the animal model (1). The threshold character was modelled with an effect of herd and an effect of year at first calving with 12 levels, which were defined by the first of January every year for 1991 to 2001. The survival trait was modelled with a time-independent effect of herd and a time-dependent effect of stage of lactation with four levels ($\beta_{SL1}, \ldots, \beta_{SL4}$). The stage of lactation effect changed at each calving and at 60, 180 and 305 days after calving in all lactations. A sire effect and a residual effect were included for both traits allowing for additive genetic and environmental correlation between the two traits. We adopted improper uniform priors for the genetic sire and residual covariance matrix.

Two independent Gibbs chains of length, 600 000 iteration each, were run. Based on visual inspection of all trace plots the first 10 000 iterations were considered as burnin and therefore discarded from the post Gibbs analysis. The interval between saved sampled values was 100, so that the total number of iterations kept was 5900 for each chain. The starting values of the parameters of the first chain were $\rho = 1.7$, $\beta_{21} = -12.0$, $G_{s11} = 0.02$, $G_{s12} = 0.006$, $G_{s22} = 0.05$ and $R_{\tilde{e}11} = 0.5$, $R_{\tilde{e}12} = 0.02$ and $R_{\tilde{e}22} = 0.2$. All of the remaining model parameters were started at zero. In the second chain, the starting values of the two Weibull parameters were changed according to $\rho = 1.2, \beta_{21} = -8.5$, whereas the remaining parameters were started as for the first chain. These starting values are to a higher degree similar to the ones used in the Danish routine evaluation of dairy cows ($\rho = 1.07$ and $\beta_{21} = -8.1$) [9].

The marginal posterior summary statistics of the first chain were very similar to those of the second chain and in the following we will therefore only give results from the second chain (Tab. II). The agreement between the two chains provides evidence that the Gibbs sampler converged and that samples can be regarded as generated from the posterior distribution of interest.

Because of large computation time in the joint mixed model (1) the convergence of the Gibbs sampler for survival parameters were prior to the joint analysis also assessed in univariate systematic analyses of longevity. This was done by varying the starting values for β_{21} (-6 to -15) and for $R_{\tilde{e}22}$ (0.02 to 0.8). In all combinations tested, the marginal posterior summary statistics were basically the same and similar to the corresponding ones obtained in the joint analysis. This further suggests satisfactory convergence of the Gibbs sampler in the joint analysis.

Table II. Posterior summary statistics for the analysis of calving difficulty and longevity: marginal posterior mode, mean, 2.5% and 97.5% percentiles, and effective sample size (N_e) of Weibull parameters (ρ, β_{21}), of time-dependent stage of lactation effects (β_{SL1}, β_{SL2}, $\beta_{SL3} = 0$, β_{SL4}), of sire variance components (G_{s11}, G_{s22}) and sire correlation (ρ_{G_s}), of residual variance components ($R_{\tilde{e}11}$, $R_{\tilde{e}22}$) and residual correlation ($\rho_{R_{\tilde{e}}}$), of heritability of calving difficulty on the liability scale ($h_1^2 = 4G_{s11}/(G_{s11} + R_{\tilde{e}11})$), and of heritability of longevity on the log-frailty scale ($h_2^2 = 4G_{s22}/(G_{s22} + R_{\tilde{e}22})$).

Parameter	Mode	Mean	2.5%	97.5%	N_e
ρ	1.69	1.70	1.65	1.75	900
β_{21}	−11.68	−11.82	−12.14	−11.51	757
β_{SL1}	−0.33	−0.4	−0.9	−0.28	3603
β_{SL2}	−0.38	−0.38	−0.43	−0.33	3319
β_{SL4}	0.24	0.24	0.19	0.28	5201
G_{s11}	0.025	0.021	0.014	0.03	8084
G_{s22}	0.038	0.037	0.025	0.051	2271
ρ_{G_s}	0.41	0.37	0.089	0.61	7177
$R_{\tilde{e}11}$	0.54	0.54	0.52	0.56	6831
$R_{\tilde{e}22}$	0.14	0.14	0.096	0.20	955
$\rho_{R_{\tilde{e}}}$	0.023	0.023	−0.040	0.086	6536
h_1^2	0.16	0.15	0.10	0.21	9080
h_2^2	0.85	0.82	0.58	0.99	2076

3. RESULTS

3.1. Simulation study

The results from the simulation study showed that parameters of the bivariate model (1) can be correctly inferred using the proposed methodology. The central posterior density (CPD) regions [6] defined by 2.5% and 97.5% percentiles covered well the parameter values used in the simulation of data (Tab. I). For example, there is 95% marginal posterior probability that the additive genetic correlation lies between 0.25 and 0.70, which covers the true value of 0.5 used in the simulation of data.

3.2. Example, calving difficulty and longevity

For calving difficulty in first lactation, the marginal posterior mean of heritability on the liability scale of 0.15 is slightly larger than previously reported heritabilities obtained from univariate threshold models (0.07 to 0.12) [17, 31, 38]. For longevity the marginal posterior mean of heritability

on the log-frailty scale ($h^2_{nor} = 4G_{s21}/(G_{s22} + R_{e22})$) is 0.82 [24]. This shows that the residual effect mainly describes additive genetic variation (*i.e.* genetic differences between dams and Mendelian segregation). Note that the heritability on the log-frailty scale for the survival trait ignores the underlying extreme value variation ($\pi^2/6$) and therefore is higher than the heritabilities most commonly reported in survival studies [11]. The reason why we prefer to define heritability on the additive linear scale is that it is the most general definition of heritability that can be used both in semi-parametric and parametric proportional hazards models, also when extended to time-dependent genetic effects [7]. Further discussion on heritabilities and their interpretation can be found in [24, 42].

The moderate positive marginal posterior mean of the additive genetic correlation (0.37) suggests that selection for improving one of the two traits (*i.e.* easier calving or reduced risk of culling) will have a beneficial effect on the other trait as well. Note that a negative sire value for longevity corresponds to low risk of culling, and that a low value for calving difficulty means less problems at calving.

The low marginal posterior mean of the residual correlation (0.023) is a bit unexpected as a calving with severe difficulties is expected to increase the risk of culling immediately after calving. A possible explanation for the low residual correlation is that the estimate obtained here represents the average residual association between calving difficulty and risk of culling in course of an animals lifetime. Therefore a high momentary residual association between calving difficulty and risk of culling will not be exploited by the model applied. Clearly this points to the need for alternative survival models that allow for a time varying residual effect.

Figure 1 shows the hazard function for the first five lactations conditional on a zero value for the systematic time-independent effects, the sire effect and the residual effect. The average lifetime at second, third, fourth and fifth calving defined the calving times. The hazard function shows that at the end of each lactation the risk of culling is elevated. This result agrees with the fact that voluntary culling, which is assumed to be the major reason for culling, mainly takes place late in lactations [13].

Figure 2 shows the three conditional survival functions corresponding to sire effects equal to zero, minus and plus two standard deviations of the estimated sire variance (−0.38, 0.38), a zero value for the time-independent systematic effect and the residual effect. These survival curves clearly illustrate that daughters from the best sire, say $s_{best} = -0.38$, have a substantial better longevity than daughters of the worst sire, say $s_{worst} = 0.38$. For example at

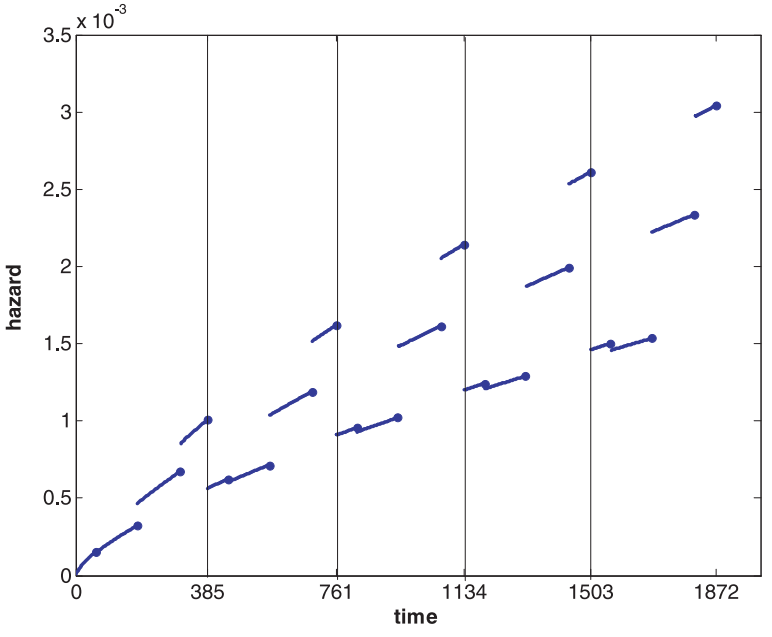

Figure 1. Hazard function based on marginal posterior mean value of ρ, β_{21}, β_{SL1}, β_{SL2}, β_{SL3}, β_{SL4} and conditional on a zero value for the systematic time-independent effects, the sire effect and the residual effect.

the time of fourth calving (day 1134) 49% of the daughters of the best sire is still alive compared to only 22% of the daughters of the worst sire.

4. DISCUSSION

We have presented a Gibbs sampler for joint Bayesian analysis of a threshold character and a survival trait. Simulation results established that the estimated marginal posterior distributions covered well the true values used in the simulation of data. In conclusion the model parameters and functions thereof including the additive genetic and environment correlations can be correctly inferred applying the method proposed.

The proposed method allows inferring the additive genetic correlation between a threshold character and a survival trait under the assumption of the additive genetic infinitesimal model [4, 14] without having to rely on approximations. The additive genetic correlation provides information about how selection for an ordered categorical trait simultaneously may affect a survival trait and vice versa. This information is essential for planning and administrating of genetic improvement programs. For example, in pig breeding the

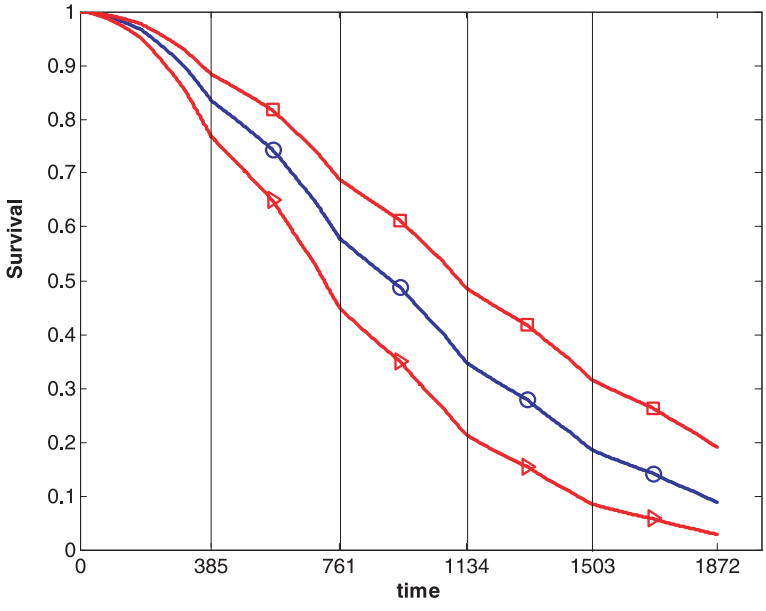

Figure 2. Survival functions based on marginal posterior mean value of ρ, β_{21}, β_{SL1}, β_{SL2}, β_{SL3}, β_{SL4} and conditional on sire effects equal to: −0.38 (□), 0.0 (O), 0.38 (△), a zero value for the time-independent fixed effect and the residual effect.

implemented breeding strategy implies that breeding sows are culled already after their first or second litter so that direct selection for improved longevity is hardly feasible. In that case the model can be used to identify and evaluate correlated threshold characters (*e.g.* locomotion traits), which can be used to select indirectly for improved longevity. The bivariate model has also been suggested as a way to improve the accuracy by which especially young sires of dairy cattle are evaluated for longevity [26, 39, 40]. The idea is to combine longevity records with, for instance, correlated ordered categorical type traits, which can be recorded early in life, and thereby increase the information on longevity.

The joint analysis of calving difficulty and longevity provides a first illustration of a bivariate analysis of a threshold character and a survival trait. We did not assess here the important step of evaluating the plausibility of the posited model [15]. However, in agreement with previous studies we found that genetic variation exists for both traits [13, 17].

For the genetic parameters of longevity, it is important to note that the marginal Weibull sire model proposed in this study is different from the univariate Weibull sire frailty model proposed by Ducrocq and Casella [11].

The model proposed here includes as an extra effect a normally distributed residual effect in log-frailty. This effect is assumed to account for unobserved individual heterogeneity due to omitted covariates, and three quarters of the genetic variation in a sire model. Moreover, if the residual effect is ignored then the univariate Weibull sire model is inconsistent with assumptions of the additive genetic infinitesimal model [2, 4, 14], and simulation results suggest that the estimated parameters will be biased towards zero [8]. This may in part explain why the marginal posterior means of the Weibull parameter $\rho = 1.70$ and the sire variance $G_{s22} = 0.037$ obtained in this study are greater than the ones estimated and used in the Danish routine genetic evaluation of sires for longevity ($\rho = 1.07$, $\sigma_s^2 = 0.030$) [10]. These parameters are obtained from an univariate Weibull sire model without a residual effect in log-frailty [11] using the Survival Kit [12]. Another part of the explanation may be that the data analyzed in this study only is a subset of the data used for the Danish routine genetic evaluations. Furthermore, the systematic part of the two models is not exactly the same. However, for the purpose of ranking of selection candidates simulation studies suggest that the difference between the two survival models (with or with a residual effect) is inconsiderable for a sire variance in the range 0.03 to 0.05 [11].

In this study, we augmented the parameter vector with the unobserved liabilities associated with the threshold trait. The advantage of this approach is that the fully conditional distributions of parameters associated with the threshold trait reduce to standard distribution, which are easy to sample from. While this approach facilitates programming it is also known to cause slow mixing properties of particular thresholds (*e.g.* [17, 37]). Joint updating of parameters is known to improve mixing in hierarchical models [28, 29, 33]. Along this line, Sorensen *et al.* [37] suggested sampling the thresholds and the liabilities jointly to improve the mixing of thresholds. In this study we were only concerned with analyzing categorical data with three categories so we were not obliged with the problem of slow mixing thresholds because they were all fixed for reason of identifiability [37].

For some of the parameters in this paper we assumed improper uniform prior distributions. A disadvantage with improper priors is that the posterior distribution may turn out to be improper as well. In the case where the posterior distribution is unavailable in closed form it is very difficult to demonstrate its property. If instead all parameters are given proper priors then the posterior distribution is guarantied to be proper [21]. A frequent used alternative to improper priors is to use bounded proper priors [37]. The two univariate models that define the bivariate model are both identifiable and that together with

the fact that the gibbs sampler fot the bivariate model worked satisfactory suggests that the associated posterior distribution is proper and therefore is valid for drawing inferences

The method described in this study can easily be extended to allow for both direct and maternal additive genetic effects, QTL effects, time dependent additive genetic effects for the survival trait, and an arbitrary baseline hazard function. Damgaard and Korsgaard [8] discussed how to implement these extensions in the context of a bivariate model of a linear Gaussian trait and a survival trait. They also sketched how this bivariate model could be generalized to an arbitrary number of ordered categorical characters, survival traits and linear Gaussian traits. This paper provides a first step towards such an analysis.

REFERENCES

[1] Albert J.H., Chib S., Bayesian analysis of binary and polychotomous response data, J. Am. Stat. Assoc. 88 (1993) 669–679.

[2] Andersen A.H., Korsgaard I.R., Jensen J., Idenfiability of parameters in - and equivalence of animal and sire models for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits, Department of Theoretical Statistics, University of Aarhus, Research reports 417 (2000) 1–36.

[3] Andersen P.K., Borgan Ø., Gill R.D., Keiding N., Statistical models based on counting processes, Springer, New York, 1992.

[4] Bulmer M.G., The mathematical theory of quantitative genetics, Clarendon Press, Oxford, 1980.

[5] Bundgaard E., Hoej S., Direct access to the cattle database with livestock registration, Annu. Rep. National Committee on Danish Cattle Husbandry, Aarhus Denmark, 2000.

[6] Carlin B.P., Louis T.A., Bayes empirical bayes methods for data analysis, Chapman and Hall, Boca Raton, 2000.

[7] Damgaard L.H., The use of survival models to infer phenotypic and genetic aspects of longevity of sows, Plant & Animal Genomes XIV Conference, 14–18 January 2006, San Diego, CA, USA.

[8] Damgaard L.H., Korsgaard I.R., A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait, Genet. Sel. Evol. 38 (2006) 45–64.

[9] Damgaard L.H., Korsgaard I.R., Simonsen J., Dalsgaard O., Andersen A.H., The effect of ignoring individual heterogeneity in Weibull log-normal sire frailty models, J. Anim. Sci. 84 (2006) 1338–1350.

[10] Danish Cattle Federation, Principles of Danish cattle breeding (2006), http://www.lr.dk/kvaeg/diverse/principles.pdf [consulted: 20 June 2006].

[11] Ducrocq V., Casella G., A Bayesian analysis of mixed survival models, Genet. Sel. Evol. 28 (1996) 505–529.

[12] Ducrocq V., Sölkner J., "The Survival Kit V3.0", a package for large analyses of survival data, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Vol 27, University of New England, Armidale, pp. 447–450.
[13] Ducrocq V., Quaas R.L., Pollak E.J., Casella G., Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation, J. Dairy Sci. 71 (1988) 3071–3079.
[14] Fisher B.A., The correlation between relatives on the supposition of Mendelian inheritance, T. Roy. Soc. Edin. 52 (1918) 399–433.
[15] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., Bayesian data analysis, Chapman and Hall, Boca Raton, 1995.
[16] Gilks W.R., Wild P., Adaptive rejective sampling for Gibbs sampling, Appl. Statist. 41 (1992) 337–348.
[17] Hansen M., Lund M.S., Pedersen J., Christensen L.G., Gestation length in Danish Holsteins has weak genetic associations with stillbirth, calving difficulty, and calf size, Livest. Prod. Sci. 91 (2004) 23–33.
[18] Hastings W.K., Monte Carlo sampling methods using Markov chains and their application, Biometrika 57 (1970) 97–109.
[19] Henryon M., Berg P., Jensen J., Andersen S., Genetic variation for resistance to clinical and subclinical diseases exists in growing pigs, Anim. Sci. 73 (2001) 375–387.
[20] Henryon M., Jokumsen A., Berg P., Lund I., Pedersen P.B., Olesen N.J., Slierendrecht W.J., Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within a farmed population of rainbow trout, Aquaculture 209 (2002) 59–76.
[21] Hobert J.P., Casella G., The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, J. Amer. Statist. Assoc. 91 (1996) 1461–1473.
[22] Jørgensen B., Andersen S., Genetic parameters for osteochondrosis in Danish Landrace and Yorkshire boars and correlations with leg weakness and production traits, Anim. Sci. 71 (2000) 427–434.
[23] Korsgaard I.R., Madsen P., Jensen J., Bayesian inference in the semi-parametric log normal frailty model using Gibbs sampling, Genet. Sel. Evol. 30 (1998) 241–256.
[24] Korsgaard I.R., Andersen A.H., Jensen J., Prediction error variance and expected response to selection is based on the best predictor – for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits, Genet. Sel. Evol. 34 (2002) 307–333.
[25] Korsgaard I.R., Lund M.S., Sorensen D., Gianola D., Madsen P., Jensen J., Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical, and binary traits using Gibbs sampling, Genet. Sel. Evol. 35 (2003) 159–183.
[26] Larroque H., Ducrocq V., An indirect approach for the estimation of genetic correlations between longevity and other traits, in: Proceedings of the 21th Interbull Meeting, May, 1999, vol. 21, Jouy-en-Josas, pp. 128–135.
[27] Little R.J.A., Rubin D.B., Statistical analysis with missing data, Wiley, New York, 1987.

[28] Liu J.S., The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem, J. Amer. Statist. Assoc. 89 (1994) 958-966.
[29] Liu J.S., Wong H.W., Kong A., Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, Biometrica 81 (1994) 27–40.
[30] Lundeheim N., Genetic analysis of osteochondrosis and leg weakness in the Swedish pig progeny testing scheme, Acta. Agr. Scand. 37 (1987) 159–173.
[31] Luo M.F., Boettcher P.J., Schaeffer L.R., Dekkers J.C.M., Estimation of genetic parameters of calving ease in first and second parities of Canadian Holsteins using Bayesian methods, Livest. Prod. Sci. 74 (2002) 175–184.
[32] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equations of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[33] Roberts G.O., Sahu S.K., Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, J. Royal Stat. Soc. B 59 (1997) 291–317.
[34] Robertson A., Lerner I.M., The heritability of all-or-none traits: Viability of poultry, Genetics 34 (1949) 395–411.
[35] Smith S.P., Quaas R.L., Productive lifespan of bull progeny groups: failure time analysis, J. Dairy Sci. 67 (1984) 2999–3007.
[36] Sorensen D., Gianola D., Likelihood, Bayesian, MCMC methods in quantitative genetics, Springer, New York, 2002.
[37] Sorensen D., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229–249.
[38] Steinbock L., Näsholm A., Berglund B., Johansson K., Philipsson J., Genetic effects on stillbirth and calving difficulty in Swedish Holsteins at first and second calving, J. Dairy Sci. 86 (2003) 2228–2235.
[39] Veerkamp R.F., Brotherstone S., Engel B., Meuwissen T.H.E., Analysis of censored survival data using random regression models, Anim. Sci. 72 (2001) 1–10.
[40] Vukasinovic N., Application of survival analysis in breeding for longevity, in: Proceedings of the 21th Interbull Meeting, May, 1999, vol. 21, Jouy-en-Josas, pp. 181–189.
[41] Wright S., An analysis of variability in number of digits in an inbred strain of guinea pigs, Genetics 19 (1934) 506–536.
[42] Yazdi M.H., Visscher P.M., Ducrocq V., Thompson R., Heritability, reliability of genetic evaluations and response to selection in proportional hazard models, J. Dairy Sci. 67 (2002) 1563–1577.