

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2009 - 21st Annual Conference Proceedings

HIERARCHICAL BAYESIAN METHODS TO MODEL HETEROGENEITY IN COW- AND HERD-LEVEL RELATIONSHIPS BETWEEN MILK PRODUCTION AND REPRODUCTION IN DAIRY COWS

Nora M. Bello

Juan P. Steibel

Robert J. Tempelman

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Bello, Nora M.; Steibel, Juan P.; and Tempelman, Robert J. (2009). "HIERARCHICAL BAYESIAN METHODS TO MODEL HETEROGENEITY IN COW- AND HERD-LEVEL RELATIONSHIPS BETWEEN MILK PRODUCTION AND REPRODUCTION IN DAIRY COWS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1076>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

HIERARCHICAL BAYESIAN METHODS TO MODEL HETEROGENEITY IN COW- AND HERD-LEVEL RELATIONSHIPS BETWEEN MILK PRODUCTION AND REPRODUCTION IN DAIRY COWS

Nora M. Bello^{*}, Juan P. Steibel^{**†} and Robert J. Tempelman^{*}

^{*} *Department of Animal Science, Michigan State University, East Lansing.*

[†] *Department of Fisheries and Wildlife, Michigan State University, East Lansing.*

ABSTRACT

Two of the most important broad classifications of phenotypes for successful dairy production are milk yield and fertility. The nature of the relationship between milk production and reproductive performance of dairy cows is uncertain due to conflicting results reported in many studies. A common deficiency in many such studies is an underappreciation of the dual dimension of the production-reproduction relationship, as defined by *herd* (random or **u**) level and *cow* (residual or **e**) level sources of (co)variation. Our overall hypothesis is that the **e**- and **u**- level relationships between milk production and reproduction in dairy cows are heterogeneous and depend upon various herd-related and management factors. Our objective is to develop hierarchical Bayesian extensions that capture heterogeneity in the relationships between traits by mixed effects modeling of **u** level and **e** level covariances between traits of interest. We specify a bivariate Bayesian model to jointly model two continuous traits and we apply a square-root free Cholesky decomposition to the variance-covariance matrices of the residuals (*cow*-level) and random effects (*herd*-level). As a result, the **e**- and **u**-level covariances among the traits are reparameterized into unconstrained and easily interpretable **e**- and **u**- regression parameters, respectively. These regression parameters specify the *cow*- and *herd*-level relationships, respectively, between the traits and can be easily modeled as functions of relevant fixed and random effects, thereby providing a mixed model extension of Pourahmadi's method. We validate our method using a simulation study and apply it to data on 305-day milk yield and calving interval of Michigan dairy cows.

KEYWORDS: dairy cow, milk production, reproduction, cow- and herd-level relationships, bivariate Bayesian modeling, Cholesky decomposition.

1. INTRODUCTION

Multivariate mixed effects models have been routinely used to investigate the architecture of relationships between two or more traits at several different levels, specifically (co)variance matrices for different sets of random (**u**) effects and residual (**e**) effects. We are specifically interested in the joint modeling of milk production and reproductive efficiency of dairy cows. These two classes of phenotypes help define the necessary foundation for a successful dairy business. Although antagonistic correlations (e.g., higher milk production leading to poorer fertility) have been generally reported, there are enough discrepancies across studies to suggest the need for modeling (co)variances as functions of covariates that characterize dairy management effects or herd environments (Laben et al., 1982; Lopez-Gatius et al., 2006; Lucy, 2001; Washburn et al., 2002). We consider the relationship between two representative traits using **u**-level (co)variances between clusters, e.g., herds, and **e**-level (co)variances between measurement units, e.g.,

cows within herds, hypothesizing that \mathbf{u} -level and \mathbf{e} -level (co)variance matrices are heterogeneous and depend upon systematic factors.

Explicit structural modeling of covariance matrices as functions of covariates requires care because of necessary positive semi-definite constraints. To facilitate this issue at the \mathbf{e} -level, Pourahmadi (1999) proposed a square root free Cholesky reparameterization of the (co)variance matrix for time ordered responses (e.g., longitudinal data) such that (co)variances are reparameterized as generalized autoregressive parameters (GARP) and innovation variances, as labeled by Pourahmadi (1999). We further extend this work by modeling sources of heterogeneity on these parameters at both the \mathbf{u} -level and \mathbf{e} -level, recognizing that (co)variance matrices between observed phenotypes (i.e., at the \mathbf{y} -level) on two or more traits could be separately affected by each of the two components. We also propose that the \mathbf{e} -level GARP be modeled not only as functions of systematic (i.e., fixed) effects, but also of exchangeable cluster-specific random effects that can be characterized by a distribution. This mixed model specification of reparameterized covariance components should facilitate efficient shrinkage estimation for clusters, e.g. herds, characterized by many levels, each with a relatively limited number of measurement units or subjects, e.g., cows.

The objectives of our study are 1) to develop a hierarchical Bayesian extension to classical bivariate mixed effects modeling of residual (\mathbf{e}) and random (\mathbf{u}) covariance matrices for the joint analysis of two phenotypes, 2) to further validate the properties of our method implemented using Markov Chain Monte Carlo (MCMC) based on a simulation study, and 3) to apply our method by jointly modeling heterogeneity in the \mathbf{u} -level and \mathbf{e} -level covariances between milk production and reproduction of first-lactation dairy cows in Michigan. We strive to choose prior density specifications that are conditionally conjugate (Gelman, 2006) in order to expedite Gibbs sampling steps in our MCMC algorithm (Gelfand and Smith, 1990)

2. METHODS

2.1. Hierarchical Bayesian Model Construction

The conventional linear mixed model. We start with the conventional bivariate linear mixed model

$$y_{ij} = \mathbf{x}_{ij}^{(1)'} \boldsymbol{\beta}_i + \mathbf{z}_j' \mathbf{u}_i + e_{ij} \quad (1)$$

where y_{ij} is the observation for trait i ($i=1, 2$) on subject j ($j=1, \dots, n$), $\boldsymbol{\beta}_i$ is a $p_i^{(1)} \times 1$ vector of unknown fixed location parameters for factors (e.g., parity, year, calving season, etc.) unique to trait i ; \mathbf{u}_i is a $q \times 1$ vector of unknown classical random effects (e.g., herd or contemporary group, etc.) unique to trait i and $e_{ij}^{(1)}$ is the corresponding residual. Also, $\mathbf{x}_{ij}^{(1)}$ and \mathbf{z}_j' are known incidence row vectors for subject j . For pedagogical reasons, we assume the same single random effects factor of clusters, e.g. herds, is common to both traits and for all subsequent random effects modeling presented thereafter.

From a Bayesian perspective, the elements of $\boldsymbol{\beta}_i$ are typically considered to be classical fixed effects (Sorensen and Gianola, 2002) whose elements would not be considered to be exchangeable random variables. Hence, we might specify subjective multivariate normal prior densities on fixed location parameters for each trait:

$\beta_i | \beta_i^0, \mathbf{V}_i^{(\beta)} \sim N(\beta_i^0, \mathbf{V}_i^{(\beta)})$, with hyperparameters β_i and $\mathbf{V}_i^{(\beta)}$ being specified as known.

Bounded uniform priors are also commonly considered (Sorensen and Gianola, 2002) as, typically, enough data is available to infer upon elements of β_i with any reasonable noninformative prior distribution in large field studies (Gelman, 2006).

Denote $\mathbf{u}_k = [u_{1,k} \ u_{2,k}]'$ where $u_{i,k}$ denotes element k of \mathbf{u}_i and is the random effect of cluster k ($1 \leq k \leq q$) for trait i . We specify independent structural bivariate normal prior densities on each \mathbf{u}_k with $E(\mathbf{u}_k) = 0$ and $\text{var}(\mathbf{u}_k) = \mathbf{G}_k$ such that:

$$\mathbf{G}_k = \begin{bmatrix} \sigma_{u_{1,k}}^2 & \sigma_{u_{12,k}} \\ \sigma_{u_{12,k}} & \sigma_{u_{2,k}}^2 \end{bmatrix}. \quad (2)$$

Independent bivariate normal densities are assumed for each subject-specific pair of residuals $\mathbf{e}_j = [e_{1,j} \ e_{2,j}]'$ on the two traits with $E(\mathbf{e}_j) = \mathbf{0}$ and $\text{var}(\mathbf{e}_j) = \mathbf{R}_j$ where

$$\mathbf{R}_j = \begin{bmatrix} \sigma_{e_{1,j}}^2 & \sigma_{e_{12,j}} \\ \sigma_{e_{12,j}} & \sigma_{e_{2,j}}^2 \end{bmatrix}. \quad (3)$$

Note that \mathbf{R}_j specifies a within-subject covariance structure between traits at the residual (\mathbf{e}) level but conditional independence is assumed between subjects.

Reparameterization of variance-covariance matrices. We implement a square-root-free Cholesky decomposition (Pourahmadi, 1999) to diagonalize each \mathbf{R}_j and \mathbf{G}_k (co)variance matrix. For the pair of residuals on subject j , this decomposition is based on the following relationship:

$$\mathbf{e}_j = \begin{bmatrix} e_{1,j} \\ e_{2,j} \end{bmatrix} = \begin{bmatrix} e_{1,j} \\ e_{1,j}\varphi_j^{(e)} + e_{2|1,j} \end{bmatrix} = \begin{bmatrix} 0 \\ e_{1,j} \end{bmatrix} \varphi_j^{(e)} + \begin{bmatrix} e_{1,j} \\ e_{2|1,j} \end{bmatrix}. \quad (4)$$

Here $\varphi_j^{(e)}$ represents the subject-specific \mathbf{e} -level regression coefficient of $e_{2,j}$ on $e_{1,j}$, such that $e_{2|1,j}$ is the conditional residual of subject j for trait 2 given trait 1. Furthermore, $e_{2|1,j}$ is independent of $e_{1,j}$ with $e_{2|1,j} \sim N(0, \sigma_{e_{2|1,j}}^2)$. Hence, we rewrite \mathbf{R}_j in Equation (2) as:

$$\mathbf{R}_j = \begin{bmatrix} \sigma_{e_{1,j}}^2 & \varphi_j^{(e)} \sigma_{e_{1,j}}^2 \\ \varphi_j^{(e)} \sigma_{e_{1,j}}^2 & \sigma_{e_{2|1,j}}^2 + (\varphi_j^{(e)})^2 \sigma_{e_{1,j}}^2 \end{bmatrix}. \quad (5)$$

Similarly, we specify the following relationship for the pair of random effects on cluster k :

$$\mathbf{u}_k = \begin{bmatrix} u_{1,k} \\ u_{2,k} \end{bmatrix} = \begin{bmatrix} u_{1,k} \\ u_{1,k}\varphi_k^{(u)} + u_{2|1,k} \end{bmatrix} = \begin{bmatrix} 0 \\ u_{1,k} \end{bmatrix} \varphi_k^{(u)} + \begin{bmatrix} u_{1,k} \\ u_{2|1,k} \end{bmatrix}. \quad (6)$$

Here $\varphi_k^{(u)}$ represents the cluster-specific \mathbf{u} -level regression coefficient of $u_{2,k}$ on $u_{1,k}$, such that $u_{2|1,k}$ is the conditional random effect on trait 2 given trait 1 corresponding to cluster k , and is independent of $u_{1,k}$ with $u_{2|1,k} \sim N(0, \sigma_{u_{2|1,k}}^2)$. Hence, we rewrite \mathbf{G}_k in Equation (3) as:

$$\mathbf{G}_k = \begin{bmatrix} \sigma_{u_{1,k}}^2 & \varphi_k^{(u)} \sigma_{u_{1,k}}^2 \\ \varphi_k^{(u)} \sigma_{u_{1,k}}^2 & \sigma_{u_{2|1,k}}^2 + \left(\varphi_k^{(u)}\right)^2 \sigma_{u_{1,k}}^2 \end{bmatrix}. \quad (7)$$

Using the conventions established by Pourahmadi (1999) and Daniels and Pourahmadi (2002), $\sigma_{u_{2|1,k}}^2$ and $\sigma_{e_{2|1,j}}^2$ might be referred to as the random effect and residual *innovation* variances on trait $i=2$ specific to cluster k and subject j , respectively. However, we prefer to use the term *conditional* variances rather than *innovation* variances for reasons that are hopefully obvious from Equations (5) and (7). With these reparameterizations, Equation (1) does not change for trait $i = 1$ since it is specified as the first trait, and hence its random or residual effects are not conditioned upon those of any other trait. However, for trait $i = 2$, Equation (1) would be rewritten as:

$$y_{2,j} = \mathbf{x}_{2,j}^{(1)'} \boldsymbol{\beta}_2 + \mathbf{z}_j' \left(\boldsymbol{\Psi}^{(u)} \mathbf{u}_1 + \mathbf{u}_{2|1} \right) + \varphi_j^{(e)} e_{1,j} + e_{2|1,j}. \quad (8)$$

where $\mathbf{u}_{2|1} = \{u_{2|1,k}\}_{k=1}^q$ is a $q \times 1$ vector of random effects on trait 2 conditional on trait 1 and

$\boldsymbol{\Psi}^{(u)}$ is a diagonal matrix with diagonal elements $\boldsymbol{\varphi}^{(u)} = [\varphi_1^{(u)} \quad \varphi_2^{(u)} \quad \dots \quad \varphi_q^{(u)}]'$. It should

then be apparent that $\varphi_k^{(u)} = \frac{\sigma_{u_{12,k}}}{\sigma_{u_{1,k}}^2}$ and $\varphi_j^{(e)} = \frac{\sigma_{e_{12,j}}}{\sigma_{e_{1,j}}^2}$. That is, $\varphi_k^{(u)}$ can be interpreted as the

conditional change in $u_{2,k(j)}$, and hence in $y_{2,j}$, for every unit change in $u_{1,k(j)}$ where $k(j)$

defines the cluster k associated with subject j . Similarly, $\varphi_j^{(e)}$ can be interpreted as the conditional change in $e_{2,j}$, and hence in $y_{2,j}$, for every unit change in $e_{1,j}$. Hence, we refer

to parameters $\varphi_k^{(u)}$ and $\varphi_j^{(e)}$ as the \mathbf{u} -level and \mathbf{e} -level *regression coefficients*, respectively, for our two trait application, rather than as GARP as in Pourahmadi (1999). Note that \mathbf{R}_j and \mathbf{G}_k are guaranteed to be positive definite for any respective values of $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ (Pourahmadi, 1999), thereby facilitating their specification as a linear function of covariates and/or random effects

Heterogeneous (co)variance modeling. We specify a linear mixed effects model on each subject-specific $\varphi_j^{(e)}$:

$$\varphi_j^{(e)} = \mathbf{x}_j^{(2)'} \boldsymbol{\gamma}_e + \mathbf{z}_j' \mathbf{m}. \quad (9)$$

Here, γ_e represents a $p^{(2)} \times 1$ vector of unknown fixed effects whereas \mathbf{m} represents a $q \times 1$ vector of unknown cluster-specific random effects as before but such that $\mathbf{m} \sim N(\mathbf{0}, \mathbf{I}\sigma_m^2)$. Furthermore, $\mathbf{x}_j^{(2) \prime}$ is a known row incidence vector. Note that the effects considered in γ_e do not necessarily need to mirror those considered for location parameters β_i ; that is, it is not necessary that $\mathbf{x}_j^{(2) \prime} = \mathbf{x}_{ij}^{(1) \prime}$ for either $i=1$ or $i=2$.

We similarly specify a linear model on each cluster-specific $\varphi_k^{(u)}$:

$$\varphi_k^{(u)} = \mathbf{x}_k^{(3) \prime} \gamma_u \tag{10}$$

where γ_u represents a $p^{(3)} \times 1$ vector of unknown fixed effects with $\mathbf{x}_k^{(3) \prime}$ being the associated known row incidence vector.

We also accommodate heterogeneity of conditional variances at the **e**-level, namely $\sigma_{e_{1,j}}^2$ and $\sigma_{e_{2|1,j}}^2$, and at the **u**-level, namely $\sigma_{u_{1,j}}^2$ and $\sigma_{u_{2|1,j}}^2$, following methodology previously described by Cardoso et al. (2005) and Kizilkaya and Tempelman (2005).

Remaining prior density specifications: In all remaining specifications, we treat all hyperparameters as known, striving to choose priors that are conditionally conjugate to facilitate Gibbs sampling. First we adopt subjectively-specified normal prior densities on the fixed effects influencing heterogeneity of the **e**-level and **u**-level regression coefficients, i.e., $\gamma_e \sim N(\boldsymbol{\mu}_\gamma^{(e)}, \mathbf{V}_\gamma^{(e)})$, $\gamma_u \sim N(\boldsymbol{\mu}_\gamma^{(u)}, \mathbf{V}_\gamma^{(u)})$, although again bounded uniform priors could be specified as well. We further specify an inverse gamma prior distribution $IG(\alpha_m, \beta_m)$ on σ_m^2 . Prior specification of parameters that characterize conditional heteroskedasticity defined at the **e**-level and **u**-level was as previously described by Cardoso et al. (2005) and Kizilkaya and Tempelman (2005).

2.2. Inference

We base our inference for the proposed hierarchical Bayesian model using MCMC. The joint posterior distribution of all unknowns as well as the FCD for these unknowns are derived and presented in the Appendix. It is further important to note that identifiability constraints are required on all fixed effects parameters, namely $\beta_1, \beta_2, \gamma_e, \gamma_u, \tau_{e_1}, \tau_{e_{2|1}}, \tau_{u_1}$, and $\tau_{u_{2|1}}$, in order to remove hypersensitivity to prior specifications. We thereby recommend and adapt the corner parameterization (Clayton, 1996; Kizilkaya and Tempelman, 2005), also known as set-to-zero restriction (Milliken and Johnson, 2009), whereby an overall intercept is always specified and the effect corresponding to one arbitrarily chose level of each fixed effects factor is “zeroed out” or removed.

3. SIMULATION STUDY

We validate our proposed model using a simulation study for which our focus was inference on γ_e, γ_u , and σ_m^2 . Two correlated response variables were simulated to mimic milk yield and calving interval for approximately 50,000 subjects (e.g., cows) distributed across 200 clusters (e.g., herds) within each replicated dataset. The number of subjects per cluster was

drawn from a gamma distribution based on the mean and variance of herd sizes observed from an actual dataset to be described later. The size of the simulated dataset was chosen to mirror that of the actual dataset (see later) and to allow for powerful inference across the highly hierarchical structure of the model. We considered three different broad scenarios or correlation architectures between traits that might be plausible for a number of disparate applications. These 3 scenarios differed in terms of general sign of the **e**-level and **u**-level regression coefficients, namely: A) *same sign*: positive **u**-level and **e**-level coefficients; B) *opposite sign*: negative **u**-level and positive **e**-level coefficients; C) *zero correlation*: zero **u**-level and **e**-level coefficients. We also considered 4 different values for the variance component σ_m^2 : I) $\sigma_m^2 = 0$; II) $\sigma_m^2 = 0.1$; III) $\sigma_m^2 = 1$; and IV) $\sigma_m^2 = 10$. Ten replicate datasets were simulated for each of the 12 possible populations as defined by the factorial of 3 different correlation architectures with 4 different values of σ_m^2 . The same two levels of a single fixed effects factor were considered, where applicable, for all location parameters, conditional residual and random effects variance components, and **e**-level and **u**-level regression coefficients. In other words, the corresponding incidence row vectors for all fixed effects terms were identical such that all covariates were cluster-specific; i.e., $\mathbf{x}_{1j}^{(1)'} = \mathbf{x}_{2j}^{(1)'} = \mathbf{x}_j^{(2)'} = \mathbf{x}_{k(j)}^{(3)'} = \mathbf{x}_{1j}^{(4)'} = \mathbf{x}_{2j}^{(4)'} = \mathbf{x}_{1,k(j)}^{(5)'} = \mathbf{x}_{2,k(j)}^{(5)'}$, with the first element set equal to 1 to specify an intercept and the second element being a Bernoulli (0,1) random draw with probability of 0.25 to partially mimic an unbalanced design structure as based on a corner parameterization. We used arbitrary 2×1 specifications for $\boldsymbol{\gamma}_e = [\gamma_{e_1} \ \gamma_{e_2}]'$ and $\boldsymbol{\gamma}_u = [\gamma_{u_1} \ \gamma_{u_2}]'$ from Equations (9) and (10) to create the intended correlation architectures such that $\boldsymbol{\gamma}_e = \boldsymbol{\gamma}_u = \mathbf{0}$ in scenario C; these specifications are provided in Table 1. We also set $\boldsymbol{\tau}_{e_1} = [\tau_{e_1,1} \ \tau_{e_1,2}] = [176 \ 220]'$ and $\boldsymbol{\tau}_{e_{2|1}} = [\tau_{e_{2|1},1} \ \tau_{e_{2|1},2}] = [9,100 \ 13,000]'$ per Equation (11) and $\boldsymbol{\tau}_{u_1} = [\tau_{u_1,1} \ \tau_{u_1,2}] = [150 \ 100]'$ and $\boldsymbol{\tau}_{u_{2|1}} = [\tau_{u_{2|1},1} \ \tau_{u_{2|1},2}] = [900 \ 600]'$ per Equation (13) for all simulated datasets. The values for $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{e_{2|1}}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ used in the simulation were arbitrarily chosen among a set of plausible values based on the subject-matter literature and a preliminary evaluation of the actual dataset to be described later.

Similarly, the same hyperparameter values $\eta_{e_1} = 8$ and $\eta_{e_{2|1}} = 4$ were used for all datasets to specify the degree of heterogeneity in conditional residual variances across clusters for traits 1 and 2, respectively. In all cases, flat unbounded priors were specified on $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, and σ_m^2 , as well as for $\boldsymbol{\beta}_i$, $i = 1, 2$ and for $\boldsymbol{\tau}_{u_i}$ and $\boldsymbol{\tau}_{e_i}$, $i = 1, 2|1$.

For the analysis of each of the 120 simulated datasets, the length of the MCMC chain was 100,000 cycles after a burn-in period of 1,000 cycles. Convergence diagnostics was monitored graphically and following Raftery and Lewis (1992). For all elements of $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, and σ_m^2 , we assessed frequentist properties based on the equal-tailed 95% posterior probability interval (PPI); i.e. the 2.5th and 97.5th percentiles of the corresponding posterior density.

We wished to validate the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002) as a means to test for the importance of σ_m^2 . Two competing models were evaluated: a full model (M_1) that included cluster-specific e-level regressions (i.e., $\sigma_m^2 > 0$) and a null model (M_0) that did not (i.e., $\sigma_m^2 = 0$). The difference between the two corresponding DIC values, respectively DIC_1 and DIC_0 , were used to draw conclusions on the importance of σ_m^2 . Smaller values of DIC are indicative of improved model fit, such that positive values of $(DIC_0 - DIC_1)$ would suggest M_1 to be the better fitting model and thus indicate evidence of non-zero σ_m^2 . Generally, DIC differences exceeding 7 are believed to indicate a decisive difference in model fit (Spiegelhalter et al., 2002).

For all 90 replicated datasets in which $\sigma_m^2 > 0$, values of $(DIC_0 - DIC_1)$ were all greater than +7, thereby always correctly selecting the full model. Moreover, as expected, the value of $(DIC_0 - DIC_1)$ increased with greater values of σ_m^2 and showed no pattern between the different correlation architectures. Ranges of $(DIC_0 - DIC_1)$ values were [11, 98] for $\sigma_m^2 = 0.1$; [522, 1378] for $\sigma_m^2 = 1.0$; and [4658, 14175] for $\sigma_m^2 = 10$. For 29 of the 30 replicated datasets where $\sigma_m^2 = 0$, the absolute values of $(DIC_0 - DIC_1)$ were less than 7, with the range being [-3.9, 5.1]. The remaining dataset had a DIC difference of 9, thereby incorrectly choosing the full model, at least based on the rule of thumb provided by Spiegelhalter et al. (2002). We then believe these results validate DIC and Spiegelhalter's rule as a reliable model choice criterion for a decision rule on σ_m^2 .

Posterior Inference on Random Regression Parameters: Table 1 presents the minimum and maximum values for each of the 2.5th and 97.5th percentiles of the posterior distribution for γ_{e_1} , γ_{e_2} , γ_{u_1} , γ_{u_2} , and σ_m^2 across the 10 replicates for each of the 12 simulation populations considered. Coverage probabilities for the e- and u-regression parameters across the entire simulation study was near frequentist expectation as the replicate-specific 95% PPI included the true parameter value in 541 out of 570 cases (based on 120 replicated datasets times 4 fixed effects parameters, namely γ_{e_1} , γ_{e_2} , γ_{u_1} , and γ_{u_2} ; plus 90 cases on σ_m^2 for datasets involving non-zero σ_m^2).

For each simulated population, posterior means (not shown) of γ_{e_1} , γ_{e_2} , σ_m^2 , γ_{u_1} , and γ_{u_2} , were evaluated for bias with respect to their true values using a one-sample non-parametric Wilcoxon Rank Sum Test and a one-sample t-test assuming normality. Based on a Type I error rate of 5% for each parameter, these tests did not support biased estimation of posterior means for any regression parameters for any of the simulated populations (not shown). As expected, posterior means of γ_{u_1} and γ_{u_2} , were more variable and their 95% PPI were wider than for γ_{e_1} and γ_{e_2} , as there is typically greater uncertainty for inferences on dispersion parameters characterizing random effects as opposed to those for residuals. Furthermore, Table 1 illustrates that increasing values of σ_m^2 had a detrimental effect on the

precision of inference on γ_{e_1} and γ_{e_2} . Nevertheless, the correlation architecture, as manifested by the three different combinations of values for γ_{e_1} , γ_{e_2} , γ_{u_1} and γ_{u_2} , did not seem to influence the width of the 95% PPI for any of those parameters.

4. APPLICATION TO DAIRY DATA

4.1. Data Description

The two traits of interest were milk yield (kg. x 100) adjusted to 305 day lactation lengths and calving interval (days) defined as the interval from the first calving to second calving in primiparous dairy cows. Data on 49,789 first-lactation cow records from 578 Michigan dairy herds from 2005 to 2007 were provided by the National Dairy Herd Improvement Association (DHIA, Raleigh, NC). Random clusters were characterized by 1,408 herd-years or contemporary groups, being defined as the cluster of animals managed within the same herd and year. All subsequent random effects modeling for this example is based on this cluster definition.

Classical fixed effects (β_1 , β_2) factors considered for both traits included the effects of 4 calving seasons (Winter: December to February; Spring: March to May; Summer: June to August; and Fall: September to November) and 3 years (2005, 2006, 2007). Additionally for β_1 (i.e., milk production), we considered the fixed effects of 3 levels of bovine somatotropin (bST) supplementation: non-users (0% of the herd enrolled), intermediate users (>0-50% of the herd enrolled), and committed users ($\geq 50\%$ of the herd enrolled), as well as the fixed effect of 2 different levels of milking frequency (2 times per day or 2X, versus 3 or more times per day or 3+X). Both of these factors are only recorded at the herd level and reflect potentially different herd management strategies

We used an ad-hoc approach (Bello et al., 2009) to select candidate sources of systematic heterogeneity to model on the **e**-level and **u**-level relationships (i.e., γ_e , γ_u) between milk production and reproductive performance although we emphasize that the chosen factors are not intended to represent a comprehensive list. We modeled $\phi_j^{(e)}$ as a function of the fixed effects (γ_e) of milking frequency in the herd whereas $\phi_k^{(u)}$ was modeled as a function of the fixed effects (γ_u) of bST supplementation. To be consistent with these specifications, the fixed effects specifications for the conditional variances were mirrored accordingly. That is, the fixed effects for the conditional residual variances were based on the herd milking frequency factor whereas the fixed effects for the conditional random effects variances were based on level of bST supplementation. Furthermore, random cluster effects were also modeled for **e**-level conditional variances. Results on modeling of heterogeneous conditional residual variances and conditional random effects variances and not shown herein due to space constraints. Prior densities for all remaining parameters were specified as indicated previously for the simulation study. Also as with the simulation study, two competing models were fitted to the data: a full model fitting herd-year as a random cluster-specific source of **e**-level heterogeneity (\mathbf{m}) with $\mathbf{m} \sim N(\mathbf{0}, \mathbf{I}\sigma_m^2)$ and a reduced model ignoring this source of heterogeneity. For each of the two competing models we ran one long MCMC chain (100,000 saved cycles after 1,000 cycles of burn-in), using the same

convergence diagnostics as described in the simulation study. For each parameter of interest, we describe the posterior density using posterior means, posterior standard deviations and 95% equal-tailed PPI. In addition, we report the effective sample size (ESS) as a measure of the number of effectively independent samples amongst the 100,000 dependent samples (Sorensen et al., 1995).

4.2. Modeling the relationship between milk production and reproduction in dairy cows: Results

Model choice: The DIC for the full model was 36.2 units less than that for the reduced model, implying that σ_m^2 or variation in cluster or herd-year effects on the e-level relationship between 305-d milk yield and calving interval among first parity cows is significant. Hence, we base all of our subsequent inference on a full model that includes a mixed model specification for each subject-specific $\varphi_j^{(e)}$.

Inferences on residual (e) and random (u) regression coefficients: Posterior means, posterior standard deviations, 95% PPI and effective number of independent samples for MCMC inference on e-level (γ_e) and u-level (γ_u) regression parameters are summarized in

Table 2. The ESS indicated sufficient number of MCMC iterations although mixing for σ_m^2 appeared to be substantially hampered relative to the other parameters. It appears that, in general, the e-level relationship between 305-d milk yield and projected calving interval differed substantially in magnitude from the u-level relationship. The average e-level relationship, based on the posterior mean of $\frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^{(\varphi_e)} \gamma_e$, was of 0.55 d longer projected

calving interval per 100 kg increase in 305-d milk yield and appeared to be significantly different from zero (95% PPI = [0.49, 0.62]). In contrast, the posterior mean of $\frac{1}{q} \sum_{k=1}^q \mathbf{x}_k^{(\varphi_u)} \gamma_u$ indicated the average u-level relationship did not depart significantly from zero (95% PPI = [-0.06, 0.37]). Hence cows with higher milk yields tended to have poorer reproductive efficiency than cows with lower milk yields, but there was no strong evidence that higher producing herds had better or worse reproductive performance than lower producing herds.

At the e-level, the estimated relationship $\gamma_{e,3^+X} = [1 \ 0] \gamma_e$ for cows in 3⁺X milking herds between the two traits was 0.45±0.05 d/100kg compared to the estimate 0.66±0.04 d/100 kg for $\gamma_{e,2X} = [1 \ 1] \gamma_e$ for cows in 2X milking herds, with estimates based on posterior means ± posterior standard deviations. A 95% PPI on their difference ($\gamma_{e,2X} - \gamma_{e,3^+X}$) was [0.08, 0.33], thereby indicating a more favorable relationship between 305-d milk yield and calving interval for cows with more frequent milking. However, at the u-level, the data did not support any evidence of bST usage influencing the relationship between the two traits, as the 95% PPI of all pairwise differences between the three levels overlapped with zero (results not shown). As also seen in the simulation study, uncertainty in inference was greater for parameters determining the between-trait correlation for random

(**u**) effects than that for residual (**e**) effects, as illustrated by the differences in width of the 95% PPI (Table 2).

Assuming that **m** is multivariate normal and that σ_m^2 is equal to its posterior mean of 0.09, one might anticipate a range of $\pm 2\sqrt{0.09} = 1.2$ d per 100 kg between the most extreme herd-year effects, using the Empirical Rule (Ott and Longnecker, 2001). Therefore, centered on an overall posterior mean of 0.55 d/100 kg as described earlier, we expect different clusters to range from -0.05 to 1.15 d of calving interval for every 100 kg increase of 305-d milk yield. Hence, it is possible for some herds to have no overall **e**-level relationship between the two traits, whereas other herds may have highly unfavorable relationships.

5. DISCUSSION:

In this study, we present a hierarchical Bayesian extension to classical bivariate mixed effects modeling that provides a general framework for investigating sources of heterogeneity for residual or subject level (**e**) and random or cluster level (**u**) (co)variances between two traits of interest. Using simulation, we validated the proposed hierarchical Bayesian model which is based on a recently developed (co)variance matrix reparameterization (Pourahmadi et al., 1999). We also validated the use of the DIC to choose between models that differ by the specification of cluster-specific random effects on the residual relationships between two traits. We then applied the model to a currently critical dairy cattle management issue as it pertains to investigating the nature of the covariance matrix architecture between milk production and reproductive fitness, specifically how herd management and environmental covariates may influence the random (i.e., herd) and residual (i.e., cow) level (co)variances.

The Cholesky-based reparameterization proposed by Pourahmadi (1999) alleviates the concern for checking positive definiteness constraints and, based on desirable orthogonality properties of the transformation (Pourahmadi, 2007), facilitates independent hierarchical modeling for each of the resulting parameters. From a multivariate applications standpoint, factors influencing $\phi_j^{(e)}$ and $\phi_k^{(u)}$ may be of greatest interest because they determine the subject and cluster specific relationships, respectively, between traits in an unconstrained and easily interpretable manner. As previously noted by Pourahmadi (1999) these two sets of parameters imply a temporal order among response variables, such that inference on the constituent fixed effects (γ_ε and γ_ν) and random effects (**m**) is also inherently order-dependent. We believe the temporal argument holds for our application based on the sequence of physiological events in a dairy cow. In a dairy production system, cows are already milking at the time reproductive management is implemented (Ensminger, 1993), thus implying milk production to be a factor potentially influencing reproductive performance. Conceptually, our model can be extended to $t > 2$ traits for more standard longitudinal data analysis applications as in Pourahmadi (1999); however, the number of different linear model components will increase to $3t + t(t-1)$ from the 8 different linear models (i.e., on $y_{1j}, y_{2j}, \sigma_{e_{1,j}}^2, \sigma_{e_{2|1,j}}^2, \sigma_{u_{1,k}}^2, \sigma_{u_{2|1,k}}^2, \phi_j^{(e)}$, and $\phi_k^{(u)}$) considered within this paper.

The results from our dairy cattle application were very intuitive. However, up until this point, we knew of no formal method to infer upon factors systematically affecting the

relationships between two traits, and, more specifically, how this relationship is differentially driven by cluster-specific random versus residual effects and their component covariate effects. Our application suggested that the antagonistic relationship (high milk production associated with poorer reproductive performance) is primarily driven at the residual or cow level, but that the degree of this relationship depended upon daily milking frequency. Specifically, 3⁺X milking, compared to 2X milking, is a dairy management practice that appears to be associated with less antagonism between the two traits. The additional mixed model extension on modeling variability (σ_m^2) in this relationship implied further that the residual relationship between the two traits is significantly heterogeneous across herds such that some herds may not have an antagonistic relationship between the two traits. These results warrant further investigation of other management practices and herd-related factors to unveil other potential sources of heterogeneity in the production-reproduction relationship across herds. Herds with inferences unusually distal to zero for their respective elements in \mathbf{m} might be investigated retrospectively to explore any potentially new important management and environmental factors that affect $\phi_j^{(e)}$. As our analysis did not consider a comprehensive set of factors, our estimates of σ_m^2 are likely to be somewhat inflated because of other potentially important covariates that were not modeled. A more comprehensive analysis based on a larger dataset and simultaneous fitting of several fixed effects is forthcoming in future animal science publications.

6. SUMMARY

Linear mixed effects modeling of (co)variances, and thus of relationships between traits of interest are possible for both random and residual effects based on a recently popularized covariance matrix decomposition. Hence, researchers should be able to further fine-tune inference on the architecture of correlations between traits by modeling (co)variances as functions of additional fixed and random effects. Using MCMC techniques, we validate the proposed methodology with a simulation study and demonstrate its applicability by addressing the question of heterogeneous relationships between milk production and reproductive performance of dairy cows.

7. ACKNOWLEDGEMENTS

We thank Dr. John Clay and his staff at the Dairy Records Management Systems, Raleigh, NC for providing the DHIA dataset used in this study. This study was partially funded by the Elwood Kirkpatrick Dairy Research Endowment, Michigan Milk Producers Association, the Department of Animal Science and the College of Agriculture and Natural Resources at Michigan State University.

8. APPENDIX: Full Conditional Densities

Write the data for the two traits on subject j as $\mathbf{y} = [\mathbf{y}_{1j} \quad \mathbf{y}_{2j}]'$ such that the entire data vector is $\mathbf{y} = [\mathbf{y}_1' \quad \mathbf{y}_2' \quad \mathbf{y}_3' \quad \dots \quad \mathbf{y}_n']'$. Furthermore, write fixed and random design matrices for the two traits specific to animal j , respectively as $\mathbf{X}_j^{(1)} = \begin{bmatrix} \mathbf{x}_{1j}^{(1)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2j}^{(1)'} \end{bmatrix}$ and $\mathbf{Z}_j = \begin{bmatrix} \mathbf{z}_j' & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_j' \end{bmatrix}; j = 1, 2, \dots, n$. Hence, the

corresponding overall design matrices can be written as $\mathbf{X}^{(1)} = [\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \dots, \mathbf{X}_n^{(1)}]'$, and $\mathbf{Z} = [\mathbf{Z}_1', \mathbf{Z}_2', \dots, \mathbf{Z}_n']'$ linking \mathbf{y} to $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2']'$ and $\mathbf{u} = [\mathbf{u}_1', \mathbf{u}_2']'$, respectively. We also specify $\text{var}(\mathbf{e}) = \boldsymbol{\Sigma}_e = \bigoplus_{j=1}^n \mathbf{R}_j$ where $\mathbf{e} = [\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3', \dots, \mathbf{e}_n']'$ and \bigoplus denotes the direct sum operator (Searle, 1982) such that it should be readily noted that $\boldsymbol{\Sigma}_e^{-1} = \bigoplus_{j=1}^n \mathbf{R}_j^{-1}$. We similarly define $\boldsymbol{\Sigma}_g = \text{var}(\mathbf{u})$ noting that $\boldsymbol{\Sigma}_g^{-1}$ can be readily determined by rearranging elements of $\bigoplus_{k=1}^q \mathbf{G}_k^{-1}$ by animals within traits rather than by traits within animals. It can then be noted using mixed model theory (Sorensen and Gianola, 2002) that the joint FCD of $\boldsymbol{\theta} = [\boldsymbol{\beta}', \mathbf{u}']'$ is multivariate Gaussian:

$$\boldsymbol{\theta} \sim N \left(\left(\mathbf{W}' \boldsymbol{\Sigma}_e^{-1} \mathbf{W} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \left(\mathbf{W}' \boldsymbol{\Sigma}_e^{-1} \mathbf{y} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\theta}_0 \right), \left(\mathbf{W}' \boldsymbol{\Sigma}_e^{-1} \mathbf{W} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \right) \quad [\text{A1}]$$

for $\mathbf{W} = [\mathbf{X}^{(1)} \quad \mathbf{Z}]$, $\boldsymbol{\Sigma}_0 = \text{diag}(\mathbf{V}_1^{(\beta)}, \mathbf{V}_2^{(\beta)}, \boldsymbol{\Sigma}_g)$, and $\boldsymbol{\theta}_0 = \text{diag}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \mathbf{0}_{2q \times 1})'$. There are a number of different alternative strategies for sampling from elements of $\boldsymbol{\theta}$, including single site or univariate Gibbs updates (Wang et al., 1994) and block sampling strategies (GarciaCortes and Sorensen, 1996) that exploit the sparsity (i.e., high frequency of zero elements) in $\left(\mathbf{W}' \boldsymbol{\Sigma}_e^{-1} \mathbf{W} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1}$. Note then that draws of $\mathbf{u}_{2|1}$ can then simply be determined as $\mathbf{u}_2 - \boldsymbol{\Psi}^{(u)} \mathbf{u}_1$ whereas draws of $\mathbf{e}_{2|1}$ can be determined as a vector with elements $\left\{ e_{2,j} - \phi_j^{(e)} e_{1,j} \right\}$.

Similar developments can be used to demonstrate that the FCD of $[\boldsymbol{\gamma}_e', \mathbf{m}']'$, is multivariate Gaussian except that one makes the following substitutions in [A1]: $\left[[\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \dots, \mathbf{x}_n^{(2)}]', [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n] \right]'$ for \mathbf{W} ,

$\text{diag} \left(\frac{\sigma_{e_{2|1,j}}^2}{e_{1,j}^2} \right)$ for $\boldsymbol{\Sigma}_e$, $\text{diag}(\mathbf{V}_\gamma^{(e)}, \mathbf{I} \sigma_m^2)$ for $\boldsymbol{\Sigma}_0$, $(\boldsymbol{\mu}_\gamma^{(e)}, \mathbf{0}_{q \times 1})'$ for $\boldsymbol{\theta}_0$, and a $n \times 1$ vector with elements $\left\{ y_{2,j} - \mathbf{x}_{2,j}^{(1)} \boldsymbol{\beta}_2 - \mathbf{z}_j' (\boldsymbol{\Psi}^{(u)} \mathbf{u}_1 + \mathbf{u}_{2|1}) \right\}$ for \mathbf{y} . Similarly, the FCD for $\boldsymbol{\gamma}_u$ is also multivariate Gaussian making the

following substitution for terms in [A1]: $[\mathbf{x}_1^{(3)}, \mathbf{x}_2^{(3)}, \dots, \mathbf{x}_n^{(3)}]'$ for \mathbf{W} , $\text{diag} \left(\frac{\sigma_{e_{2|1,j}}^2}{(\mathbf{z}'_j \mathbf{u}_1)^2} \right)$ for $\boldsymbol{\Sigma}_e$, $\mathbf{V}_\gamma^{(u)}$ for

$\boldsymbol{\Sigma}_0$, and $\left\{ y_{2,j} - \mathbf{x}_{2,j}^{(1)} \boldsymbol{\beta}_2 - \mathbf{z}_j' \mathbf{u}_{2|1} - \phi_j^{(e)} e_{1,j} \right\}$ for \mathbf{y} . The FCD for σ_m^2 can be readily demonstrated to be inverse gamma with parameters $\frac{q}{2} + \alpha_m$ and $\frac{1}{2} \mathbf{m}' \mathbf{m} + \beta_m$.

Conditional Variances: The FCD for parameters affecting the \mathbf{e} -level and \mathbf{u} -level variances was implemented as previously described (Cardoso et al., 2005; Kizilkaya and Tempelman, 2005).

9. REFERENCES

Bello, N. M., R. J. Erskine, and R. J. Tempelman. 2009. Heterogeneous relationship between milk production and reproduction in dairy cows: Preliminary evidence. In: Proceedings of the Annual Joint Meeting of

- the American Dairy Science Association and the American Society of Animal Science, Montreal, Quebec, Canada
- Cardoso, F. F., G. J. Rosa, and R. J. Tempelman. 2005. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. *Journal of Animal Science* 83: 1766-1779.
- Clayton, D. G. 1996. Generalized linear mixed models. In: W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.) *Markov chain monte carlo in practice*. p 275-302. Chapman and Hall, London.
- Daniels, M. J., and M. Pourahmadi. 2002. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* 89: 553-566.
- Ensminger, M. E. 1993. *Dairy cattle science*. Third ed. Interstate Publishers Inc.
- GarciaCortes, L. A., and D. Sorensen. 1996. On a multivariate implementation of the gibbs sampler. *Genetics Selection Evolution* 28: 121-126.
- Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398-409.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1: 515-533.
- Kizilkaya, K., and R. J. Tempelman. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetic Selection and Evolution* 37: 31-56.
- Laben, R. L., R. Shanks, P. J. Berger, and A. E. Freeman. 1982. Factors affecting milk-yield and reproductive-performance. *Journal of Dairy Science* 65: 1004-1015.
- Lopez-Gatius, F. et al. 2006. Screening for high fertility in high-producing dairy cows. *Theriogenology* 65: 1678-1689.
- Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? *Journal of Dairy Science* 84: 1277-1293.
- Milliken, G. A., and D. E. Johnson. 2009. *Analysis of messy data - volume 1: Designed experiments*. Second ed. Chapman and Hall/CRC Press.
- Ott, R. L., and M. Longnecker. 2001. *An introduction to statistical methods and data analysis*. Fifth ed. Duxbury, United States of America.
- Pourahmadi, M. 1999. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 86: 677-690.
- Pourahmadi, M. 2007. Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika* 94: 1006-1013.
- Raftery, A. E., and S. Lewis (Editors). 1992. *How many iterations in the gibbs sampler?* Bayesian statistics. Oxford University Press, Oxford, UK, 763-773 pp.
- Searle, S. R. 1982. *Matrix algebra useful for statistics*. First ed. Wiley-Interscience.
- Sorensen, D., and D. Gianola. 2002. *Likelihood, bayesian and mcmc methods in quantitative genetics*. Springer Verlag, New York.
- Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian-inference in threshold models using gibbs sampling. *Genetics Selection Evolution* 27: 229-249.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series B* 64: 1-34.
- Wang, C. S., J. J. Rutledge, and D. Gianola. 1994. Bayesian-analysis of mixed linear-models via gibbs sampling with an application to litter size in iberian pigs. *Genetics Selection Evolution* 26: 91-115.
- Washburn, S. P., W. J. Silvia, C. H. Brown, B. T. McDaniel, and A. J. McAllister. 2002. Trends in reproductive performance in southeastern holstein and jersey dhi herds. *Journal of Dairy Science* 85: 244-251.

Table 1. Minimum and maximum of the 2.5th and 97.5th percentiles of the posterior distribution (PPct), as well as true values used for simulation, for parameters defining heterogeneity in the residual (**e**) level and random (**u**) level regression coefficients, namely γ_e , σ_m^2 and γ_u , across 10 replicates for each of 12 simulation populations defined by all factorial combinations of 3 correlation architectures (A, B, C) with 4 different values of σ_m^2 (I, II, III, IV) as indicated in text.

Scenarios	A. Same sign: positive u -level and e -level coefficients				B. Opposite sign: negative u -level and positive e -level coefficients				C. Zero correlation: zero u -level and e -level coefficients				
	Regression Parameters †	True	2.5 th PPct [Min, Max]	97.5 th PPct [Min, Max]	True	2.5 th PPct [Min, Max]	97.5 th PPct [Min, Max]	True	2.5 th PPct [Min, Max]	97.5 th PPct [Min, Max]	True	2.5 th PPct [Min, Max]	97.5 th PPct [Min, Max]
I. $\sigma_m^2 = 0$	γ_{u1}	0.3	[-1.80, 0.15]	[0.54, 2.23]	-0.3	[-2.17, -0.80]	[-0.23, 0.68]	0	[-1.46, -0.08]	[0.15, 1.74]	0	[-1.46, -0.08]	[0.15, 1.74]
	γ_{u2}	0.1	[-0.63, 0.08]	[0.35, 1.02]	-0.5	[-1.18, -0.32]	[-0.16, 0.59]	0	[-0.82, -0.24]	[0.05, 0.54]	0	[-0.82, -0.24]	[0.05, 0.54]
	γ_{e1}	0.6	[0.44, 0.58]	[0.60, 0.74]	0.6	[0.44, 0.57]	[0.60, 0.73]	0	[-0.10, 0.01]	[0.05, 0.17]	0	[-0.10, 0.01]	[0.05, 0.17]
	γ_{e2}	0.5	[0.37, 0.48]	[0.53, 0.65]	0.5	[0.37, 0.47]	[0.53, 0.63]	0	[-0.16, -0.04]	[0.00, 0.13]	0	[-0.16, -0.04]	[0.00, 0.13]
	σ_m^2	0	-	-	0	-	-	0	-	-	0	-	-
II. $\sigma_m^2 = 0.1$	γ_{u1}	0.3	[-1.03, 0.19]	[0.73, 1.80]	-0.3	[-1.50, -0.63]	[0.04, 0.98]	0	[-1.47, 0.45]	[0.22, 2.44]	0	[-1.47, 0.45]	[0.22, 2.44]
	γ_{u2}	0.1	[-0.46, -0.04]	[0.48, 0.86]	-0.5	[-1.37, -0.38]	[-0.38, 0.52]	0	[-0.89, -0.06]	[0.10, 0.81]	0	[-0.89, -0.06]	[0.10, 0.81]
	γ_{e1}	0.6	[0.35, 0.57]	[0.55, 0.80]	0.6	[0.40, 0.54]	[0.61, 0.76]	0	[-0.19, -0.03]	[0.00, 0.19]	0	[-0.19, -0.03]	[0.00, 0.19]
	γ_{e2}	0.5	[0.30, 0.45]	[0.53, 0.69]	0.5	[0.31, 0.53]	[0.52, 0.77]	0	[-0.20, -0.02]	[0.00, 0.20]	0	[-0.20, -0.02]	[0.00, 0.20]
	σ_m^2	0.1	[0.03, 0.09]	[0.11, 0.24]	0.1	[0.02, 0.08]	[0.12, 0.22]	0.1	[0.02, 0.06]	[0.11, 0.22]	0.1	[0.02, 0.06]	[0.11, 0.22]
III. $\sigma_m^2 = 1$	γ_{u1}	0.3	[-0.87, 0.04]	[0.39, 1.44]	-0.3	[-1.59, -0.59]	[-0.03, 0.98]	0	[-1.51, -0.20]	[0.15, 1.78]	0	[-1.51, -0.20]	[0.15, 1.78]
	γ_{u2}	0.1	[-0.94, -0.14]	[0.05, 0.71]	-0.5	[-1.44, -0.41]	[-0.43, 0.63]	0	[-0.89, -0.27]	[0.10, 0.65]	0	[-0.89, -0.27]	[0.10, 0.65]
	γ_{e1}	0.6	[0.08, 0.52]	[0.48, 0.89]	0.6	[0.15, 0.65]	[0.51, 1.04]	0	[-0.42, 0.08]	[-0.01, 0.47]	0	[-0.42, 0.08]	[-0.01, 0.47]
	γ_{e2}	0.5	[-0.01, 0.45]	[0.38, 0.82]	0.5	[0.04, 0.47]	[0.40, 0.86]	0	[-0.47, 0.07]	[-0.06, 0.46]	0	[-0.47, 0.07]	[-0.06, 0.46]
	σ_m^2	1	[0.59, 0.88]	[1.11, 1.55]	1	[0.53, 0.79]	[1.00, 1.39]	1	[0.43, 0.84]	[0.82, 1.50]	1	[0.43, 0.84]	[0.82, 1.50]
IV. $\sigma_m^2 = 10$	γ_{u1}	0.3	[-1.51, 0.17]	[0.69, 2.10]	-0.3	[-1.71, -0.70]	[0.13, 0.86]	0	[-1.27, -0.30]	[0.52, 1.48]	0	[-1.27, -0.30]	[0.52, 1.48]
	γ_{u2}	0.1	[-0.91, -0.15]	[-0.08, 0.79]	-0.5	[-1.34, -0.94]	[-0.41, 0.02]	0	[-0.85, 0.24]	[0.10, 1.05]	0	[-0.85, 0.24]	[0.10, 1.05]
	γ_{e1}	0.6	[-0.15, 0.58]	[0.84, 1.49]	0.6	[-0.24, 0.68]	[0.79, 1.67]	0	[-0.96, -0.05]	[0.02, 0.88]	0	[-0.96, -0.05]	[0.02, 0.88]
	γ_{e2}	0.5	[-0.29, 0.44]	[0.71, 1.35]	0.5	[-0.30, 0.52]	[0.71, 1.51]	0	[-0.95, -0.07]	[0.04, 0.86]	0	[-0.95, -0.07]	[0.04, 0.86]
	σ_m^2	10	[6.3, 9.8]	[10.3, 15.1]	10	[6.9, 10.7]	[10.9, 16.7]	10	[6.6, 8.5]	[10.6, 13.3]	10	[6.6, 8.5]	[10.6, 13.3]

Table 2. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% posterior probability intervals (PPI) reported as [2.5th, 97.5th] posterior percentiles, and effective sample size (ESS) on residual (**e**) level (namely, γ_e and σ_m^2) and random (**u**) level (namely, γ_u) regression parameters between milk yield at 305 days-in-lactation and calving interval in Michigan first lactation dairy cows.

Regression parameters †	PMEAN	PSD	PPI	ESS
$\gamma_{u,0\%bST}$, d/100 kg	0.16 ^x	0.17	[-0.17, 0.49]	28 549
$\gamma_{u,>0-50\%bST}$, d/100 kg	0.17 ^x	0.20	[-0.22, 0.56]	28 959
$\gamma_{u,>50\%bST}$, d/100 kg	0.15 ^x	0.19	[-0.22, 0.51]	28 409
$\gamma_{e,2X}$, d/100 kg	0.66 ^a	0.04	[0.57, 0.74]	79 573
$\gamma_{e,3^+X}$, d/100 kg	0.45 ^b	0.05	[0.36, 0.54]	61 687
σ_m^2 , (d/100 kg) ²	0.09	0.03	[0.04,0.16]	612

(^x) and (^{a,b}) Letters indicate significant differences (two-tailed Bayesian P-value < 0.05) between management practices within the **u**-level and **e**-level regression parameters, respectively,

† $\gamma_{u,0\%bST} = [1 \ 1 \ 0]\gamma_u$, $\gamma_{u,>0-50\%bST} = [1 \ 0 \ 1]\gamma_u$ and $\gamma_{u,>50\%bST} = [1 \ 0 \ 0]\gamma_u$ are the random (**u**) level regression parameters between milk yield at 305 days-in-lactation and calving interval for herds that had 0%, >0 to 50% and >50% of their cows enrolled for supplementation with bovine somatotropin (bST), respectively.

$\gamma_{e,2X} = [1 \ 1]\gamma_e$ and $\gamma_{e,3^+X} = [1 \ 0]\gamma_e$ are the residual (**e**) level regression parameters between milk yield at 305 days-in-lactation and calving interval for cows in herds with twice a day and three times a day (or greater) milking frequency, respectively.

σ_m^2 is the parameter defining random between-herd heterogeneity among the **e**-level regression parameters.