Fitting linear mixed models under nonstandard assumptions: a Bayesian approach

Aline S. Damascena[1], **Francisco M.M. Rocha**[2] and **Julio M. Singer**[3]

[1,3]Departamento de Estatística, Universidade de São Paulo

[2]Escola Paulista de Política Economia e Negócios, Universidade Federal de São Paulo

[1]damacena@ime.usp.br [2]f.marcelo.rocha@gmail.com [3]jmsinger@ime.usp.br

Abstract

The standard assumptions considered for fitting Linear Mixed Models (LMM) to longitudinal data include Gaussian distributions and homoskedastic conditional independence. In many situations, however, these assumptions may not be adequate. From a frequentist point of view, adopting different distributions along with more general within subjects covariance structure can be nontrivial because the integrals involved in the estimation process do not allow analytical expressions. Under a Bayesian approach, however, the estimation process can be facilitated by using posterior conditional distributions, which are generally more treatable than posterior marginal distributions. We use diagnostic tools to show that a Gaussian distribution for the random effects is not acceptable for fitting a LMM with AR(1) structure for the within subjects covariance matrix to a dataset involving lactation of dairy cows. We consider alternative Bayesian models adopting t distributions with different degrees of freedom for the random effects. The results indicate that the fixed effects are not considerably affected by the different models, but the corresponding standard errors are smaller when heavier tailed distributions are adopted.

Keywords: linear mixed models, residual analysis, Bayesian analysis, Wood function, Incomplete Gamma, milk production, dairy cow.

1 Introduction

The current standard approach to analyze repeated measurements data considers linear mixed models (LMM) under the frequentist paradigm, for which the theory has been well documented (Henderson, 1984; Laird & Ware, 1982; Verbeke & Lesaffre, 1987) and software is widely available (Pinheiro et al., 2015). Usual adopted assumptions include Gaussian distributions for the random effects as well as homoskedastic independent Gaussian random errors. In practice such assumptions may not be adequate, leading to vulnerable inference. This is where diagnostic procedures play an important role [Nobre & Singer (2007); Singer et al. (2017)].

Under the frequentist paradigm, assuming other distributions for the random effects and/or for the random errors may not be trivial, since the integrals involved in the estimation process of the model parameters generally do not allow analytical expression. Under a Bayesian approach, the estimation process can be facilitated by the possibility of using full conditional posterior distributions that are generally more treatable than marginal posterior distributions.

Several authors have obtained full conditional posterior distributions for the model parameters assuming different distributions for the errors and/or for random effects. Seltzer (1993), Wakefield et al. (1994) and Seltzer et al. (1996), consider t-distributions for the random effects while Strandén & Gianola (1998) consider t-distributions for both errors and random effects. Rosa et al. (2003), propose univariate and multivariate versions of the t, Slash and Contaminated Gaussian distributions for the random errors, and Rosa et al. (2004) extend the results to random errors and random effects. Arellano-Valle et al. (2007) propose asymmetric Gaussian distributions for the random errors and random effects. Jara et al. (2008), assuming that the error covariance matix is known, consider multivariate asymmetric elliptic distributions (which include t-Symmetric, Asymmetric Gaussian, t and Gaussian distributions as special cases) for both errors and random effects. In all these models, however, the authors consider that the within unit observations are independent and homoskedastic.

In the literature there are few applications with models that flexibilize the distributions of the errors and/or of the random effects and also use more general structures for the within unit covariance matrix (Lin & Lee, 2007). Our purpose is to show how Bayesian methods can be easily adapted for the analysis of longitudinal data for which the random effects do not follow Gaussian distributions and for which a conditional independence homoskedastic model is not adequate.

In Section 2 we describe the study that motivated our approach. In Section 3 we present the models under investigation. Results are detailed in Section 4. We consider a discussion in Section 5. Details are given in the Appendix.

2 The study

We considered data from a study conducted at the Universidad Nacional de Rosario, Argentina (Garcia, Rapelli, Koegel, 2009, paper presented at the Workshop on Mixed Models, Tucuman, Argentina, 2010) with the objective of investigating the effects of season and parity in milk production. A total of 2250 weekly milk production records from 150 Dutch dairy cows was collected along 15 weeks *postpartum*. For each cow, parity (first, second and third or more) and season of calving (fall and spring) were collected. Different cows were observed at each combination of parity and season. The lactation curves according to season or parity are displayed in Figure 1.

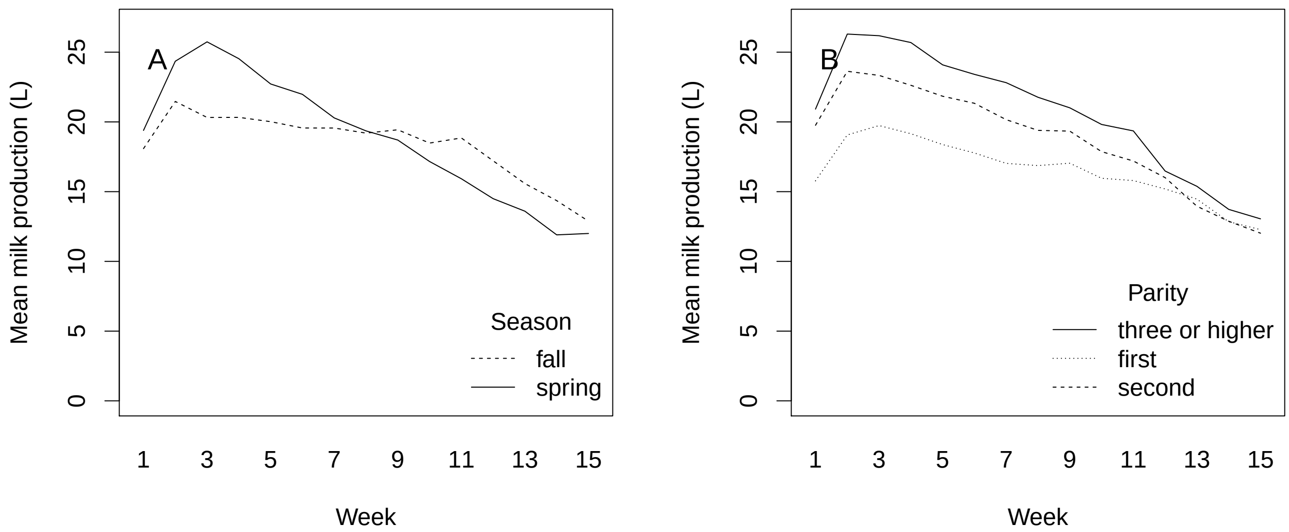

Figure 1: Lactation curves of Dutch dairy cows along 15 weeks according to season (A) and parity (B).

3 The model

The most used function for modeling lactation curves is the incomplete gamma function (Figure 2) also known as Wood model (Wood, 1967) in the context of dairy science, expressed as

$$y(t) = \phi_1 t^{\phi_2} \exp^{-\phi_3 t} \tag{1}$$

where $y(t)$ is the milk production at time t, $\phi_1 > 0$ is a parameter related to the initial milk production, $0 < \phi_2 < 1$ and $0 < \phi_3 < 1$ respectively represent the rate of increase in milk production before and after the peak production (occurring at $t = \phi_2/\phi_3$).

The specification of a multiplicative version of the Wood mixed model for the data under investigation may be expressed as

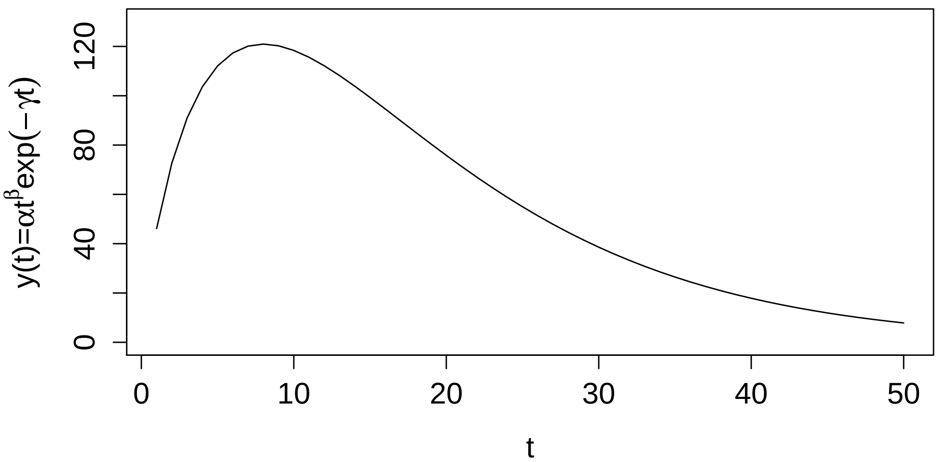

Figure 2: Wood's function with parameters $\phi_1 = 51$, $\phi_2 = 0.8$ e $\phi_3 = 0.1$

$$
\begin{aligned}
y_{ij}|\phi_{ij} &= \phi_{1i} t_{ij}^{\phi_{2i}} \exp\left(-\phi_{3i} t_{ij}\right) e_{ij}, \quad i = 1, \ldots, n \;\text{ and }\; j = 1, \ldots, m_i \\
\phi_{1i} &= \exp(\beta_1 + \beta_2 E_{ij} + \beta_3 P_{2ij} + \beta_4 P_{3ij} + \beta_5 EP_{2ij} + \beta_6 EP_{3ij} + b_{1i}) \\
\phi_{2i} &= \beta_7 + \beta_8 E_{ij} + \beta_9 P_{2ij} + \beta_{10} P_{3ij} + \beta_{11} EP_{2ij} + \beta_{12} EP_{3ij} + b_{2i} \\
\phi_{3i} &= \beta_{13} + \beta_{14} E_{ij} + \beta_{15} P_{2ij} + \beta_{16} P_{3ij} + \beta_{17} EP_{2ij} + \beta_{18} EP_{3ij} + b_{3i}
\end{aligned} \tag{2}
$$

where

y_{ij} is the milk production of the i-th cow on the j-th week;

ϕ_{1i} is the initial milk production parameter for the i-th cow ($\phi_{1i} > 0$);

ϕ_{2i} denotes the rate of increase in production before peak production for the i-th cow ($0 < \phi_{2i} < 1$);

ϕ_{3i} denotes the rate of decrease in production after peak production for the i-th cow ($0 < \phi_{3i} < 1$);

t_{ij} corresponds to the week where the j-th milk production for the i-th cow was observed;

E_{ij} assumes the value 0 when j-th observation for the i-th occurs in the fall or 1 when it occurs in spring;

P_{kij}: assumes the value 1 when j-th observation for the i-th occurs for cows at the k-th parity level, $k = 2, 3$ and 0 otherwise, ;

EP_{2ij} and EP_{3ij} are Season $\times$ Parity interaction terms;

$\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i})^\top$ is a random effect vector related to the i-th cow;

e_{ij} is a random error related to the j-th milk production outcome of the i-th cow.

For the data under evaluation we have $n = 150$ and $m_i = m = 15$; the model includes $q = 3$ random effects. Under this formulation we are assuming that, at each season, the initial production, the rate of increase (decrease) in production before (after) peak production depends on parity and season.

Model (2) may be linearized, yielding

$$\begin{aligned} y^*_{ij}|\phi_{ij} &= \phi^*_{1i} + \phi_{2i}t^*_{ij} - \phi_{3i}t_{ij} + e^*_{ij} \\ \phi^*_{1i} &= \beta_1 + \beta_2 E_{ij} + \beta_3 P_{2ij} + \beta_4 P_{3ij} + \beta_5 EP_{2ij} + \beta_6 EP_{3ij} + b_{1i} \\ \phi_{2i} &= \beta_7 + \beta_8 E_{ij} + \beta_9 P_{2ij} + \beta_{10} P_{3ij} + \beta_{11} EP_{2ij} + \beta_{12} EP_{3ij} + b_{2i} \\ \phi_{3i} &= \beta_{13} + \beta_{14} E_{ij} + \beta_{15} P_{2ij} + \beta_{16} P_{3ij} + \beta_{17} EP_{2ij} + \beta_{18} EP_{3ij} + b_{3i} \end{aligned} \quad (3)$$

where $y^*_{ij} = log(y_{ij})$, $\phi^*_{1i} = \log(\phi_{1i})$, $t^*_{ij} = \log(t_{ij})$ and $e^*_{ij} = \log(e_{ij})$.

Letting $\mathbf{y}^*_i = [\log(y_{i1}), \dots, \log(y_{im_i})]^\top$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_{18}]^\top$, $\mathbf{v} = [\log(t_{i1}), \dots, \log(t_{im_i})]^\top$, $\mathbf{t} = [-t_{i1}, \dots, -t_{im_i}]^\top$, $\mathbf{b}_i = [b_{1i}, b_{2i}, b_{3i}]^\top$, $\mathbf{e}^*_i = [\log(e_{i1}), \dots, \log(e_{im_i})]^\top$,

$$\mathbf{W}_i = \begin{bmatrix} 1 & E_{i1} & P_{2i1} & P_{3i1} & EP_{2i1} & EP_{3i1} \\ 1 & E_{i2} & P_{2i2} & P_{3i2} & EP_{2i2} & EP_{3i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & E_{im_i} & P_{2im_i} & P_{3im_i} & EP_{2im_i} & EP_{3im_i} \end{bmatrix},$$

$$\mathbf{X}^*_i = [\mathbf{W}_i \circ \mathbf{1}_{15} \quad \mathbf{W}_i \circ \mathbf{v} \quad \mathbf{W}_i \circ \mathbf{t}],$$

where $\mathbf{A} \circ \mathbf{c}$ indicates that each element of the i-th row of the matrix $\mathbf{A}$ is multiplied by the corresponding element in the vector $\mathbf{c}$,

$$\mathbf{Z}_i = [\mathbf{1}_{15} \quad \mathbf{v} \quad \mathbf{t}], \quad \mathbf{G} = \begin{bmatrix} \sigma^2_{b_1} & \sigma_{b_1 b_2} & \sigma_{b_1 b_3} \\ \sigma_{b_1 b_2} & \sigma^2_{b_2} & \sigma_{b_2 b_3} \\ \sigma_{b_1 b_3} & \sigma_{b_2 b_3} & \sigma^2_{b_3} \end{bmatrix} \text{ and } \mathbf{R}_i = \sigma^2 \mathbf{C}_i$$

where $\mathbf{C}_i$ is a ($m_i \times m_i$) matrix with elements

$$c_{ijj'} = \begin{cases} 1 & \text{if } j = j', \\ \rho^{|j-j'|} & \text{if } j \neq j' \end{cases}, \quad j, j' = 1, \dots, m_i. \quad (4)$$

we can write model (3) as

$$\mathbf{y}^*_i = \mathbf{X}^*_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}^*_i, \quad i = 1, \dots n \quad (5)$$

where the random effects $\mathbf{b}_i$ follow independent Gaussian distributions with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{G}$, the $\mathbf{e}^*_i$ follow independent Gaussian distributions with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}_i$ and $\mathbf{b}_i$ and $\mathbf{e}^*_i$ are all independent. According to this model, a first order autorregressive structure [AR(1)] with correlation coefficient ρ is assumed for the conditional errors.

Model (5) may be classified as a standard linear mixed model and may be expressed as a two-stage hierarchical model (Laird & Ware, 1982) as

$$\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{b}_i = \mathbf{X}_i^*\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i^* \tag{6}$$

$$\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{G}). \tag{7}$$

In the first stage, for fixed random effects $\mathbf{b}_i$, we have $\mathbf{e}_i^* \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i)$. In the second stage, we let $\mathbf{b}_i$ be random, mutually independent and independent of $\mathbf{e}_i^*$.

Under the frequentist paradigm, model (5) may be fitted via standard linear mixed models methodology outlined in Dey et al. (2000) or Demidenko (2013), for example. In the context of Bayesian analysis, estimation of model (5) parameters is based on the joint posterior distribution, $p(\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}, \mathbf{b}|\mathbf{y})$, where $\mathbf{y}^* = [\mathbf{y}_1^{*\top}, \ldots, \mathbf{y}_n^{*\top}]^\top$ that can be decomposed as

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}, \mathbf{b}|\mathbf{y}^*) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}, \mathbf{b}_i) \prod_{i=1}^{n} p(\boldsymbol{\beta})p(\mathbf{R}_i)p(\mathbf{b}_i|\mathbf{G})p(\mathbf{G}) \tag{8}$$

where $p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}, \mathbf{b}_i)$ is the likelihood function of $\mathbf{y}_i^*$ and $p(\cdot)$ denote prior density functions specified for the different parameters.

Marginal posterior distributions of the parameters, $p(\boldsymbol{\theta}|\mathbf{y})$, may be obtained by integrating (8). Metropolis-Hastings (Hastings, 1970; Metropolis et al., 1953) or Gibbs sampler algorithms (Gelfand & Smith, 1990; Geman & Geman, 1984), two of the most popular Markov Chain Monte Carlo (MCMC) methods may be employed to generate (8). The idea is to generate samples of conditional posterior distributions (conditioned on the data, $\mathbf{y}$, and on the other parameters), a generally easier process than to generate samples directly from the marginal posterior distribution.

Assuming that the errors and random effects follow Gaussian distributions with $\mathbf{R}_i = \sigma^2\mathbf{I}_{m_i}$ and that the remaining components of variance are known, Lindley & Smith (1972) and Fearn (1975) have shown analytically that the fixed effects also follow Gaussian distributions. When the components of variance are unknown, computation of the conditional posterior distribution can be found in Seltzer (1993), Seltzer et al. (1996), Sorensen (2002) or Congdon (2010), for example.

Bayesian approaches to LMM with other covariance structures for $\mathbf{R}_i$ are scarce in the literature (Lee & Lien, 2001; Lee et al., 2005). Fitting models that combine different covariance structures for $\mathbf{G}$ and $\mathbf{R}_i$ is generally a challenge under either the frequentist or the Bayesian paradigms (Lesaffre & Lawson, 2012). Under the Bayesian approach, to generate samples from posterior distributions with more complicated models, we may use software such as WinBUGS (Lunn et al., 2000; Spiegelhalter et al., 2003), JAGS (Plummer, 2015) or STAN (Stan Development Team, 2014) in which the most popular MCMC algorithms are implemented. When conjugate prior distributions are specified, the full conditional distributions are known and the software will use standard sampling algorithms to generate samples from the posterior distributions. Otherwise, the software will use one of the following algorithms: Gibbs sampler (Geman & Geman, 1984), adaptive rejection (Gilks & Wild, 1992), slice sampler (Neal, 1997) or Metropolis-Hastings (Hastings, 1970; Metropolis et al., 1953).

4 Results

First, model (5) was fitted via standard LMM methodology using the package *nlme* (Pinheiro et al., 2015) in the R software (R Core Team, 2014). Due to a software limitation for fitting the Bayesian models (to be discussed later), we did not consider a random effect associated with the rate of increase in production before peak production in model (9). The strategy used to reduce model (5) was as follows:

i) Test whether the AR(1) structure for the conditional errors could be replaced by an independence structure in the model, using a likelihood ratio test (Pinheiro & Bates, 2000).

ii) Test the significance of the interaction effects, that is, test whether $\beta_5 = \beta_6 = \beta_{11} = \beta_{12} = \beta_{17} = \beta_{18} = 0$, using conditional F tests (Pinheiro & Bates, 2000).

iii) In the absence of interaction effects, test the significance of Season and Parity main effects, that is, test whether $\beta_2 = \beta_3 = \beta_4 = 0$, $\beta_8 = \beta_9 = \beta_{10} = 0$ and $\beta_{14} = \beta_{15} = \beta_{16} = 0$, using conditional F tests.

iv) Fit a model that incorporates the conclusions obtained in i)-iii).

The linearized version of the reduced model

$$\begin{aligned} y_{ij} &= \phi_{1i} t_{ij}^{\phi_{2i}} exp(-\phi_{3i} t_{ij}) e_{ij} \\ \phi_{1i} &= exp(\beta_1 + \beta_2 E_{ij} + \beta_3 P_{2ij} + \beta_4 P_{3ij} + b_{1i}) \\ \phi_{2i} &= \beta_7 + \beta_8 E_{ij} \\ \phi_{3i} &= \beta_{13} + \beta_{14} E_{ij} + \beta_{15} P_{2ij} + \beta_{16} P_{3ij} + \beta_{17} EP_{2ij} + \beta_{18} EP_{3ij} + b_{3i} \end{aligned} \tag{9}$$

was fitted to the data and the results are displayed in Table 1.

Table 1: Estimates (standard errors - SE) of the fixed effects and 95% confidence intervals (CI) for the covariance components of model (9)

Interpretation	Parameter	Estimate (SE) or CI
initial milk production (intercept)	$^*\beta_1$	14.507 (0.709)
Spring effect	$^*\beta_2$	1.187 (0.060)
Parity=2 effect	$^*\beta_3$	1.344 (0.079)
Parity$\geq$ 3 effect	$^*\beta_4$	1.465 (0.086)
Rate of increase before peak production (intercept)	β_7	0.305 (0.029)
Spring effect	β_8	0.168 (0.044)
rate of decreasing after peak production (intercept)	β_{13}	-0.059 (0.007)
Spring effect	β_{14}	-0.066 (0.011)
Parity=2 effect	β_{15}	-0.033 (0.007)
Parity$\geq$ 3 effect	β_{16}	-0.026 (0.007)
Spring $\times$ Parity=2 interaction	β_{17}	0.002 (0.001)
Spring $\times$ Parity$\geq$ 3 interaction	β_{18}	0.001 (0.001)
Covariance components	σ_{b1}	[0.173; 0.303]
	σ_{b3}	[0.011; 0.025]
	$\sigma_{b1.b3}$	[-0.655; 0.657]
	σ_e	[0.207; 0.236]
	ρ	[0.416; 0.553]

*: Estimates computed as $exp(\beta_i)$; SE computed via the Delta method

The initial milk production tend to increase in spring as well as when cows have more calvings. Only Season affects the rate of increase in production before peak production. The Parity effect on the rate of decrease in production after peak production depends on the season (Table 1).

To evaluate if the assumption of Gaussian distributions for random errors or random effects in model (9) is adequate, we considered the diagnostic plots described in (Singer et al., 2017). A plot of the standardized least counfounded residuals displayed in Figure (3) does not suggest violation of the Gaussian assumption for the random errors. On the other hand, the QQ-plot of the Mahalanobis Distance (Figure 4) suggests that the assumption of Gaussian distribution for the random effects is not reasonable.

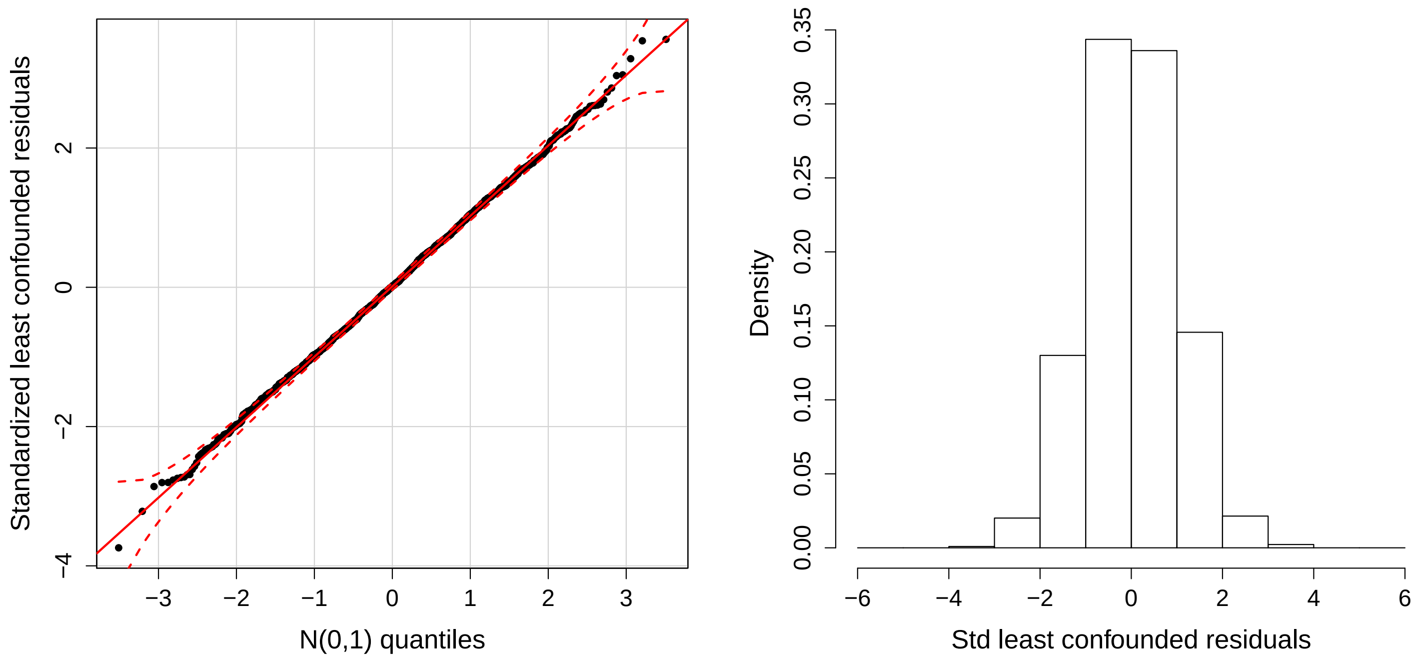

Figure 3: QQ-plot and histogram for standardized least confounded residuals

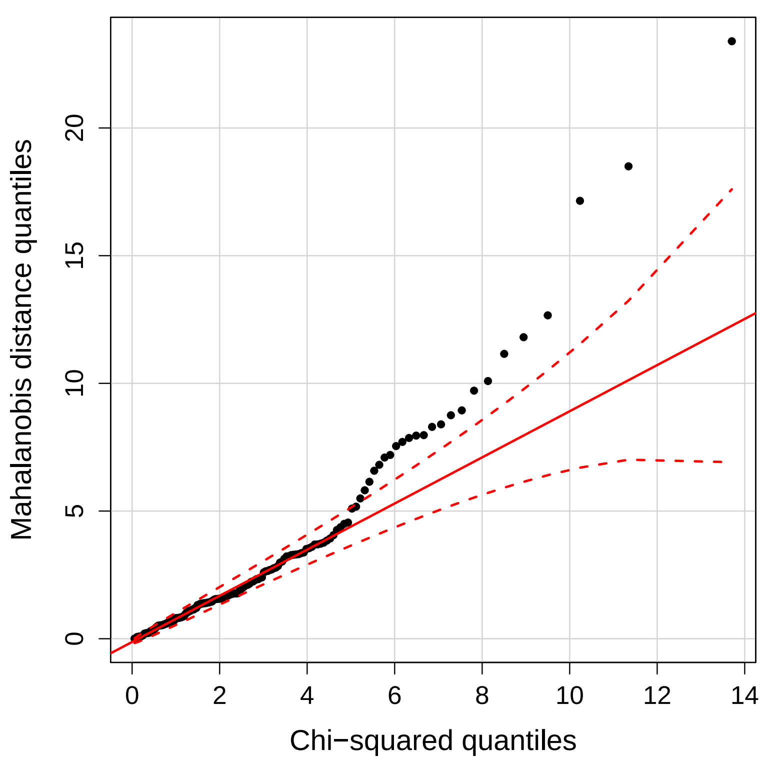

Figure 4: Chi-squared QQ plot for Mahalanobis distance

Considering a Bayesian approach we fitted model (9) assuming multivariate t-distributions $\mathbf{t}_q(\mathbf{0}, \boldsymbol{\Delta}, \nu)$ for the random effects, where

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta^2_{b_1} & \delta_{b_1 b_3} \\ \delta_{b_1 b_3} & \delta^2_{b_3} \end{bmatrix}$$

is the corresponding dispersion matrix and ν denotes the degrees of freedom. Under that assumption, it is not difficult to obtain samples of the conditional posterior distributions of parameters of model (9), since the multivariate t-distribution can be viewed as a mixture of Gaussian distributions. In this context, samples from a multivariate t-distribution with dimension q and parameters $\boldsymbol{\Delta}$ and ν can be generated in two steps: the first consists in generating w from a $Gamma(\nu/2, \nu/2)$ distribution and the second, in generating a vector $\mathbf{x}$ from a $N_q(\boldsymbol{\mu}, \boldsymbol{\Delta}/w)$ distribution (Gamerman & Lopes, 2006). With this in mind, model (5) can be rewritten as

$$\begin{aligned} \mathbf{y}^*_i|\boldsymbol{\beta}, \mathbf{b}_i &= \mathbf{X}^*_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}^*_i \\ \mathbf{b}_i|\boldsymbol{\Delta} &\sim \mathbf{t}_q(\mathbf{0}, \boldsymbol{\Delta}, \nu). \end{aligned} \tag{10}$$

In this case the joint posterior distribution is given by

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \boldsymbol{\Delta}, \mathbf{b}, \mathbf{w}|\mathbf{y}^*) \propto \prod_{i=1}^{n} p(\mathbf{y}^*_i|\boldsymbol{\beta}, \mathbf{R}_i, \boldsymbol{\Delta}, \mathbf{b}_i) \prod_{i=1}^{n} p(\boldsymbol{\beta})p(\mathbf{R}_i)p(\mathbf{b}_i|\boldsymbol{\Delta}, w_i)p(\boldsymbol{\Delta}|wi)p(w_i) \tag{11}$$

Samples of the $\mathbf{t}_q(\mathbf{0}, \boldsymbol{\Delta}, \nu)$ distribution are generated as follows

$$\begin{aligned} w_i &\sim Gamma(\nu/2, \nu/2) \\ \mathbf{b}_i|(\boldsymbol{\Delta}, w_i) &\sim N_q(\mathbf{0}, \boldsymbol{\Delta}/w_i). \end{aligned} \tag{12}$$

Model (9) was fitted considering four distributions for the random effects, namely, M_1: $N_2(\mathbf{0}, \mathbf{G})$, M_2: $t_2(\mathbf{0}, \boldsymbol{\Delta}, \nu = \mathbf{3})$, M_3: $t_2(\mathbf{0}, \boldsymbol{\Delta}, \nu = \mathbf{7})$ and M_4: $t_2(\mathbf{0}, \boldsymbol{\Delta}, \nu = \mathbf{30})$. In all models, locally uniform prior distributions were assigned to the parameters associated with initial milk production, rate of increase/decrease in production before/after peak production and to the parameters associated with the elements of the covariance structure. The full conditional posterior distributions either under Gaussian or multivariate t-distributions for the random effects with an AR(1) covariance structure for the random errors are presented in Appendix.

We used the JAGS software (Plummer, 2015) through the interface available in the *runjags* package (Denwood, In Review) available in the R package (R Core Team, 2014). Three chains of length 20000 with different initial values were generated and the first 10000 iterations (*burn-in*) of each chain were excluded. The estimates were based on the remaining 30000 iterations. In order to evaluate the convergence of the chains for each parameter, traceplots were constructed and the Gelman-Rubin statistic (Gelman & Rubin, 1992) was computed. A laptop computer (Intel Core i5 processor with 8GB of RAM) was used to fit the models and the computing time ranged from 4 to 4.5 hours.

It was not possible to include more than two random effects in the models. Software like WinBUGS, JAGS or STAN use the precision matrix (inverse of the covariance matrix) during the MCMC estimation process and the corresponding matrix with dimension higher than three is not always invertible. This justification was found in online forums for JAGS users and no solution for this limitation was found.

Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), Watanabe-Akaike's Information Criterion (WAIC) (Watanabe, 2010), Leave-One-Out Cross Validation Criterion (LOO_{CV}) (Gelman et al., 2014) and Pseudo-Marginal Likelihood (Ntzoufras, 2011) computed during the MCMC estimation process were used to compare models $M1$-$M4$. Models with smaller values of DIC, WAIC, LOOcv and higher value of LVPM have a better fit.

Estimates and the HPD 95% credibility intervals for the parameters of models M_1-M_4 are presented in Table 2. Estimates associated with the initial milk production and the rates of increase/decrease in production before/after peak production did not differ much across the models, suggesting that the inference about these parameters is robust to the non-normality of random effects. Estimates associated with the elements of the random effects covariance structure differed very little when models M_2 and M_3 are compared with model M_1.

A small decrease in the DIC, WAIC and LOO_{CV} values were observed when models M_2-M_4 are compared with model M_1, indicating an improvement of fit when multivariate t-distributions are assumed for the random effects. The same can be concluded when comparing the LVPM values, since a small increase in the values of this statistic was observed.

5 Discussion

We analyze weekly milk production data of Dutch dairy cows. This type of data is a typical example of repeated measurements data and the linear mixed models are, in general, the standard option in statistical analysis. Evaluate assumptions of the model fitted to data is extremely important since under incorrect specifications the inference is impaired. In particular, when the assumption of normality of either the random effects and/or random errors is not suitable, a Bayesian approach can be considered to fit the model. With this in mind, we developed a hierarchical Bayesian model under the assumption that the random effects follow a multivariate t-distribution. In addition we consider a first order autorregressive correlation structure for the within sample units covariance matrix, more suitable to longitudinal data.

Our results showed that under the assumption of a multivariate t-distribution for the random effects there was a improvement of fit to the data. The estimates of the fixed effects did present considerable differences among models but there was an improvement in the estimates of the covariance structure elements.

Table 2: Estimates and 95% credibility HPD intervals for the parameters of models M_1-M_4

Interpretation	Parameter	$M1: \mathbf{b} \sim Gaussian$	$M2: \mathbf{b} \sim t_{\nu=3}$	$M3: \mathbf{b} \sim t_{\nu=7}$	$M4: \mathbf{b} \sim t_{\nu=30}$
Initial milk production	$^*\beta_1$	14.77 (13.37; 16.15)	15.08 (13.69; 16.44)	14.91 (13.63; 16.28)	14.75 (13.30; 16.25)
	$^*\beta_2$	1.17 (1.05; 1.29)	1.16 (1.05; 1.27)	1.17 (1.06; 1.28)	1.17 (1.06; 1.30)
	$^*\beta_3$	1.30 (1.16; 1.46)	1.32 (1.18; 1.47)	1.32 (1.18; 1.46)	1.31 (1.16; 1.48)
	$^*\beta_4$	1.45 (1.29; 1.62)	1.47 (1.31; 1.63)	1.46 (1.30; 1.63)	1.46 (1.30; 1.63)
Rate of increase before peak production	β_7	0.31 (0.246; 0.357)	0.31 (0.25; 0.36)	0.31 (0.25; 0.37)	0.31 (0.25; 0.37)
	β_8	0.13 (0.05; 0.21)	0.13 (0.05; 0.21)	0.13 (0.04; 0.21)	0.13 (0.05; 0.21)
Rate of decrease after peak production	β_{13}	-0.06 (-0.073; -0.047)	-0.06 (-0.07; -0.05)	-0.06 (-0.07; -0.05)	-0.06 (-0.07; -0.05)
	β_{14}	-0.06 (-0.08; -0.04)	-0.06 (-0.08; -0.04)	-0.06 (-0.08; -0.04)	-0.06 (-0.08; -0.04)
	β_{15}	-0.03 (-0.05; -0.02)	-0.03 (-0.04; -0.02)	-0.03 (-0.04; -0.02)	-0.03 (-0.05; -0.02)
	β_{16}	-0.03 (-0.04; -0.01)	-0.03 (-0.04; -0.01)	-0.03 (-0.04; -0.01)	-0.03 (-0.04; -0.01)
	β_{17}	0.02 (0.00; 0.04)	0.02 (0.01; 0.04)	0.02 (0.00; 0.04)	0.02 (0.00; 0.04)
	β_{18}	0.01 (-0.01; 0.03)	0.01 (-0.00; 0.03)	0.01 (-0.01; 0.03)	0.01 (-0.01; 0.03)
Covariance structure elements	σ_e	0.22 (0.21; 0.24)	0.22 (0.21; 0.247)	0.22 (0.21; 0.24)	0.22 (0.21; 0.24)
	σ_{b1} (δ_{b1})	0.23 (0.19; 0.28)	0.16 (0.12; 0.20)	0.20 (0.15; 0.24)	0.22 (0.18; 0.27)
	σ_{b3} (δ_{b3})	0.02 (0.01; 0.02)	0.01 (0.01; 0.02)	0.01 (0.01; 0.02)	0.02 (0.01; 0.02)
	$\sigma_{b1.b3}$ ($\delta_{b1.b3}$)	-0.06 (-0.46; 0.31)	-0.04 (-0.43; 0.32)	-0.01 (-0.39; 0.37)	-0.02 (-0.40; 0.32)
	ρ	0.50 (0.44; 0.56)	0.50 (0.44; 0.56)	0.49 (0.44; 0.56)	0.50 (0.44; 0.56)
Goodness of fit	LVPM	2273.2	2375.0	2282.4	2276.8
	DIC	-4889.7	-4899.2	-4898.1	-4892.8
	WAIC	-4647.7	-4671.2	-4661.8	-4652.3
	LOO_{CV}	-4648.9	-4665.1	-4661.0	-4654.8

*: computed as $exp(\beta_i)$

Appendix

Results based on multivariate Gaussian distributions for the random effects

Let $\mathbf{y}^* = [\mathbf{y}_1^{*\top}, \ldots, \mathbf{y}_n^{*\top}]^\top$, $\mathbf{X}^* = [\mathbf{X}_1^{*\top}, \ldots, \mathbf{X}_n^{*\top}]^\top$, $\mathbf{Z} = \oplus \mathbf{Z}_i$, $\mathbf{b} = [\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top]^\top$, $\mathbf{R}_i = \mathbf{R}(\sigma^2, \rho) = \sigma^2 \mathbf{C}_i$ where $\mathbf{C}_i$ is defined in (4) and $\mathbf{G} = \mathbf{G}(\sigma^2_{b1}, \sigma^2_{b2}, \sigma_{b1b2})$. Also consider the joint posterior distribution of the model (8) given by

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}, \mathbf{b}|\mathbf{y}^*) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{b}_i) p(\mathbf{b}_i|\mathbf{G}) \prod_{i=1}^{n} p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}). \tag{A1}$$

where

$$p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{b}_i) \propto \frac{1}{\sigma^2} |\mathbf{C}_i|^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_i^* - \mathbf{X}_i^*\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \mathbf{X}_i^*\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\right], \tag{A2}$$

$$p(\mathbf{b}_i|\mathbf{G}) \propto |\mathbf{G}|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{b}_i^\top \mathbf{G}^{-1}\mathbf{b}_i\right], \tag{A3}$$

$p(\boldsymbol{\beta})$, $p(\mathbf{R}_i) = p(\sigma^2, \rho)$, $p(\mathbf{G}) = p(\sigma^2_{b1}, \sigma^2_{b3}, \sigma_{b1b3})$ and $p(\mathbf{b}_i|\mathbf{G})$ are the prior distributions for $\boldsymbol{\beta}$, $\mathbf{R}_i$, $\mathbf{G}$ e $\mathbf{b}_i$, respectively. We consider independent uniform prior distributions for $\boldsymbol{\beta}$, $\mathbf{R}_i$ as well as for the elements of $\mathbf{G}$.

1) Full conditional distribution for $\boldsymbol{\beta}$: $p(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{b}, \mathbf{y}^*)$.

   Given $\sigma^2, \rho, \mathbf{b}$ and $\mathbf{y}*$, it follows that $p(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{b}, \mathbf{y}^*)$ is proportional to the product of independent Gaussian distributions with unknown means and known variances.

$$p(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{b}, \mathbf{y}^*) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{b}_i) p(\boldsymbol{\beta}) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{b}_i) \tag{A4}$$

   We want to find the posterior distribution of $\boldsymbol{\beta}$, vector of parameters of the regression of $\mathbf{y}^* = \mathbf{y} - \mathbf{Z}\mathbf{b}$ in $\mathbf{X}^*$. We know that the likelihood estimator of $\boldsymbol{\beta}$, namely, $\widehat{\boldsymbol{\beta}}_{MV}$, is a sufficient statistic for $\boldsymbol{\beta}$ and follows a $N_p[\boldsymbol{\beta}, Var(\widehat{\boldsymbol{\beta}}_{MV})]$ distribution. Following Lemma 1.4.1 in Box & Tiao (1973), we have

$$p(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{b}, \mathbf{y}) \propto N_p[\widehat{\boldsymbol{\beta}}_{MV}, Var(\widehat{\boldsymbol{\beta}}_{MV})] \tag{A5}$$

   where

$$\widehat{\boldsymbol{\beta}}_{MV} = \left(\sum_{i=1}^{n} \mathbf{X}_i^{*\top} \mathbf{R}_i^{-1} \mathbf{X}_i^*\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^{*\top} \mathbf{R}_i^{-1} \mathbf{y}_i^* \tag{A6}$$

   and

$$Var(\widehat{\boldsymbol{\beta}}_{MV}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^{*\top} \mathbf{R}_i^{-1} \mathbf{X}_i^*\right)^{-1} \tag{A7}$$

2) Full conditional distribution of σ^2: $p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*)$.

Given $\boldsymbol{\beta}, \rho, \mathbf{b}$, $\mathbf{y}^*$ and letting $\boldsymbol{\theta}_i = \mathbf{X}_i^* \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, we have

$$\begin{aligned} p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*) &\propto \prod_{i=1}^n p(\mathbf{y}_i^*|\boldsymbol{\beta}, \rho, \mathbf{b}_i) p(\sigma^2) p(\rho) \propto \prod_{i=1}^n p(\mathbf{y}_i^*|\boldsymbol{\beta}, \rho, \mathbf{b}_i) \\ &= \prod_{i=1}^n \left(\frac{1}{\sigma^2}\right)^{1/2} |\mathbf{C}_i|^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)\right] \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (\mathbf{y}_i^* - \boldsymbol{\theta}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)\right] \\ &= \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[\frac{-\sum_{i=1}^n (\mathbf{y}_i^* - \boldsymbol{\theta}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)}{2\sigma^2}\right] \end{aligned} \tag{A8}$$

From (A8), disregarding a constant term, $1/\sigma^2$ follows a Gamma distribution. Applying the Jacobian method, it follows that

$$p(1/\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*) \propto p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*) \left|\frac{\partial\,(1/\sigma^2)}{\partial\,(1/\sigma^2)}\right| \tag{A9}$$

where $\sigma^2 = g(1/\sigma^2) = (1/\sigma^2)^{-1}$. Hence

$$\begin{aligned} p(1/\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*) &\propto p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*) \left(1/\sigma^2\right)^{-2} \\ &= \left(\frac{1}{\sigma^2}\right)^{(n/2-1)-1} \exp\left[\frac{-\sum_{i=1}^n (\mathbf{y}_i^* - \boldsymbol{\theta}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)}{2\sigma^2}\right] \end{aligned} \tag{A10}$$

that is a *Gamma* $\left[n/2 - 1, 2/\sum_{i=1}^n (\mathbf{y}_i^* - \boldsymbol{\theta}_i)^\top \mathbf{C}_i^{-1}(\mathbf{y}_i^* - \boldsymbol{\theta}_i)\right]$ distribution. Inverting the values obtained from $p(1/\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*)$ is equivalent to sampling values from $p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*)$.

3) Full conditional distribution of $\mathbf{G}$: $p(\mathbf{G}|\mathbf{b})$

Given $\mathbf{b}$, we have

$$\begin{aligned} p(\mathbf{G}|\mathbf{b}) &\propto \prod_{i=1}^n p(\mathbf{b}_i|\mathbf{G}) p(\mathbf{G}) \propto \prod_{i=1}^n |\mathbf{G}^{-1}|^{1/2} \exp\left[-\frac{1}{2}\mathbf{b}_i^\top \mathbf{G}^{-1}\mathbf{b}_i\right] \\ &= |\mathbf{G}^{-1}|^{n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{G}^{-1}\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\top\right)\right] \end{aligned} \tag{A11}$$

where $\mathrm{tr}(\mathbf{A})$ represents the trace of $\mathbf{A}$. From (A11), disregarding a constant term, we have that $\mathbf{G}^{-1}$ follows a Wishart distribution. Applying the Jacobian method,

$$p(\mathbf{G}^{-1}|\mathbf{b}) \propto p(\mathbf{G}|\mathbf{b}) \left|\frac{\partial g(\mathbf{G}^{-1})}{\partial\,(\mathbf{G}^{-1})}\right| \tag{A12}$$

From Box and Tiao (1973, p.426), it follows that

$$\left|\frac{\partial g(\mathbf{G}^{-1})}{\partial\left(\mathbf{G}^{-1}\right)}\right| = \left|\mathbf{G}^{-1}\right|^{-(q+1)}. \tag{A13}$$

Hence

$$\begin{aligned} p(\mathbf{G}^{-1}|\mathbf{b}) &\propto p(\mathbf{G}|\mathbf{b})\left|\mathbf{G}^{-1}\right|^{-(q+1)} = |\mathbf{G}^{-1}|^{\frac{n}{2}-q-1}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{G}^{-1}\sum_{i=1}^{n}\mathbf{b}_i\mathbf{b}_i^\top\right)\right] \\ &= |\mathbf{G}^{-1}|^{\frac{1}{2}[(n-q-1)-q-1]}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{G}^{-1}\sum_{i=1}^{n}\mathbf{b}_i\mathbf{b}_i^\top\right)\right], \end{aligned} \tag{A14}$$

a $Wishart\left(n-q-1, \sum_{i=1}^{n}\mathbf{b}_i\mathbf{b}_i^\top\right)$ distribution. Note that the process of inverting the values of $p(\mathbf{G}^{-1}|\mathbf{b},\mathbf{y}^*)$ is equivalent to sampling from $p(\mathbf{G}|\mathbf{b},\mathbf{y}^*)$.

4) Full conditional distribution for $\mathbf{b}$: $p(\mathbf{b}|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{G},\mathbf{y}^*)$.

   Given $\boldsymbol{\beta},\sigma^2,\rho,\mathbf{G}$ e $\mathbf{y}^*$, we have

$$\begin{aligned} p(\mathbf{b}|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{G},\mathbf{y}^*) &\propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{b}_i)p(\mathbf{b}_i|\mathbf{G}) \\ &\propto \prod_{i=1}^{n}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_i^*-\mathbf{X}_i^*\boldsymbol{\beta}-\mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{C}_i^{-1}(\mathbf{y}_i^*-\mathbf{X}_i^*\boldsymbol{\beta}-\mathbf{Z}_i\mathbf{b}_i)\right] \\ &\quad \times\exp\left[-\frac{1}{2}\mathbf{b}_i^\top\mathbf{G}^{-1}\mathbf{b}_i\right] \end{aligned} \tag{A15}$$

   which an unspecified distribution.

5) Full conditional distribution for ρ: $p(\rho|\boldsymbol{\beta},\sigma^2,\mathbf{b},\mathbf{y}^*)$.

   Given $\boldsymbol{\beta},\sigma^2,\mathbf{b}$ and $\mathbf{y}^*$, we have

$$\begin{aligned} p(\rho|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{b},\mathbf{y}^*) &\propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{b}_i)p(\rho) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta},\sigma^2,\mathbf{b}_i) \\ &= \prod_{i=1}^{n}\left(\frac{1}{\sigma^2}\right)^{1/2}|\mathbf{C}_i|^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_i^*-\boldsymbol{\theta}_i)^\top\mathbf{C}_i^{-1}(\mathbf{y}_i^*-\boldsymbol{\theta}_i)\right] \\ &\propto \prod_{i=1}^{n}|\mathbf{C}_i|^{-1/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i^*-\boldsymbol{\theta}_i)^\top\mathbf{C}_i^{-1}(\mathbf{y}_i^*-\boldsymbol{\theta}_i)\right] \end{aligned} \tag{A16}$$

   which has an unspecified distribution.

The MCMC algorithm to fit the proposed models is:

a) Draw $\boldsymbol{\beta}$ from $p^{(t)}(\boldsymbol{\beta}|\sigma^{2(t-1)}, \rho^{(t-1)}, \mathbf{b}^{(t-1)}, \mathbf{y}^*)$ defined in (A5).

b) Draw $1/\sigma^2$ from $p^{(t)}(\sigma^2|\rho^{(t-1)}, \mathbf{b}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{y}^*)$ defined in (A10).

c) Draw $\mathbf{G}^{-1}$ from $p^{(t)}(\mathbf{G}^{-1}|\mathbf{b}^{(t-1)})$ defined in (A14).

d) Draw $\mathbf{b}$ from $p^{(t)}(\mathbf{b}|\boldsymbol{\beta}^{(t)}, \sigma^{2(t)}, \mathbf{G}^{(t)}, \rho^{(t-1)}, \mathbf{y}^*)$ defined in (A15), using one of the following three algorithms: Rejection-adaptive, Sliced Sampler or Metropolis-Hastings.

e) Draw ρ from $p^{(t)}(\rho|\mathbf{b}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma^{2(t)}, \mathbf{y}^*)$ defined in (A16), using one of the aforementioned algorithms.

Results based on multivariate t distributions for the random effects

Let $\mathbf{y}^* = [\mathbf{y}_1^{*\top}, \dots, \mathbf{y}_n^{*\top}]^\top$, $\mathbf{X}^* = [\mathbf{X}_1^{*\top}, \dots, \mathbf{X}_n^{*\top}]^\top$, $\mathbf{Z} = \oplus \mathbf{Z}_i$, $\mathbf{b} = [\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top]^\top$, $\mathbf{R}_i = \mathbf{R}(\sigma^2, \rho) = \sigma^2 \mathbf{C}_i$ where $\mathbf{C}_i$ is defined in (4) and $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\delta_{b_1}^2, \delta_{b_1 b_3}, \delta_{b_3}^2)$. Also consider the joint posterior distribution of the model (11) given by

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \boldsymbol{\Delta}, \mathbf{b}|\mathbf{y}^*) \propto \prod_{i=1}^{n} p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{b}_i) p(\mathbf{b}_i|\boldsymbol{\Delta}, w_i) \prod_{i=1}^{n} p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\boldsymbol{\Delta}|w_i) p(w_i). \tag{A17}$$

where

$$p(\mathbf{y}_i^*|\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{b}_i) \propto \frac{1}{\sigma^2} |\mathbf{C}_i|^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^\top \mathbf{C}_i^{-1} (\mathbf{y}_i^* - \mathbf{X}_i^* \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)\right], \tag{A18}$$

$$p(\mathbf{b}_i|\boldsymbol{\Delta}, w_i) \propto w_i^{q/2} |\boldsymbol{\Delta}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{b}_i^\top \boldsymbol{\Delta}^{-1} w_i \mathbf{b}_i\right], \tag{A19}$$

$p(\boldsymbol{\beta})$, $p(\mathbf{R}_i) = p(\sigma^2, \rho)$, $p(\boldsymbol{\Delta}) = p(\delta_{b1}^2, \delta_{b3}^2, \delta_{b1b3})$ and $p(\mathbf{b}_i|\boldsymbol{\Delta}, w_i)$ are the prior distributions for $\boldsymbol{\beta}$, $\mathbf{R}_i$, $\boldsymbol{\Delta}$ e $\mathbf{b}_i$, respectively. We consider independent uniform prior distributions for $\boldsymbol{\beta}$, $\mathbf{R}_i$ as well as for the elements of $\boldsymbol{\Delta}$. Like in (12), $p(w_i) \sim Gamma(\nu/2, \nu/2)$.

1) Full conditional distribution for $\boldsymbol{\beta}$: $p(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{b}, \mathbf{y}^*)$.

   The form of this distribution is identical to (A5).

2) Full conditional distribution of σ^2: $p(\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{b}, \mathbf{y}^*)$.

   The form of this distribution is identical to (A10).

3) Full conditional distribution of $\boldsymbol{\Delta}$: $p(\boldsymbol{\Delta}|\mathbf{b}, w_i)$

In this case, (A14) is expressed as

$$|\boldsymbol{\Delta}^{-1}|^{\frac{1}{2}((n-q-1)-q-1)} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Delta}^{-1}\sum_{i=1}^{n} w_i \mathbf{b}_i \mathbf{b}_i^{\top}\right)\right], \tag{A20}$$

a $Wishart\left(n-q-1, \sum_{i=1}^{n} w_i \mathbf{b}_i \mathbf{b}_i^{\top}\right)$ distribution.

4) Full conditional distribution for $\mathbf{b}$: $p(\mathbf{b}|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{G}, w_i, \mathbf{y}^*)$.

In this case, (A15) is proportional to

$$\prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_i^* - \mathbf{X}_i^*\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^{\top}\mathbf{C}_i^{-1}(\mathbf{y}_i^* - \mathbf{X}_i^*\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\right] \exp\left[-\frac{1}{2}\mathbf{b}_i^{\top}\boldsymbol{\Delta}^{-1}w_i\mathbf{b}_i\right] \tag{A21}$$

which is also an unspecified distribution.

5) Full conditional distribution for ρ: $p(\rho|\boldsymbol{\beta}, \sigma^2, \mathbf{b}, \mathbf{y}^*)$.

The form of this distribution is identical to (A16).

6) Full conditional distribution for $\mathbf{w}$: $p(\mathbf{w}|\mathbf{b}, \boldsymbol{\Delta}, \mathbf{y}^*)$.

Given $\mathbf{b}$ e $\boldsymbol{\Delta}$, we have

$$\begin{aligned} p(\mathbf{w}|\mathbf{b}, \boldsymbol{\Delta}) &\propto \prod_{i=1}^{n} p(\mathbf{b}_i|\boldsymbol{\Delta}, w_i)p(w_i) \\ &= \prod_{i=1}^{n} w_i^{q/2} \exp\left[-\frac{1}{2}\mathbf{b}_i^{\top}\boldsymbol{\Delta}_i^{-1}w_i\mathbf{b}_i\right] w_i^{\nu/2-1} \exp\left[-\frac{\nu}{2}w_i\right] \\ &= \prod_{i=1}^{n} w_i^{(q/2+\nu/2)-1} \exp\left[-\frac{w_i}{2}\left(\mathbf{b}_i^{\top}\boldsymbol{\Delta}_i^{-1}\mathbf{b}_i + \nu\right)\right] \\ &= \prod_{i=1}^{n} w_i^{(q/2+\nu/2)-1} \exp\left[-\frac{w_i}{2/\mathbf{b}_i^{\top}\boldsymbol{\Delta}_i^{-1}\mathbf{b}_i + \nu}\right], \end{aligned} \tag{A22}$$

a $Gamma\left[(\nu+q)/2, 2/(\mathbf{b}_i^{\top}\boldsymbol{\Delta}_i^{-1}\mathbf{b}_i + \nu)\right]$ distribution.

The MCMC algorithm to fit the proposed models is:

a) Draw $\boldsymbol{\beta}$ from $p^{(t)}(\boldsymbol{\beta}|\sigma^{2^{(t-1)}}, \rho^{(t-1)}, \mathbf{b}^{(t-1)}, \mathbf{y}^*)$ defined in (A5).

b) Draw $1/\sigma^2$ from $p^{(t)}(\sigma^2|\rho^{(t-1)}, \mathbf{b}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{y}^*)$ defined in (A10).

c) Draw $\boldsymbol{\Delta}^{-1}$ from $p^{(t)}(\boldsymbol{\Delta}^{-1}|\mathbf{b}^{(t-1)}, w_i^{(t-1)})$ defined in (A20).

d) Draw $\mathbf{b}$ from $p^{(t)}(\mathbf{b}|\boldsymbol{\beta}^{(t)}, \sigma^{2^{(t)}}, \boldsymbol{\Delta}^{(t)}, \rho^{(t-1)}, w_i^{(t-1)}, \mathbf{y}^*)$ defined in (A21), using one of the following three algorithms: Rejection-adaptive, Sliced Sampler or Metropolis-Hastings.

e) Draw ρ from $p^{(t)}(\rho|\mathbf{b}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma^{2^{(t)}}, \mathbf{y}^*)$ defined in (A16), using one of the aforementioned algorithms.

f) Draw $\mathbf{w}$ from $p^{(t)}(\mathbf{w}|\mathbf{b}^{(t)}, \boldsymbol{\Delta}^{(t)})$ defined in (A22).

References

Arellano-Valle, R., Bolfarine, H. and Lachos, V. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* **34**, 663–682.

Box, G. E. and Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading: Addison-Wesley Series.

Congdon, P. D. (2010). *Applied Bayesian hierarchical methods*. Boca Raton : CRC Press.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. New Jersey: John Wiley & Sons.

Denwood, M. J. (In Review). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*.

Dey, D. K., Ghosh, S. K. and Mallick, B. K. (2000). *Generalized linear models: A Bayesian perspective*. New York : Marcel Dekker.

Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62**, 89–100.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo stochastic simulation for Bayesian inference*. Boca Raton: Taylor & Francis.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation usinf multiple sequences. *Statistical Science* **7**, 457–511.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2014). *Bayesian data analysis*. London: Chapman & Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **PAMI-6**, 721–741.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Henderson, C. R. (1984). *Applications of linear models in animal breeding*. Ontario: University of Guelph Press.

Jara, A., Quintana, F. and San Martín, E. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Computational statistics & data analysis* **52**, 5033–5045.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lee, J. C. and Lien, W. (2001). Bayesian analysis of a growth curve model with power transformation, random effects and AR (1) dependence. *Journal of Applied Statistics* **28**, 223–238.

Lee, J. C., Lin, T. I., Lee, K. J. and Hsu, Y. L. (2005). Bayesian analysis of Box–Cox transformed linear mixed models with ARMA (p, q) dependence. *Journal of statistical Planning and Inference* **133**, 435–451.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*. West Sussex: John Wiley & Sons.

Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* **137**, 484–495.

Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 1–41.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092.

Neal, R. M. (1997). Markov chain Monte Carlo methods based on 'slicing' the density function. Technical report. Technical Report 9722, University of Toronto, Department of Statistics.

Nobre, J. S. and Singer, J. M. (2007). Residual Analysis for Linear Mixed Models. *Biometrical Journal* **38**, 1063–1072.

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*. Vol. 698. New Jersey: John Wiley & Sons.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-119.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Plummer, M. (2015). *JAGS Version 3.1. 0 user manual*.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rosa, G., Gianola, D. and Padovani, C. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics* **31**, 855–873.

Rosa, G., Padovani, C. R. and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal* **45**, 573–590.

Seltzer, M. H. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational and Behavioral Statistics* **18**, 207–235.

Seltzer, M. H., Wong, W. H. and Bryk, A. S. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics* **21**, 131–167.

Singer, J. M., Rocha, F. M. M. and Nobre, J. S. (2017). Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* **85**, 290–324.

Sorensen, D. e Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York: Springer.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS user manual .

Stan Development Team (2014). *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0.*

Strandén, I. and Gianola, D. (1998). Mixed effects linear models with t-distributions for quantitative genetic analysis: a Bayesian approach. **25**, 25–42.

Verbeke, G. and Lesaffre, E. (1987). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541–556.

Wakefield, J., Smith, A., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics* **43**, 201–221.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research* **11**, 3571–3594.

Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature* **216**, 164–165.