Bayesian Inference in Nonlinear Mixed–Effects Models Using Normal Independent Distributions

Victor H. Lachos*[a], Luis M. Castro[b], Dipak K. Dey[c],

[a]Department of Statistics, Campinas State University, Brazil
[b]Departament of Statistics, Universidad de Concepción, Chile
[c]Department of Statistics, University of Connecticut, USA

Abstract

Nonlinear mixed-effects (NLME) models are popular in many longitudinal studies, including human immunodeficiency virus (HIV) viral dynamics, pharmacokinetic analyses, and studies of growth and decay. Generally, the normality of the random effects is a common assumption in NLME models but it may, sometimes, be unrealistic, obscuring important features of among-subjects variation. In this article, we utilize normal/independent distributions as a tool for robust modeling of NLME models under a Bayesian paradigm. The normal/independent distributions is an attractive class of symmetric heavy-tailed distributions that includes the normal distribution, the generalized Student-t, Student-t, slash and the contaminated normal distributions as special cases, providing an appealing robust alternative to the routine use of normal distributions in this type of models. In order to examine the robust aspects of this flexible class, against outlying and influential observations, we present a Bayesian case deletion influence diagnostics based on the q-divergence measure. Further, some discussions on model selection criteria are given. These analysis are computationally possible due to an important result that approximating the likelihood function of a NLME model with normal/independent distributions for a simple normal/independent distribution with specified parameters. The new methodologies are exemplified through simulated and a real data set of AIDS/HIV infected patients that was initially analyzed using a normal NLME model, illustrating the usefulness of the proposed methodology.

Key words: Gibbs Algorithms; MCMC; Metropolis-Hastings; Nonlinear mixed-effects models; Normal/independent distributions.

1. Introduction

Mixed-effects models have become a popular tool for the analysis of grouped data that arise in many areas as diverse as clinical trials, epidemiology, and sociology. Examples of grouped data include longitudinal data, repeated measures, and multilevel data. Linear mixed-effects (LME) (Laird & H.Ware, 1982) and nonlinear mixed-effects (NLME) models (Pinheiro & Bates, 2000) are typically used to describe grouped data for which the random effects and the error term are assumed

* *Address for correspondence*: Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: hlachos@ime.unicamp.br.

to follow a normal distribution. But normal Linear mixed-effects (N-LME) and normal nonlinear mixed-effects (N-NLME) models suffer from the same lack of robustness against departures from distributional assumptions as other statistical models based on the Gaussian distribution and may be too restrictive to provide an accurate representation of the structure that is present in the data. To deal with this problem, some proposals have been made in the literature by replacing the assumption of normality by a class of elliptical distributions that cover both light-and heavy-tailed distributions such as Student-t, logistic and exponential power family. Bayesian works for nonlinear models (NLM) with elliptical distributions for the error term can be found, for instance, in Osiewalski & Steel (1993), Osiewalski (1999). For LME models, Rosa *et al.* (2003) advocate the use of a subclass of elliptical distributions, called normal/independent (NI) distributions (Liu, 1996) and adopted a Bayesian framework to carry out posterior analysis for thick–tailed LME models. Interestingly, this rich class includes distributions such as the generalized Student-t (GT), Student-t (T), slash (SL) and the contaminated normal (CN) ones. From a frequentist perspective, recent references in LME models with elliptical distributions are the works by Savalli *et al.* (2006) and Osorio *et al.* (2007). For NLME models with elliptical distributions, the work by Russo *et al.* (2009) can be cited with the limitation that the nonlinearity is incorporated only in the fixed-effects (see also, Meza *et al.*, 2012). Although, some works to NLME models with elliptical distributions has recently appeared in the literature, there are no studies on Bayesian inference for NLME models in the elliptical family and certainly not in the NI class. In this paper we propose a robust parametric modeling of NLME models based on NI distributions so that the normal/independent nonlinear mixed effects (NI-NLME) model is defined and a fully Bayesian approach considering the MCMC method is developed to carry out posterior analysis.

On the other hand, after fitting the model, it is important to check the model assumptions and conduct sensitivity analysis to detect possible influential or extreme observations that can cause distortions on the results of the analysis. Following the pioneering work by Cook (1986), case-deletion and local influence diagnostics have been widely applied to many regression models in order to assess the effect of perturbations in the model and/or the data on the parameter estimates. Several authors have applied these methods to nonlinear regression models different than the normal case; see Galea *et al.* (2005). However, to the best of our knowledge, there are neither studies on Bayesian inference for NLME models in the NI family and nor on influence diagnostics related to this topic. Thus, we believe that the research to develop statistical tools with nonstandard assumptions in NLME models is a significant contribution to the field. In this paper, we discuss influence diagnostic from a Bayesian perspective where the objective is to develop diagnostic measures based on the q-divergence measures as proposed by Csiszár (1967), Peng & Dey (1995) and more recently by Vidal & Castro (2010). This analysis is computationally possible due to an important result that approximating the likelihood function of a NI-NLME model for a simple NI distribution with specified parameters.

The rest of the paper is organized as follows. In Section 2, after a brief introduction of NI distributions, the NI–NLME model is presented as well as some inferential results. In Section 3, we carry out Bayesian inference for NI–NLME models. Further, some measures for Bayesian model selection are discussed. In Section 4 we introduce Bayesian case influence diagnostics based on the q-divergence measure. In Section 5, a simulation study is conducted in order to illustrate the

performance of our proposed methods and the robust aspect of this flexible class against outlying and influential observations. The advantage of the proposed methodology is illustrated through the analysis of a longitudinal HIV viral load in AIDS study in Section 6. Finally, a brief discussion is given in Section 7.

2. Nonlinear mixed-effects models with NI distributions

2.1. Normal/independent distributions

An element of the normal/independent family (Lange & Sinsheimer, 1993; Liu, 1996; Rosa *et al.*, 2003) is defined as the distribution of the p-variate random vector

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{1}$$

where $\boldsymbol{\mu}$ is a location vector, $\mathbf{Z}$ is a normal random vector with mean vector $\mathbf{0}$, variance–covariance matrix $\boldsymbol{\Sigma}$ and U is a mixing positive random variable with cumulative distribution function (cdf) $H(u|\boldsymbol{\nu})$ and probability density function (pdf) $h(u|\boldsymbol{\nu})$, independent of $\mathbf{Z}$, where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U. Note that given U, $\mathbf{Y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $u^{-1}\boldsymbol{\Sigma}$. In other words, the NI distributions are scale mixtures of the normal distribution, where the distribution of the scale factor U is the mixing distribution. Hence, the pdf of $\mathbf{Y}$ is given by $\mathrm{NI}(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) = \int_0^\infty \phi_p(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})dH(u|\boldsymbol{\nu})$, where $\phi_p(.;\boldsymbol{\mu},\boldsymbol{\Sigma})$ stands for the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$. We use the notation $\mathrm{NI}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},H)$ when $\mathbf{Y}$ has distribution in the NI class. Three scale mixtures of multivariate normal distributions are commonly used for robust estimations: the multivariate Generalized Student-t, multivariate slash and multivariate contaminated multivariate normal distributions.

- *The multivariate generalized Student–t distribution*, $Gt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2)$, where ν is called the degrees of freedom, can be derived from the mixture model (1), where U is distributed as $Gamma(\nu_1/2,\nu_2/2)$, with $\nu_l > 0,\ l = 1,2$. The pdf of $\mathbf{Y}$ takes the following form:

$$GT(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) = \frac{\Gamma(\frac{p+\nu_1}{2})}{\Gamma(\frac{\nu_1}{2})\pi^{p/2}}\nu_2^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{d}{\nu_2}\right)^{-(p+\nu_1)/2}, \quad \mathbf{y}\in\mathbb{R}^p,$$

  where $\Gamma(.)$ is the standard gamma function. Particular cases of the generalized Student-t distribution are the Student-t (T) distribution when $\nu_1 = \nu_2 = \nu$ and the Cauchy one, when $\nu_1 = \nu_2 = 1$. Also, when $\nu_1,\nu_2 \to \infty$, one gets the normal distribution as the limiting case.

- *The multivariate slash distribution*, $SL_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$, arises when the distribution of U is $Beta(\nu,1)$, with $u\in(0,1)$ and $\nu > 0$. Its pdf is given by

$$SL(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \nu\int_0^1 u^{\nu-1}\phi_p(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})du, \quad \mathbf{y}\in\mathbb{R}^p$$

  and can be evaluated using the R function *integrate* (R Development Core Team, 2009), for example. The slash distribution reduces to the normal distribution when $\nu\to\infty$. For an overview and recent discussion of this distribution we refer Gomez *et al.* (2007).

- *The multivariate contaminated normal distribution*, $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \gamma)$, where $\nu, \gamma \in (0, 1)$. Here, U is a discrete random variable taking one of two states and with probability function given by

$$h(u|\boldsymbol{\nu}) = \nu \mathbb{I}_{\{\gamma\}}(u) + (1-\nu)\mathbb{I}_{\{1\}}(u),$$

where $\boldsymbol{\nu} = (\nu, \gamma)$ and $\mathbb{I}_{\{\tau\}}(.)$ is the indicator function of the set $\{\tau\}$. The associated density is

$$CN(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \nu\phi_p(\mathbf{y}; \boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) + (1-\nu)\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Parameter ν can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor. The contaminated normal distribution reduces to the normal one when $\gamma \to 1$. Applications as well as discussions of this distribution are given, for example, in Lange *et al.* (1989).

2.2. Model specification

We propose the following general mixed-effects model in which the random terms are assumed to follow NI distributions within the class defined in (1). The novel model, denoted by NI-NLME, is defined as follows

$$\mathbf{Y}_i \quad = \quad \eta(\boldsymbol{\phi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \;\; \boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \;\; \text{with} \tag{2}$$

$$(\mathbf{b}_i, \boldsymbol{\epsilon}_i) \overset{ind.}{\sim} \mathrm{NI}_{n_i+q}\left(0, \mathrm{Diag}(\mathbf{D}, \sigma_e^2\mathbf{I}_{ni}), H\right), \;\; i = 1, \ldots, n, \tag{3}$$

where the subscript i, denotes the subject index, $\mathbf{Y}_i = (y_{i1}, \cdots, y_{in_i})^\top$ is a $n_i \times 1$ vector of observed n_i continuous responses for subject i, η represents a nonlinear differentiable function of a vector-valued mixed-effects parameters $\boldsymbol{\phi}_i$, $\mathbf{X}_i$ is a $n_i \times r$ matrix of covariates, $\mathbf{A}_i$ is a design matrix that possibly depends on elements of $\mathbf{X}_i$ and $\mathbf{B}_i$ is a design matrix allowing to incorporate, for instance, "time-varying" covariates in the random effects. $\mathrm{Diag}(\mathbf{R}, \mathbf{S})$ denotes a block diagonal whose elements are the square matrices $\mathbf{R}$ and $\mathbf{S}$, matrix $\mathbf{I}_p$ denotes the $p \times p$ identity matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\mathbf{b}_i$ is a q-dimensional random effects vector associates with the i-th group and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors. The dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$, e.g., the upper triangular elements in the unstructured case; and $H = H(\cdot|\boldsymbol{\nu})$ is the cdf-generator that determines the specific NI model we are consider. It is important to stress that the model in (2) will be nonlinear when η is a nonlinear function of the individual mixed-effects parameters $\boldsymbol{\phi}_i$.

Using the definition of a NI random vector and (3), it follows that marginally

$$\mathbf{b}_i \overset{ind}{\sim} \mathrm{NI}_q(0, \mathbf{D}, H) \quad \text{and} \quad \boldsymbol{\epsilon}_i \overset{ind.}{\sim} \mathrm{NI}_{n_i}(\mathbf{0}, \sigma_e^2\mathbf{I}_{n_i}, H), \quad i = 1, \ldots, n \tag{4}$$

and they are uncorrelated, since $Cov(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathrm{E}\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top\} = \mathrm{E}\{\mathrm{E}\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top|U_i\}\} = \mathbf{0}$. Our model can be seen as an extension of the elliptical NLME model proposed by Russo *et al.* (2009), where the nonlinearity is incorporated only in the fixed-effects. Moreover, when the $U_i \overset{iid}{\sim} Gamma(\nu/2, \nu/2)$ and $\eta(.)$ is a linear function of the individual random parameters $\boldsymbol{\phi}_i$, the NI–NLME model reduces to a slight modification of the (hierarchical) Student–t LME model proposed by Pinheiro *et al.* (2001).

A key feature of this model is that it can be formulated in a flexible hierarchical representation that is useful for writing easily BUGS codes and for analytical derivations. It follows from (2) and (4) that

$$\mathbf{Y}_i|\mathbf{b}_i, U_i = u_i \overset{\text{ind.}}{\sim} N_{n_i}(\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i), u_i^{-1}\sigma_e^2\mathbf{I}_{n_i}), \tag{5}$$

$$\mathbf{b}_i|U_i = u_i \overset{\text{ind.}}{\sim} N_q(\mathbf{0}, u_i^{-1}\mathbf{D}), \tag{6}$$

$$U_i \overset{\text{iid.}}{\sim} H(u_i|\boldsymbol{\nu}), \;\; i = 1, \ldots, n. \tag{7}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\nu}^\top)^\top$, then classical inference on the parameter vector $\boldsymbol{\theta}$ is based on the marginal distribution for $\mathbf{Y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)$ (Pinheiro & Bates, 1995), that in this case is given by

$$f(\mathbf{Y}|\boldsymbol{\theta}) \;\; = \;\; \prod_{i=1}^{n} \int_0^\infty \int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i), u_i^{-1}\sigma_e^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i; 0, u_i^{-1}\mathbf{D})d\mathbf{b}_i dH(u_i|\boldsymbol{\nu}),$$

which generally does not have a closed form expression because the model function is not linear in the random effect. In the normal case, various approximations (viz. first-order Taylor series expansion of the model function around the conditional mode of $\mathbf{b}_i$) have been proposed to achieve tractable numerical optimizations (Lindstrom & Bates, 1990). Most algorithms for computing the approximate MLE $\widehat{\boldsymbol{\theta}}$ and empirical Bayes estimators (predictors) for the random effects $\widehat{\mathbf{b}}_i$ considers iterative maximization of the approximate log-likelihood functions $\ell(\boldsymbol{\theta}, \widetilde{\mathbf{b}}) = \sum_{i=1}^n \log f(\mathbf{y}_i|\boldsymbol{\theta}, \widetilde{\mathbf{b}}_i)$. Following Taylor series expansions, we have the following Theorem, whose proof is given in the Appendix A.

Theorem 1. *Let $\widetilde{\mathbf{b}}_i$ be an expansion point in a neighborhood of $\mathbf{b}_i$, then under the NI–NLME model as (2)–(3), the marginal distribution of $\mathbf{Y}_i$, for $i = 1, \ldots, n$, can be approximated as*

$$\mathbf{Y}_i \dot{\sim} NI_{ni}(\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i, \widetilde{\mathbf{V}}_i, \boldsymbol{\nu}), \tag{8}$$

where $\widetilde{\mathbf{V}}_i = \widetilde{\mathbf{H}}_i\mathbf{D}\widetilde{\mathbf{H}}_i^\top + \sigma_e^2\mathbf{I}_{n_i}$, $\widetilde{\mathbf{H}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ and $\dot{\sim}$ denotes distributed approximately.

The estimates obtained by maximizing the approximate log-likelihood functions $\ell(\boldsymbol{\theta}, \widetilde{\mathbf{b}}) = \sum_{i=1}^n \log f(\mathbf{y}_i|\boldsymbol{\theta}, \widetilde{\mathbf{b}}_i)$ are thus approximate maximum likelihood estimates (MLEs). An algorithm to obtain an estimator of $\boldsymbol{\theta}$ based on the result given in Theorem 1 is as follows: Let the estimate of $\boldsymbol{\theta}$ at the k-th iteration be $\widehat{\boldsymbol{\theta}}^{(k)}$, for $k = 0, 1, 2, \ldots$. Then we follow the following steps:

Step 1. With $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$ to obtain the random effect estimates $\widetilde{\mathbf{b}}^{(k)}$,

Step 2. Maximize the approximate joint log-likelihood $\ell(\boldsymbol{\theta}, \widetilde{\mathbf{b}}^{(k)})$ to obtain a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$.

Repeat the two steps until the sequence of $\{\boldsymbol{\theta}^{(k)} \;:\; k = 0, 1, 2, \ldots, \}$ converges. To perform the *step 2*, there are many optimization procedures available in standard softwares, such as *fmincon()* and *optim()* in Matlab and R, respectively. In the k-th iteration, we can use the empirical Bayesian estimates (or minimum mean-squared error) of the random effects $\widetilde{\mathbf{b}}^{(k)}$, which can be approximately obtained by

$$\widetilde{\mathbf{b}}_i^{(k)} = \mathrm{E}\{\mathbf{b}_i|\mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k-1)}\} \approx \widehat{\boldsymbol{\Lambda}}_i^{(k-1)}\widetilde{\mathbf{H}}_i^{(k-1)}\left(\mathbf{y}_i - \eta(\mathbf{A}_i\widehat{\boldsymbol{\beta}}^{(k-1)} + \mathbf{B}_i\widetilde{\mathbf{b}}_i^{(k-1)}, \mathbf{X}_i) + \widetilde{\mathbf{H}}_i^{(k-1)}\widetilde{\mathbf{b}}_i^{(k-1)}\right), \tag{9}$$

where $\widehat{\mathbf{\Lambda}}_i^{(k-1)} = (\widehat{\mathbf{D}}^{(k-1)-1} + \frac{1}{\widehat{\sigma}_e^{2(k-1)}}\widetilde{\mathbf{H}}_i^{(k-1)\top}\widetilde{\mathbf{H}}_i^{(k-1)})^{-1}$. Note that (see Appendix A) the distribution of $\mathbf{b}_i|\mathbf{y}_i$ is approximately symmetric, and thus $\widetilde{\mathbf{b}}_i^{(k)}$ is also the mode of the distribution.

However, there are difficult challenges for likelihood-based inference, such as finding their standard errors in multidimensional problems and having samples of small to moderate size where the asymptotic theory of MLE may not apply. Hence, in this paper, we develop a Bayesian method for inference. Our approach relies on the Markov chain Monte Carlo (MCMC) algorithms to obtain posterior inference for the parameters. It allows for full parameter uncertainty and Bayesian inference does not depend on asymptotic results (see Gelman *et al.*, 2006). Interval estimates for model parameters or functions of model parameters can be easily obtained directly from the MCMC output.

3. Bayesian Inference

3.1. Prior distributions and joint posterior density

In this section, we implement the Bayesian methodology using MCMC techniques for NI-NLME models. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, it follows from (5)–(7) that the complete likelihood function associated with $(\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$, is given by

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{u}) \quad \propto \quad \prod_{i=1}^{n}[\phi_{n_i}(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i), \sigma_e^2 u_i^{-1}\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i; \mathbf{0}, u_i^{-1}\mathbf{D})h(u_i|\boldsymbol{\nu})]. \qquad (10)$$

Now, to complete the Bayesian specification of the model we need to consider prior distributions to all the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\nu}^\top)^\top$. A popular choice to ensure posterior propriety in a LMM is to consider proper (but diffuse) conditionally conjugate priors (Hobert & Casella, 1996). In general, we have $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_\beta)$, $\sigma_e^2 \sim IGamma(\tau_o/2, T_o/2)$ and $\mathbf{D} \sim IWish_q(\mathbf{M}_o, l)$, where $IGamma(a, b)$ is the inverse gamma distribution with mean $b/(a-1)$, $a > 1$, and $IWish_q(\mathbf{M}, l)$ is the inverse Wishart distribution with mean $\mathbf{M}/(l-q-1)$, $l > q+1$, where $\mathbf{M}$ is a $q \times q$ known positive definite matrix. For the specific models, the prior for $\boldsymbol{\nu}$ was chosen accordingly as follows.

i) For the Student-t model. $\nu \sim Texp(\lambda_o/2, (2, \infty))$, where $Texp(\lambda, (a, b))$ denotes the exponential distribution with mean $1/\lambda$, truncated in the interval (a, b). This truncation point was chosen to assure finite variance;

ii) For the slash model. A $Gamma(a, b)$ distribution with small positive values of a and b ($b \ll a$) is adopted as a prior distribution for ν.

iii) For the contaminated normal model. A Beta(ν_0, ν_1) distribution is used as a prior for ν, and an independent Beta(ρ_0, ρ_1) is adopted as prior for γ.

Assuming elements of the parameter vector to be independent we consider that the joint prior distribution of all unknown parameters have density given by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma_e^2)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\nu}), \qquad (11)$$

where $\boldsymbol{\alpha}$ is the upper triangular elements of the unstructured scale matrix $\mathbf{D}$. Combining the likelihood function (10) and the prior distribution (11), the joint posterior density of all unobservable is then:

$$\pi(\boldsymbol{\theta}, \mathbf{b}, \mathbf{u}|\mathbf{y}) \quad \propto \quad \prod_{i=1}^{n} \left[\phi_{n_i}(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i), u_i^{-1}\mathbf{I}_{n_i}\sigma_e^2)\phi_q(\mathbf{b}_i; \mathbf{0}, u_i^{-1}\mathbf{D})h(u_i|\boldsymbol{\nu})\right] \pi(\boldsymbol{\theta}). \quad (12)$$

Distribution (12) is analytically intractable but MCMC methods such as the Gibbs sampler and Metropolis-Hastings algorithm can be used to draw samples, from which features of marginal posterior distribution of interest can be inferred. Our Bayesian model allows us a straightforward construction of a Gibbs sampler through the hierarchical representation (5)-(7). In order to do this, it is necessary to obtain the full conditional posterior distributions, defined by the conditional distribution of one variable given values of all the remaining – $\mathbf{y}$ included. Given $\mathbf{u}$, all conditional posterior distributions are as in a standard N-NLME model and have the same form for any element of the NI family. These are given by

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\mathbf{y}, \mathbf{b}, \mathbf{u}, \boldsymbol{\theta}_{(-\beta)}) &\propto (\sigma_e^2)^{-N/2} \exp\{-\frac{1}{2}(s/\sigma_e^2 + (\boldsymbol{\beta} - \boldsymbol{\beta}_o)^\top S_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_o))\}; \\
\sigma_e^2|\mathbf{y}, \mathbf{b}, \mathbf{u}, \boldsymbol{\theta}_{(-\sigma_e^2)} &\sim IGamma\left(\frac{N + \tau_o}{2}, \frac{T_o + s}{2}\right); \\
\mathbf{D}|\mathbf{y}, \mathbf{b}, \mathbf{u}, \boldsymbol{\theta}_{(-\alpha)} &\sim IWish_q\left((\mathbf{M} + \sum_{i=1}^{n} u_i\mathbf{b}_i\mathbf{b}_i^\top), l + n\right); \\
\pi(\mathbf{b}_i|\mathbf{y}, u_i, \boldsymbol{\theta}) &\propto \phi_{n_i}\left(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i), u_i^{-1}\sigma^2\mathbf{I}_{n_i}\right)\phi_q\left(\mathbf{b}_i; \mathbf{0}, u_i^{-1}\mathbf{D}\right),
\end{aligned}$$

where $s = \sum_{i=1}^{n} u_i\lambda_{1i}$ and $N = \sum_{i=1}^{n} n_i$, with $\lambda_{1i} = (y_i - \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i))^\top(y_i - \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i))$, $i = 1, \ldots, n$.

Note that the full conditional for $\mathbf{b}_i$ and $\boldsymbol{\beta}$ cannot be directly sampled from and the Metropolis-Hasting algorithm within Gibbs iterations needs to be considered (Chib & Greenberg, 1995). To generate samples from $\mathbf{b}_i|\mathbf{y}, u_i, \boldsymbol{\theta}$, $i = 1, \ldots, n$, at the r-th iteration in the chain, we generate a sample $\mathbf{b}_i^*$ from $N_p(\widetilde{\mathbf{b}}_i^{(r)}, \mathbf{V})$, where $\widetilde{\mathbf{b}}_i^{(r)}$ is given in (9), and a random number W from the uniform distribution on the interval $[0, 1]$, then set the new value $\mathbf{b}_i^{(r+1)}$ to be either $\mathbf{b}_i^*$ or $\boldsymbol{\phi}_i^{(r)}$ depending on whether

$$W < \frac{\phi_{n_i}\left(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta}^{(r)} + \mathbf{B}_i\mathbf{b}_i^*, \mathbf{X}_i), u_i^{-1(r)}\sigma^{2(r)}\mathbf{I}_{n_i}\right)\phi_q\left(\mathbf{b}_i^{(r)}; \mathbf{0}, u_i^{-1(r)}\mathbf{D}^{(r)}\right)}{\phi_{n_i}\left(\mathbf{y}_i; \eta(\mathbf{A}_i\boldsymbol{\beta}^{(r)} + \mathbf{B}_i\mathbf{b}_i^{(r)}, \mathbf{X}_i), u_i^{-1(r)}\sigma^{2(r)}\mathbf{I}_{n_i}\right)\phi_q\left(\mathbf{b}_i^{(r)}; \mathbf{0}, u_i^{-1(r)}\mathbf{D}^{(r)}\right)}, \quad (13)$$

or not. The matrix $\mathbf{V}$ is a proposal variance-covariance matrix that could be assumed as $\sigma_p\mathbf{I}_p \equiv (\psi^2/p)\mathbf{I}_p$. The choice of the constant ψ should be done such that ensures an acceptable rejection rate for the Metropolis step $\approx 50\%$ (see Gelman *et al.*, 1996; Gamerman & Lopes, 2006). A similar Metropolis-Hasting algorithm can be implemented to drawn samples from $\boldsymbol{\beta}$.

To complete the Gibbs sampling specifications, we need the full conditional posterior distributions of $\mathbf{u}$ and parameter $\boldsymbol{\nu}$, which depends on the density $h(.|\boldsymbol{\nu})$. The general form for u_i is $\pi(u_i|\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) \propto u_i^{(n_i+q)/2}h(u_i|\boldsymbol{\nu})\exp\{-\frac{u_i}{2}(\frac{\lambda_{1}i}{\sigma_e^2} + \mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i)\}$, $i = 1, \ldots, n$. For $\boldsymbol{\nu}$, the density is $\pi(\boldsymbol{\nu}|\mathbf{y}, \mathbf{u}, \mathbf{b}, \boldsymbol{\theta}_{-\nu}) \propto \pi(\boldsymbol{\nu})\prod_{i=1}^{n} h(u_i|\boldsymbol{\nu})$. The forms of u_i and $\boldsymbol{\nu}$ depends on the specific NI distribution adopted and also on the prior for $\boldsymbol{\nu}$ (see Appendix B).

3.2. Model comparison criteria

There exist a variety of methodologies to compare several competing models for a given data set and to select the one that best fits the data. One of the most used in applied works is derived from the conditional predictive ordinate (CPO) statistic. For a detailed discussion on the CPO statistic and its applications to model selection, see Gelfand *et al.* (1992). Let $\boldsymbol{\mathcal{D}}$ be the full data and $\boldsymbol{\mathcal{D}}^{(-i)}$ denote the data with the i-th observation deleted. We denote the posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{\mathcal{D}}^{(-i)}$ by $\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}^{(-i)})$, for $i = 1, \ldots, n$. For the i-th observation, the CPO_i can be written as

$$CPO_i = \int_{\Theta} g(\mathbf{y}_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}^{(-i)})d\boldsymbol{\theta} = \left\{ \int_{\Theta} \frac{\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})}{g(\mathbf{y}_i|\boldsymbol{\theta})} d\boldsymbol{\theta} \right\}^{-1}, \ i = 1, \ldots, n,$$

where $g(\mathbf{y}_i|\boldsymbol{\theta})$ is the likelihood function of the observed data.

The CPO_i can be interpreted as the height of the marginal density at $\mathbf{y}_i$. Thus, large values of CPO_i imply a better fit of the model. For the proposed model a closed form of the CPO_i is not available. However, a Monte Carlo estimate of the CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$. Let $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q$ be a sample of size Q of $\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$ after the burn-in. A Monte Carlo approximation of the CPO_i (Dey *et al.*, 1997) is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^{Q} \frac{1}{g(\mathbf{y}_i|\boldsymbol{\theta}_q)} \right\}^{-1}.$$

A summary statistic of the CPO_i's is the *log pseudo marginal likelihood*, given by LPML $= \sum_{i=1}^{n} \log(\widehat{CPO}_i)$. The larger values of LMPL indicates better fit. Moreover using LPML, is easy to compute pseudo Bayes factors (PBF) (see, *e.g.*, Geisser & Eddy, 1979; Gelfand & Dey, 1994) for models comparison. Although the harmonic-mean identity provides a convenient and simplified practical implementation of the CPO statistic, it is susceptible to instability (Raftery *et al.*, 2007) for very small values of the likelihood. Although several other alternative approaches (Raftery *et al.*, 2007; Gelfand & Dey, 1994; Dey *et al.*, 1997) have been prescribed, they can be computationally challenging. As suggested by one referee, here we consider a more pragmatic route and compute the CPO (and associated LPML) statistics using 500 non-overlapping blocks of the Markov chain each of size 2000 post-convergence (i.e. after discarding the initial burn-in samples), and report the expected LPML and Monte Carlo standard errors computed over the 500 blocks. If the HM identity is stable, we expect to have small Monte Carlos sd of the LPMLs. Congdon (2005) also suggests that the HM estimate is stable as long as the individual log-likelihoods exceed -10 or -20 in value.

It is important to stress that in our model, for an observed data we have from (8) that $g(\mathbf{y}_i|\boldsymbol{\theta}) = \int_0^{\infty} \int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i; \eta(\boldsymbol{\phi}_i, \mathbf{X}_i), u_i^{-1}\sigma_e^2 \mathbf{I}_{n_i})\phi_q(\boldsymbol{\phi}_i; \mathbf{A}_i\boldsymbol{\beta}, u_i^{-1}\mathbf{D})d\boldsymbol{\phi}_i dH(u_i|\boldsymbol{\nu})$, which can be approximate by the result given in Theorem 1, i.e., $g(\mathbf{y}_i|\boldsymbol{\theta}) \approx NI_{ni}(\mathbf{y}_i|\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i, \widetilde{\mathbf{V}}_i, \boldsymbol{\nu})$.

The deviance information criterion (DIC) proposed by Spiegelhalter *et al.* (2002) is another Bayesian measure of goodness-of-fit and complexity for model selection. The criterion is based on the posterior mean of the deviance, which is also a measure of goodness-of-fit. It can be approximated by $\overline{D} = \sum_{q=1}^{Q} D(\boldsymbol{\theta}_q)/Q$, where $D(\boldsymbol{\theta}) = -2\sum_{i=1}^{n} \log\left[g(\mathbf{y}_i|\boldsymbol{\theta})\right]$. The criterion can be

estimated using the MCMC output by $\widehat{\text{DIC}} = \overline{D} + \widehat{\rho_D} = 2\overline{D} - \widehat{D}$, where ρ_D is the effective number of parameters, which is defined as $E\{D(\boldsymbol{\theta})\} - D\{E(\boldsymbol{\theta})\}$, where $D\{E(\boldsymbol{\theta})\}$ is the deviance evaluated at the posterior mean. Finally, $D\{E(\boldsymbol{\theta})\}$ can be estimated as

$$\widehat{D} = D\left(\frac{1}{Q}\sum_{q=1}^{Q}\boldsymbol{\beta}_q, \frac{1}{Q}\sum_{q=1}^{Q}\sigma_{eq}^2, \frac{1}{Q}\sum_{q=1}^{Q}\boldsymbol{\alpha}_q, \frac{1}{Q}\sum_{q=1}^{Q}\boldsymbol{\nu}_q\right).$$

Given the comparison of two alternative models, the model that better fits a data set is the model with the smallest value of the DIC. The other properties of the DIC can be found in Spiegelhalter *et al.* (2002). Note that it is important to integrate out all latent variables in the deviance calculation as this yields a more appropriate penalty term $\widehat{\rho_D}$; see Kim *et al.* (2008). Once again this can be achieved efficiently through the result given in Theorem 1.

In Section 5 we will present a simulation study in order to investigate if the measurement used for model comparison (LPML and DIC), works well when the likelihood function is approximated by the result given in Theorem 1.

4. Bayesian case influence diagnostics

Since regression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. Cook (1986) uses this idea to motivate his assessment of influence analysis. He suggests that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook & Weisberg, 1982) in which the effects are studied of completely removing cases from the analysis. For our model, we will consider a case-deletion scheme based on the use of perturbation functions.

Perturbation functions were introduced by Kass *et al.* (1989), and Weiss (1996). These functions tackle the problem of assessment of the influence of model M assumptions on the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y}, M)$. Suppose that $\pi(\boldsymbol{\theta}|\mathbf{y}, M_1)$ is the posterior pdf of $\boldsymbol{\theta}$ under the model M_1 and $\pi(\boldsymbol{\theta}|\mathbf{y}, M_2)$ is the posterior pdf of the same parameter, but under the model M_2. Therefore, the perturbation function is defined by

$$p(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta}|\mathbf{y}, M_2)}{\pi(\boldsymbol{\theta}|\mathbf{y}, M_1)}.$$

Let us consider a subset I with k elements of the set $\{1, \ldots, n\}$. When the subset I is deleted from the data $\mathbf{y}$, we will denote $\mathbf{y}_I$ the eliminated data and $\mathbf{y}_{(-I)}$ the remaining data. Then, the perturbation function for deletion cases is $p(\boldsymbol{\theta}) = \pi\left(\boldsymbol{\theta}|\mathbf{y}_{(-I)}\right)/\pi(\boldsymbol{\theta}|\mathbf{y})$. After some straightforward algebraic manipulations we obtain the perturbation function for the NI–NLME model as follows

$$p(\boldsymbol{\theta}) = \frac{\left[\prod_{i\in I} g(\mathbf{y}_i|\boldsymbol{\theta})\right]^{-1}}{E_{\boldsymbol{\theta}|\mathbf{y}}\left\{\left[\prod_{i\in I} g(\mathbf{y}_i|\boldsymbol{\theta})\right]^{-1}\right\}}, \tag{14}$$

where $g(\mathbf{y}_i|\boldsymbol{\theta})$ represents the likelihood function.

The perturbation function for the parameters of the NI–NLME model for the deleted cases can be approximated by using the Theorem 1 and MCMC techniques by sampling from the posterior distribution. In fact, when subset $I = \{i\}$ is considered and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q$ is a sample of size Q of

$\pi(\boldsymbol{\theta} \mid \mathbf{y})$ after the burn-in, the approximation of the perturbation function $p(\boldsymbol{\theta})$ is given by (see Theorem 1)

$$\widehat{p(\boldsymbol{\theta})} = \widehat{\mathrm{CPO}_i} \left[NI_{ni}(\mathbf{y}_i|\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i, \widetilde{\mathbf{V}}_i, \boldsymbol{\nu}) \right]^{-1}, \tag{15}$$

where $NI_{ni}(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ is the pdf of a NI distribution with location vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and parameter $\boldsymbol{\nu}$.

Divergence measures between posterior distributions with and without a given subset of data is a common way of quantifying influence. The q-divergence measure between two densities π_1 and π_2 for $\boldsymbol{\theta}$ is defined by Csiszár (1967) as follows

$$d_q(\pi_1, \pi_2) = \int q\left(\frac{\pi_1(\boldsymbol{\theta})}{\pi_2(\boldsymbol{\theta})}\right) \pi_2(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{16}$$

where q is a convex function such that $q(1) = 0$. Specific divergence measures are obtained by considering particular $q(\cdot)$ functions. For example, the Kullback-Leibler divergence is obtained when $q(z) = -\log(z)$, the J-distance (symmetric version of Kullback-Leibler divergence) is obtained when $q(z) = (z-1)\log(z)$ and the L_1-distance arise by taking $q(z) = |z-1|$.

The q-influence of the data $\mathbf{y}_I$ on the posterior distribution of $\boldsymbol{\theta}$, $d_q(I) = d_q(\pi_1, \pi_2)$, is obtained by considering $\pi_1(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta}|\mathbf{y}_{(-I)})$ and $\pi_2(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{y})$ in Equation (16), and can be written as

$$d_q(I) = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}} \left\{ q(p(\boldsymbol{\theta})) \right\}, \tag{17}$$

where the expected value is taken with respect to the unperturbed posterior distribution. These influence measures have been already used by Peng & Dey (1995) and Weiss (1996) and more recently by Vidal & Castro (2010).

Note that the influence measure $d_q\,(I)$ do not determine when an observation is influential. It is necessary to define a cutoff point to determine if a small subset of observations is influential. In this context, we will use the proposal given by Peng & Dey (1995).

Thus, we consider the probability function of a biased coin, which is given by $\pi_1(x \mid p) = p^x(1-p)^{1-x}$, with $x = 0, 1$, while the probability function of an unbiased coin is given by $\pi_2(x \mid p) = 0.5$. From (16), the q-divergence between a biased and an unbiased coin is given by

$$d_q(p) = \frac{q(2p) + q(2(1-p))}{2},$$

where $d_q(p)$ increases as p moves away from 0.5, is symmetric around $p = 0.5$ and achieves its minimum value at $p = 0.5$. Also, if $d_q(0.5) = 0$ then $\pi_1 = \pi_2$. Consequently, if we consider $p \geq 0.90$ (or $p \leq 0.10$) as a strong bias in a coin, then, $d_{L_1}(0.90) = 0.90$ and we can indicate an influential observation when $d_{L_1}(i) \geq 0.90$, $i = 1, \ldots, n$. Similarly, for the Kullback-Leibler divergence we have $d_{KL}(0.90) = 0.83$, and for the J-distance $d_J(0.90) = 1.33$. These cutoff values will be used in this work.

In the following section, results from simulation studies are presented in order to illustrate the performance of the developed methodology.

5. Simulated Data

In this section, results from simulation studies are presented in order to illustrate the performance of the proposed methodology.

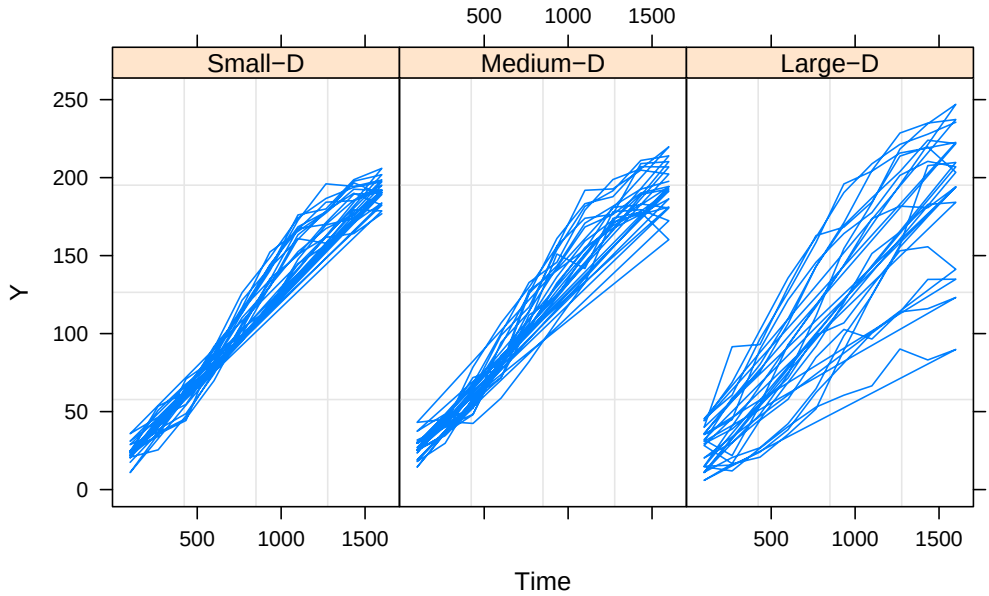

Figure 1: Simulated logistic curves under Student-t distribution for three sizes of the scale matrix of the random effects.

5.1. Frequentist properties

The goals of this simulation study are: 1) to show the good behavior of the Bayesian estimates based on the empirical mean squared error (MSE) and the empirical mean (Mean); 2) to investigate the consequences for population inferences when the normality assumption is inappropriate, and finally 3) to investigate if the measurement used for model comparison (DIC, LPML) works well when the result given in Theorem 1 is used. We performed the simulation with the following nonlinear growth-curve:

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp\left(-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3\right)} + \epsilon_{ij}, \;\; i = 1, \ldots, 15, \;\; j = 1, \ldots, 10, \tag{18}$$

where $t_{ij} = 100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600$ for all i. The random effects $\mathbf{b}_i = (b_{i1}, b_{i2})^\top$ and the $\boldsymbol{\epsilon}_i = (\epsilon_{i1} \ldots, \epsilon_{i10})^\top$ are non-correlated with

$$\boldsymbol{\varphi}_i = \begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} NI_{12}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \mathbf{I}_{10} \end{pmatrix}, H \right), \; i = 1, \ldots, 15, \tag{19}$$

where $U_i = 1, \; \forall i$, in situation 1 (Normal) and $U_i \sim Gamma(2, 2)$, in situation 2 (Student-t with 4 degrees of freedom). We set $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \beta_3) = (200, 700, 350)$,

$$\mathbf{D} = \begin{bmatrix} 4 & -2 \\ -2 & 25 \end{bmatrix},$$

and $\sigma_e^2 = 25$. In this case, the matrices $\mathbf{A}_i$, $\mathbf{B}_i$ are, respectively, given by

$$\mathbf{A}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{B}_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

			Situation 1: Simulated normal data					
Setting	Fit		β_1	β_2	β_3	ν	MC LPML	MC DIC
Small-**D**	Normal	MC Mean	199.08	699.68	298.27	-	-460.15	915.39
		MC Sd	(1.59)	(7.18)	(5.5)	-		
	Student-t	MC Mean	199.10	699.79	298.29	20.83	-461.03	917.32
		MC Sd	(1.58)	(7.15)	(5.5)	(11.7)		
Medium-**D**	Normal	MC Mean	198.98	698.33	298.37	-	-486.58	965.6
		MC Sd	(3.07)	(8.32)	(5.80)	-		
	Student-t	MC Mean	199.02	698.35	298.45	19.98	-487.4	967.54
		MC Sd	(3.04)	(8.32)	(5.76)	(11.36)		
Large-**D**	Normal	MC Mean	199.01	695.41	298.32	-	-506.85	1003.42
		MC Sd	(5.67)	(14.93)	(5.86)			
	Student-t	MC Mean	198.75	696.10	298.41	19.92	-508.19	1006.14
		MC Sd	(5.43)	(14.40)	(5.82)	(10.97)		

Table 1: Monte Carlo results based on 100 simulated normal samples. MC mean and MC Sd (in parenthesis) are the respective posterior mean average and the posterior standard deviations average from fitting NLME models with Student-t and Normal assumptions. MC DIC and MC LPML are the arithmetic average of the measurement used for model comparison.

with covariates $\mathbf{X}_i = (t_{i1}, \ldots, t_{in_i})^\top$.

Additional simulations are created by using the same values of $(\boldsymbol{\beta}, \sigma_e^2)$ and multiplying each element of $\mathbf{D}$ by 25 and 100. That is done to verify if the approximation is reliable to different settings of the scale matrix $\mathbf{D}$. Therefore six different simulation runs are performed with 100 simulated data sets for each setting. Once both the data are simulated we fit the NLME model assuming that $\boldsymbol{\varphi}_i$ have a normal and Student-t distributions. Figure 1 plots example profiles for each of the three sizes of variance components under the Student-t setting. The adjectives small, medium, and large are with reference to the values assumed for $\mathbf{D}$. Note that the variability increases with the mean for this particular model and also the response curves are more variable as the variance components become larger. A similar behavior can be observed in Wolfinger & Lin (1997) under the normal distribution.

For each of the 600 simulated data set, we fit the model (18) assuming normal and Student-t distributions with vague prior distribution. The following independent priors are considered to perform the Gibbs sampler. $\beta_k \sim N_1(\mathbf{0}, 10^3)$, $k = 1, 2, 3$, $\sigma_e^2 \sim IGamma(0.1, 0.01)$, $\mathbf{D} \sim IWish_2(\mathbf{H}, 2)$ with $\boldsymbol{H} = 0.01\mathbf{I}_2$ and $\nu \sim Texp(0.3, (2, \infty))$ for the Student-t model. Considering these prior densities, we generated two parallel independent runs of the Gibbs sampler chain with size 250,000 for each parameter, disregarding the first 50,000 iterations to eliminate the effect of the initial values and to avoid correlation problems, we considered a spacing of size 20, obtaining a sample of size 10,000. For each sample the posterior mean of the parameter and the LPML and DIC are recorded. These criteria are calculated by approximation, using the result given in Theorem 1.

Simulation summary statistics for the fixed-effects parameters assuming Normal and Student-t distributions, crossed with the three settings of the covariance parameter, are given in Tables 1 and

			Situation 2: Simulated Student-t data					
Simulation	Fit		β_1	β_2	β_3	ν	MC LPML	MC DIC
Small-**D**	Normal	MC Mean	198.60	694.89	296.64	-	-509.93	1009.39
		MC Sd	(2.21)	(10.13)	(7.60)			
	Student-t	MC Mean	199.24	699.05	298.88	4.06	-492.92	979.82
		MC Sd	(1.77)	(8.07)	(6.19)	(2.53)		
Medium-**D**	Normal	MC Mean	198.44	691.37	296.50	-	-535.55	1058.28
		MC Sd	(4.45)	(13.7)	(7.91)			
	Student-t	MC Mean	199.21	698.33	299.02	4.99	-519.35	1030.12
		MC Sd	(3.37)	(9.46)	(6.46)	(2.51)		
Large-**D**	Normal	MC Mean	199.33	683.03	296.47	-	-556.67	1098.85
		MC Sd	(7.72)	(20.82)	(8.22)			
	Student-t	MC Mean	199.54	694.2	298.99	4.91	540.63	1069.79
		MC Sd	(6.10)	(16.25)	(6.59)	(2.50)		

Table 2: Monte Carlo results based on 100 simulated Student-t samples. MC mean and MC Sd (in parenthesis) are the respective posterior mean average and the posterior standard deviations average from fitting NLME models with Student-t and Normal assumptions. MC DIC and MC LPML are the arithmetic average of the measurement used for model comparison.

2. In these tables, MC Mean denotes the arithmetic average of the 100 posterior mean estimates given by $\sum_{j=1}^{100} \hat{\theta}_{kj}/100$ and MC Sd is the arithmetic average of the 100 posterior standard deviations given by $\sum_{j=1}^{100} sd(\hat{\theta}_{kj})/100$.

For the simulated normal dataset, the results in Table 1 show little difference in the parameter estimates irrespective of the distribution and the values assumed to **D**. However, note that the standard deviation for the fixed-effects tend to increase as the variance components become larger. As expected, there is no evidence of heavy tails ($\nu > 19$). For the simulated Student-t dataset, the results in Table 2 show that the fitting Student-t distributions detects heavy-tail behavior in the random terms once the estimated robustness parameter $\boldsymbol{\nu}$ is small. We notice from this table that the bias and standard deviations of the parameters estimates under Student-t model are smaller than those of the normal model, indicating that models with longer-than-normal tails to produce more accurate Bayesian estimates. The inferences for the variance components are not comparable for the two fitted models, since they are in different scales, thus are not presented here.

In Tables 1 and 2 we also present the arithmetic averages (MC LPML and MC DIC) of the two

		Criteria	
Setting	Simulated data	MC LPML	MC DIC
Small-**D**	N-NLME	83%	90%
	T-NLME	99%	99%
Medium-**D**	N-NLME	77%	88%
	T-NLME	98%	98%
Large-**D**	N-NLME	79%	85%
	T-NLME	94%	93%

Table 3: Percentage (%) of samples that the LPML and DIC choose the true model properly under differents setting of the scale matrix **D**.

model comparison measures mentioned earlier. We notice that all these measures favored the true (simulated) model. In addition, Table 3 presents the percentage (%) of samples when these criteria chooses the true model. We can see from this table that the % are quite high, though it gets a little less efficient as the generated data is normal and the variance components increase. Thus, our primary conclusion based on both situations is that the result given in Theorem 1 seems to lead a plausible CPO and DIC values. In fact, this approximation is needed to make the calculation of the CPO and DIC computationally feasible (and easy).

Fit	β_1	β_2	β_3	σ_e^2	ν	LPML	DIC
Normal$^{(*)}$	201.7	700.2	300.5	30.7	-	-491.50	983.67
	(2.794)	(9.426)	(6.136)	(3.944)			
Normal	208.8	684.9	309.1	42.93		-521.90	1033.59
	(4.104)	(16.480)	(7.300)	(5.426)			
Student-t	202.0	702.5	301.6	32.5	9.3	-512.76	1029.10
	(2.910)	(10.230)	(6.621)	(6.431)	(4.780)		
Slash	202.0	703.3	301.6	21.9	2.12	-504.87	989.63
	(2.791)	(10.241)	(6.485)	(4.765)	(0.958)		
Cont. Normal	201.8	702.6	302.0	32.2	$\nu = 0.15$ $\gamma = 0.22$	-510.41	1034.94
	(2.944)	(10.640)	(6.561)	(4.691)	(0.097) (0.089)		

Table 4: Parameter estimates, standard errors (in parenthesis) and DIC and LPML criteria from fitting NLME models under four NI distributions to perturbed simulated normal data. $^{(*)}$ indicates the model has been fitted to the original data

5.2. Influence of a single outlier

One of our main goals in this work is to show the necessity of robust models to deal with the presence of outliers in the data. In order to do that, we take the normal simulated data with moderate variance components and introduce one outlier by adding $\Delta = 30$ in the second component of the 11-th observation. Then we fit the model (18) assuming the normal (N-NLME), Student-t (T-NLME), slash (SL-NLME) and the contaminated normal (CN-NLME) distributions. The MCMC computations were done similar to those in the last section and further we consider $\nu \sim$ Gamma$(0.01, 0.01)$ for the slash model, $\nu \sim Beta(1,1)$ and $\gamma \sim Beta(1,1)$, for the contaminated normal model. In Table 4 we report the posterior estimates for the original and perturbed data under the four NI distributions. As expected, the estimates for $\boldsymbol{\beta}$ and σ_e^2 are less affected than those under N-NLME model. Note also from this table that ν can be treated as an outlier accommodation parameter. We also report the Monte Carlo estimates of LPML and DIC. Once again we can see that the NI distributions with heavy tails demonstrate a better fit when outlying observations are present in the dataset.

Considering the samples from the posterior parameter distributions, the d_{KL} divergence measure were computed under four models using (17) with function $q(\cdot) = -\log(\cdot)$, i.e., the Kullback-Leibler divergence. In Figure 4 we have plotted d_{KL}, for $i = 1, \ldots, 15$, for the data set before perturbation and for N–NLME, T–NLME and SL-NLME models after perturbation. Clearly we can see that d_{KL} performed well for identifying influential case(s) providing larger d_{KL} for perturbed cases. Note however that, as expected, for the N–NLME model, after perturbation, d_{KL}

increased a lot and for the NI–NLME models with heavy tails these measures varied little, indicating that NI models with heavy tails are more robust than the N–NLME counterpart in the presence of outlying observations.

6. Data Analysis

In this application we re-analyze the HIV viral load data from clinical trial ACTG 315 (Wu, 2002). In this study, 46 HIV-1infected patients were treated with a potent antiretroviral regimen consisting of protease inhibitor and reverse transcriptase inhibitor drugs. Viral load was repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168, and 196 after initiation of treatment, with a total of 361 observations. Covariates such as CD4 counts were also measured throughout the study on similar scheme. Figure 3 shows the measurements of viral load in natural log10 scale and CD4 cell count for the four randomly selected patients. Note that the rate change in viral load appears to vary substantially across patients, reflecting both biological variation and systematic associations with subject-level covariate (CD4).

As in Wu (2002), we propose the following NLME models or hierarchical two-level nonlinear

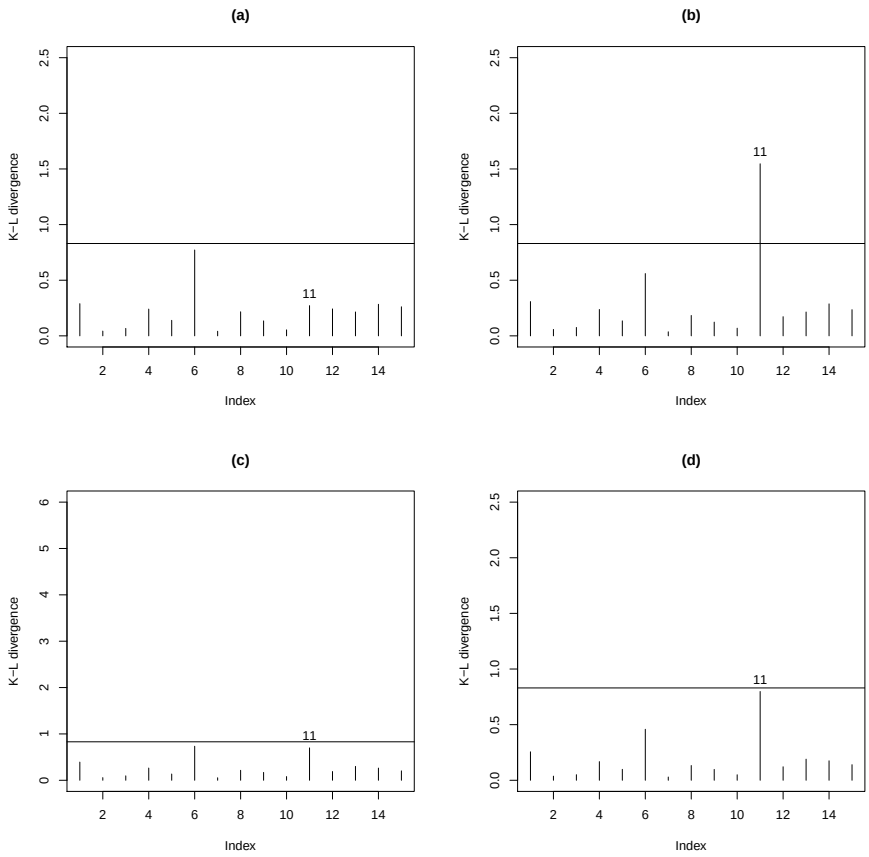

Figure 2: Simulated normal data. Index plots of d_{KL} in the NLME model under: (a) normal to the original data (b) normal to the perturbed data (c) Student-t to the perturbed data (d) slash to the perturbed data.

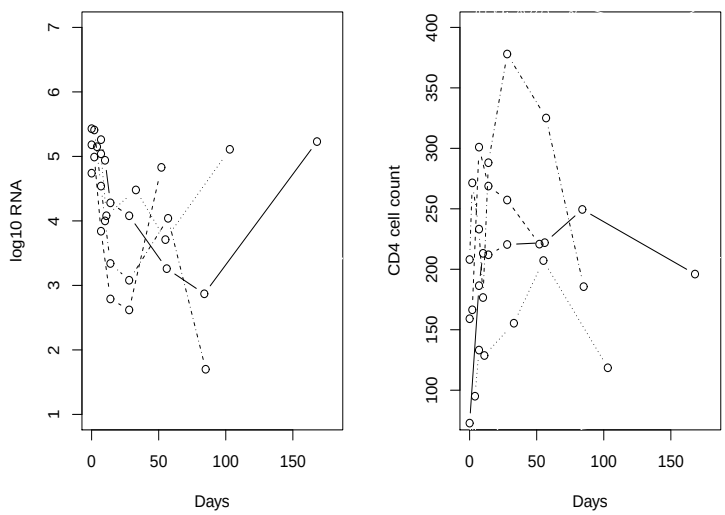

Figure 3: ACTG 315 data. Profiles of viral load (response) in natural $\log_{10}$ scale and CD4 cell count (covariate) for four randomly selected patients.

Table 5: ACTG 315 data. Comparison between Normal and NI–NLME models with heavy-tails by using different model selection criteria and posterior parameter estimates. The numbers in parentheses are the Monte Carlo standard errors of the LPML statistics computed using 500 blocks of the Markov chain, each of size 2000.

criterion	N–NLME		T–NLME		SL–NLME		CN–NLME	
LPML	-307.3104		-296.131		-297.3779		-294.8974	
	(1.25)		(1.17)		(1.22)		(1.12)	
DIC	608.9779		590.8812		594.6343		586.8106	
Parameter estimates	Mean	SD	Mean	SD	Mean	SD	Mean	SD
β_1	11.654	0.220	11.560	0.201	11.602	0.199	11.621	0.200
β_2	30.914	1.940	30.254	1.701	30.490	1.791	30.710	1.688
β_3	6.457	0.273	6.608	0.283	6.538	0.290	6.585	0.278
β_4	-1.897	0.682	-1.229	0.779	-1.486	0.743	-1.204	0.680
β_5	0.663	0.202	0.576	0.210	0.626	0.209	0.566	0.198
σ_e^2	0.15	0.017	0.101	0.016	0.067	0.012	0.078	0.014
ν	-	-	6.001	1.828	1.715	0.437	0.415	0.146
γ	-	-	-	-	-	-	0.298	0.110

model for inference:

$$\begin{aligned} y_{ij} &= \log_{10}(e^{\phi_{1i}-\phi_{2i}t_{ij}} + e^{\phi_{3i}-\phi_{4ij}t_{ij}}) + \epsilon_{ij}, \\ \beta_{1ij} &= \phi_{1i} = \beta_1 + b_{1i}, \;\; \beta_{3ij} = \phi_{3i} = \beta_3 + b_{3i}, \\ \beta_{2ij} &= \phi_{2i} = \beta_2 + b_{2i}, \;\; \beta_{4_{ij}} = \phi_{4ij} = \beta_4 + \beta_5 CD4_{ij} + b_{4i}, \end{aligned} \tag{20}$$

where y_{ij} is the $\log_{10}$-transformation of the viral load for the ith subject at time t_{ij} ($i = 1, 2, \ldots, n, j = 1, 2, \ldots, n_i$) and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{in_i})^\top$ represents within-individual random error; $CD4_{ij}$ indicates a summary of the unobserved CD4 values up to time t_{ij} ; $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \beta_{2ij}, \beta_{3ij}, \beta_{4ij})^\top$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_5)^\top$ are individual parameters for the i-th subject at time t_{ij} and population pa-

rameters, respectively, $\mathbf{b}_i = (b_{1i}, \ldots, b_{4i})^\top$ is individual random effects. The parameters ϕ_{2i} and ϕ_{4i} are called the first and the second phase viral decay rates, which may represent the minimum turnover rate of productivity infected cells and that of latently or long-lived infected cells, respectively. Other covariates such as CD8 cell counts may also be incorporated into the model 20), but this covariate was found to be insignificant in the ACTG 315 study (Wu, 2002). For this NLME model $\boldsymbol{\phi}_i = (\phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i1}, \ldots, \phi_{4in_i})^\top$, the matrices $\mathbf{A}_i$, and $\mathbf{B}_i$ are, respectively, given by

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_i} & \mathbf{cd4}_i \end{pmatrix}, \quad \mathbf{B}_i = \begin{pmatrix} \mathbf{I}_3 & 0 \\ 0 & \mathbf{1}_{n_i} \end{pmatrix},$$

with $\mathbf{cd4}_i = (CD4_{i1}, \ldots, CD4_{in_i})^\top$ and $\mathbf{X}_i = (t_{i1}, \ldots, t_{in_i})^\top$. Wu (2002) analyzed the same data set by fitting a similar N-LMEC from a frequentist perspective. We revisit the ACTG 315 data with the aim of providing robust inferences by using NI distributions. In our analysis, we assume Normal (N-LMEC), Student–t (T-LMEC), slash (SL-LMEC) and contaminated normal (CN-LMEC) distributions from the NI class for comparisons.

For choice of priors, we have, $\beta_j \sim N_1(0, 10^3)$, $j = 1, \ldots, 5$, $\sigma_e^2 \sim IGamma(0.1, 0.01)$, $\mathbf{D} \sim IWish_4(\mathbf{T}, 4)$ with $\boldsymbol{T} = 0.01\mathbf{I}_4$, $\nu \sim \text{TExp}(0.3; (2, \infty))$ for the Student–t model, $\nu \sim \text{Gamma}(0.1, 0.01)$, for the slash model and $\nu \sim Beta(1, 1)$ and $\rho \sim \text{Beta}(2, 2)$, for the contaminated normal model. We generated 4 parallel independent MCMC runs of size 250,000 with widely dispersed initial values for each parameter for all the 4 sub-classes, where the first 50,000 iterations were discarded as burn-in samples. To eliminate potential problems due to auto-correlation, we considered a spacing of size 10. The convergence of the MCMC chains were monitored using trace plots, auto-correlation (ACF) plots and Gelman-Rubin $\hat{R}$ diagnostics. Following Gelman *et al.* (2006), we considered a sensitivity analysis on the routine use of the inverse-gamma prior on the variance components and found that the results are fairly robust under different choices of the priors.

Table 5 compares among the four sub-classes of NI models using the model selection criteria discussed in Section 3.2. Notice that all the 3 members of the NI–NLME class (with heavy tails) perform significantly better than the N-NLME, with the CN-NLME outperforming all the rest. About 92% of the individual log-likelihoods exceed -10 for all the models and the reported standard deviations for the LPML statistics are small posing minimal threat on the stability of the HM identity for CPO and LPML computations. For the T-NLME and SL-NLME models, as ν (the t degrees of freedom) $\to \infty$, it approaches the N-NLME model as a limiting case. For both these models, the estimated value of ν is small, indicating the lack of adequacy of the normal assumption for the ACTG 315 data. In Table 5, we also report the posterior mean and standard deviations for the model parameters from the four fitted NI models. Note that the posterior estimates of β_1, β_2, β_3 and β_5 for the four NI-NLME models are quite close to each other. However, the 95% posterior confidence interval (CI) to β_5 under the 3 members of the NI–NLME class (with heavy tails) includes zero, this is in contrast to the normal model, where the 95% CI does not include 0. The estimates of the between-subject variance-covariance components $\mathbf{D}$ (omitted for brevity) and within-subject scale parameter (σ_e^2) for the NI models (with heavy tails) are slightly smaller as compared to the N-NLME model.

To determine possible influential observations, we computed the q- divergence measures for all

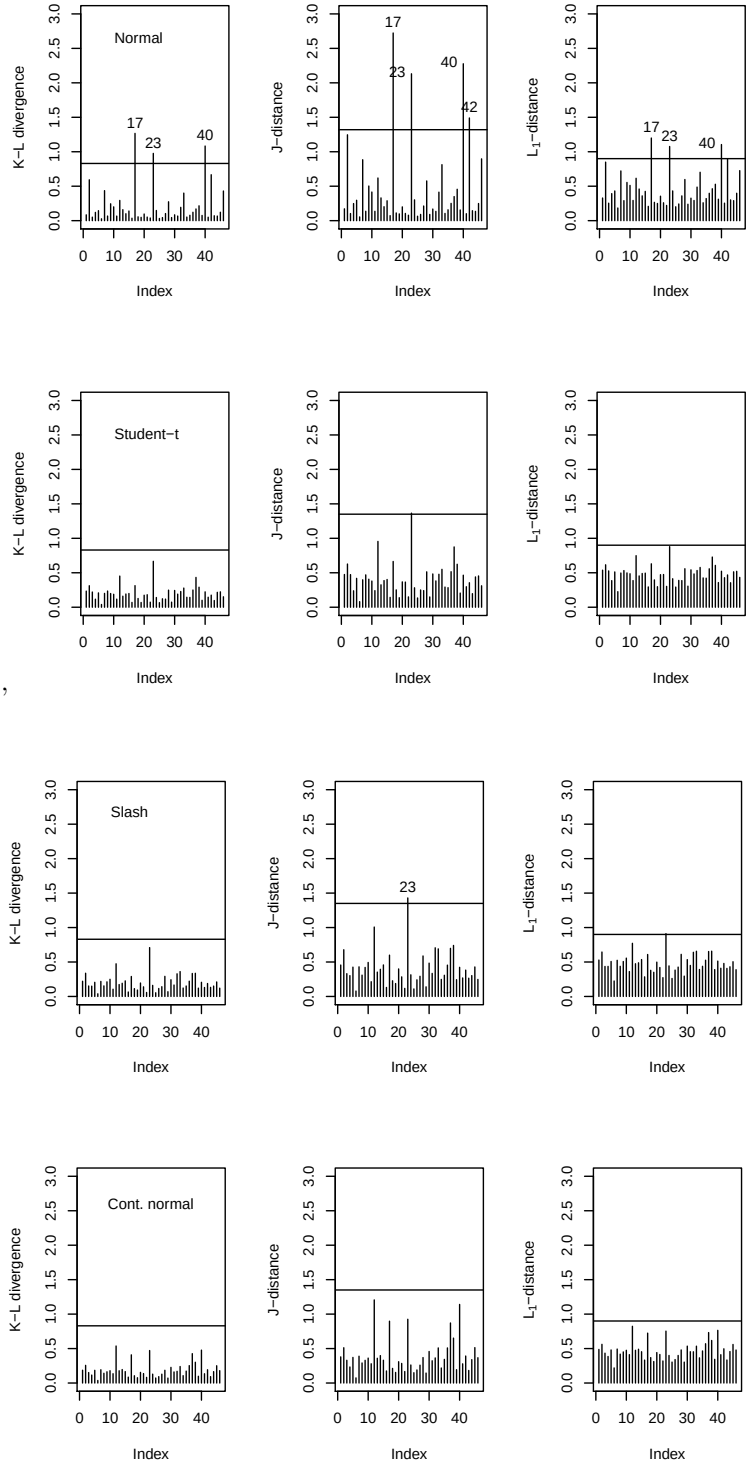

Figure 4: ACTG 315 data. Index plots of q-divergence measures in the four NI-NLME models. The horizontal lines are the cutoff values to determine an influential observation.

the 4 competing models. From Figure 4 (upper panels) for the N-NLME, cases 17, 23, 40 and 42 have larger $d_q(.)$ as compared to the Student–t, slash and contaminated-normal models. As expected, the effect of these cases on the posterior estimates of the mean of $\boldsymbol{\beta}$ were attenuated using the NI class, and hence is a robust alternative for censored viral loads with possible influential observations. This finding is also observed in Figure 5, where the presence of these outliers might have overestimate the predicted mean curve for the N-NLME model as compared to the other three NI-NLME models with heavy tails. However, the fits from these three models are not very distinguishable.

The estimated results presented in Table 5 based on the (best) model (CN-NLME) indicate that the first and second population decay rates of change in viral load may be approximated by $\widehat{\phi_2}(t) = 30.710$ and $\widehat{\phi}_4(t) = -1.204 + 0.566\text{z}(t)$, where $z(t)$ are the CD4 values at time t. The population viral load process may be approximated by $\widehat{V}(t) = \exp\{11.621 - \widehat{\phi}_2(t)t\} + \exp\{6.585 - \widehat{\phi}_4(t)t\}$. Since the second-phase viral decay rate ($\phi_4(t)$) is significantly associated with the observed CD4 values, this suggests that the viral load change $V(t)$ may be significantly associated with the CD4 values. The CD4 process at time t has a significantly positive effect on the second-phase viral decay rate; this finding confirms that the CD4 cell process has different programs for the expansion phase Huang & Dagne (2010). In particular, the CD4 covariate is a more significant predictor on viral decay rate during the late stage. More rapid increase in CD4 cell count may be associated with faster viral decay in late stage. This may be explained by the fact that higher CD4 cell count suggests a higher turnover rate of lymphocyte cells, which may cause a positive correlation between viral decay and the CD4 cell count.

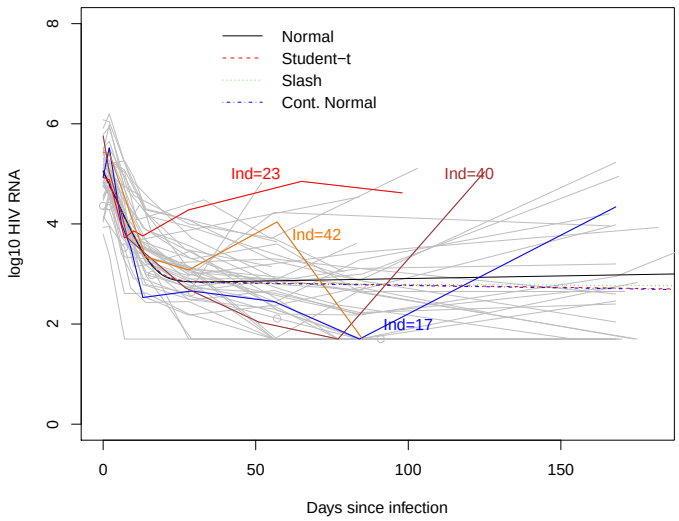

Figure 5: Individual profiles and overall mean (in $\log_{10}$ scale) using the four NI distributions for HIV viral load. The trajectories for the influential individuals are numbered.

7. Conclusions

This article discusses a Bayesian implementation of some robust alternatives to the nonlinear mixed effects model using MCMC technique. The Gaussian distribution of the random terms are replaced by multivariate thick-tailed processes, known as normal/independent distributions. Three specific cases studied were the Student-t, the slash, and the contaminated normal distributions. Bayesian case influence diagnostics based on the q-divergence were also used in order to study the sensitivity of the Bayesian estimates under perturbations in the model/data. This analysis is computationally possible due to an important result which approximates the likelihood function of a NI-NLMM model for a NI distribution with specified parameters. Results obtained by application of the methodology to simulated and HIV data in a medical study, illustrated how the procedures could be used to evaluate assumptions, to identify outliers and to obtain more robust estimation. As can be seen from our data analysis, this family provides a robustness that the multivariate normal distribution does not. We suspect that it may even be difficult to find a data set for which this class of distributions do not provide an improved fit over the multivariate normal distribution. Finally, the proposed algorithms has been coded and implemented in the R package (R Development Core Team, 2009) and is available from the authors upon request.

On the other hand, this does not imply that multivariate NI distributions are the ideal solution to all problems in nonlinear mixed effects models. Other forms of models will also be necessary, such as those currently being developed through of the skew-normal/independent distributions (Lachos *et al.*, 2009). Another solution to obtain more robust estimation is to consider semi parametric models based on Dirichlet process priors (DPP). For example, Wang (2012) proposes two efficient approaches to deal with the difficulty in computing the intractable integrals when implementing Gibbs sampling in the nonlinear mixed effects model (NLMM) based on Dirichlet processes (DP). This approach performs well in density estimation and prediction and can detect important model features, such as skewness and heavy tails.

Acknowledgement

The authors thank the Editor, Associate Editor and two referees whose constructive comments led to a far improved presentation. V.H.Lachos acknowledges support from CNPq-Brazil (Grant 2011/201384-6) and from FAPESP-Brazil (Grant 305054/2011-2). The research of L. M. Castro is supported by Grant FONDECYT 11100076 from the Chilean government.

Appendix A: Proof of Theorem 1

For simplicity we omit the subindex i. Thus, for $\boldsymbol{\beta}$ fixed and based on first-order Taylor expansion of the function η around $\widetilde{\mathbf{b}}$, we have that

$$\mathbf{y} - \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}, \mathbf{X}) \approx \mathbf{y} - [\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\widetilde{\mathbf{b}}, \mathbf{X}) + \widetilde{\mathbf{H}}\mathbf{b} - \widetilde{\mathbf{H}}\widetilde{\mathbf{b}}].$$

Then from (2) and (4)

$$\mathbf{y} - [\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\widetilde{\mathbf{b}}, \mathbf{X}) + \widetilde{\mathbf{H}}\mathbf{b} - \widetilde{\mathbf{H}}\widetilde{\mathbf{b}}] \mid \mathbf{b} \dot{\sim} \mathrm{NI}_n(\mathbf{0}, \sigma_e^2\mathbf{I}, H)$$

and the approximate conditional distribution of $\mathbf{y}$ is

$$\mathbf{y} \mid \mathbf{b} \dot{\sim} \mathrm{NI}_n(\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\widetilde{\mathbf{b}}, \mathbf{X}) - \widetilde{\mathbf{H}}\widetilde{\mathbf{b}} + \widetilde{\mathbf{H}}\mathbf{b}, \sigma_e^2\mathbf{I}, H), \tag{21}$$

or equivalently

$$\mathbf{y} \mid \mathbf{b}, u \dot{\sim} \mathrm{N}_n(\eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\widetilde{\mathbf{b}}, \mathbf{X}) - \widetilde{\mathbf{H}}\widetilde{\mathbf{b}} + \widetilde{\mathbf{H}}\mathbf{b}, u^{-1}\sigma_e^2\mathbf{I}).$$

The rest of the proof follows by noting that

$$f(\mathbf{y} \mid \boldsymbol{\theta}, \widetilde{\mathbf{b}}) \approx \int_0^\infty \int_{\mathbb{R}^q} \phi_n(\mathbf{y}; \eta(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\widetilde{\mathbf{b}}, \mathbf{X}) - \widetilde{\mathbf{H}}\widetilde{\mathbf{b}} + \widetilde{\mathbf{H}}\mathbf{b}, u^{-1}\sigma_e^2\mathbf{I})\phi_q(\mathbf{b}; \mathbf{0}, u^{-1}\mathbf{D})d\mathbf{b}dH(u|\boldsymbol{\nu}),$$

which can be easily accomplished by using successively Lemma 2 given in Arellano-Valle *et al.* (2005).

Appendix B: Conditional distributions for special cases

i) Student-t.

$$\begin{aligned} u_i|\mathbf{y}, \mathbf{b}, \boldsymbol{\theta} &\sim Gamma((n_i + q + \nu)/2; \nu/2 + \lambda_i/2), \\ \pi(\nu|\mathbf{y}, \mathbf{b}, \mathbf{u}, \boldsymbol{\theta}_{(-\nu)}) &\propto \frac{(\nu/2)^{n\nu/2}}{(\Gamma(\nu/2))^n} \exp\left\{-\nu\left[(1/2)\sum_{i=1}^n (u_i - \log(u_i)) + \lambda_o\right]\right\} \mathbb{I}_{\{(2,\infty)\}}(\nu), \end{aligned}$$

where $\lambda_i = (\lambda_{1i}/\sigma_e^2 + \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i)$. As $\pi(\nu|.)$ do not have a standard form, we can generate a sample from this distribution using the Metropolis-Hastings path with a *log-normal* proposal density.

ii) Slash.

$$u_i|\mathbf{y}, \mathbf{b}, \boldsymbol{\theta} \sim TG((n_i + q + 2\nu)/2, \lambda_i/2, (0, 1)), \tag{22}$$

$$\nu|\mathbf{y}, \mathbf{b}, \mathbf{u}, \boldsymbol{\theta}_{(-\nu)} \sim Gamma(n + a,\, b - \sum_{i=1}^n \log u_i), \tag{23}$$

where $TG(a, b, (0, 1))$ denotes the Gamma distribution with parameters a and b truncated in the interval $(0, 1)$.

iii) Contaminated normal.

$$\mathrm{P}(u_i = \gamma|\cdots) = 1 - \mathrm{P}(u_i = 1|\cdots) = \eta_i/(\eta_i + \zeta_i),$$

where

$$\eta_i = \nu\gamma^{(n_i+q)/2} \exp\{-(1/2)\lambda_i\gamma\} \text{ and } \zeta_i = (1 - \nu)\exp\{-(1/2)\lambda_i\},$$

$$\nu|\cdots \sim \mathrm{Beta}(\nu_0 + m_\gamma, n - m_\gamma + \nu_1),$$

where

$$m_\gamma = \frac{n - \sum_{i=1}^n u_i}{1 - \gamma}$$

is the cardinality of the set $\{i;\ u_i = \gamma\}$. The full conditional for γ is given by

$$\pi(\gamma|\cdots) \propto \nu^{m_\gamma}(1 - \nu)^{n-m_\gamma}\gamma^{\rho_0-1}(1 - \gamma)^{\rho_1-1},$$

which does not have a closed form. An interesting Metropolis–Hastings method to update from γ is described in Rosa *et al.* (2003).

References

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**, 415–438.

Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, **49**, 327–335.

Congdon, P. (2005). *Bayesian models for Categorical Data*. John Wiley & Sons Inc.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B,*, **48**, 133–169.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, **2**, 299–318.

Dey, D. K., Chen, M. H. & Chang, H. (1997). Bayesian approach for the nonlinear random effects models. *Biometrics*, **53**, 1239–1252.

Galea, M., Paula, G. A. & Cysneiros, F. J. A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, **73**(4), 459–467.

Gamerman, D. & Lopes, H. (2006). *Markov Chain Monte Carlo*. Chapman & Hall/CRC, Boca Raton, FL.

Geisser, S. & Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.

Gelfand, A. E. & Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, **56**, 501–514.

Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4 (Peñíscola, 1991)*, pages 147–167. Oxford Univ. Press, New York.

Gelman, A., Roberts, G. & Gilks, W. (1996). Efficient Metropolis jumping rules. In J. Bernardo, J. Berger, A. Dawid, & A. Smith, editors, *Bayesian Statistics 5*, pages 599–607. Oxford University Press.

Gelman, A., Carlin, J. & Rubin, D. (2006). *Bayesian Data Analysis*. Chapman & Hall/CRC, New York, NY.

Gomez, H., Quintana, F. & Torres, F. (2007). A new family of slash-distributions with elliptical contours'. *Statistics and Probability Letters*, **77**, 717–725.

Hobert, J. & Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, **91**(436), 1461–1473.

Huang, Y. & Dagne, G. (2010). A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics*, **67**, 260–269.

Kass, R., Tierney, L. & Kadane, J. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*, **76**, 663–674.

Kim, S., Chen, M. H. & Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika*, **95**, 93–106.

Lachos, V. H., Dey, D. K. & Cancho, V. G. (2009). Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *Journal of Statistical Planning and Inference*, **139**, 4098–4110.

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

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

Lindstrom, M. & Bates, D. (1990). Nonlinear mixed-effects models for repeated-measures data. *Biometrics*, **46**.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, **91**.

Meza, C., Osorio, F. & de la Cruz, R. (2012). Estimation in non-linear mixed effects models using heavy tailed distributions. *Statistics and Computing*, **22**, 121–139.

Osiewalski, J. (1999). Bayesian analysis of nonlinear regression with equicorrelated elliptical errors. *Test*, **8**, 339–344.

Osiewalski, J. & Steel, M. F. J. (1993). Robust Bayesian-inference in elliptic regression models. *Journal of Econometrics*, **57**, 345–363.

Osorio, F., Paula, G. A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354–4368.

Peng, F. & Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics*, **23**, 199–213.

Pinheiro, J. & Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.

Pinheiro, J. C. & Bates, Douglas, M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY.

Pinheiro, J. C., Liu, C. H. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics*, **10**, 249–276.

R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raftery, A., Newton, M., Satagopan, J. & Krivitsky, P. (2007). *Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion)*, volume 8, pages 1–45. Oxford University Press.

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

Russo, C. M., Paula, G. A. & Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis*, **53**(12), 4143–4156.

Savalli, C., Paula, G. A. & Cysneiros, F. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, **6**, 59–76.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**(4), 583–639.

Vidal, I. & Castro, L. M. (2010). Influential observations in the independent Student-t measurement error model with weak nondifferential error. *Chilean Journal of Statistics*, **1**(2), 17–34.

Wang, J. (2012). Dirichlet processes in nonlinear mixed effects models. *Communication in Statistics - Simulation and Computation*, (39), 539–556.

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society, Series B*, **58**, 739–750.

Wolfinger, R. D. & Lin, X. (1997). Two taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis*, **25**, 465–490.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association*, **97**(460), 955–964.