Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads

Victor H. Lachos[*1], **Dipankar Bandyopadhyay** [2], **Dipak K. Dey** [3], and **Luis M. Castro** [4]

[1] Departamento de Estatística, IMECC, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

[2] Division of Biostatistics and Epidemiology, Medical University of South Carolina, Charleston, SC, U.S.A

[3] Department of Statistics, University of Connecticut, Storrs, CT, U.S.A.

[4] Departamento de Estadística, Universidad de Concepcion, Concepción, Chile

Often in biomedical studies, the routine use of linear mixed-effects models (based on Gaussian assumptions) can be questionable when the longitudinal responses are skewed in nature. Skew-normal/elliptical models are widely used in those situations. Often, those skewed responses might also be subjected to some upper and lower quantification limits (viz. longitudinal viral load measures in HIV studies), beyond which they are not measurable. In this paper, we develop a Bayesian analysis of censored linear mixed models replacing the Gaussian assumptions with skew-normal/independent (SNI) distributions. The SNI is an attractive class of asymmetric heavy-tailed distributions that includes the skew-normal, the skew-t, skew-slash and the skew-contaminated normal distributions as special cases. The proposed model provides flexibility in capturing the effects of skewness and heavy tail for responses which are either left- or right-censored. For our analysis, we adopt a Bayesian framework and develop a MCMC algorithm to carry out the posterior analyses. The marginal likelihood is tractable, and utilized to compute not only some Bayesian model selection measures but also case-deletion influence diagnostics based on the Kullback-Leibler divergence. The newly developed procedures are illustrated with a simulation study as well as a HIV case study involving analysis of longitudinal viral loads.

Key words: Bayesian inference; Detection limit; Gibbs Sampler; HIV viral load; Linear mixed models; skewness; Skew-normal/independent distribution;

1 Introduction

In AIDS research, HIV1-RNA (or viral load) measures are collected longitudinally (Wu, 2010) over a period of treatment to assess reduction in viral load, which is a primary end-point in clinical trials of antiretroviral (ARV) theraphy (Jacqmin-Gadda et al., 2000). Mixed-effects models are routinely used (Vaida et al., 2007; Vaida and Liu, 2009a; Qiu and Wu, 2010) to estimate viral load trajectories as well as to quantify within-subject and between-subject variations in viral load measurements. Although viral load has been widely recognized as the best prognostic marker with CD4 + cell counts (Mellors et al., 1996), its measurement comes with some additional complications, viz., the measures may be subjected to some upper and lower detection limits, below or above which they are not quantifiable. As a result, the viral load responses are either left or right censored depending on the diagnostic assays used. However, the proportion

*Corresponding author: e-mail: hlachos@ime.unicamp.br

of censored data in these studies may not be trivial (Hughes, 1999) and considering crude/adhoc methods viz., substituting threshold value or some arbitrary point such as mid-point between zero and cut-off for detection (Huang et al., 2001) might lead to biased estimates of fixed effects and variance components (Gray et al., 2004).

Our motivating data in this paper is from an AIDS clinical trial study (Saitoh et al., 2008), which measures HIV-1 viral loads after unstructured treatment interruption (UTI). In this study, about 7% observations lie below the limits of the quantification assay, and hence are considered to be left-censored. As alternatives to the crude imputation methods described above, Hughes (1999) proposed a likelihood-based Monte Carlo EM algorithm (MCEM) for linear mixed effects (LME) with censored response (LMEC). Vaida et al. (2007) proposed a hybrid EM using a more efficient implementation of Hughes algorithm, including numerical computation at the E-step for clusters with one or two censored observations. They also extended the algorithm to NLME with censored data (NLMEC). Their MCEM improves the simulation at the E-step, the numerical implementation at the M-step, and includes automatic monitoring and stopping of the algorithm. Recently, Vaida and Liu (2009a) proposed an exact EM algorithm for LMEC/NLMEC, which uses closed-form expressions at the E-step, as opposed to Monte Carlo simulations. A common feature of all these methods is the fidelity to the 'Gaussian' paradigm for the random-effects and within-subject random errors. Although the assumptions of normality provides a much simplistic framework for statistical analysis in guiding the development of treatment strategies and clinical decision, it may lack robustness against departures from normality and/or outliers (Sahu et al., 2003) and thus statistical inference and analysis with normal assumption may lead to misleading results. Viral-load measurements are often highly skewed, and even log-transformations do not render normality (Ghosh et al., 2007). This characteristic further complicates analysis of mixed-effects models, because one (or both) of the (within-subject) random error and (between-subject) random effects might contribute to the 'shift from normality'. For example, Figure 1 (panels a and b) display the density histogram and associated Q-Q plots for (repeated and non-censored) viral load measurements (in natural $\log_{10}$ scale) from the above study, which reveals some degree of left skewness. Panels (c and d) present the same for empirical Bayes estimates (Laird and Ware, 1982) of random effects b_i and panels (e and f) for the residuals, all obtained after fitting a NLMEC model to the UTI data using the R package *lmec* (Vaida and Liu, 2009b). These plots reveal left-skewed nature of subject-specific intercepts at the level of random-effects, but symmetric tail behavior for the random errors.

To deal with the problem of departure from normality in LME models for complete data and eliminate the need of *ad hoc* data transformations (Azzalini and Capitanio, 1999), several proposals have been considered in the literature by replacing the normal random effects using finite normal mixtures (Verbeke and Lesaffre, 1996), smoothing (Ghidey et al., 2004), a semi-nonparametric density (Zhang and Davidian, 2001), a skew-t LME model (Ho and Lin, 2010) etc. Much of the recent frequentist and Bayesian advances in regression problems revolve around the attractive and popular skew-normal/independent (SNI) distributions (Lachos et al., 2010). Starting with the multivariate SN density (Azzalini and Dalla Valle, 1996), SNI linear mixed effects models (SNI-LME) was proposed in Lachos et al. (2010), primarily using EM-type algorithm for maximum likelihood (ML) estimation. Bayesian analysis in the context of non-linear regression with multivariate skew-elliptical (skew-normal and skew-t) errors were considered in De la Cruz and Branco (2009). There had been some recent Bayesian proposals for LME models with SNI distributions

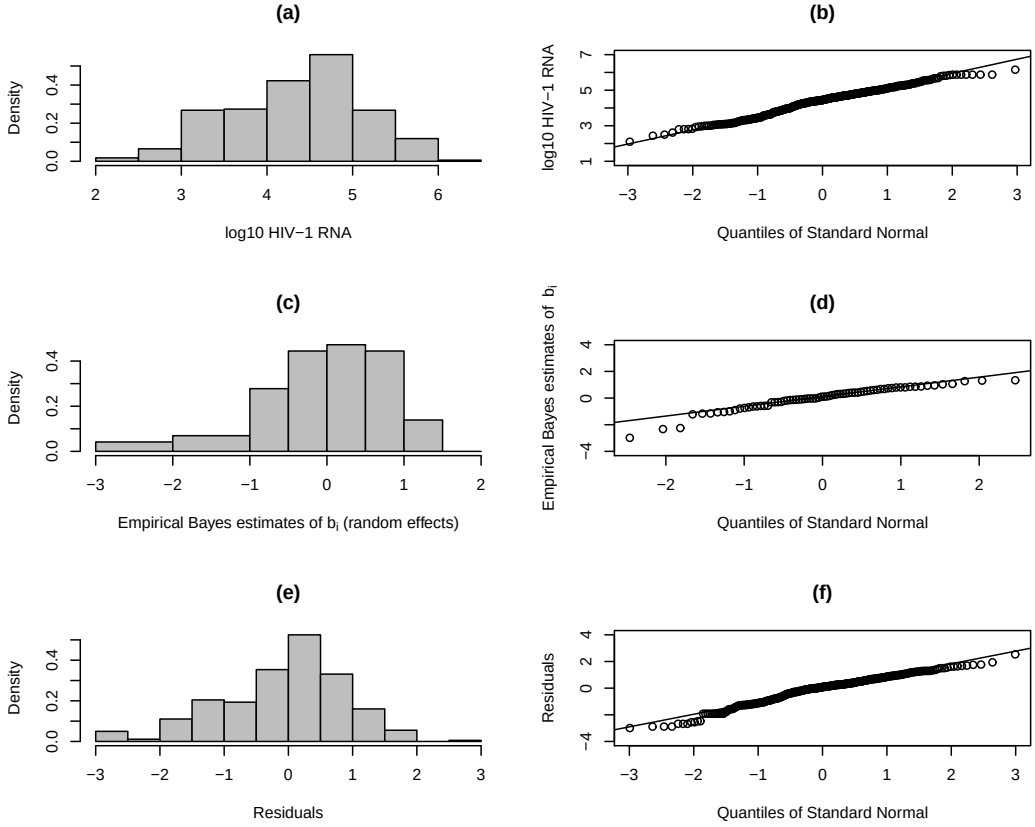

Figure 1 UTI Data: Density histogram and corresponding Q-Q plots for raw HIV viral load measures (in $\log_{10}$ scale) (Panels a and b), empirical Bayes estimates of random effects (Panels c and d), and model residuals (Panels e and f) respectively, after fitting a NLMEC model using R package *lmec*

(Jara et al., 2008; Lachos et al., 2009), but to the best of our knowledge, there are no reported study exploring a robust unified Bayesian framework simultaneously for asymmetrical and heavy-tailed responses for LMEC models using the family of SNI distributions.

In this paper, we proceed with our robust parametric LMEC model assuming a SNI distribution for the random effects and a NI distribution for the random errors, so that the SNI-LMEC model is defined. The marginal density of the observed quantities is obtained analytically. Our main contribution aims at providing a Bayesian treatment to censored version of SNI-LME models and studying various Bayesian model selection and related case-deletion influence diagnostics based on Kullback-Leibler (K-L) divergence measures as proposed in Peng and Dey (1995) and Cancho et al. (2010). The multivariate SNI distribution used in our proposition is developed primarily from the multivariate SN density proposed in Azzalini and Dalla Valle (1996). Although other propositions (Sahu et al., 2003) exist for the multivariate SN density, the differences between these are only due to the various parameterizations (Arellano-Valle and Azzalini, 2006) used. An unification of all skew-normal variants is presented in Arellano-Valle and Genton (2005).

The rest of the paper is organized as follows. In Section 2, we introduce the multivariate skew-normal and SNI distributions and state some propositions which are crucial for our methodology developments. Section 3 formulates our SNI-LMEC and proposes a Bayesian approach for inference. In Section 4, we discuss relevant Bayesian model selection criteria and influence diagnostic measures. The advantage of the proposed methodology is illustrated through analysis of the longitudinal HIV viral load in a AIDS study in Section 5. Section 6 presents a simulation study to compare the finite sample performance of the various sub-classes of our proposed model. Finally, Section 7 concludes with some discussions and citing avenues for future research.

2 Preliminaries

2.1 The multivariate skew-normal distribution

We begin with a review of the multivariate SN distribution and a study of some related properties. Some versions, extensions and unifications of the SN family are carefully surveyed in works like Azzalini (2005) and Arellano-Valle et al. (2006). A random vector $\mathbf{Y}$ has multivariate skew-normal distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, if its density is given by

$$SN(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) \quad = \quad 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{1}$$

where $\phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote respectively the pdf and the cdf of the p-variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$, respectively and $\boldsymbol{\Sigma}^{-1/2}$ is such that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. The univariate standard normal cdf will be denoted by $\Phi(.)$. In usual notation, we shall write $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ for a random vector with density (1). Note that if $\boldsymbol{\lambda} = \mathbf{0}$, then the density of $\mathbf{Y}$ reduces to the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density. Observe that (1) is the conditional density of $\mathbf{Y}|X > 0$, where

$$\begin{pmatrix} \mathbf{Y} \\ X \end{pmatrix} \sim \mathrm{N}_{p+1}\left[\begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta} \\ \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2} & 1 \end{pmatrix} \right].$$

This distribution belongs to the class of *fundamental skew-normal (FUSN) distributions* proposed in Arellano-Valle and Genton (2005). It is worth mentioning that although belonging to the FUSN family, the classical versions introduced by Azzalini and Dalla Valle (1996) and Sahu et al. (2003) are different from (1), the differences are only due to the various parameterizations (Arellano-Valle and Azzalini, 2006) used as mentioned earlier. Using version (1) above allows us to develop an EM-type and Gibbs algorithms for parameter estimation in SN-LME models. Indeed, it has been used as a powerful tool to extend some traditional normal based models, see Lachos et al. (2009) and Lachos et al. (2010), for more details. For computational purposes, an useful parametrization is given by

$$\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, \;\; \boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{1/2}(\mathbf{I} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \tag{2}$$

where $\mathbf{I}$ denotes the identity matrix of appropriate dimension and $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}$. We recover $\boldsymbol{\lambda}$ and $\boldsymbol{\Sigma}$ using the expressions

$$\boldsymbol{\lambda} = \frac{(\boldsymbol{\Gamma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^\top)^{-1/2}\boldsymbol{\Delta}}{[1 - \boldsymbol{\Delta}^\top(\boldsymbol{\Gamma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^\top)^{-1}\boldsymbol{\Delta}]^{1/2}}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \tag{3}$$

Under this parametrizations and following Lachos et al. (2010), the multivariate SN distribution has a convenient stochastic representation

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta} T + \boldsymbol{\Gamma}^{1/2}\mathbf{T_1}, \tag{4}$$

where $T = |T_0|$, $T_0 \sim N_1(0,1)$ and $\mathbf{T_1} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ are independent, with $\mathbf{I}_p$ being the $p \times p$ identity matrix and $|\cdot|$ denoting absolute value. It should also be noted crucially that (4) offers a stochastic representation of $\mathbf{Y}$, which is useful for random number generation and for theoretical purposes. From (4) the mean and covariance of $\mathbf{Y}$ are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Delta},\ Var[\mathbf{Y}] = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \tag{5}$$

2.2 The multivariate skew-normal/independent (SNI) distribution

The idea of the SNI distributions originated from an early work by Branco and Dey (2001), which included the skew-normal distribution as a special case. A SNI distribution is defined by a $p-$dimensional random vector

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{6}$$

where U is a positive random variable with the cdf $H(u;\boldsymbol{\nu})$ and pdf $h(u;\boldsymbol{\nu})$, and independent of the random vector $\mathbf{Z}$, following $SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Here $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of the mixing scale factor U. Given $U = u$, $\mathbf{Y}$ follows a multivariate skew–normal distribution with location vector $\boldsymbol{\mu}$, scale matrix $u^{-1}\boldsymbol{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda}$. Then, from (1), the marginal pdf of $\mathbf{Y}$ is

$$f(\mathbf{y}) \quad = \quad 2\int_0^\infty \phi_p(\mathbf{y};\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))dH(u;\boldsymbol{\nu}) \tag{7}$$

The notation $\mathbf{Y} \sim \text{SNI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$ will be used when $\mathbf{Y}$ has pdf (7).

The asymmetrical class of SNI distributions includes the skew–t, the skew–slash, and the skew–contaminated normal distributions. All these distributions have heavier tails than the skew-normal and can be used for robust inferences. Some of these distributions are described subsequently. When $\boldsymbol{\lambda} = \mathbf{0}$, the SNI distributions reduces to the normal–independent (NI) class (Lange and Sinsheimer, 1993), i.e., the class of scale–mixtures of the normal distribution represented by the pdf $f_0(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y};\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dH(u;\boldsymbol{\nu})$. We use the notation $\mathbf{Y} \sim \text{NI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H)$ when $\mathbf{Y}$ has distribution in the NI class. We refer to Lachos et al. (2010) for details and additional properties related to this SNI class.

The following results show that the cdf of a SNI random vector and its conditional distribution can be obtained from a mixture representation of the SNI class. These result are crucial for our methodology development, in particular essential for computing the likelihood function of a SNI-LMEC. The proofs are given in the Appendix.

Proposition 2.1 *If* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$*, then for any* $\mathbf{y} \in \mathbb{R}^p$

i) *the cdf of* $\mathbf{Y}$ *is given by*

$$P(\mathbf{Y} \le \mathbf{y}) = \int_0^\infty P(\mathbf{Y} \le \mathbf{y}|u)dH(u;\boldsymbol{\nu}), \tag{8}$$

where $P(\mathbf{Y} \leq \mathbf{y}|u) = \Phi_{p+1}\left((\mathbf{z}^\top, 0)^\top; \mathbf{0}, u^{-1}\boldsymbol{\Omega}\right)$, $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & -u^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda} \\ -u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{1/2} & u(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}) \end{pmatrix}$ *and* $\mathbf{z} = \mathbf{y} - \boldsymbol{\mu}$;

ii) *the conditional cdf of* $\mathbf{Y}_2|\mathbf{Y}_1$ *is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = \frac{\int_0^\infty P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1, u)P(\mathbf{Y}_1 = \mathbf{y}_1|u)dH(u;\boldsymbol{\nu})}{\int_0^\infty P(\mathbf{Y}_1 = \mathbf{y}_1|u)dH(u;\boldsymbol{\nu})}. \tag{9}$$

Proposition 2.2 *If* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$ *then, under the notations of Lemma 1 (see Appendix), we have that:*

i) *for* H *such that* $P(U = 1) = 1$, *then the conditional cdf is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = \frac{\Phi_{p_2+1}\left((\mathbf{z}^\top, z_1)^\top; \mathbf{0}, \boldsymbol{\Omega}_N\right)}{\Phi(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}, \tag{10}$$

where $\mathbf{z} = \mathbf{y}_2 - \boldsymbol{\mu}_{2.1}$, $z_1 = (\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2)^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\boldsymbol{\Omega}_N = \begin{pmatrix} \boldsymbol{\Sigma}_{22.1} & -\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2 \\ -\boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1} & 1 + \boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2 \end{pmatrix}$, *with* $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \boldsymbol{\mu}_1)$;

ii) *for* $H \equiv Gamma(\boldsymbol{\nu}/2, \boldsymbol{\nu}/2)$, *then the conditional cdf is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = \frac{T_{p_2+1}\left((\mathbf{z}^\top, \widetilde{\tau})^\top; \mathbf{0}, \boldsymbol{\Omega}_T, \boldsymbol{\nu} + p_1\right)}{T(\widetilde{\tau}; \boldsymbol{\nu} + p_1)}, \tag{11}$$

where $\mathbf{z} = \widetilde{\boldsymbol{\Sigma}}_{22.1}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1})$, $\widetilde{\tau} = \tau_{2.1}\sqrt{1 + \boldsymbol{\lambda}_{2.1}^\top\bar{\boldsymbol{\Sigma}}_{22.1}\boldsymbol{\lambda}_{2.1}}$ *and* $\boldsymbol{\Omega}_T = \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{22.1} & -\boldsymbol{\delta} \\ -\boldsymbol{\delta} & 1 \end{pmatrix}$, *with* $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \boldsymbol{\mu}_1)$, $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\frac{\boldsymbol{\nu}+Q(\mathbf{y}_1)}{\boldsymbol{\nu}+p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\lambda}_2$, $\tau_{2.1} = \sqrt{\frac{\boldsymbol{\nu}+p_1}{\boldsymbol{\nu}+Q(\mathbf{y}_1)}}\widetilde{\boldsymbol{\lambda}}_1\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{Y}_1 - \boldsymbol{\mu}_1)$, $\widetilde{\boldsymbol{\lambda}}_1 = \boldsymbol{\lambda}_1 + \bar{\boldsymbol{\Sigma}}_{11}^{-1}\bar{\boldsymbol{\Sigma}}_{12}\boldsymbol{\lambda}_2$, $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}_{2.1}}{\sqrt{1+\boldsymbol{\lambda}_{2.1}^\top\bar{\boldsymbol{\Sigma}}_{22.1}\boldsymbol{\lambda}_{2.1}}}$, $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}$, $Q(\mathbf{y}_1) = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, *and* $T_k(.;\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ *(*$T(.;\boldsymbol{\nu})$*) denotes the cdf of the* k*-variate (univariate) Student-t distribution with parameters* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\nu}$ *(*$0, 1$ *and* $\boldsymbol{\nu}$*).*

For the SSL case, we can obtain it by using (9) along with the routine *integrate()* in R For the SCN, calculation of $P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1)$ is direct from (10).

3 Model formulation and Bayesian approach

We consider the following general LME model in which the random effects are assumed to follow a SNI distribution. Simultaneously, the model can be written in matrix form as (Lachos et al., 2010)

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{12}$$

with the assumption that

$$\mathbf{b}_i \overset{\text{iid}}{\sim} SNI_q(c\boldsymbol{\Delta}, \mathbf{D}, \boldsymbol{\lambda}, H) \text{ and } \boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} NI_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, H),\ i = 1, \ldots, n, \tag{13}$$

where $c = -(2/\pi)^{1/2}E[U^{-1/2}]$, the subscript i is the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is a $n_i \times 1$ vector of observed continuous responses for sample unit i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors, and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. Note that (13) in conjunction with (5) gives $E[\mathbf{b}_i] = E[\boldsymbol{\epsilon}_i] = \mathbf{0}$. Thus, this model considers the within-subject errors $\boldsymbol{\epsilon}_i$ to be symmetrically distributed, while the distribution of random effects to be asymmetric with mean zero. An interesting property of this formulation is that $E[\mathbf{Y}_i] = \mathbf{X}_i\boldsymbol{\beta}$ and hence the fixed effects estimates of SNI-LME model and N-LME model are all comparable, an important characteristic not introduced in previous developments by Lachos et al. (2007), Lin and Lee (2008) and Lachos et al. (2009). In the present formulation, we consider the case where the response Y_{ij} is not fully observed for all i, j (Vaida and Liu, 2009a). The observed data for the i-th subject is $(\mathbf{Q}_i, \mathbf{C}_i)$, where $\mathbf{Q}_i$ represents the vector of uncensored reading or censoring level, and $\mathbf{C}_i$ the vector of censoring indicators, such that

$$\begin{aligned} y_{ij} &\leq Q_{ij} \quad \text{if } C_{ij} = 1, \\ y_{ij} &= Q_{ij} \quad \text{if } C_{ij} = 0. \end{aligned} \tag{14}$$

For simplicity we will assume that the data are left-censored and thus the SNI-LMEC is defined. The extensions to arbitrary censoring are immediate. For N-LMEC, an EM algorithm was proposed by Hughes (1999), with computational improvements considered in Vaida et al. (2007) and Vaida and Liu (2009a).

3.1 The log-likelihood function

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$ is based on the marginal distribution for $\mathbf{y}_i$ (Vaida et al., 2007). For complete data, we have from Lachos et al. (2007), that marginally

$$\mathbf{Y}_i \sim SNI_{n_i}(\widetilde{\mathbf{X}}_i\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, H),\ i = 1, \ldots, n,$$

where

$$\widetilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_i & c\mathbf{Z}_i \end{pmatrix},\ \widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\Delta} \end{pmatrix},\ \boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top,\ \bar{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{\Psi}_i^{-1/2}\mathbf{Z}_i\mathbf{D}\boldsymbol{\zeta}}{\sqrt{1 + \boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}},$$

with $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{Z}_i/\sigma^2)^{-1}$ and $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$. For responses with censoring pattern as defined in (14), we have that $\mathbf{Y}_i \sim TSNI_{n_i}(\widetilde{\mathbf{X}}_i\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, H; \mathbb{A})$, where $TSNI_{n_i}(.; \mathbb{A})$ denotes the truncated SNI distribution on the interval $\mathbb{A}_i = A_{i1} \times \ldots, \times A_{ini}$, with A_{ij} as the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and $(-\infty, Q_{ij}]$ if $C_{ij} = 1$. Specifically, a p-dimensional vector $\mathbf{X} \sim TSNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H; \mathbb{A})$ if its density is given by $TSNI_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H; \mathbb{A}) = \{SNI_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)/\prod_{r=1}^p \int_{-\infty}^{a_r} SNI_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)d\mathbf{x}\}\mathbb{I}_{\{\mathbb{A}\}}(\mathbf{x})$, where the notation $\prod_{r=1}^p \int_{-\infty}^{a_r} = \int_{-\infty}^{a_1} \ldots \int_{-\infty}^{a_r}$ stand for the abbreviation of multiple integrals. When $\boldsymbol{\lambda} = 0$ we will use the notation $\mathbf{X} \sim TNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H; \mathbb{A})$ and its density by $TNI_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, H; \mathbb{A})$

For computing the likelihood function, the first step is to treat separately the observed and censored components of $\mathbf{Y}_i$. We partition $\mathbf{Y}_i$ into the observed and censored parts as $\mathbf{Y}_i = vec(\mathbf{Y}_i^o, \mathbf{Y}_i^c)$, such that, $C_{ij} = 0$ for all elements in $\mathbf{Y}_i^o$, and 1 for all elements in $\mathbf{Y}_i^c$. Define $\mathbf{Q}_i = vec(\mathbf{Q}_i^o, \mathbf{Q}_i^c)$, where $vec(.)$

denote the function which stacks vectors or matrices of the same number of columns. From Lemma 1 (see Appendix), we can write $\mathbf{Y}_i^o \sim SNI_{n_i^o}(\widetilde{\mathbf{X}}_i^o\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\Sigma}^{oo}{}_i^{1/2}\widetilde{\boldsymbol{\upsilon}}_i, H)$, where

$$\widetilde{\boldsymbol{\upsilon}}_i = \frac{\boldsymbol{\upsilon}_{1i} + \boldsymbol{\Sigma}_i^{oo\,-1}\boldsymbol{\Sigma}_i^{oc}\boldsymbol{\upsilon}_{2i}}{\sqrt{1 + \boldsymbol{\upsilon}_{2i}^\top \boldsymbol{\Sigma}_i^{cc.o}\boldsymbol{\upsilon}_{2i}}},$$

with $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{co\,-1}\boldsymbol{\Sigma}_i^{oc}$, $\boldsymbol{\upsilon}_i = \boldsymbol{\Sigma}_i^{-1/2}\bar{\boldsymbol{\lambda}}_i = (\boldsymbol{\upsilon}_{1i}^\top, \boldsymbol{\upsilon}_{2i}^\top)^\top$ and $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}$.

Let us denote $\alpha_i = P(\mathbf{y}_i^c \leq \mathbf{Q}_i^c | \mathbf{Y}_i^o = \mathbf{Q}_i^o, \boldsymbol{\theta})$, which can be directly obtained from Proposition 2.2. Now, following Vaida and Liu (2009a), the likelihood for cluster i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}) &= P(\mathbf{y}_i^c \leq \mathbf{Q}_i^c | \mathbf{Y}_i^o = \mathbf{Q}_i^o, \boldsymbol{\theta}) P(\mathbf{Y}_i^o = \mathbf{Q}_i^o | \boldsymbol{\theta}) \\ &= \alpha_i\, SNI_{n_i^o}(\mathbf{Q}_i^o | \widetilde{\mathbf{X}}_i^o\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\Sigma}^{oo}{}_i^{1/2}\widetilde{\boldsymbol{\upsilon}}_i, H) \end{aligned} \tag{15}$$

Therefore, the log-likelihood function for the observed data is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{\log \alpha_i + \log SNI_{n_i^o}(\mathbf{Q}_i^o | \widetilde{\mathbf{X}}_i^o\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\Sigma}^{oo}{}_i^{1/2}\widetilde{\boldsymbol{\upsilon}}_i, H)\},$$

which can be evaluated without much computational burden through the routines *integrate* and *mvtnorm()* available in R (Genz et al., 2008; R Development Core Team, 2009) along with proposition 2.1 and 2.2. This log-likelihood is used for Bayesian model selection and to develop a case-deletion influence diagnostics based on the Kullback-Leibler divergence as discussed in Cancho et al. (2010).

Although one might use standard EM-type algorithms for maximum-likelihood (ML) inference, we choose a Bayesian route primarily for computational simplicity. The 'M' step in the EM routine, and the high-dimensional integrals in our likelihood function can make the ML estimation quite complicated. Our Bayesian inference relies on the recent developments in Markov chain Monte Carlo (MCMC) algorithms which facilitates easy and straightforward implementation in conventional software like `WinBUGS`. The Bayesian proposition allows for full parameter uncertainty and does not depend on asymptotic results (Gelman et al., 2006). Interval estimates for model parameters or functions of model parameters can be easily obtained directly from the MCMC output.

3.2 Prior and posterior specifications

A key feature of this model is that it can be formulated in a flexible hierarchical representation as follows:

$$\mathbf{Y}_i|\mathbf{b}_i, \mathbf{C}_i, \mathbf{Q}_i, T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} TN_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma^2\mathbf{I}_{n_i}; \mathbb{A}), \tag{16}$$

$$\mathbf{b}_i|T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N_q(\boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma}), \tag{17}$$

$$T_i|U_i = u_i \overset{\text{ind}}{\sim} TN(0, u_i^{-1}; (0, \infty)) \tag{18}$$

$$U_i \overset{\text{ind}}{\sim} H(.; \boldsymbol{\nu}), \tag{19}$$

where the observed data for the i-th subject is $(\mathbf{Q}_i, \mathbf{C}_i)$, for $i = 1, \ldots, n$. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, $\mathbf{Q} = vec(\mathbf{Q}_1, \ldots, \mathbf{Q}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$.

It follows that the complete likelihood function associated with $(\mathbf{y}, \mathbf{b}, \mathbf{Q}, \mathbf{C}, \mathbf{t})$, is given by

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{Q}, \mathbf{C}, \mathbf{t}) \propto \prod_{i=1}^{n} \Bigg[\int_0^{\infty} TN_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma_e^2\mathbf{I}_{n_i}; \mathbb{A})\phi_q(\mathbf{b}_i; \boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma}) \times TN_1(t_i|0, u_i^{-1}, (0, \infty))dh(u_i; \boldsymbol{\nu}) \Bigg]. \quad (20)$$

In order to complete the Bayesian specification, we need to consider prior distributions to all the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda})^\top$. A popular choice to ensure posterior propriety in a LMM is to consider proper (but diffuse) conditionally conjugate priors (Hobert and Casella, 1996; Zhao et al., 2006). Following Lachos et al. (2009), we have

$$\begin{aligned} \boldsymbol{\beta} &\sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_\beta), \\ \sigma^2 &\sim IGamma(q_0/2, \lambda_0/2), \\ \boldsymbol{\Gamma} &\sim IWish_q(\boldsymbol{\Lambda}_0^{-1}, \nu_0), \\ \boldsymbol{\Delta} &\sim N_q(\boldsymbol{\Delta}_0, \mathbf{S}_\Delta), \end{aligned}$$

where $IGamma(a, b)$ denotes the inverse gamma distribution with mean $b/(a-1)$, $a > 1$, and $IWish_q(\mathbf{M}^{-1}, \nu_0)$ denotes the inverse Wishart distribution with mean $\mathbf{M}^{-1}/(\nu_0 - q - 1)$, $\nu_0 > q + 1$, where $\mathbf{M}$ is a $q \times q$ known positive definite matrix. Assuming elements of the parameter vector to be independent we consider that the joint prior distribution of all unknown parameters have density given by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\Gamma})\pi(\boldsymbol{\Delta}). \quad (21)$$

For the specific NI models, the prior for $\boldsymbol{\nu}$ was chosen accordingly as follows.

(i) *Skew-t model (ST-LMEC):* Here $\nu \sim \text{TExp}(\frac{\gamma}{2}; (2, \infty))$, i.e., the degrees of freedom parameter ν has a truncated exponential prior distribution on the interval $(2, \infty)$. This truncation point was chosen to assure finite variance.

(ii) *Skew-slash model (SSL-LMEC):* 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) *Skew-contaminated normal model (SCN-LMEC):* 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 ρ to achieve conjugacy.

Combining the likelihood function (20) and the prior distribution, the joint posterior density of all unobservable quantities is given by

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\Delta}, \mathbf{b}, \mathbf{t}, \mathbf{y}, \mathbf{u}|\mathbf{Q}, \mathbf{C}) \propto \prod_{i=1}^{n} \left[TN_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma_e^2\mathbf{I}_{n_i}; \mathbb{A})\phi_q(\mathbf{b}_i; \boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma}) TN_1(t_i|0, u_i^{-1}, (0, \infty))H(u_i|\boldsymbol{\nu}) \right] \pi(\boldsymbol{\theta}). \quad (22)$$

Distribution is analytically intractable, but MCMC methods such as Gibbs sampler and Metropolis-Hastings algorithm can be used to draw samples and derive inference from marginal posterior densities of interest. Given $\mathbf{b}$ and $\mathbf{u}$, all conditional posteriors are in the standard SN-LMEC form, and have the same form for any element of the SNI class. An outline of the conditional posterior densities are given in the Appendix.

4 Bayesian model selection and influence diagnostics

4.1 Model comparison

There exists a wide variety of model selection/assessment measures within the Bayesian toolbox. One of the most widely used criterion is the conditional predictive ordinate (CPO) statistic, derived from the posterior predictive distribution (Gelfand et al., 1992). Let $\boldsymbol{\mathcal{D}}$ be the full data and $\boldsymbol{\mathcal{D}}^{(-i)}$ denote the data with the ith 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 $CPO_i = \int_\Theta f(\mathbf{y}_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}^{(-i)})d\boldsymbol{\theta} = \left\{\int_\Theta \{\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})/f(\mathbf{y}_i|\boldsymbol{\theta})\}d\boldsymbol{\theta}\right\}^{-1}$. For our proposed models, a closed form of the CPO_i is not available. However, a Monte Carlo estimate of CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$ using a harmonic-mean (HM) identity (Dey et al., 1997b) given by $\widehat{CPO_i} = \{1/Q\sum_{q=1}^{Q} 1/f(\mathbf{y}_i|\boldsymbol{\theta}_q)\}^{-1}$, where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q$ is a post burn-in sample of size Q from $\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$. A summary statistic of the CPO_i's is the log pseudo-marginal likelihood (LPML), defined by $LPML = \sum_{i=1}^{n} \log(\widehat{CPO_i})$. Larger values of $LPML$ indicates better fit. 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 and Dey, 1994; Dey et al., 1997a) have been prescribed, they can be computationally challenging. As suggested by the Editor, 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.

Some other measures, like the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002b), the expected Akaike information criterion (EAIC) by Brooks (2002), and the expected Bayesian (or Schwarz) information criterion (EBIC), given in Carlin and Louis (2001) can also be used. These are based on the posterior mean of the deviance, which can be approximated as $\overline{D} = \sum_{q=1}^{Q} D(\boldsymbol{\theta}_q)/Q$, where $D(\boldsymbol{\theta}) = -2\sum_{i=1}^{n} \log\left[f(\mathbf{y}_i|\boldsymbol{\theta})\right]$. The DIC criterion can be estimated using the MCMC output as $\widehat{DIC} = \overline{D} + \widehat{\rho_D}$, where ρ_D is the effective number of parameters, defined as $E\{D(\boldsymbol{\theta})\} - D\{E(\boldsymbol{\theta})\}$, where $D\{E(\boldsymbol{\theta})\}$ is the deviance evaluated at the posterior mean. Similarly, the EAIC and EBIC can be estimated as $\widehat{EAIC} = \overline{D} + 2\#(\boldsymbol{\vartheta})$ and $\widehat{EBIC} = \overline{D} + \#(\boldsymbol{\vartheta})\log(n)$, where $\#(\boldsymbol{\vartheta})$ is the number of model parameters. Note that for all these criteria, the evaluation of the marginal likelihood $f(\mathbf{y}_i|\boldsymbol{\theta})$ is a key aspect, however for our proposed models (SN-LMEC) it can be easily computed from the results given in Subsection 3.1. Note that the CPO statistics (and associated LPML) compare between competing models marginally by averaging out all unknown parameters, which is fundamentally different to that of DIC, EAIC and EBIC which compares models conditional on all unknown parameters (including all fixed and random effects), penalizing models with more complexity. In this paper, we examine all of them to accommodate multiple viewpoints (Spiegelhalter et al., 2002a).

4.2 Bayesian case influence diagnostics

Our proposed regression models might be sensitive to the underlying model assumptions, so it is of interest to determine which subjects/observations might be influential for the analysis. Let $K(P, P_{(-i)})$ denote the K–L divergence between P and $P_{(-i)}$, defined as $K(P, P_{(-i)}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) \log \left[\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})/\pi(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}^{(-i)})\right] d\boldsymbol{\theta}$, where P denotes the posterior distribution of $\boldsymbol{\theta}$ for full data, and $P_{(-i)}$ denotes the posterior distribution of $\boldsymbol{\theta}$ without the i-th case. As pointed by Peng and Dey (1995) and Cancho et al. (2010), $K(P, P_{(-i)})$ can be expressed as a posterior expectation $K(P, P_{(-i)}) = \log E_{\boldsymbol{\theta}|\mathcal{D}} \left\{[f(\mathbf{Y}_i|\boldsymbol{\theta})]^{-1}\right\} + E_{\boldsymbol{\theta}|\mathcal{D}} \left\{\log\left[f(\mathbf{y}_i|\boldsymbol{\theta})\right]\right\} = -\log(CPO_i) + E_{\boldsymbol{\theta}|\mathcal{D}} \left\{\log\left[f(\mathbf{y}_i|\boldsymbol{\theta})\right]\right\}$, where $E_{\boldsymbol{\theta}|\mathcal{D}}(\cdot)$ denotes the expectation with respect to the joint posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$. A Monte Carlo estimate of the K-L divergence (see Cancho et al., 2010), is given by

$$\widehat{K(P, P_{(-i)})} = -\log(\widehat{CPO}_i) + \frac{1}{Q}\sum_{q=1}^{Q} \log\left[f(\mathbf{y}_i|\boldsymbol{\theta}_q)\right], \; i = 1, \ldots, n. \tag{23}$$

5 Applications: UTI Data

We illustrate the proposed methods with the analysis of the HIV UTI data previously analyzed using normal LMEC model. This is a study of 72 perinatally HIV-infected children (Saitoh et al., 2008; Vaida and Liu, 2009a). The data set is available in the R package $lmec$ (Vaida and Liu, 2009b). Primarily due to treatment fatigue, unstructured treatment interruptions (UTI) is common in this population. Suboptimal adherence can lead to ARV resistance and diminished treatment options in the future. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI were studied. The number of observations from baseline (month 0) to month 24 are 71, 62, 58, 57, 43, 34, 24, and 13, respectively. Out of 362 observations, 26 (7%) observations were below the detection limits (50 or 400 copies/mL) and were left-censored at these values. The individual profiles of viral load at different followup times after UTI is presented in Figure 2. Following Vaida and Liu (2009a), we consider a profile LME model with random intercepts b_i given by

$$y_{ij} = b_i + \beta_j + \epsilon_{ij}, \tag{24}$$

where y_{ij} is the log_{10} HIV RNA for subject i at time t_j, $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $t_4 = 6$, $t_5 = 9$, $t_6 = 12$, $t_7 = 18$, $t_8 = 24$. Vaida and Liu (2009) analyzed the same data set by fitting a N-LMEC from a frequentist perspective, but from Figure 1 it is clear that inference based on normality assumptions can be questionable. In our analysis, we assume a SNI-LMEC as defined in (12), (13) and (14). As prior choices, we have $\beta_j \sim N_1(\mathbf{0}, 10^3)$, $j = 1, \ldots, 8$, $\sigma^2 \sim IGamma(0.1, 0.1)$, $\Gamma \sim IGamma(0.1, 0.1)$, $\Delta \sim N(0, 0.001)$. Additionally, for the ST model we have $\nu \sim$ TExp$(0.1; (2, \infty))$, for the SSL model we have $\nu \sim$ Gamma$(0.1, 0.01)$ and for the SCN model, we have $\nu \sim \mathbf{Beta}(\mathbf{1}, \mathbf{1})$ and $\rho \sim$ Beta$(2, 2)$. We generated two parallel independent MCMC runs of size 100,000 with widely dispersed initial values, where the first 20,000 iterations (burn-in samples) were discarded for computing posterior estimates. To eliminate potential problems due to auto-correlation, we considered a spacing of size 40. The convergence of the MCMC chains were monitored using trace plots, auto-correlation (ACF) plots and Gelman-Rubin $\hat{R}$ diagnostics. The SSL and SCN models require much larger number of iterations as compared to the N, SN and ST models to converge. 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 prior.

Table 1 UTI data: Comparison between N-LMEC and SNI-LMEC models using various Bayesian model selection criteria. 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-LMEC	SN-LMEC	ST-LMEC	SSL-LMEC	SCN-LMEC
LPML	-423.90	-418.62	-375.26	-368.27	**-366.27**
	(1.37)	(3.14)	(1.81)	(1.04)	(1.07)
DIC	2491.82	2459.69	2206.11	2172.61	**2149.57**
EAIC	853.80	844.99	762.71	751.36	**746.45**
EBIC	892.71	887.81	809.41	798.06	**797.05**

Table 1 presents comparison of the four sub-classes with the N-LMEC using various model selection criteria discussed in Section 4. About 90% 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. Note that DIC picks the SCN-LMEC to be far superior (difference of about 23 units) to the nearest competitor (SSL-LMEC), although the LPML, EAIC and EBIC values were close. All the heavy tailed models (viz. ST, SSL and SCN) provide much better fit than the SN-LMEC for all the measures. The fit of N-LMEC is the worst providing a lot of concern for normality assumptions with this dataset. For the SNI-LMEC models, the estimated values of λ and the corresponding 95% credible intervals (CI) (as in Tables 2 and 5) justifies the left-skewed nature of the UTI data. For the ST- and SSL-LMEC models, as ν (the t degrees of freedom) $\rightarrow \infty$, they approach the SN-LMEC model as the limiting case. Because the estimate of ν is small, it indicates lack of adequacy of the normality (and skew-normality) assumptions.

Table 2 reports the posterior mean, standard deviations (sd) and 95% credible intervals (CI) of the model parameters after fitting the SCN-LMEC model (our best model). The posterior estimates for the other competing models are presented in Table 5 in the Appendix. Note that the posterior estimates of $\beta_1 - \beta_8$ (the slope parameters corresponding to the time points) for the SNI-LMEC models with heavy tails are quite close to those from the N and SN versions, however the 95% posterior CI of $\boldsymbol{\beta}$ are tighter (due to smaller estimated standard deviations), indicating that the three heavy-tailed models seem to produce more precise estimates. As in Vaida and Liu (2009a), our dropout (censored) model does not bias the inference regarding the mean of β_j. The mean viral load $E(y_{ij}) = \beta_j$ increases gradually throughout 24 months for all the models. In particular for the SCN-LMEC, it increases from 3.86 at the time of UTI to 4.68 at 24 months. The estimates of the between-subject variance $(Var(b))$ and within-subject variance (σ^2) (in $\log_{10}$ scale) are 1.23 and 0.13 respectively. In addition to $\lambda < 0$ indicating left-skewness, the (small) posterior mean estimates of ν and ρ reveal presence of tail behavior.

To determine presence of possible influential observations, K-L divergence measures (23) are presented in Figure 2 for the competing models. Although it appears from Figure 2 that there might be more 'outlying' subjects/observations, the individual profiles for subject # 20 and 42 were quite different from the

Table 2 Posterior parameter estimates obtained after fitting the SC-LMEC model to the UTI data.'sd' denotes standard deviation and 'CI' denotes credible intervals.

Parameter	Mean	sd	95% CI
β_1	3.86	0.119	[3.61 ; 4.09]
β_2	4.18	0.119	[3.93 ; 4.41]
β_3	4.21	0.122	[3.95 ; 4.46]
β_4	4.37	0.120	[4.14 ; 4.61]
β_5	4.53	0.128	[4.25 ; 4.76]
β_6	4.51	0.132	[4.25 ; 4.77]
β_7	4.55	0.145	[4.25 ; 4.83]
β_8	4.68	0.164	[4.37 ; 4.99]
σ^2	0.13	0.015	[0.10 ; 0.16]
$Var(b)$	1.23	0.335	[0.74 ; 1.94]
λ	-7.09	2.827	[-12.47 ; -2.27]
ν	0.121	0.031	[0.07 ; 0.19]
ρ	0.127	0.024	[0.08 ; 0.18]

fitted (mean) profiles for the various distributions. Our outlier analysis was designed to detect the 'most' influential observations, hence we choose cases 20 and 42 which have larger $K(P, P_{(-i)})$ for N-LMEC models as compared to the others. For the SN-LMEC, case 20 no longer remains an outlier. In particular, the effect of these 2 cases (and possibly others) nullified when considering our SNI class, in particular the SSL and SCN densities. As expected, the posterior estimates of $E(\beta)$ were attenuated when the heavy-tailed SNI distributions were considered, suggesting that the SNI-LMEC model is a robust alternative for censored viral load data in presence of asymmetry. This is also observed in Figure 3 where the presence of these outliers might have overestimated the predicted mean curve for the N-LMEC model as compared to the SNI-LMEC. The fitted viral load trajectories for some randomly chosen subjects are presented in Figure 4. To summarize, these results suggest that our proposed class of SNI-LMEC models provide precise posterior parameter estimates for our motivating dataset on HIV viral load that exhibits departure from the traditional normality assumptions due to skewness and (or) presence of heavy tails.

6 Simulation studies

In this section, we conduct a simulation study to illustrate the performance of our proposed methodology. The goal of this simulation study is to investigate the consequences on parameter inference when the normality assumption is inappropriate as well as to investigate whether the model comparison measures, viz., LPML, DIC, EAIC and EBIC determines the best-fitting model to the simulated data. We assume the following linear mixed-effects model with random slope and intercept, given by:

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_{0i} + b_{1i} t_{ij} + e_{ij},\ i = 1, \ldots, 100,\ \ j = 1, \ldots, 5, \tag{25}$$

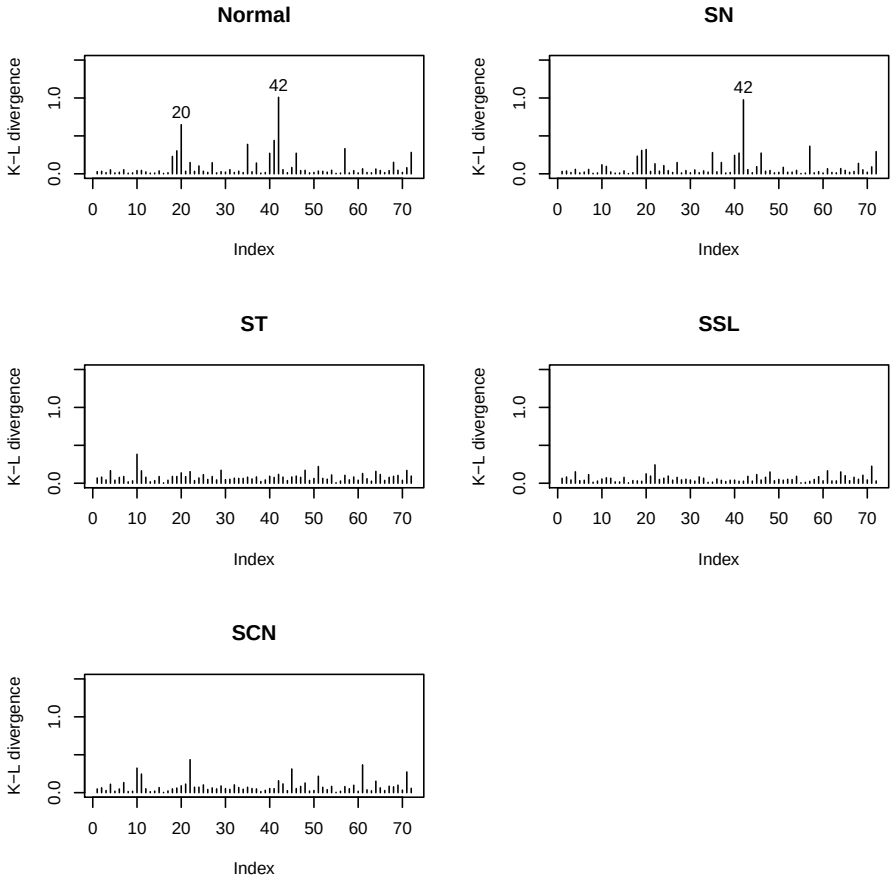

Figure 2 UTI data: Estimated KL-divergence measures $K(P, P_{(-i)})$

where $\mathbf{b}_i = (b_{0i}, b_{1i}) \overset{\text{iid}}{\sim} p_1 \mathrm{N}_2(\cdot|\boldsymbol{\xi}_1, \boldsymbol{\Omega}_1) + p_2 t_2(\cdot|\boldsymbol{\xi}_2, \boldsymbol{\Omega}_2, \nu)$, $\quad \epsilon_{ij} \sim t(\cdot|0, \sigma^2, \nu)$, with

$$\boldsymbol{\xi}_1 = (2,2),\ \ \boldsymbol{\xi}_2 = (0,0),\ \ \sigma^2 = 2,\ \ \nu = 3,\ \ \boldsymbol{\Omega}_1 = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix},\ \ \boldsymbol{\Omega}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$(p_1, p_2) = (0.25, 0.75),\ \ \boldsymbol{\beta} = (0.05, 0.1).$$

Thus, the joint density of the random slope and the random intercept vector $\mathbf{b}_i$ comes from a mixture of a bivariate normal and a bivariate $t-$ distribution with 3 degrees of freedom. This yields a skewed thick-tailed unimodal distribution inspired by an example presented in Ho and Hu (2008). For each subject $i = 1, \ldots, 100$ and fixed time-point vector $\mathbf{t}_i = (1, 2, 3, 4, 5)^\top$, we set $x_{1ij} = 1$ if $j \geq 3$ and 0 otherwise – an indicator variable corresponding to an 'intervention (treatment)' effect at $t = 3$. The co-variate x_{2ij} is a random sample from a $\mathrm{N}(3, 3)$ distribution. To study the effect of the level of censoring on the posterior estimates, we choose various settings of censoring proportions, say 10%, 20% and 40%. Once 500 simulated datasets are generated for each of these 3 settings, we fit the N-LMEC, SN-LMEC and ST-LMEC models using *R2WinBUGS* package available in R. For each of the settings, the following independent

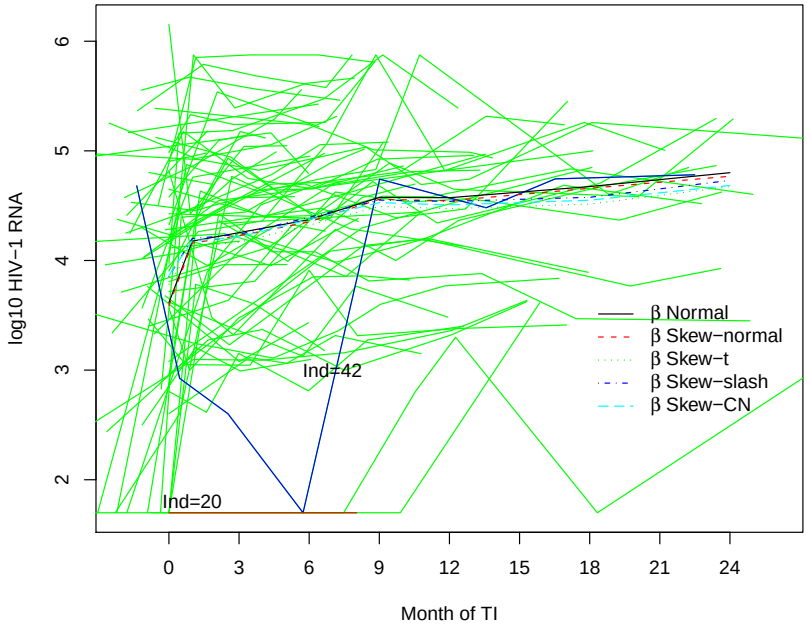

Figure 3 Individual profiles and overall mean (in $\log_{10}$ scale) using the 4 members of the SNI class and the Normal model at different follow-up times post-UTI. The trajectories of the influential observations are numbered.

priors are considered for the MCMC sampling: $\beta_k \sim N_1(\mathbf{0}, 10^3)$, $k = 1, 2$ $\sigma^2 \sim IGamma(0.1, 0.1)$, $\mathbf{\Gamma} \sim IWish_2(\mathbf{H}^{-1}, 2)$ with $\boldsymbol{H} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$ and $\Delta_i \sim N(0, 0.001)$, $i = 1, 2$ and in addition $\nu \sim \text{TExp}(0.1; (2, \infty))$ for the ST model. The MCMC scheme follows exactly as in Section 5. We compute the 'Relative Bias' (RelBias) and 'Mean Squared Error' (MSE) for each parameters over the 500 samples under the 3 different settings. They are defined as:

$$\text{RelBias}(\gamma) = \frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\gamma}^{(i)}/\widehat{\gamma} - 1\right) \text{ and } \text{MSE}(\gamma) = \frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\gamma}^{(i)} - \gamma\right)^2,$$

where $\boldsymbol{\gamma} = (\beta_1, \beta_2, \sigma^2)$ and $\widehat{\gamma}^{(i)}$ is the posterior estimate of $\boldsymbol{\gamma}$ for the ith sample. In addition, we also estimate the Power of the intervention effect β_1, i.e. the proportion of times the 95% credible interval of β_1 excludes 0.

From Table 3, we observe that the ST-LMEC model has the smallest RelBias and MSE for β_1 and β_2 for all levels of censoring, however both the RelBias and MSE increases with increase in the censoring proportion as expected. The ST model also detects the right-skewed-heavy tailed feature of the simulated data with posterior estimates of ν to be small, however the estimates drifts farther with increasing proportion of censoring. Thus, the ST model (accommodating both skewness and thick tails) produces more

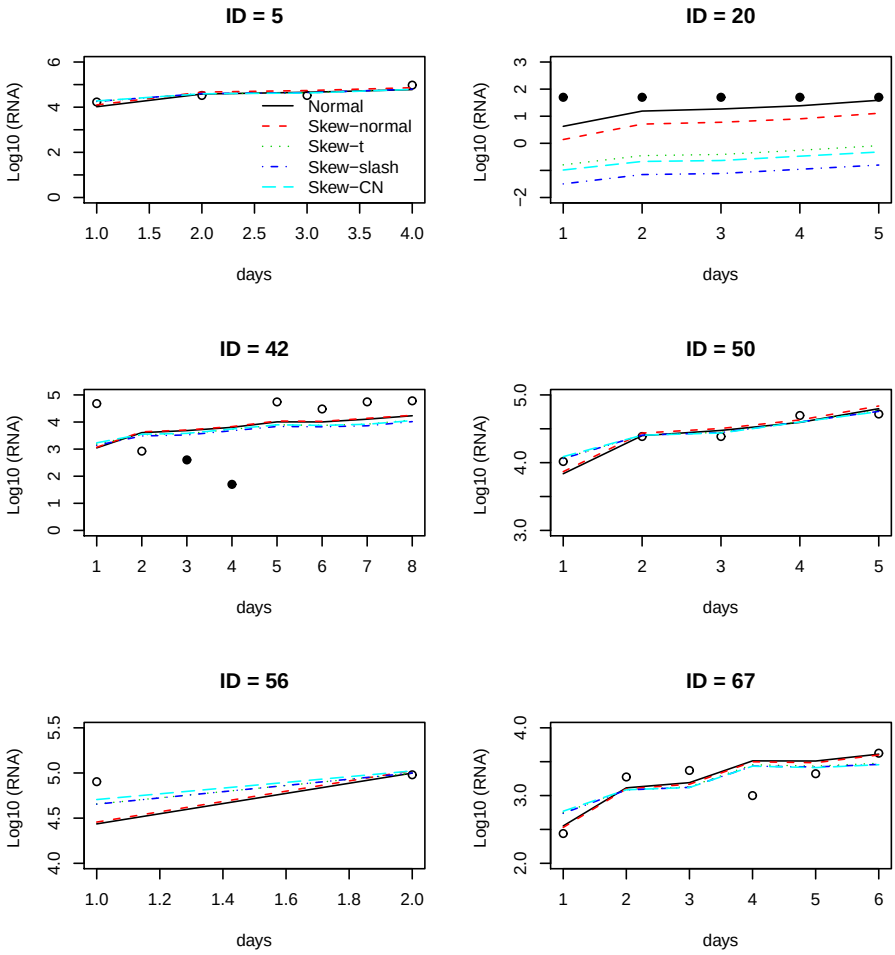

Figure 4 UTI Data. Individual viral load trajectory estimates for six randomly chosen subjects after fitting the 4 SNI models. The bold line indicates the Normal fit. Censored observations are represented by filled circles, whereas the observed ones by not-filled circles.

accurate Bayesian estimates in the context of censored data; the degree and direction of the bias in fixed effects depends both on the relative proportions of censoring as well as model assumptions. However, the power to detect the intervention effect (posterior of β_1) was higher for the SN model as compared to the N and ST models across all the 3 scenarios, however the power decreases as the proportion of censoring increases. In Table 4, we present the arithmetic averages (MC LPML, MC DIC, MC EAIC and MC EBIC) of the various model comparison measures mentioned earlier. We notice that all these measures favored the ST-LMEC model for our (true) simulated data demonstrating the ability of these Bayesian selection methods to detect an obvious departure from normality. The % of samples (out of the 500 samples) when these criteria chooses the ST-LMEC model also remains high.

Table 3 Monte Carlo simulation results based on 500 simulated datasets comparing the N, SN and ST models for various levels of censoring.

			Posterior estimates of parameters				
Censored	Fit		β_1	β_2	σ^2	ν (mean and s.d.)	
10%	Normal	RelBias	-0.053	0.0121	0.024	-	-
		MSE	2.869	8.547	2.524		
		Power	0.28				
	Skew-Normal	Bias	-0.046	0.0123	0.0369		
		MSE	2.882	8.547	2.516		
		Power	0.34				
	Skew-T	Bias	-0.034	0.011	0.0383	6.759 (0.508)	
		MSE	2.776	8.542	2.511		
		Power	0.31				
20%	Normal	Bias	-0.077	0.016	0.029	-	-
		MSE	2.906	8.562	2.544		
		Power	0.26				
	Skew-Normal	Bias	-0.069	0.016	0.036		
		MSE	2.893	8.563	2.536		
		Power	0.31				
	Skew-T	Bias	-0.043	0.015	0.042	6.819 (0.807)	
		MSE	2.819	8.554	2.532		
		Power	0.29				
40%	Normal	Bias	-0.087	0.021	0.035	-	-
		MSE	2.923	8.607	2.550		
		Power	0.13				
	Skew-Normal	Bias	-0.074	0.021	0.029		
		MSE	2.908	8.606	2.566		
		Power	0.22				
	Skew-T	Bias	-0.055	0.018	0.036	6.921(1.009)	
		MSE	2.831	8.391	2.564		
		Power	0.16				

7 Conclusions

This article proposes Bayesian implementation of a robust alternative to the linear mixed-effects model with censored response, where the Gaussian distribution of the random terms are replaced by the skew-normal/independent distribution. We apply our methodology to a recent AIDS study (freely downloadable from R) to illustrate how the procedure developed can be used to evaluate model assumptions, identify outliers and obtain robust parameter estimates. Depending on assay quantifications, censoring can be both left or right. Our application is based on right-censoring, considerations for right-censoring is immediate

Table 4 Monte Carlo estimates of various model comparison measures. 'perc' is the % of samples that the criteria choose a ST-LMEC properly.

		Criteria			
Censored	Fit	MC B	MC DIC	MC EAIC	MC EBIC
10%	Normal	-2502.27	9970.12	4995.58	5029.93
	Skew-Normal	-2680.12	10453.78	5240.28	5284.45
	Skew-T	-2484.16	9868.17	4948.54	4997.62
	perc	82%	89%	84%	78%
20%	Normal	-2303.73	9091.89	4556.34	4605.42
	Skew-Normal	-2610.53	10057.94	5039.18	5083.35
	Skew-T	-2267.035	9026.95	4524.01	n4558.37
	perc	83%	87%	85%	76%
40%	Normal	-1963.02	7681.79	3864.34	3913.42
	Skew-Normal	-2544.04	9159.87	4573.77	4617.94
	Skew-T	-1808.25	7193.08	3607.05	3641.40
	perc	85%	87%	87%	77%

and follows from (14) by reversing the role of y_{ij} and Q_{ij}. It is worth emphasizing that recent papers by Huang and Dagne (2011, 2010) provides an ad-hoc treatment to the censoring phenomenon in HIV viral load studies; they replaced censored data with half the value of the quantification limit (QL), which might lead to bias estimates. This paper provides a first attempt to incorporate censoring in the context of skew-normal/independent linear mixed-effects models (SNI-LMEC).

Our model assumes the censoring mechanism to be 'missing-at-random (MAR)', hence conditional on a correct model the estimation of the mean viral loads remains unbiased (Vaida and Liu, 2009a). Our method provides improvement over results from Vaida and Liu (2009a), who considered analysis of this dataset using normal linear mixed-effects models. Simulation studies reveal gain in efficiency and accuracy for parameter estimates as well as performance of various model selection techniques to pick the best-fitting model, where typical assumptions of normality are questionable.

In our SNI proposition, the random error terms follow homoscedastic Normal/Independent (Lange and Sinsheimer, 1993) distribution typically used for thick tails, with the premise that modeling of any associated skewness (in the response) is delegated to the random effects term through appropriate skewed versions of various densities that are members of the SNI class. As pointed out by a reviewer, introduction of serially (or spatially) correlated random error terms or a doubly-skewed framework (with skewed random errors in addition) is certainly possible but comes with additional complications in its estimation, interpretation as well as (latent) identifiability issues as to how the skewness parameter in the random error term might be related to the skewness in the random effects. It can be further complicated due to unevenly spaced data and also the choice of an appropriate time-series/ARIMA structure for the within-subject time profile. Exploring all these is beyond the current scope of this paper and will be pursued elsewhere.

The models considered in this paper can be fitted using standard available software packages, like `R` and `WinBUGS` (code available as supplementary material associated with this paper), and this makes our approach quite powerful and accessible to practitioners in the field and thus provide guidance on using appropriate statistical models for handling complicated HIV viral load responses.

Acknowledgements D. Bandyopadhyay acknowledges funding support from Grants P20RR017696-06, R03DE020114 and R03DE021762 from the US National Institutes of Health (NIH). The research of V.H.Lachos was supported by Grant 2008/201384-6 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and Grant 2008/11455-0 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). The research of L. M. Castro is supported by Grant FONDECYT 11100076 from the Chilean government.

Conflict of Interest
The authors have declared no conflict of interest.

Appendix

Appendix A: Proof of Lemma 1 and Propositions 2.1 and 2.2

Lemma 1 Let $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$ and $\mathbf{Y}$ is partitioned as $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ with dimensions p_1 and p_2 ($p_1 + p_2 = p$), respectively. Let

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \ \ \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{1/2}$, $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. Then, the marginal density of $\mathbf{Y}_1$ is $SNI_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}}, H)$, where

$$\widetilde{\boldsymbol{\upsilon}} = \frac{\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}.$$

with $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\upsilon} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} = (\boldsymbol{\upsilon}_1^\top, \boldsymbol{\upsilon}_2^\top)^\top$.

Proof. See Lachos et al. (2010). □

Proof of Proposition 2.1: (i) From (1) and from Arellano-Valle and Genton (2005), we have

$$\begin{aligned} P(\mathbf{Y} \leq \mathbf{y}|u) &= 2\int_{\mathbf{w} \leq \mathbf{0}} \int_{V \leq 0} \phi(\mathbf{w}| - \mathbf{y} + \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\phi(V + \boldsymbol{\lambda}^\top u^{-1/2}\boldsymbol{\Sigma}^{1/2}(\mathbf{w} + \mathbf{y} - \boldsymbol{\mu}))dV d\mathbf{w} \\ &= 2P(\mathbf{W} \leq 0, V \leq 0|u), \end{aligned}$$

where $\mathbf{W}|U = u \sim N(-\mathbf{y} + \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})$ and $V|\mathbf{W} = \mathbf{w}, U = u \sim N(-\boldsymbol{\lambda}^\top u^{-1/2}\boldsymbol{\Sigma}^{1/2}(\mathbf{w} + \mathbf{y} - \boldsymbol{\mu}), 1)$. The proof follows from the fact that

$$\begin{pmatrix} \mathbf{W} \\ V \end{pmatrix} \Big| U = u \Bigg) \sim N_{p+1}\left(\begin{pmatrix} -\mathbf{y} + \boldsymbol{\mu} \\ 0 \end{pmatrix}, u^{-1} \begin{pmatrix} \boldsymbol{\Sigma} & -u^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda} \\ -u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{1/2} & u(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}) \end{pmatrix} \right)$$

and that $P(\mathbf{Y} \leq \mathbf{y}) = \int_0^\infty P(\mathbf{Y} \leq \mathbf{y}|u)dH(u; \boldsymbol{\nu})$.

The proof of (ii) follows from the mixture representation of the SNI class.

Proof of Proposition 2.2: i) It is easy to note that, if $U = 1$, we are in the skew–normal case and then, the conditional distribution of $\mathbf{Y}_2|\mathbf{Y}_1$ is given by

$$\begin{aligned} f(\mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1, \boldsymbol{\theta}) &= \phi_{p_2}(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}) \frac{\Phi(\boldsymbol{\upsilon}^\top(\mathbf{y} - \boldsymbol{\mu}))}{\Phi(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}, \\ &= \phi(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}) \frac{\Phi(z_1 + \boldsymbol{\upsilon}_2^\top(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))}{\Phi(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}, \end{aligned}$$

where $z_1 = (\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2)^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. Then, the proof follows similarly as in Proposition 2.1, part i). In addition from Lemma 2 given in Lachos et al. (2010), we have

$$\begin{aligned} E[\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1, \boldsymbol{\theta}] &= \boldsymbol{\mu}_{2.1} + W_\Phi(\bar{c})\mathbf{G}, \\ E[\mathbf{Y}_2\mathbf{Y}_2^\top|\mathbf{Y}_1 = \mathbf{y}_1, \boldsymbol{\theta}] &= \boldsymbol{\mu}_{2.1}\boldsymbol{\mu}_{2.1}^\top + \boldsymbol{\Sigma}_{22.1} + W_\Phi(\bar{c})\left\{\boldsymbol{\mu}_{2.1}\mathbf{G}^\top + \mathbf{G}\boldsymbol{\mu}_{2.1}^\top - \bar{c}\mathbf{G}\mathbf{G}^\top\right\}, \end{aligned}$$

where $\mathbf{G} = \dfrac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}$, with $W_\Phi(\bar{c}) = \frac{\phi(\bar{c})}{\Phi(\bar{c})}$ and $\bar{c} = \widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{Y}_1 - \boldsymbol{\mu}_1)$.

ii) From Lee et al. (2010), $\mathbf{Y}_2|\mathbf{Y}_1 \sim EST(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \boldsymbol{\nu} + p_1)$, where EST is the extended skew–t distribution proposed by Arrellano-Valle and Genton (2010). The proof follows by using the form of the EST cdf also proposed by Arrellano-Valle and Genton (2010).

Appendix B: Outline of the conditional distributions

Our Bayesian model allows a straightforward construction of a Gibbs sampler through the hierarchical representation given in (12)-(14). To proceed, it is necessary to obtain the conditional distribution of one variable given values of all the remaining - $(\mathbf{C}_i, \mathbf{Q}_i)$ included. We have the following expressions:

1. $\mathbf{y}_i|\mathbf{b}_i, u_i, \mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\theta} \sim f(\mathbf{y}_i|\mathbf{b}_i, u_i, \mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\theta})$. Thus, conditional on $(\mathbf{b}_i, u_i)$, $\mathbf{y}_i$ is a vector of independent observations, whose distributions are truncated normal, each with untruncated variance $u_i^{-1}\sigma^2$ and untruncated mean $\mathbf{x}_{ij}^\top\boldsymbol{\beta} + \mathbf{z}_{ij}^\top\mathbf{b}_i$, on the interval $y_{ij} \leq Q_{ij}$, i.e. $TN_1(\mathbf{x}_{ij}^\top\boldsymbol{\beta} + \mathbf{z}_{ij}^\top\mathbf{b}_i, u_i^{-1}\sigma^2; (-\infty, Q_{ij}))$.

2. $\mathbf{b}_i|\mathbf{y}_i, u_i, t_i, \mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\theta} \equiv \mathbf{b}_i|\mathbf{y}_i, u_i, t_i, \boldsymbol{\theta} \sim f(\mathbf{b}_i|\mathbf{Y}_i, t_i, \boldsymbol{\theta})$. This distribution is multivariate normal with mean $\widehat{\mathbf{b}}_i = \boldsymbol{\Lambda}_i(\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) + t_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})$. and variance $u_i^{-1}\boldsymbol{\Lambda}_i$ with $\boldsymbol{\Lambda}_i = (\boldsymbol{\Gamma}^{-1} + \mathbf{Z}_i^\top\mathbf{Z}_i/\sigma^2)^{-1}$. Note that the entire vector $\mathbf{Y}_i$ is used for sampling from $\mathbf{b}_i$.

3. $T_i|\mathbf{y}_i, u_i, \mathbf{b}_i, \mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\theta} \equiv T_i|\mathbf{y}_i, u_i, \mathbf{b}_i, \boldsymbol{\theta} \sim TN_1(A_{t_i}a_{t_i}, A_{t_i}; (c, \infty))$, where $A_{t_i} = u_i^{-1}(1 + \boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1}$ and $a_{t_i} = (\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\mathbf{b}_i + c)$.

4. $U_i|\mathbf{y}_i, \mathbf{b}_i, t_i, \mathbf{C}_i, \mathbf{Q}_i, \boldsymbol{\theta} \equiv \pi(u_i|\mathbf{y}_i, \mathbf{b}_i, t_i, \boldsymbol{\theta}) \propto u_i^{(n_i+1+q)/2} \exp\{-\frac{u_i}{2}(\frac{\boldsymbol{\epsilon}_i^\top\boldsymbol{\epsilon}_i}{\sigma^2} + (\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i) + t_i^2\}h(u_i|\boldsymbol{\nu})$, with $\boldsymbol{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i$.

5. Now, by observing that $\boldsymbol{\theta}_1|\mathbf{y}, u_i, \mathbf{C}, \mathbf{Q}, \mathbf{b}_i, t_i, \boldsymbol{\theta}_{(-\boldsymbol{\theta}_1)}$ and $\boldsymbol{\theta}_1|\mathbf{y}, u_i, \mathbf{b}_i, t_i, \boldsymbol{\theta}_{(-\boldsymbol{\theta}_1)}$ are two equivalent process, we have:

$$\begin{aligned}
\boldsymbol{\beta}|\mathbf{y}, \mathbf{u}, \mathbf{b}, \mathbf{t}, \boldsymbol{\theta}_{(-\boldsymbol{\beta})} &\sim N\left(\mathbf{A}_\beta \boldsymbol{\mu}_\beta,\ \mathbf{A}_\beta\right), \\
\sigma^2|\mathbf{y}, \mathbf{u}, \mathbf{b}, \mathbf{t}, \boldsymbol{\theta}_{(-\sigma^2)} &\sim IGamma(\frac{q_0+N}{2}, \frac{\lambda_0+s}{2}), \\
\boldsymbol{\Gamma}|\mathbf{y}, \mathbf{u}, \mathbf{b}, \mathbf{t}, \boldsymbol{\theta}_{(-\boldsymbol{\alpha})} &\sim IWish_q(\boldsymbol{\Lambda}^{-1}, \nu_0+n), \\
\boldsymbol{\Delta}|\mathbf{y}, \mathbf{u}, \mathbf{b}, \mathbf{t}, \boldsymbol{\theta}_{(-\boldsymbol{\lambda})} &\sim N_q(\mathbf{A}_\Delta \mathbf{a}_\Delta, \mathbf{A}_\Delta),
\end{aligned}$$

where $\boldsymbol{\mu}_\beta = (\mathbf{S}_\beta^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^n u_i \mathbf{X}_i^\top(\mathbf{y}_i - \mathbf{Z}_i\mathbf{b}_i))$, $\mathbf{A}_\beta = (\mathbf{S}_\beta^{-1} + \sum_{i=1}^n u_i \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i)^{-1}$, $N = \sum_{i=1}^n n_i$, $s = \sum_{i=1}^n u_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$, $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0 + \sum_{i=1}^n u_i(\mathbf{b}_i - \boldsymbol{\Delta}t_i)(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top$, $\mathbf{A}_\Delta = (\mathbf{S}_\Delta^{-1} + \boldsymbol{\Gamma}^{-1}\sum_{i=1}^n u_i t_i^2)^{-1}$ and $\mathbf{a}_\Delta = \mathbf{S}_\Delta^{-1}\boldsymbol{\Delta}_0 + \boldsymbol{\Gamma}^{-1}\sum_{i=1}^n u_i t_i \mathbf{b}_i$.

5. To complete the Gibbs sampling specifications, we need the full conditional posterior distributions of $\boldsymbol{\nu}$. This density is $\pi(\boldsymbol{\nu}|\mathbf{y}, \mathbf{b}, \mathbf{t}, \mathbf{u}, \boldsymbol{\theta}_{(-\boldsymbol{\nu})}) \propto \pi(\boldsymbol{\nu}) \prod_{i=1}^n h(u_i|\boldsymbol{\nu})$. These conditionals have been given in Lachos et al. (2009).

Appendix C: Posterior estimates

Table 5 Posterior parameter estimates from fitting the N, SN, ST and SSL sub-classes of the LMEC model to the UTI data. 'sd' denotes standard deviation and 'CI' denotes credible intervals.

	N-LMEC			SN-LMEC		
Parameter	Mean	sd	95% CI	Mean	sd	95% CI
β_1	3.65	0.131	[3.39 ; 3.91]	3.61	0.130	[3.34 ; 3.85]
β_2	4.17	0.133	[3.91 ; 4.44]	4.14	0.133	[3.87 ; 4.40]
β_3	4.24	0.135	[3.98 ; 4.51]	4.21	0.135	[3.94 ; 4.47]
β_4	4.36	0.136	[4.10 ; 4.63]	4.33	0.132	[4.05 ; 4.59]
β_5	4.56	0.142	[4.29 ; 4.85]	4.54	0.141	[4.25 ; 4.82]
β_6	4.58	0.151	[4.29 ; 4.88]	4.54	0.151	[4.24 ; 4.84]
β_7	4.69	0.171	[4.36 ; 5.03]	4.65	0.169	[4.30 ; 4.98]
β_8	4.80	0.207	[4.41 ; 5.21]	4.76	0.206	[4.36 ; 5.18]
σ^2	0.33	0.030	[0.28 ; 0.40]	0.33	0.030	[0.28 ; 0.40]
$Var(b)$	0.78	0.153	[0.53 ; 1.15]	0.84	0.192	[0.54 ; 1.27]
λ	-	-	-	-10.31	3.422	[-16.29 ; -3.66]
	ST-LMEC			SSL-LMEC		
Parameter	Mean	sd	95% CI	Mean	sd	95% CI
β_1	3.78	0.129	[3.53 ; 4.04]	3.84	0.123	[3.56 ; 4.07]
β_2	4.12	0.128	[3.88 ; 4.39]	4.19	0.125	[3.89 ; 4.41]
β_3	4.17	0.129	[3.91 ; 4.41]	4.23	0.128	[3.93 ; 4.47]
β_4	4.32	0.130	[4.08 ; 4.58]	4.38	0.122	[4.11 ; 4.59]
β_5	4.49	0.139	[4.21 ; 4.78]	4.53	0.129	[4.25 ; 4.77]
β_6	4.47	0.144	[4.20 ; 4.75]	4.53	0.134	[4.24 ; 4.77]
β_7	4.51	0.152	[4.21 ; 4.81]	4.56	0.148	[4.26 ; 4.82]
β_8	4.67	0.179	[4.29 ; 4.98]	4.71	0.168	[4.33 ; 5.05]
σ^2	0.16	0.026	[0.11 ; 0.21]	0.098	0.0151	[0.07 ; 0.13]
$Var(b)$	1.29	0.681	[1.02 ; 2.25]	1.27	0.480	[0.67 ; 2.32]
λ	-7.21	3.172	[-13.11 ; -2.03]	-6.13	2.280	[-10.49 ; -2.06]
ν	4.67	0.785	[3.44 ; 6.36]	1.65	0.196	[1.32 ; 2.12]

References

Arellano-Valle, R. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**, 561–574.

Arellano-Valle, R. and Genton, M. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis* **96**, 93–116.

Arellano-Valle, R. B., Branco, M. D. and Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics* **34**.

Arrellano-Valle, R. and Genton, M. (2010). Multivariate extended skew-t distributions and related families with applications. *Metron, Special issue on "Skew-symmetric and flexible distributions"* **68**, 201–234.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**, 159–188.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**, 579–602.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002) **64**, 616–618.

Cancho, V., Dey, D., Lachos, V. and Andrade, M. (2010). Bayesian nonlinear regression models with scale mixtures of skew normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis* **55**, 588–602.

Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2nd edition.

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

De la Cruz, R. and Branco, M. (2009). Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. *Biometrical Journal* **51**, 588–609.

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

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

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

Gelfand, A. E., Dey, D. K. and 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., Carlin, J. and Rubin, D. (2006). *Bayesian Data Analysis*. Chapman & Hall/CRC, New York, NY.

Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F. and Scheipl, F. (2008). mvtnorm: Multivariate Normal and t Distribution. *R package version 0.9-2, URL http://CRAN. R-project. org/package= mvtnorm* .

Ghidey, W., Lesaffre, E. and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60**, 945–953.

Ghosh, P., Branco, M. and Chakraborty, H. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Statistics in Medicine* **26**, 1255–1267.

Gray, L., Cortina-Borja, M. and Newell, M. (2004). Modelling HIV-RNA viral load in vertically infected children. *Statistics in Medicine* **23**, 769–781.

Ho, H. and Lin, T. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical Journal* **52**, 449–469.

Ho, R. K. W. and Hu, I. (2008). Flexible modelling of random effects in linear mixed modelsa bayesian approach. *Computational Statistics & Data Analysis* **52**, 1347–1361.

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

Huang, W., De Gruttola, V., Fischl, M., Hammer, S., Richman, D., Havlir, D., Gulick, R., Squires, K. and Mellors, J. (2001). Patterns of plasma human immunodeficiency virus type 1 RNA response to antiretroviral therapy. *The Journal of infectious diseases* **183**, 1455–1465.

Huang, Y. and Dagne, G. (2010). Skew-normal Bayesian nonlinear mixed-effects models with application to AIDS studies. *Statistics in Medicine* **29**, 2384–2398.

Huang, Y. and Dagne, G. (2011). A Bayesian Approach to Joint Mixed-Effects Models with a Skew-Normal Distribution and Measurement Errors in Covariates. *Biometrics* **67**, 260–269.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics* **55**, 625–629.

Jacqmin-Gadda, H., Thiebaut, R., Chene, G. and Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* **1**, 355–368.

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

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. and Montenegro, L. C. C. (2007). Likelihood-Based Inference for Multivariate Skew-Normal Regression Models. *Communications in Statistics-Theory and Methods* **36**, 1769–1786.

Lachos, V. H., Dey, D. K. and 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.

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010). Likelihood based inference for skew–normal independent linear mixed models. *Statistica Sinica* **20**, 303–322.

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

Lange, K. L. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat* **2**, 175–198.

Lee, S., Genton, M. and Arellano-Valle, R. (2010). Perturbation of numerical confidential data via skew-t distributions. *Management Science* **56**, 318–333.

Lin, T. and Lee, J. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in medicine* **27**, 1490–1507.

Mellors, J., Rinaldo Jr, C., Gupta, P., White, R., Todd, J. and Kingsley, L. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167–1170.

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

Qiu, W. and Wu, L. (2010). HIV Viral Dynamic Models With Censoring and Informative Dropouts. *Statistics in Biopharmaceutical Research* **2**, 220–228.

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

Raftery, A., Newton, M., Satagopan, J. and 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.

Sahu, S., Dey, D. and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* **31**, 129–150.

Saitoh, A., Foca, M., Viani, R., Heffernan-Vacca, S., Vaida, F., Lujan-Zilbermann, J., Emmanuel, P., Deville, J. and Spector, S. (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics* **121**, e513.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002b). Bayesian measures of model complexity and fit **64**, 583–639.

Vaida, F., Fitzgerald, A. and DeGruttola, V. (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational statistics & data analysis* **51**, 5718–5730.

Vaida, F. and Liu, L. (2009a). Fast Implementation for Normal Mixed Effects Models With Censored Response. *Journal of Computational and Graphical Statistics* **18**, 797–817.

Vaida, F. and Liu, L. (2009b). *lmec: Linear Mixed-Effects Models with Censored Responses*. R package version 1.0.

Verbeke, G. and Lesaffre, E. (1996). A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association* **91**, 217–221.

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, Boca Raton, FL.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57**, 795–802.

Zhao, Y., Staudenmayer, J., Coull, B. and Wand, M. (2006). General design Bayesian generalized linear mixed models. *Statistical Science* **21**, 35–51.