Robust Linear Mixed Models with Skew–Normal Independent Distributions from a Bayesian Perspective

Victor H. Lachos^{a1}**, Dipak K. Dey**^{b} **and Vicente G. Cancho**^{c}

^{a} Departamento de Estatística, Universidade Estadual de Campinas, Brazil

^{b} Department of Statistics, University of Connecticut, USA

^{c} Departamento de Matemática Aplicada e Estatística, Universidade de São Paulo, Brazil

Abstract

Linear mixed models were developed to handle clustered data and have been a topic of increasing interest in statistics for the past fifty years. Generally, the normality (or symmetry) of the random effects is a common assumption in linear mixed models but it may, sometimes, be unrealistic, obscuring important features of among-subjects variation. In this article, we utilize skew-normal/independent distributions as a tool for robust modeling of linear mixed models under a Bayesian paradigm. The skew-normal/independent distributions is an attractive class of asymmetric heavy-tailed distributions that includes the skew-normal distribution, the skew-t distribution, the skew-slash distribution and the skew contaminated normal distribution as special cases, providing an appealing robust alternative to the routine use of symmetric distributions in this type of models. The methods developed are illustrated using a real data set from Framingham cholesterol study.

Keywords and phrases: Gibbs Algorithms; Linear mixed models; MCMC; Metropolis-Hastings; Skew-normal/independent distribution

1 Introduction

Linear mixed model (LMM; Laird and Ware, 1982) has become the most frequently used analytic tool for longitudinal data analysis with continuous repeated measures. A linear mixed model consists of a fixed effects and random effects. The random effects account for the between–subject variation. In a linear mixed model framework

---

^{1}*Address for correspondence*: Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, Caixa Postal 6065, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: hlachos@ime.unicamp.br.

it is routinely assumed that the random effects and the within–subject measurement error have a normal distribution. While this assumption makes the model easy to apply in widely used softwares such as SAS, the accuracy of this assumptions is difficult to check and the routine use of normality is recently questioned by many authors (Verbeke and Lesaffre, 1997; Pinheiro, 2001; Zhang and Davidian, 2001; Ghidey et al., 2004; Lin and Lee, 2007). Normality assumption is too restrictive as it suffers from the lack of robustness against departures from the normal distribution, particularly when data show multimodality and skewness, and thus may not provide an accurate estimation of between–subject variation. For example, Zhang and Davidian (2001) showed that the estimated subject–specific intercept from the Framingham heart study data was not normally distributed and thus use of normal distribution in this scenario may obscure important features of between–subject variation. Thus, it is of practical interest to develop statistical model with considerable flexibility in the distributional assumptions of the random effects as well as measurement error.

There has been considerable work in this direction. Verbeke and Lesaffre (1996) introduce a heterogeneous linear mixed model where random effects distribution is relaxed using a finite mixture of normal. Pinheiro et al. (2001) proposed a multivariate t–linear mixed model and showed that it would perform well in the presence of outliers. Zhang and Davidian (2001) proposed an LMM in which the random effects follow the so–called seminonparametric (SNP) distribution. Rosa et al. (2003) adopted a Bayesian framework to carry out posterior analysis in LMM with the thick–tailed class of normal/independent (NI) distributions (Lange and Sinsheimer, 1993). Ma et al. (2004) consider a generalized flexible skew–elliptical distribution for the random effects density and proposed algorithms for maximum likelihood (ML) estimation and Bayesian inference via Markov Chain Monte Carlo (MCMC). Lachos et al. (2007) and Jara et al. (2008) propose a Bayesian approach for drawing inferences in skew–normal LMM (SN-LMM) and skew-t LMM proposed by Sahu et al. (2003), respectively, and noted that these robust models can be easily fitted in freely available software with equivalent computational effort than the one necessary to fit the normal (and symmetric) version of the model.
Recently, from a frequentist perspective, Lachos et al. (2008) proposed the skew–normal/independent linear mixed model (SNI–LMM) based on the class of scale mixtures of skew-normal distributions introduced by Branco and Dey (2001). This generalization has the potential to make the inference robust to departures from

a normal distribution, once an entire class of distributions is defined. As a special case, this new class contains, the skew-normal distribution (Azzalini and Dalla–Valle, 1996), skew-t distribution (Azzalini and Capitanio, 2003; Gupta, 2003), skew-slash distribution (Wang and Genton, 2006), the skew-contaminated normal distribution and all the symmetric family of NI distribution studied by Lange and Sinsheimer (1993). To our knowledge, although Bayesian analysis to skewed LMM has appeared in the literature, the Bayesian approach to SNI-LMM has never been studied. Thus, the objective of this paper is to propose a Bayesian approach for drawing inferences in SNI-LMM, which is closely related to the one proposed by Rosa et al. (2003) in a symmetric context. Our motivating data set is from the Framingham cholesterol study whose distribution of the random effects has been found to be non–normal and positively skewed by Zhang and Davidian (2001), Lin and Lee (2007), Lachos et al. (2008), among others.

The rest of the article is organized as follows. After a brief introduction of SNI distributions in Section 2, the SNI–LMM is presented in Section 3 as well as the Bayesian formulation of the model. Prior distributions and joint posterior density are also discussed. The measures of model selection are included in Section 4. The advantage of the proposed methodology is illustrated in Section 5 using the Framingham cholesterol data and finally, some concluding remarks are presented in Section 6.

2 Skew-normal/independent distribution

A SNI distribution is a process of generating a $p-$dimensional random vector of the form

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

where $\boldsymbol{\mu}$ is a location vector, U is a positive random variable with cumulative distribution function (cdf) $H(u|\boldsymbol{\nu})$ and probability density function (pdf) $h(u|\boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or vector of parameters indexing the distribution of U, which is a positive value and $\mathbf{Z}$ is a multivariate skew–normal random vector (Arellano–Valle et al., 2005) with location vector $\mathbf{0}$, scale matrix $\boldsymbol{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda}$. In usual notation, we write $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Given U, $\mathbf{Y}$ follows a multivariate skew–normal distribution with location vector $\mathbf{0}$, scale matrix $u^{-1}\boldsymbol{\Sigma}$ and skewness

parameter vector $\boldsymbol{\lambda}$, i.e., $\mathbf{Y}|U = u \sim \text{SN}_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. The marginal pdf of $\mathbf{Y}$ is:

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

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 $\boldsymbol{\Sigma}$, $\Phi(.)$ represents the cdf of the standard univariate normal distribution. We will use the notation $\mathbf{Y} \sim \text{SNI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$. When $\boldsymbol{\lambda} = \mathbf{0}$, the SNI distributions reduces to the normal–independent (NI) class, i.e., the class of scale–mixture 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 will use the notation $\mathbf{Y} \sim \text{NI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H)$ when $\mathbf{Y}$ has distribution in the NI class. Some of these distributions are described subsequently.

2.1 Multivariate skew–t distribution

The multivariate ST distribution (Azzalini and Capitanio, 2003; Gupta, 2003) with ν degrees of freedom, $\text{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, can be derived from the mixture model (2), by taking U to be distributed as $Gamma(\nu/2, \nu/2)$, $u > 0$, $\nu > 0$. The pdf of $\mathbf{Y}$ is:

$$f(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)T\left(\sqrt{\frac{p + \nu}{d + \nu}}\text{A}, \nu + p\right), \quad \mathbf{y} \in \mathbb{R}^p,$$

where, $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T(\cdot, \nu)$ denote, respectively, the pdf of the p–variate Student–t distribution, namely $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and the cdf of the standard univariate t–distribution. The mean and the variance of this skew-t distribution are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\boldsymbol{\Delta}, \ \ \nu > 1$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma} - \frac{\nu}{\pi}(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})})^2\boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \ \nu > 2.$$

where $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. A particular case of the ST distribution is the skew–Cauchy distribution, when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the SN distribution as the limiting case. Applications of the ST distribution to robust estimation can be found in Lin et al. (2007) and Azzalini and Genton (2007).

2.2 Multivariate skew–slash distribution

Another SNI distribution, termed as multivariate skew–slash distribution denoted by $\text{SSL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, arise when the distribution of U is $Beta(\nu, 1)$, $0 < u < 1$ and

$\nu > 0$. Its pdf is given by

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \Phi(u^{1/2}A) du, \quad \mathbf{y} \in \mathbb{R}^p.$$

From (1), it can be shown that

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \frac{2\nu}{2\nu - 1} \boldsymbol{\Delta}, \ \nu > 1/2, \ \text{and}$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu - 1}\boldsymbol{\Sigma} - \frac{2}{\pi}(\frac{2\nu}{2\nu - 1})^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \ \nu > 1.$$

The SL distribution reduces to the SN distribution when $\nu \uparrow \infty$. Applications of the SL distribution can be found in Wang and Genton (2006).

2.3 Multivariate skew–contaminated normal distribution

The Multivariate SCN distribution is denoted by $SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$, $0 \le \nu \le 1$, $0 < \gamma \le 1$. Here, U is a discrete random variable taking one of two states. The probability function of U, given the parameter vector $\boldsymbol{\nu} = (\nu, \gamma)^\top$, is denoted by

$$h(u|\boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \ \ 0 < \nu < 1, 0 < \gamma \le 1,$$

It follows then that $f(\mathbf{y}) = 2\{\nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma})\Phi(\gamma^{1/2}A) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A)\}$. In this case, we have

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}(\frac{\nu}{\gamma^{1/2}} + 1 - \nu)\boldsymbol{\Delta},$$

$$Var[\mathbf{Y}] = (\frac{\nu}{\gamma} + 1 - \nu)\boldsymbol{\Sigma} - \frac{2}{\pi}(\frac{\nu}{\gamma^{1/2}} + 1 - \nu)^2\boldsymbol{\Delta}\boldsymbol{\Delta}^\top.$$

The SCN distribution reduces to the SN one when $\gamma = 1$.

3 The SNI–LMM

Following Lachos et al. (2008), we consider a generalization of the classical N–LMM in which the random errors are assumed to have a NI distribution and the random effects are assumed to have a multivariate SNI distributions within the class defined in (1). Simultaneously, the heavy-tail asymmetric model can be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{3}$$

with the assumption that

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \stackrel{ind.}{\sim} SNI_{n_i+q} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda}_b \\ \mathbf{0} \end{pmatrix}, H \right), i = 1, \ldots, n, \quad (4)$$

where the subscript i is the subject index, $\mathbf{Y}_i$ 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 $\mathbf{b}_i$, and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors. The matrices $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\gamma})$, $i = 1, \ldots, n$, are dispersion matrices, corresponding to the within and between subjects variability, and depend on unknown and reduced parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively. Finally, as was indicated in the previous section, $H = H(\cdot; \boldsymbol{\nu})$ is the cdf-generator that determines the specific SNI model that we are assuming. From now on, we assume that $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}) = \sigma_e^2 \mathbf{R}_i$, with $\mathbf{R}_i$ known, thus $\boldsymbol{\gamma} = \sigma_e^2$ is a scalar parameter. We refer to Lachos et al. (2008) for details and additional interesting properties on this proposed model.

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$, is based on the marginal distribution for $\mathbf{Y}_i$, which is given by (Lachos et al., 2008)

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = 2 \int_0^\infty \phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta}, u_i^{-1}\boldsymbol{\Psi}_i)\Phi\left(u_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right) dH(u_i|\boldsymbol{\nu}), \quad (5)$$

i.e., $\mathbf{Y}_i \stackrel{ind.}{\sim} \mathrm{SNI}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}_i, \bar{\boldsymbol{\lambda}}_i; H)$, $i = 1, \ldots, n$, where $\boldsymbol{\Psi}_i = \sigma_e^2 \mathbf{R}_i + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$,

$$\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \sigma_e^{-2}\mathbf{Z}_i^\top \mathbf{R}_i^{-1}\mathbf{Z}_i)^{-1} \text{ and } \bar{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{\Psi}_i^{-1/2}\mathbf{Z}_i \mathbf{D}\boldsymbol{\zeta}}{\sqrt{1 + \boldsymbol{\zeta}^\top \boldsymbol{\Lambda}_i \boldsymbol{\zeta}}},$$

with $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$. It follows that the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is given by

$$\ell(\boldsymbol{\theta}) \propto -\sum_{i=1}^n \log\left[\int_0^\infty \phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta}, u_i^{-1}\boldsymbol{\Psi}_i)\Phi\left(u_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right) dH(u_i; \boldsymbol{\nu})\right],$$
$$(6)$$

The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta}$ can be obtained by direct maximization of (6) or alternatively by using the EM–type algorithm proposed in Lachos et al. (2008), while inference for the parameters will be based on the asymptotic normality of the MLEs (Cox and Hinkley, 1974). Instead, 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.

Bayesian hierarchical modeling is very attractive due to its flexibility. It allows for full parameter uncertainty and Bayesian inference 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. We note that conditional on u_i, some derivations are common to all the members of the SNI-LMM (see Appendix A). In the next section, we discuss aspects of model (3) and (4) that are shared by all members of the family considered, and in the Appendix we address aspects relevant to specific cases.

3.1   Prior distributions and joint posterior density

In this section, we implement the Bayesian methodology using MCMC techniques for the SNI-LMM. A key feature of this model, which allows writing easily BUGS codes, is that it can be formulated in a flexible hierarchical representation as follows:

$$\mathbf{Y}_i|\mathbf{b}_i, U_i = u_i \overset{\text{ind.}}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma_e^2\mathbf{R}_i), \tag{7}$$

$$\mathbf{b}_i|T_i = t_i, U_i = u_i \overset{\text{ind.}}{\sim} N_q(\boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma}), \tag{8}$$

$$T_i|U_i = u_i \overset{\text{ind.}}{\sim} HN_1(0, u_i^{-1}), \tag{9}$$

$$U_i \overset{\text{iid.}}{\sim} H(u_i|\boldsymbol{\nu}), \tag{10}$$

$i = 1, \ldots, n$, where $HN_1(0, \sigma^2)$ is the half-$N_1(0, \sigma^2)$ distribution, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\delta}$ and $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, with $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$ and $\mathbf{D}^{1/2}$ being the square root of $\mathbf{D}$ containing $q(q+1)/2$ distinct elements. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$ it follows that the complete likelihood function associated with $(\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$, is given by

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t}) \propto \prod_{i=1}^{n}[\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma_e^2\mathbf{R}_i)\phi_q(\mathbf{b}_i|\boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma})$$
$$\times \phi_1(t_i|0, u_i^{-1})h(u_i|\boldsymbol{\nu})]. \tag{11}$$

Now, to complete the Bayesian specification of the model we need to consider prior distribution to all the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$ Since we have no prior information from historical data or from previous experiment, we assign conjugate but weakly informative prior distributions to the parameters. Note that since we have assumed informative (but weakly) prior distribution, posterior is well defined and proper. Assuming elements of the parameter vector to be independent we consider that the joint prior distribution of all unknown parameters have density

given by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma_e^2)\pi(\mathbf{\Gamma})\pi(\mathbf{\Delta})\pi(\boldsymbol{\nu}). \tag{12}$$

The prior on $\boldsymbol{\beta}$ is taken to be normal with known hyperparameters $\boldsymbol{\beta}_0$ and $\mathbf{S}_\beta$. The scale parameter σ_e^2 is taken to be $IG(\frac{\tau_e}{2}, \frac{T_e}{2})$ (inverted Gamma) with density function given by

$$\pi(\sigma_e^2) = \frac{\left(\frac{T_e}{2}\right)^{\frac{\tau_e}{2}}}{\Gamma(\frac{\tau_e}{2})} \left(\frac{1}{\sigma_e^2}\right)^{\frac{\tau_e}{2}+1} \exp\left\{-\frac{T_e}{2\sigma_e^2}\right\}.$$

The family of the random effects were chosen mainly for computational simplicity. The inverted Wishart ($IW_\tau(\mathbf{T})$) distribution was used as prior for the matrix $\mathbf{\Gamma} = \mathbf{D} - \mathbf{\Delta}\mathbf{\Delta}^\top$. The prior for $\mathbf{\Delta}$ is taken to be Normal with known hyperparameters $\mathbf{\Delta}_0$ and $\mathbf{S}_\Delta$. Finally, the prior distribution of $\boldsymbol{\nu}$, with density $\pi(\boldsymbol{\nu})$, depends on the particular SNI distribution we use. Combining the likelihood function (11) and the prior distribution, the joint posterior distribution for $\boldsymbol{\theta}$ is given by,

$$
\begin{aligned}
\pi(\boldsymbol{\beta}, \sigma_e^2, \mathbf{\Gamma}, \mathbf{\Delta}, \mathbf{b}, \mathbf{u}, \mathbf{t}|\mathbf{y}) \quad \propto \quad & \prod_{i=1}^{n} \Big[\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\mathbf{R}_i\sigma_e^2) \\
& \times \phi_q(\mathbf{b}_i|\mathbf{\Delta}t_i, u_i^{-1}\mathbf{\Gamma})\phi_1(t_i|0, u_i^{-1})h(u_i|\boldsymbol{\nu})\Big] \pi(\boldsymbol{\theta}). \tag{13}
\end{aligned}
$$

Distribution (13) 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. The Gibbs sampler works by drawing samples iteratively from conditional posterior distributions deriving from (13). Given $\mathbf{u}$, all conditional posterior distributions are as in a standard SN-LMM and have the same form for any element of the SNI class. These have closed form and will be present in the following Proposition:

Proposition 1. *Under the full model as described in (13), given $\mathbf{u}$, the full conditional distribution of $\boldsymbol{\beta}$, σ_e^2, $\mathbf{\Delta}$, $\mathbf{\Gamma}$, $\mathbf{b}_i$, t_i, $i = 1, \ldots, n$, are given by*

$$\boldsymbol{\beta}|\mathbf{b}, \mathbf{u}, \mathbf{t}, \sigma_e^2, \mathbf{\Delta}, \mathbf{\Gamma} \quad \sim \quad N_p(\mathbf{A}_\beta^{-1}\mathbf{a}_\beta, \ \mathbf{A}_\beta^{-1}), \tag{14}$$

where $\mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \frac{1}{\sigma_e^2}\sum_{i=1}^{n} u_i \mathbf{X}_i^\top \mathbf{R}_i^{-1}\mathbf{X}_i$ and $\mathbf{a}_\beta = \mathbf{S}_\beta^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma_e^2}\sum_{i=1}^{n} u_i \mathbf{X}_i^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{Z}_i\mathbf{b}_i)$;

$$\sigma_e^2|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}, \mathbf{\Delta}, \mathbf{\Gamma} \quad \sim \quad IG(\frac{N + \tau_e}{2}, \ \frac{T_e + \sum_{i=1}^{n} u_i \boldsymbol{\mu}_i^\top \mathbf{R}^{-1}\boldsymbol{\mu}_i}{2}), \tag{15}$$

where $N = \sum_{i=1}^{n} n_i$ and $\boldsymbol{\mu}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i$;

$$\mathbf{\Delta}|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}, \sigma_e^2, \mathbf{\Gamma} \quad \sim \quad N(\Sigma_\Delta^{-1}\boldsymbol{\mu}_\Delta, \ \Sigma_\Delta^{-1}), \tag{16}$$

where $\boldsymbol{\mu}_\Delta = \mathbf{S}_\Delta^{-1}\boldsymbol{\Delta}_o + \boldsymbol{\Gamma}^{-1}\sum_{i=1}^n u_i \mathbf{b}_i$, $\boldsymbol{\Sigma}_\Delta = \boldsymbol{\Gamma}^{-1}\sum_{i=1}^n u_i t_i^2 + \mathbf{S}_\Delta^{-1}$;

$$\boldsymbol{\Gamma}|\mathbf{b},\mathbf{u},\mathbf{t},\boldsymbol{\beta},\sigma_e^2,\boldsymbol{\Delta} \ \sim\ IW_{\tau_b+n}((\mathbf{T}_b^{-1} + \sum_{i=1}^n u_i(\mathbf{b}_i - \boldsymbol{\Delta}t_i)(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top)^{-1}); \qquad (17)$$

$$\mathbf{b}_i|\boldsymbol{\beta},\sigma_e^2,\boldsymbol{\Delta},\boldsymbol{\Gamma},\mathbf{u}_i,\mathbf{t}_i \ \sim\ N_q(\mathbf{A}_{bi}^{-1}\mathbf{a}_i,\ u_i^{-1}\mathbf{A}_{bi}^{-1}), \qquad (18)$$

where $\mathbf{A}_{bi} = (\frac{1}{\sigma_e^2}\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i + \boldsymbol{\Gamma}^{-1})$ *and* $\mathbf{a}_i = \frac{1}{\sigma_e^2}\mathbf{Z}_i^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) + t_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}$, $i = 1,\ldots,n$;

$$T_i|\boldsymbol{\beta},\sigma_e^2,\boldsymbol{\Gamma},\boldsymbol{\Delta},\mathbf{b}_i,u_i \ \sim\ N(A_t^{-1}a_{ti},\ u_i^{-1}A_t^{-1})\mathbb{I}\{T_i > 0\}, \qquad (19)$$

where $A_t = (1 + \boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})$, $a_{ti} = \mathbf{b}_i^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}$, $i = 1,\ldots,n$. *Finally*

$$\mathbf{D} = \boldsymbol{\Gamma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^\top \quad and \quad \boldsymbol{\lambda} = \mathbf{D}^{-1/2}\boldsymbol{\Delta}/(1 - \boldsymbol{\Delta}^\top\mathbf{D}^{-1}\boldsymbol{\Delta})^{1/2}.$$

Proof. All of the full conditional distributions are straightforward to derive by working the complete joint posterior and using Lemma 2 as described in Arellano–Valle et al. (2005). □

To complete the specifications for a Gibbs sampling scheme, we need the full conditional posterior distributions of $\mathbf{u}$ and parameter $\boldsymbol{\nu}$. For each element of $\mathbf{u}$, the density is:

$$\begin{aligned}
\pi(u_i|\boldsymbol{\theta},\mathbf{y},\mathbf{b},\mathbf{t}) \ \propto\ & u_i^{n_i/2+q/2+1/2}h(u_i|\boldsymbol{\nu}) \\
& \times \exp\left\{-\frac{1}{2\sigma_e^2}u_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\right. \\
& \left. -\frac{1}{2}u_i(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i) - \frac{1}{2}u_i t_i^2\right\}, \qquad (20)
\end{aligned}$$

$i = 1,\ldots,n$. For $\boldsymbol{\nu}$, the density is:

$$\pi(\boldsymbol{\nu}|\boldsymbol{\theta},\mathbf{y},\mathbf{b},\mathbf{u},\mathbf{t}) \propto \pi(\boldsymbol{\nu})\prod_{i=1}^n h(u_i|\boldsymbol{\nu}). \qquad (21)$$

The form of (20) depends on the specific SNI distribution adopted and, in the case of (21), on the form of prior distribution of $\boldsymbol{\nu}$ (see appendix A).

4 Models comparison

Model diagnostics and comparison measures based on the posterior predictive densities are often easier to work with in MCMC model fitting settings. MCMC methods

are able to produce these measures without much extra effort. See for example Gelfand (1996). In the discussion below, first we describe one criterion to perform Bayesian model choice and then, we describe some other Bayesian criteria.

Let $\mathbf{y}_{obs}$ with components $y_{i,obs}$; $i = 1, \ldots, n$ denote the set of observed values of $\mathbf{y}$. Similarly we use the notation $\mathbf{y}_{rep}$ with components $y_{i,rep}$ to denote a future set of observations under the assumed model (here obs and rep are the abbreviations for the observation and replicate, respectively). Let $\boldsymbol{\theta}$ denote the set of parameters of the current model.

The posterior predictive density, $\pi(\mathbf{y}_{rep}|\mathbf{y}_{obs})$, is the predictive density of a new independent set of observable, $\mathbf{y}_{rep}$, under the assumed model given the actual data set of observables, $\mathbf{y}_{obs}$. By marginalizing $\pi(\mathbf{y}_{rep}|\mathbf{y}_{obs})$ we obtain the posterior predictive density of one observation $y_{i,rep}$, $i = 1, \ldots, n$, as follows,

$$\pi(y_{i,rep}|\mathbf{y}_{obs}) = \int \pi(y_{i,rep}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}. \tag{22}$$

Let $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ denote the posterior predictive mean and covariance of $y_{i,rep}$ under the density (22). We can easily estimate $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ by Monte Carlo integration as follows. Suppose that $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ denote R Gibbs samples from $\pi(\boldsymbol{\theta}|\mathbf{y}_{obs})$. Then, a random sample $y_{i,rep}^{(r)}$ drawn from $\pi(y_{i,rep}|\boldsymbol{\theta}^{(r)})$, is a sample from the predictive density (22). See for example Gelfand (1996). To perform model choice, first, we considered the Bayesian criteria called L- measure proposed by Laud and Ibrahim (1995). It is defined as the expected squared Euclidean distance between the vector of observations, $\mathbf{y}_{obs}$, and the vector of future observations, $\mathbf{y}_{rep}$, i.e., $L = E\left[(\mathbf{y}_{rep} - \mathbf{y}_{obs})^\top(\mathbf{y}_{rep} - \mathbf{y}_{obs})\right]$, where the expectation is taken with respect to the posterior predictive distribution given in (22). Straightforward algebra shows that L measure can be written as

$$L = \sum_{i=1}^{n} tr(\boldsymbol{\Sigma}_i) + \sum_{i=1}^{n} (\boldsymbol{\mu}_i - y_{i,obs})^\top(\boldsymbol{\mu}_i - y_{i,obs}),$$

and thus L can be written as a sum of two terms, one involving the predictive variances and the other term is like a bias term involving the squared difference between the predictive means and the observed data.

Many other Bayesian criteria have been proposed in the literature. In this paper we are also going to consider some of these Bayesian model selection criteria: the The Deviance Information Criterion (DIC; Spiegelhalter et al., 2002), the Expected Akaike Information Criteria (EAIC), the Expected Bayesian Information Criteria

(EBIC) as proposed by Carlin and Louis (2000) and Brooks (2002), and the conditional predictive ordinate (CPO) (Gelfand, Dey and Chang 1992). Each of these model criteria is simple to compute as the relevant quantities can be calculated directly from the MCMC output.

The criteria DIC, EAIC and EBIC are based on the posterior mean of the deviance, i.e., $E\left[D(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})\right]$ which is also a measure of fit and can be approximated by using the MCMC output, considering the value of

$$\bar{D} = \frac{1}{B} \sum_{b=1}^{B} D(\boldsymbol{\beta}^{(b)}, (\sigma_e^2)^{(b)}, \boldsymbol{\alpha}^{(b)} \boldsymbol{\lambda}^{(b)}, \boldsymbol{\nu}^{(b)}),$$

where B represents the number of iterations, and

$$\begin{aligned} D(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= -2log(f(\mathbf{y}|\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})) \\ &= -2\sum_{i=1}^{n} log(f(\mathbf{y}_i|\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})), \end{aligned}$$

where $f(\mathbf{y}_i|\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is given in (5). The criteria EAIC, EBIC and DIC can be estimated using MCMC output by considering

$$\widehat{EAIC} = \bar{D} + 2p, \quad \widehat{EBIC} = \bar{D} + plog(N) \text{ and } \widehat{DIC} = \bar{D} + \widehat{\rho_D} = 2\bar{D} - \widehat{D},$$

respectively, where p is the number of parameters in the model, N is the total number of observations. The ρ_D, is the effective number of parameters as described in (Spiegelhalter et al., 2002), and is defined as

$$\rho_D = E\left[D(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})\right] - D(E[\boldsymbol{\beta}], E[\sigma_e^2], E[\boldsymbol{\alpha}], E[\boldsymbol{\lambda}], E[\boldsymbol{\nu}]).$$

The term $D(E[\boldsymbol{\beta}], E[\sigma_e^2], E[\boldsymbol{\alpha}], E[\boldsymbol{\lambda}], E[\boldsymbol{\nu}])$ is the deviance of posterior mean obtained when considering the mean values of the generated posterior means of the model parameters, which is estimated by

$$\widehat{D} = D\left(\frac{1}{B}\sum_{b=1}^{B}\boldsymbol{\beta}^b, \frac{1}{B}\sum_{b=1}^{B}(\sigma_e^2)^b, \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{\alpha}^b, \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{\lambda}^b, \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{\nu}^b\right).$$

Smaller value of the DIC, EBIC and EAIC, implies better fit of the model. The CPO is a cross-validated predictive approach i.e., predictive distributions conditioned on the observed data with a single data point deleted. Chen, Shao, and Ibrahim (2000, chapter, 10) show in detail how to obtain Monte Carlo estimates of the CPO statistic. For model comparison we use the *log pseudo marginal likelihood*

(LPML) defined by $\text{LPML} = \sum_i \log \widehat{CPO}_i$ where $\widehat{CPO}_i$ is the Monte Carlo estimate of the i th subjects CPO statistic. Models with greater LPML values will indicate a better fit.

Figure 1: Histogram and normal Q-Q plots of empirical bayes estimates of: In the first row subject–specific intercepts and the second row the subject–specific slopes.
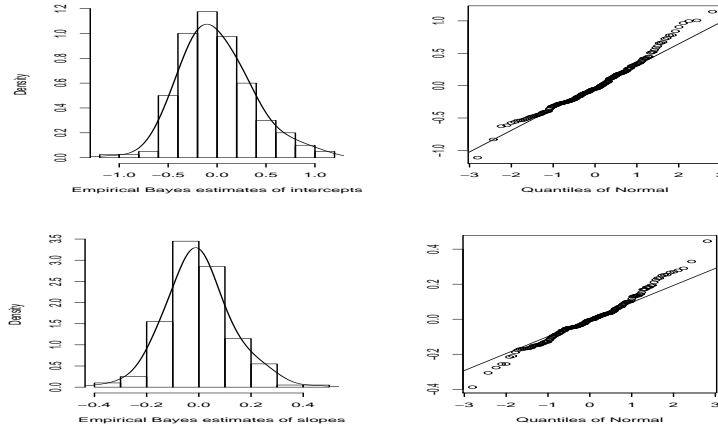
5   Illustrative example

The Framingham heart study has examined the role of serum cholesterol as a risk factor for the evolution of cardiovascular disease. Zhang and Davidian (2001) proposed a semiparametric approach to analyze a subset of the Framingham cholesterol data, which consist of gender, baseline age and cholesterol levels measured at the beginning of the study and then every two years over a period of ten years, for 200 randomly selected participants. Lachos at al. (2008), analyzed the same data set by fitting a SNI–LMM from a frequentis perspective. In this section, we revisit the Framingham cholesterol data with the aim of providing additional inferences by using MCMC methods.

Assuming a linear growth model with subject–specific random intercept and slopes, we fit a LMM model to the data as specified by Zhang and Davidian (2001)

$$Y_{ij} = \beta_o + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \beta_3 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij} \, , \tag{23}$$

where Y_{ij} is the cholesterol level, divided by 100, at the j-th time for subject i; t_{ij} is $(\text{time} - 5)/10$, with time measured in years from the start of the study; age_i is age at the start of the study; sex_i is the gender indicator ($0 = \text{female}, 1 = \text{male}$). Thus, $\mathbf{x}_{ij} = (1, \text{sex}_i, \text{age}_i, t_{ij})^\top$, $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$ and $\mathbf{Z}_{ij} = (1, t_{ij})^\top$. To verify the

existence of skewness in the random effects, we start by fitting an ordinary N–LMM. Figure 1 depicted histograms and corresponding envelopes of the empirical Bayes estimates of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i = \mathbf{DZ}_i^\top \mathbf{\Psi}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$ and shows that there are no apparent non-normal pattern for subject–specific slopes. However, the subject–specific intercept are positively skewed and therefore the suggested Gaussian model did not fit well. Moreover, the QQ – plots clearly support the use of heavy-tailed distributions.

We reanalysed the data within a Bayesian perspective, adopting the model in (23) using SNI distributions. The following independent priors were considered to perform the Gibbs sampler. $\beta_k \sim N_1(\mathbf{0}, 10^3)$, $k = 0, 1, 2, 3$, $\Delta_k \sim N_1(0, 10^3)$, $k = 1, 2$, $1/\sigma_e^2 \sim Gamma(0.1, 0.01)$, $\mathbf{\Gamma} \sim IW_3(\mathbf{T})$ with $\boldsymbol{T} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$ and $\nu \sim$ Exp$(0.1)I(3,)$ for the skew-t model, $\nu \sim$ Gamma$(0.01, 0.001)$ for the skew-slash model, $\nu \sim U(0, 1)$ and $\gamma \sim U(0, 1)$, for the skew-contaminated normal model. Considering these prior densities we generated two parallel independent runs of the Gibbs sampler chain with size 25000 for each parameter, disregarding the first 5000 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 2000. To monitor the convergence of the Gibbs samples we used the methods recommended by Cowles and Carlin (1996).

Table 1: Comparison between SNI–LMM by using different Bayesian criteria.

criterion	SN–LMM	ST–LMM	SSL–LMM	SCN–LMM
Measure L	90.9629	90.7765	91.0485	90.4635
LPML	-167.6699	-139.6163	-144.0634	-136.0166
DIC	343.7494	285.7520	297.1779	278.5508
EAIC	353.3398	294.3327	304.1269	290.0132
EBIC	407.7987	348.8517	358.5858	344.4721

Several statistical models with differing distribution in the unobserved covariate and random errors are compared. These models are:

Model 1: Skew-normal distribution for the random effects and normal distribution for the random error (SN-LMM).

Model 2: Skew-t distribution for the random effects and Student-t distribution for the random error (ST-LMM).

Model 3: Skew-slash distribution for the random effects and slash distribution for

the random error (SL-LMM).

Model 4: Contaminated skew-normal distribution for the random effects and contaminated-normal distribution for the random error (SCN-LMM).

Table 1 presents the comparison among the five different models using the model selection criteria discussed in Section 4. Notice that the asymmetric heavy–tailed SNI–LMM improves the corresponding SN-LMM in all the criteria displayed in Table 1, specifically the SCN-LMM presents the best fit. In Table 2 we report posterior mean and standard deviation for all the SNI-LMM. We note from Table 2 that the intercept and slope estimates are similar among the four fitted models, however the standard errors of the ST–LMM , SCN–LMM and SSL–LMM are smaller than those in the SN model, indicating that the three models with longer tails than SN seem to produce more accurate estimates. The estimates for the variance components are not comparable since they are in different scale.

On the other hand, the normal distribution is a limiting case of the skew-t, skew-slash and skew-contaminated normal distributions. The approximate posterior densities of the parameter $\boldsymbol{\nu}$ are presented in the Figure 2. Note that for the ST-LMM and SSL-LMM, the densities are concentrated around small values of ν, indicating the lack of adequacy of the normality assumption for the model. A similar picture emerged considering the SCN–LMM. These results are in accordance with the result present in Lachos et al. (2008). In Figure 3, we showed the box-plots for skewness parameter estimates for the four models. Note that the credible interval for λ_1 does not include zero for all models, which confirms the asymmetry of the data.

Influence of a single outlier

The robustness of the SNI models can be studied through the influence of a single outlying observation on the posterior distribution of the parameters. We study the influence of a change of ∇ units in a single individuals on the posterior mean and 95% credible interval of β_k, $k = 2, 3$. We replace a single observation y_{1j} for $y_{1j}(\nabla) = y_{1j} + \nabla$, $j = 1, \ldots, 6$ with ∇ between -5 and 5.

In Figure 4 and 5, we depicted the posterior mean and 95% credible interval of β_2 (sex) and β_3 (age) respectively, for the SN, ST, SSL and SCN models. Note also that heavy–tailed models are less affected by variations of ∇ than the SN model. In the SN model, the outlying observation has also considerable more impact on the size of the credible interval for both β_2 and β_3.

14

Table 2: Summary results from the posterior distribution, mean, standard deviation (SD), of the parameter of the SNI distributions to the Framingham cholesterol data set. (d_{11}, d_{12}, d_{22}), are the distinct elements of the matrices $\mathbf{D}$.

Parameter	SN–LMM		ST–LMM		SCN–LMM		SSL–LMM	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
β_0	1.3782	0.1346	1.3985	0.1278	1.4057	0.1369	1.3838	0.1340
β_1	-0.0571	0.0489	-0.0621	0.0454	-0.0634	0.0451	-0.0575	0.0460
β_2	0.0135	0.0033	0.0138	0.0031	0.0138	0.0034	0.0143	0.0033
β_3	0.3324	0.1305	0.3412	0.0913	0.3649	0.1102	0.3241	0.1207
σ_e^2	0.0437	0.0023	0.0332	0.0025	0.0265	0.0028	0.0228	0.0023
d_{11}	0.3299	0.0502	0.2277	0.0411	0.1757	0.0340	0.1549	0.0298
d_{12}	-0.0054	0.0874	-0.0039	0.0449	-0.0141	0.0450	-0.0030	0.0416
d_{22}	0.0589	0.0259	0.0436	0.0174	0.0417	0.0205	0.0330	0.0135
λ_1	7.1600	3.0030	5.2750	2.2610	5.2091	2.2960	4.4970	2.0350
λ_2	-2.8200	2.6051	-2.2970	1.8100	-2.7231	2.2070	-1.8700	1.9710
ν	-	-	8.4684	1.8738	0.3399	0.0888	2.0219	0.3105
γ	-	-	-	-	0.3437	0.0391	-	-

6 Concluding remarks

In this article, we have developed a new class of linear mixed models to handle clustered and clumped data. By considering various types skewed distributions we developed a class of asymmetric, heavy-tailed distributions of the random effects in the linear mixed models. Consequently, we developed robust and very flexible class of models. Due to the complexity of the model, Bayesian paradigm was used through Markov chain Monte Carlo method. The proposed methodology is exemplified using a real data set from Framingham cholesterol study. This methodology can be further extended to modeling categorical and survival data analysis, which will be pursued in future research.
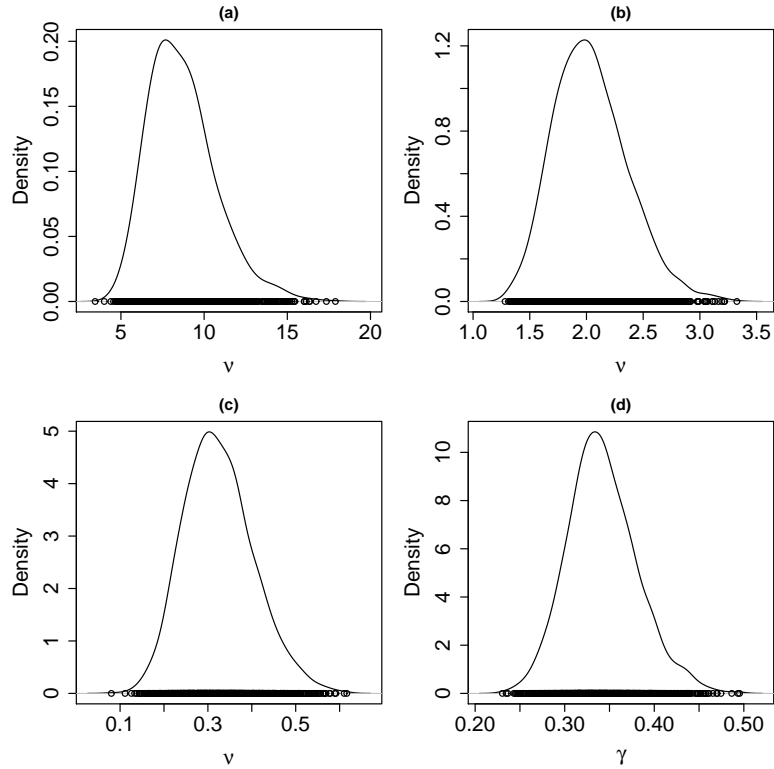
Acknowledgements

Figure 2: Marginal posterior densities of parameters $\boldsymbol{\nu}$ for each tick-tailed distribution, (a) degrees of freedom - skew-t distribution, (b) Parameter ν - skew-slash distribution and (c) Parameter ν- skew-contaminated distribution (d) Parameter γ-skew-contaminated normal distribution.

Appendix A: Conditional posterior distributions for specific skew-normal/independent cases

• **Skew-t**

Here $\boldsymbol{\nu}$ is a scalar parameter and when $\nu = 1$, the models is skew-Cauchy. We adopt a truncated exponential prior for ν of the form $E(\frac{\varrho}{2})\mathbb{I}_{(2,\infty)}$. The density of the conditional posterior distribution in (20) takes the form:

$$u_i|\boldsymbol{\theta}, \mathbf{y}, \mathbf{b}, \mathbf{t} \sim Gamma((n_i + q + \nu + 1)/2; \nu/2 + C_i/2),$$

where $C_i = \frac{1}{\sigma_e^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) + (\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top \boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i) + t_i^2$.
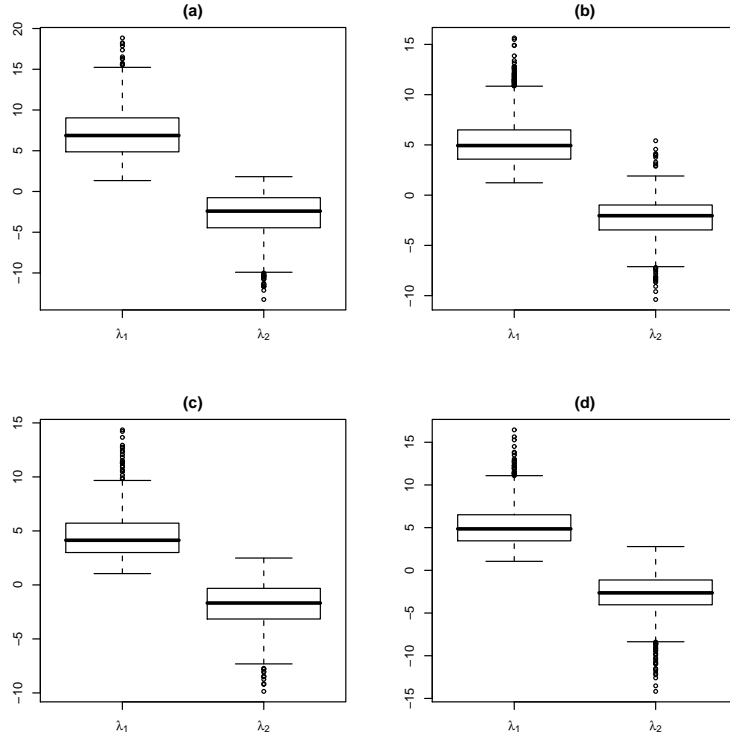
16

Figure 3: Boxplots of parameters $\boldsymbol{\lambda}$ for each tick–tailed distribution, (a) skew–normal distribution, (b) skew-t distribution (c) Parameter ν skew–slash distribution and (d) skew–contaminated normal distribution .

The full conditional posterior density of ν is:

$$\pi(\nu|\boldsymbol{\theta}_{(-\nu)},\mathbf{y},\mathbf{b},\mathbf{u},\mathbf{t}) \propto (2^{\nu/2}\Gamma(\nu/2))^{-n}\nu^{n\nu/2} \times$$
$$\exp\left(-\frac{\nu}{2}\left[\sum_{i=1}^{n}(u_i - \log u_i) + \varrho\right]\right)\mathbb{I}_{(2,\infty)}. \qquad (24)$$

It is seen that $\pi(\nu|\boldsymbol{\theta}_{(-\nu)},\mathbf{y},\mathbf{b},\mathbf{u},\mathbf{t}) \propto \pi_1(\nu) \times Gamma(\frac{n\nu}{2}-1,\frac{1}{2}\sum_{i=1}^{n}(u_i-\log u_i))\mathbb{I}_{(2,\infty)}$, where $\pi_1(\nu) = (2^{\nu/2}\Gamma(\nu/2))^{-n}$. Notice that(24) does not have a closed form but a Metropolis-Hastings or rejection sampling step can be embedded in the MCMC scheme to obtain draws for ν.

• **Skew-slash**

A $Gamma(a,b)$ distribution with small positive values of a and b ($b \ll a$) can be adopted as a prior distribution for ν; this is convenient, because of conjugacy. In this case, the fully conditional posterior density of each u_i is:

$$u_i|\boldsymbol{\theta}_{(-\nu)},\mathbf{y},\mathbf{b},\mathbf{t} \sim Gamma((n_i+q+2\nu+1)/2;C_i/2)\mathbb{I}\{0 < u_i < 1\}.$$

17

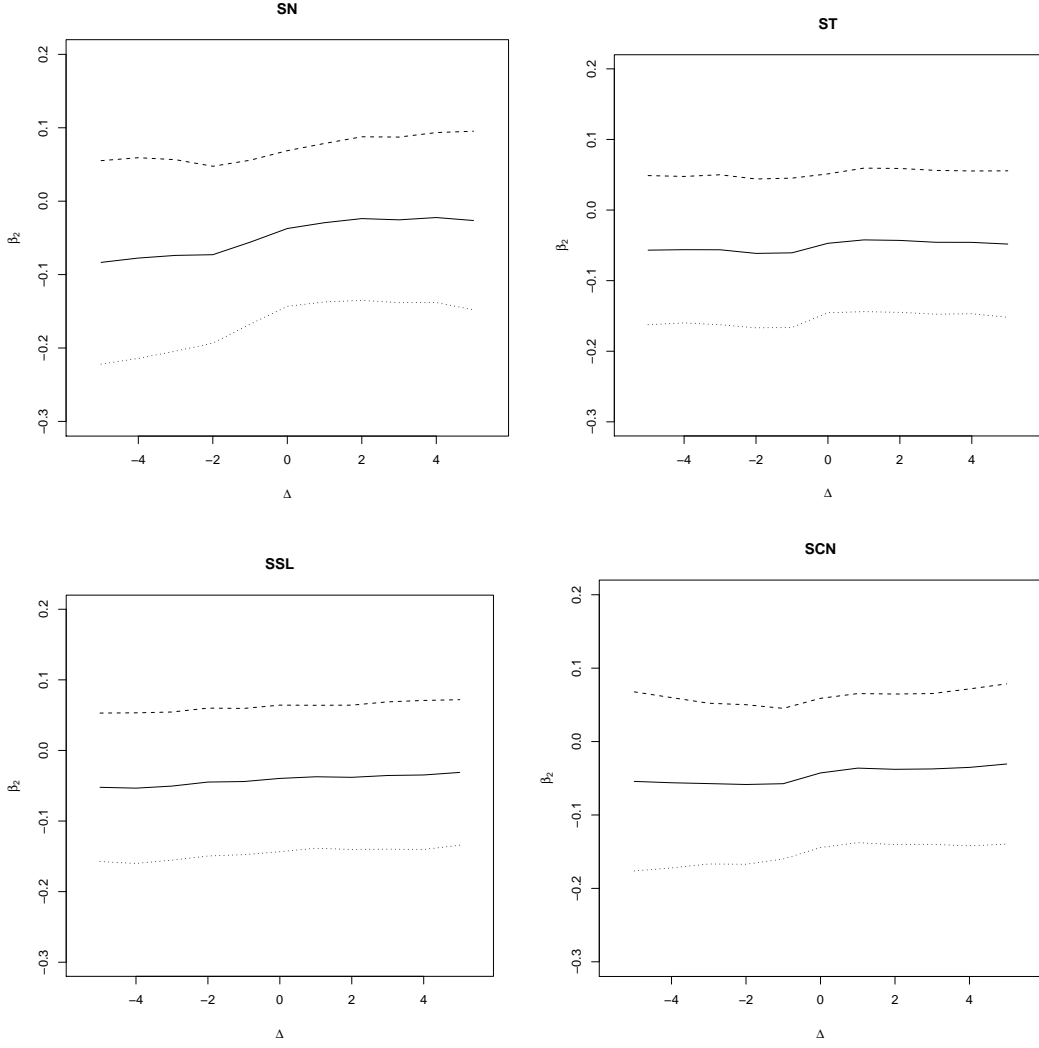Figure 4: Posterior mean (dashed line) and 95% credible interval (solid line) for β_2 of fitting SNI models for different contaminations of ∇ of a single observation

Further, the conditional posterior density of ν is

$$\pi(\nu|\boldsymbol{\theta}_{(-\nu)}, \mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t}) \quad \propto \quad \nu^{n+a-1} \times$$
$$\exp\left(-\nu\left[b - \sum_{i=1}^{n} \log u_i\right]\right), \qquad (25)$$

that is, the conditional density of ν is $\nu|\boldsymbol{\theta}_{(-\nu)}, \mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t} \sim Gamma(n + a, b - \sum_{i=1}^{n} \log u_i)$.

• **Skew contaminated normal distribution**

The possible states of the "weights" u_i are either γ or 1, with $\boldsymbol{\nu} = (\nu, \gamma)^{\top}$. A $U(0, 1)$ distribution is used as a prior for ν, and an independent $Beta(a, b)$ is adopted
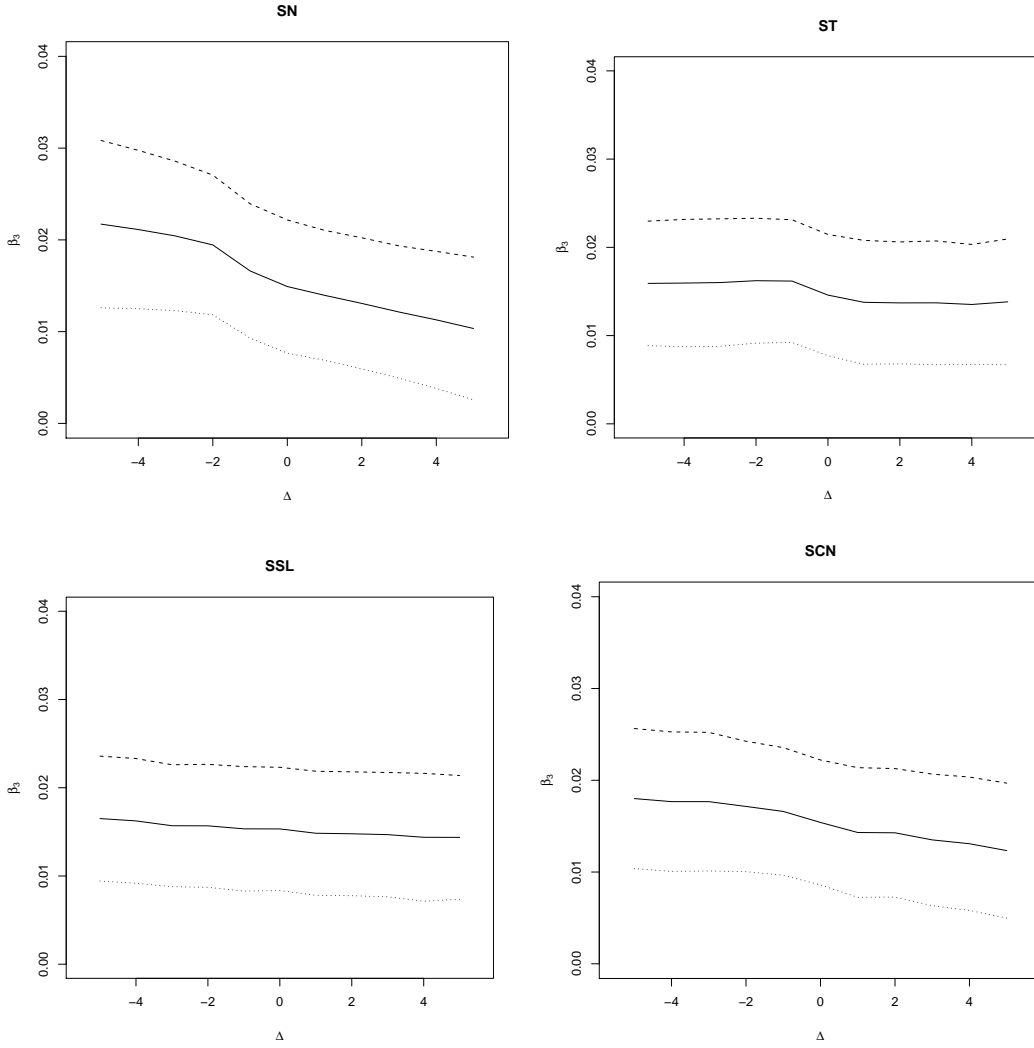
18

Figure 5: Posterior mean (dashed line) and 95% credible interval (solid line) for β_3 of fitting SNI models for different contaminations of ∇ of a single observation

as prior for γ, because of conjugacy. The fully conditional distribution of each u_i is proportional to

$$
\begin{cases}
\nu \gamma^{\frac{n_i+q+1}{2}} \exp\{-\frac{1}{2}\gamma C_i\}, & \text{if } u_i = \gamma, \\
(1-\nu) \exp\{-\frac{1}{2}C_i\}, & \text{if } u_i = 1,
\end{cases}
\tag{26}
$$

and the conditional probabilities are arrived at readily by suitable normalisation.

The full conditional posterior density of the proportion of outliers ν is:

$$
\pi(\nu|\boldsymbol{\theta}_{(-\nu)}, \mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t}) \quad \propto \quad \nu^{(a + \frac{n - \sum_{i=1}^n u_i}{1-\gamma} - 1)} \times (1-\nu)^{(b + \frac{\sum_{i=1}^n u_i - n\gamma}{1-\gamma} - 1)}.
$$

19

It follows that this distribution is:

$$\nu|\boldsymbol{\theta}_{(-\nu)}, \mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t} \sim Beta\left(a + \frac{n - \sum_{i=1}^{n} u_i}{1 - \gamma}; b + \frac{\sum_{i=1}^{n} u_i - n\gamma}{1 - \gamma}\right).$$

The conditional posterior density of γ is:

$$\pi(\gamma|\boldsymbol{\theta}_{(-\gamma)}, \mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t}) \propto \nu^{\left(\frac{n - \sum_{i=1}^{n} u_i}{1 - \gamma}\right)} \times (1 - \nu)^{\left(\frac{\sum_{i=1}^{n} u_i - n\gamma}{1 - \gamma}\right)}.$$

An interesting Metropolis–Hastings method to update from γ is described in Rosa et al. (2003).

References

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

[2] Azzalini, A., Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on the multivariate skew-t distribution. *Journal of the Royal Statistical Society*, Series B, **65**, 367-389.

[3] Azzalini, A. and Genton, M. (2007). Robust likelihood methods based on the skew–t and related distributions. *International Statistical Review* (In press).

[4] Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis,* **79**, 93-113.

[5] Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002). *Journal of the Royal Statistical Society Series B,* **64**, 3,616-618.

[6] Carlin, B.P. and Louis, T.A. (2000). Bayes and Empirical Bayes Methods for Data Analysis Second edition. New York: Chapman & Hall/CRC.

[7] Chen, M-H., Shao, Q-M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

[8] Cowles, M. K., Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association,* **434**, 883-904.

[9] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* New York: Chapman & Hall.

[10] Gelfand, A.E. (1996). Model determination using sampling based methods. In Markov Chain Monte Carlo in Practice (Eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter), Chapman and Hall: London, 145-161.

[11] Gelman, A., Carlin, J. B., Stern, H., and Rubin, D. (2006). *Bayesian Data Analysis*, Second Edition, Chapman & Hall/CRC.

[12] Gelfand, A. E., Dey, D. K. and Chang H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with Discussion). In *Bayesian Statistics 4*, Bernardo, J. M., Berger J. O., Dawid A. P., Smith AFM (eds). Oxford University Press: Oxford, 1992.

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

[14] Gupta A. K. (2003). Multivariate skew t-distributions. *Statistics,* **37**, 359-363.

[15] Jara, A., Quintana, F. and San Martin, E. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach . *Computational Statistics and Data Analysis*, **52**, 5033–5045.

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

[17] Lachos, V. H., Ghosh, P. and Arellano–Valle, R. B. (2008). Likelihood based inference for skew–normal/independent linear mixed models. *Statistica Sinica*, to appear.

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

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

[20] Laud, P. W. and Ibrahim, J. G. (1995). Predictive model selection. *Journal Royal Statistical Society, B*, **57**, 247-262.

[21] Lin, T. I. and Lee, J. C. (2007). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine.* Available online early view.

[22] Lin, T. I., Lee, J. C. and Hsieh, W. J. (2007). Roubust mixture modeling using the skew t distribution. *Statistics and Computing.* **17**, 81-92.

[23] Ma, Y., Genton, M. G., and Davidian, M. (2004). Linear mixed models with flexible generalized skew-elliptical random effects. In *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality,* Genton, M. G., Ed., Chapman & Hall/CRC, Boca Raton, FL, pp. 339-358.

[24] Pinheiro, J. C., Liu, C. H. and 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.

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

[26] Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with aplications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129-150.

[27] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. V. (2002). Bayesian measures of model complexity and *Journal Royal Statistical Society, B* **64**, 583-639.

[28] 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.

[29] Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541-556.

[30] Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209–220.

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