UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

BAYESIAN ANALYSIS FOR A SKEW EXTENSION OF THE MULTIVARIATE NULL INTERCEPT MEASUREMENT ERROR MODEL

Vicente G. Cancho
Reiko Aoki
V. H. Lachos

Nº 284

RELATÓRIOS TÉCNICOS

São Carlos – SP
Nov./2006

Bayesian Analysis for a Skew Extension of the Multivariate Null Intercept Measurement Error Model

Vicente G. Cancho [‡] *ICMC-USP*

Reiko Aoki[§] *ICMC-USP*

V.H. Lachos[¶] *IMECC-UNICAMP*

Abstract

The skew-normal distribution is a class of distributions which includes the normal distributions as a special case. In this paper, we explore the use of Markov Chain Monte Carlo (MCMC) methods to develop a Bayesian analysis in multivariate null intercept measurement error model (Aoki et al., 2003b) where the unobserved value of the covariate (latent variable) follows a skew-normal distribution. The results and methods are applied to a real dental clinical trial presented in Hadgu and Koch (1999).

Key Words: *Skew-normal distribution; Gibbs algorithm; Metropolis-Hasting; Skewness; Multivariate null intercepts model; Measurement error.*

1 Introduction

Recent statistical literature is broadly related to the skew-normal distribution, which represents a superset of the normal family and has a shape parameter that defines the direction of the asymmetry of the distribution. Motivation is originated from real data sets presenting clear indication of skewness in diverse areas, such as, medicine, psychology, engineering, agriculture, among others. In this paper we use a real data set from a dental clinical trial in which the outcome measurements are typically not symmetric so that it seems more adequate to consider a skew-normal distribution to describe the behavior of these measurements.

As discussed in Lachos et al. (2006), a k-dimensional random vector $\mathbf{Y}$ has a multivariate skew-normal distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Psi}$ and skewness parameter $\boldsymbol{\lambda}$, namely $\mathbf{Y} \sim SN_k(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f(\mathbf{y}) = 2\phi_k(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Psi})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^k, \tag{1}$$

where as usual, $\phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Psi})$ and $\Phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Psi})$ denote, respectively, the probability density function (pdf) and cumulative distribution function (cdf) of the $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$ distribution and $\boldsymbol{\Psi}^{-1/2}$

[‡]Address: ICMC, Universidade de São paulo, São Carlos, São Paulo, Brasil. E-mail: garibay@icmc.usp.br
[§]Address: ICMC, Universidade de São paulo, São Carlos, São Paulo, Brasil. E-mail: reiko@icmc.usp.br
[¶]Address: Universidade de Campinas, Campinas, São Paulo, Brasil. E-mail: vlachos@ime.unicamp.br

satisfies $\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}^{-1/2} = \boldsymbol{\Psi}^{-1}$. The stochastic representation, which can be used to simulate random realizations from $\mathbf{Y}$, is given by

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_k - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{2}$$

where $T_0 \sim N_1(0,1)$ and $\mathbf{T}_1 \sim N_k(\mathbf{0}, \mathbf{I}_k)$ are independent, and " $\stackrel{d}{=}$ " meaning "distributed as". For more details on this approach, see Lachos et al. (2006).

Measurement error models constitute an attractive option for modeling many practical experimental problems, specially when the same response is observed for the same individuals under different experimental conditions. One special case is the situation in which the same experimental unit is measured using two or more measuring devices, the analysis of such a model considering the null intercept measurement error model was discussed in Aoki et al. (2001, 2003a). An extensive bibliography for such models in general, can be found in Fuller (1987) and Cheng and Van Ness (1999). Recently, Aoki et al. (2003b) discussed a multivariate normal (symmetric) null intercept measurement error model (N-IMEM), with a dependence structure between the response variables within the same group, applicable to longitudinal data studies and proposed a Bayesian methodology with the estimation performed via MCMC. In that study 105 volunteers were randomized to two experimental mouth rinses (A or B) or a control mouth rinse and evaluated under these three experimental conditions (mouth rinses), with respect to the dental plaque index at the beginning of the study, after three months and after six months from the beginning of the study with the use of these mouth rinses. This model, in a general context, can be written as

$$\mathbf{X}_i = \mathbf{x}_i + \boldsymbol{\delta}_i, \tag{3}$$

$$\mathbf{y}_{k_i} = \mathbf{x}_i\beta_{k_i} + \boldsymbol{\epsilon}_{k_i}, \tag{4}$$

$i = 1,\ldots,p$, $k = 1,\ldots,m$, where $\mathbf{X}_i = (X_{i1},\ldots,X_{in_i})^\top$ is a vector of observed covariates, $\mathbf{y}_{k_i} = (y_{k_{i1}},\ldots,y_{k_{in_i}})^\top$ is a vector of observed responses, $\mathbf{x}_i = (x_{i1},\ldots,x_{in_i})^\top$ is the unobserved covariates vector (latent variables), β_{k_i} is the regression parameter, and $\boldsymbol{\epsilon}_{k_i} = (\epsilon_{k_{i1}},\ldots,\epsilon_{k_{in_i}})^\top$, $\boldsymbol{\delta}_i = (\delta_{i1},\ldots,\delta_{in_i})^\top$ are the related vector of random errors. In the case of the pretest/posttest data set just described, we have i=1,2,3 representing the control mouth rinse, experimental mouth rinse A and experimental mouth rinse B, respectively and k=1,2 representing three months and six months from the baseline, respectively. Moreover the covariate, $\mathbf{X}_i$, is the dental plaque index at the baseline, $\mathbf{y}_{1_i}$ ($\mathbf{y}_{2_i}$) represents the dental plaque index after three months from the baseline (six months from the baseline). Typically, it is assumed that $\delta_{ij} \stackrel{iid}{\sim} N_1(0,\sigma^2)$, $\epsilon_{kij} \stackrel{ind}{\sim} N_1(0,\sigma^2_{c_i})$, δ_{ij} and ϵ_{kij} are not correlated and independent of the unobserved quantities $x_{ij} \stackrel{iid}{\sim} N_1(\mu_x,\sigma^2_x)$, $i = 1,\ldots,p$, $j = 1,\ldots,n_i$ $k = 1,\ldots,m$. Although the assumption of normality (or symmetry) is reasonable in many situations, it may not be appropriate when the data exhibit some departure from this assumption. In this case, a more flexible model that includes the normal distribution as a special case, allowing continuous variation from normality to non-normality, and accommodates skewness may be of great interest. In this paper, we consider the skew normal distribution defined in (1) to extend the model defined in (3) and (4), considering that the true unobserved value of the covariate x_{ij} follows a univariate skew-normal distribution, implicating that the observation vector

$\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, follows a multivariate skew-normal distribution. The skewness is regulated by the parameter λ and the assumption of normality can be checked on this parameter, as $\lambda = 0$ yields the normal model.

The remainder of the paper is structured as follows. In Section 2 we present the multivariate skew-normal null intercept measurement error model (SN-IMEM) with a dependence structure between the response variables within the same group appropriate for longitudinal data analysis. In Section 3 we discuss the Bayesian formulation of the model. The methodology proposed for SN-IMEM models is illustrated in Section 4 considering the real pretest/posttest data set and finally, some concluding remarks are presented in Section 5.

2 The model

To specify the multivariate null intercept measurement error model in the multivarite skew-normal class, first, notice that we can write the linear model (3)-(4) as,

$$\mathbf{z}_{ij} = \boldsymbol{\beta}_{0i} x_{ij} + \boldsymbol{\zeta}_{ij}, \tag{5}$$

where the observations vector $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$; the regression parameters vector $\boldsymbol{\beta}_{0i} = (1, \boldsymbol{\beta}_i^\top)^\top$, with $\boldsymbol{\beta}_i = (\beta_{1_i}, \ldots, \beta_{m_i})^\top$ and the random errors vector $\boldsymbol{\zeta}_{ij} = (\delta_{ij}, \epsilon_{1_{ij}}, \ldots, \epsilon_{m_{ij}})^\top$ are all of dimension $(m+1) \times 1$. We refer to this model as the SN-IMEM if

$$\boldsymbol{\zeta}_{ij} \overset{ind}{\sim} N_{m+1}(\mathbf{0}, \mathrm{D}(\boldsymbol{\phi}_i)) \quad \text{and} \quad x_{ij} \overset{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \tag{6}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, with $\mathrm{D}(\boldsymbol{\phi}_i)$ denoting a diagonal matrix with the diagonal elements given by $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i = (\sigma_u^2, \sigma_{e_i}^2, \ldots, \sigma_{e_i}^2)^\top$ has dimension $(m+1) \times 1$. The above model is considering, for instance, that in the case of the Hadgu and Koch (1999) data set the index plaque may not be symmetrically distributed in the population. On the other hand, the errors $\boldsymbol{\zeta}_{ij}$, are related to measurement errors so that it is expected to be normally distributed. The asymmetry parameter λ_x incorporates skewness in the latent variable x_{ij} and consequently in the observed quantities $\mathbf{z}_{ij}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, which can be shown to have marginally a $(m+1)$-variate skew-normal distribution. If $\lambda_x = 0$, then the asymmetric model reduces to the N-IMEM. Note from (2) that, the regression set up defined in (5)-(6) can be written hierarchically as

$$\mathbf{z}_{ij} \mid x_{ij} \overset{ind}{\sim} N_{m+1}(\boldsymbol{\beta}_{0i} x_{ij}, \mathrm{D}(\boldsymbol{\phi}_i)), \tag{7}$$

$$x_{ij} \mid T_{ij} = t_{ij} \overset{ind}{\sim} N_1(\mu_x + \phi_x^{1/2} \delta_x t_{ij}, \phi_x(1 - \delta_x^2)), \tag{8}$$

$$T_{ij} \overset{iid}{\sim} \mathrm{HN}_1(0, 1), \tag{9}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, all independent, where $\mathrm{HN}_1(0, 1)$ denote the standardized univariate half-normal distribution and $\delta_x = \lambda_x / (1 + \lambda_x^2)^{1/2}$. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}_e^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x)^\top \in \mathbb{R}^{(m+1)p+4}$, with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top)^\top$, $\boldsymbol{\sigma}_e^2 = (\sigma_{e_1}^2, \ldots, \sigma_{e_p}^2)^\top$ is based on the marginal distribution for the response $\mathbf{z}_{ij}$, which is given by (see Lachos et al., 2006)

$$f(\mathbf{z}_{ij}|\boldsymbol{\theta}) = 2\phi_{m+1}(\mathbf{z}_{ij}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\Phi_1(\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)), \tag{10}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$ i.e., $\mathbf{z}_{ij} \overset{ind}{\sim} SN_{m+1}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i)$, with $\boldsymbol{\mu}_i = \boldsymbol{\beta}_{0i}\mu_x$, $\boldsymbol{\Sigma}_i = \mathrm{D}(\boldsymbol{\phi}_i) + \phi_x\boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top$, $\bar{\boldsymbol{\lambda}}_i = \dfrac{\lambda_x\phi_x\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\beta}_{0i}}{\sqrt{\phi_x + \lambda_x^2\Lambda_i}}$, where $\Lambda_i = \dfrac{\phi_x}{1 + \phi_x\boldsymbol{\beta}_{0i}^\top\mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}}$. It follows that the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{z} = (\mathbf{z}_{11}^\top, \ldots, \mathbf{z}_{1n_1}^\top, \ldots, \mathbf{z}_{p1}^\top, \ldots, \mathbf{z}_{pn_p}^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) \propto -\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \frac{1}{2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i) - \log(K_{ij})\right), \tag{11}$$

where $K_{ij} = \Phi_1(\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i))$ and $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, $\bar{\boldsymbol{\lambda}}_i$ as in (10). The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta}$ can be obtained by direct maximization of (11), while large samples inference for the parameters can be based on the asymptotic normality of the MLEs (Cox and Hinkley, 1974). However, in the dental clinical data set the slopes represent the percentage of dental plaque remaining after brushing, so these parameters must lie in the $[0, 1]$ interval and the ordinary regularity conditions for the asymptotic normality and consistency do not include restricted parameter spaces. Moreover, when we have small or moderate samples sizes, the asymptotic theory can not be very accurate. Thus in this paper we consider a Bayesian approach in which such restrictions is easily incorporated.

3 Bayesian analysis

In the following we implement the Bayesian methodology using Markov chain Monte Carlo techniques for the SN-IMEM using double augmentation by considering that $(\mathbf{x}, \mathbf{t})$ are missing data, with $\mathbf{x} = (\mathbf{x}_1^\top, \ldots, \mathbf{x}_p^\top)^\top$ and $\mathbf{t} = (\mathbf{t}_1^\top, \ldots, \mathbf{t}_p^\top)^\top$. Thus, under the hierarchical representation (7)-(9), with $\lambda_x = \tau/\upsilon$ and $\phi_x = \tau^2 + \upsilon^2$, it follows that the complete likelihood function associated with $(\mathbf{z}, \mathbf{x}, \mathbf{t})$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{t}) \propto\ & (\upsilon^2)^{-n/2}\prod_{i=1}^{p}(|D(\boldsymbol{\phi}_i)|)^{-n_i/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i}x_{ij})^\top D^{-1}(\boldsymbol{\phi}_i)(\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i}x_{ij})\right. \\ & \left. -\frac{1}{2\upsilon^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \mu_x - \tau t_{ij})^2 - \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}t_{ij}^2\right\}, \end{aligned} \tag{12}$$

where $n = \sum_{j=1}^{p} n_i$.

Now, we consider the joint prior density for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top, \sigma_{e_1}^2, \ldots, \sigma_{e_p}^2, \sigma_u^2, \mu_x, \upsilon^2, \tau)^\top$ of the form

$$\pi(\boldsymbol{\theta}) = \left(\prod_{i=1}^{p}\prod_{k=1}^{m}\pi(\beta_{k_i})\right)\left(\prod_{i=1}^{p}\pi(\sigma_{e_i}^2)\right)\pi(\sigma_u^2)\pi(\mu_x)\pi(\upsilon^2)\pi(\tau), \tag{13}$$

where $\mu_x \sim N(m, \sigma_{\mu_x}^2)$, $\tau \sim N(\mu_\tau, \sigma_\tau^2)$, $\upsilon^2 \sim IG(\frac{a}{2}, \frac{b}{2})$, $\beta_{k_i} \sim N(a_{k_i}, b_{k_i})I(0, 1)$, $k = 1, \ldots, m$, $i = 1, \ldots, p$, $\sigma_{e_i}^2 \sim IG(\frac{a_i}{2}, \frac{b_i}{2})$, $i = 1, \ldots, p$, $\sigma_u^2 \sim IG(\frac{a_u}{2}, \frac{b_u}{2})$, with $N(a, b)I(0, 1)$ denoting the truncated Normal distribution in the interval $[0, 1]$, $IG(a, b)$ denoting the Inverse Gamma

distribution with shape parameter $a > 0$ and scale $b > 0$. We assume that the hiperparameters are specified.

Combining the likelihood function (12) and prior specification (13), the joint posterior distribution for $\boldsymbol{\theta}$ is given by,

$$\pi(\boldsymbol{\theta}, \mathbf{t}, \mathbf{x}|\mathbf{z}) \propto (\upsilon^2)^{-n/2}(\sigma_u^2)^{-n/2}\prod_{i=1}^{p}(\sigma_{e_i}^2)^{-n_i/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\frac{(\mathbf{y}_{ij}-\boldsymbol{\beta}_i x_{ij})^\top(\mathbf{y}_{ij}-\boldsymbol{\beta}_i x_{ij})}{\sigma_{e_i}^2}\right.$$
$$\left.-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\frac{(X_{ij}-x_{ij})^2}{\sigma_u^2}-\frac{1}{2\upsilon^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\mu_x-\tau t_{ij})^2-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}t_{ij}^2\right\}\pi(\boldsymbol{\theta}), \qquad (14)$$

where $\mathbf{y}_{ij} = (y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$. The marginal posterior densities of the parameters in the equation (14) are not easily obtained. This is because the integration of the joint posterior density is difficult to perform. An alternative, is the use of the Gibbs sampler algorithm. In this direction, we first obtain the full conditional distribution of each unknown. These distributions have closed form and are given by

(*i*) $T_{ij}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{x} \sim N\left(\frac{\tau(x_{ij}-\mu_x)}{\tau^2+\upsilon^2}, \frac{\upsilon^2}{\tau^2+\upsilon^2}\right) I(t_{ij} > 0),\ i = 1, \ldots, p,\ j = 1, \ldots, n_i;$

(*ii*) $x_{ij}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{t} \sim N\left(\mu_{x_{ij}}, \sigma_{x_{ij}}^2\right),\ i = 1, \ldots, p,\ j = 1, \ldots, n_i,$ with

$$\mu_{x_{ij}} = \frac{\upsilon^2\sigma_u^2\sum_{k=1}^{m}\beta_{k_i}y_{k_{ij}} + \sigma_{e_i}^2\sigma_u^2(\mu_x+\tau t_{ij})}{\sigma_{e_i}^2(\sigma_u^2+\upsilon^2)+\upsilon^2\sigma_u^2\sum_{k=1}^{m}\beta_{k_i}^2},\ \sigma_{x_{ij}}^2 = \frac{\upsilon^2\sigma_u^2\sigma_{e_i}^2}{\sigma_{e_i}^2(\sigma_u^2+\upsilon^2)+\upsilon^2\sigma_u^2\sum_{k=1}^{m}\beta_{k_i}^2};$$

(*iii*) $\beta_{k_i}|\boldsymbol{\theta}_{(-\beta_{k_i})}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim N\left(\frac{b_{k_i}\sum_{j=1}^{n_i}x_{ij}y_{k_{ij}}+a_{k_i}\sigma_{e_i}^2}{b_{k_i}\sum_{j=1}^{n_i}x_{ij}^2+\sigma_{e_i}^2}, \frac{b_{k_i}\sigma_{e_i}}{b_{k_i}\sum_{j=1}^{n_i}x_{ij}^2+\sigma_{e_i}^2}\right) I(0,1)\ i = 1, \ldots, p,\ k = 1, \ldots, m;$

(*iv*) $\sigma_{e_i}^2|\boldsymbol{\theta}_{(-\sigma_{e_i}^2)}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim IG\left(\frac{n_i+a_i}{2}; \frac{1}{2}\left(\sum_{j=1}^{n_i}(\mathbf{y}_{ij}-\boldsymbol{\beta}_i x_{ij})^\top(\mathbf{y}_{ij}-\boldsymbol{\beta}_i x_{ij})+b_i\right)\right), i = 1, \ldots, p;$

(*v*) $\sigma_u^2|\boldsymbol{\theta}_{(-\sigma_u^2)}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim IG\left(\frac{n+a_u}{2}, \frac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-x_{ij})^2+b_u\right)\right);$

(*vi*) $\mu_x|\boldsymbol{\theta}_{(-\mu_x)}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim N\left(\frac{\sigma_{\mu_x}^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\tau t_{ij})^2+\upsilon^2 m}{n\sigma_{\mu_x}^2+\upsilon^2}, \frac{\upsilon\sigma_{\mu_x}^2}{n\sigma_{\mu_x}^2+\upsilon^2}\right);$

(*vii*) $\upsilon^2|\boldsymbol{\theta}_{(-\upsilon^2)}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim IG\left(\frac{n+a}{2}, \frac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\mu_x-\tau t_{ij})^2+b\right)\right);$

(*viii*) $\tau|\boldsymbol{\theta}_{(-\tau)}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim N\left(\frac{\kappa_1\kappa_2\sigma_\tau^2+\upsilon^2\mu_\tau}{\kappa_1\sigma_\tau^2+\upsilon^2}, \frac{\upsilon^2\sigma_\tau^2}{\kappa_1\sigma_\tau^2+\upsilon^2}\right)$ with $\kappa_1 = \sum_{i=1}^{p}\sum_{j=1}^{n_i}t_{ij}^2,$

$\kappa_2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i}t_{ij}(x_{ij}-\mu_x)$

Note that it is easy to simulate and sample from these full conditional distributions. The shape and scale parameters of the latent variable x_{ij} can be obtained by noting that $\lambda_x = \tau/\upsilon$ and $\phi_x = \tau^2 + \upsilon^2$

3.1 Predictive model selection

Model diagnostic 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 we develop one criterion to perform Bayesian model choice.

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 observables, $\mathbf{y}_{rep}$ 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,

$$\pi(y_{i,rep}|\mathbf{y}_{obs}) = \int \pi(y_{i,rep}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}. \tag{15}$$

Let $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ denote the posterior predictive mean and covariance of $y_{i,rep}$ under the density (15). 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 sampled 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 above predictive density. See for example Gelfand (1996). To perform model choice, we considered the Bayesian criteria called measured L 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 (15). Straightforward algebra shows that L 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 has been proposed in the literature, for instance, the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), the Expected Akaike Information Criterion (EAIC, Carlin and Louis, 2000) and the Expected Schwarz Information Criterion (EBIC, Brooks, 2002). For further details see Ibrahim, et al. (2001).

4 Application

In this section, we apply the methodology discussed in this work to a real data set, previously analyzed in Hadgu and Koch (1999) and Aoki et al. (2003b). The main objective of this

study was to compare the efficiency of the experimental mouth rinses A ($i = 2, n_2 = 33$) and B ($i = 3, n_3 = 36$), with the control mouth rinse ($i = 1, n_1 = 36$). As the covariate (observed plaque index) is measured imprecisely, one way of analyzing the data is to fit a measurement error model. We do not include intercepts in the model as we consider it reasonable to assume that a zero dental plaque index in the beginning of the trial should imply a zero expected post-test value; that is, we assume that the dental plaque index should not increase after the use of the mouth rinses. To account for the possible structure of dependency of the within-subjects measurements, the structural model is considered. The response variable of interest in this context is the true dental plaque index.

The following independent priors were considered to perform the Gibbs sampler. $\beta_{k_i} \sim N(0, 100)I_{(0,1)}(\beta_{k_i}), k = 1, 2,\ i = 1, 2, 3,\ \sigma^2_{e_i} \sim IG(0.01, 0.01),\ i = 1, 2, 3\ \sigma^2_u \sim IG(0.01, 0.01)$, $\mu_x \sim N(0, 1000),\ v^2 \sim IG(0.01, 0.01),\ \tau \sim N(0, 1000)$, so that we have a vague prior distribution. 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 10, obtaining a sample of size 2000 from each chain. To monitor the convergence of the Gibbs samples we used the between and within sequence information, following the approach developed in Gelman and Rubin (1992) to obtain the potential scale reduction, $\hat{R}$. In all cases, these values were close to one, indicating the convergence of the chain. The histogram with the approximate posterior density of the parameters are presented in Figure (1). In order to compare the two models the N-IMEM ($\lambda_x = 0$) was also fitted to the data set. In Table 1 we report posterior summaries for the parameters of the N-IMEM and SN-IMEM.

Table 1: Bayesian estimates. Posterior summary results of fitting SN-IMEM and N-IMEM to the dental plaque index data set. SD, 2,5% and 97.5% represents standard deviation and percentiles from the posterior distributions of parameters, respectively.

	N-IMEM				SN-IMEM			
Parameter	Estimate	SD	2,5%	97.5%	Estimate	SD	2,5%	97.5%
β_{1_1}	0.7032	0.0352	0.6341	0.7724	0.7021	0.0354	0.6327	0.7712
β_{1_2}	0.5367	0.0458	0.4475	0.6265	0.5368	0.0456	0.4471	0.6261
β_{1_3}	0.5084	0.0328	0.4440	0.5731	0.5086	0.0330	0.4440	0.5732
β_{2_1}	0.6869	0.0354	0.6169	0.7569	0.6860	0.0353	0.4471	0.6261
β_{2_2}	0.5022	0.0456	0.4130	0.5919	0.5024	0.0458	0.4123	0.5929
β_{2_3}	0.4135	0.0328	0.3488	0.4781	0.4140	0.0327	0.3502	0.4785
$\sigma^2_{e_1}$	0.2912	0.0517	0.2067	0.4081	0.2952	0.0525	0.2101	0.4143
$\sigma^2_{e_2}$	0.4595	0.0848	0.3211	0.6504	0.4573	0.0834	0.2101	0.4143
$\sigma^2_{e_3}$	0.2389	0.0420	0.1703	0.3342	0.2406	0.0423	0.1713	0.3371
σ^2_u	0.0117	0.0074	0.0030	0.0301	0.0040	0.0036	0.0005	0.0142
μ_x	2.5340	0.0334	2.4680	2.5990	2.1140	0.0348	2.0480	2.1850
ϕ_x	0.1057	0.0172	0.0754	0.1430	0.2981	0.0511	0.2065	0.4071
λ_x	-	-	-	-	9.5390	5.0120	3.1680	21.9660

To compare the SN-IMEM and N-IMEM fits we also obtained the values of DIC, EAIC, EBIC and the L measure. These information criteria furnish the values given in Table 2.

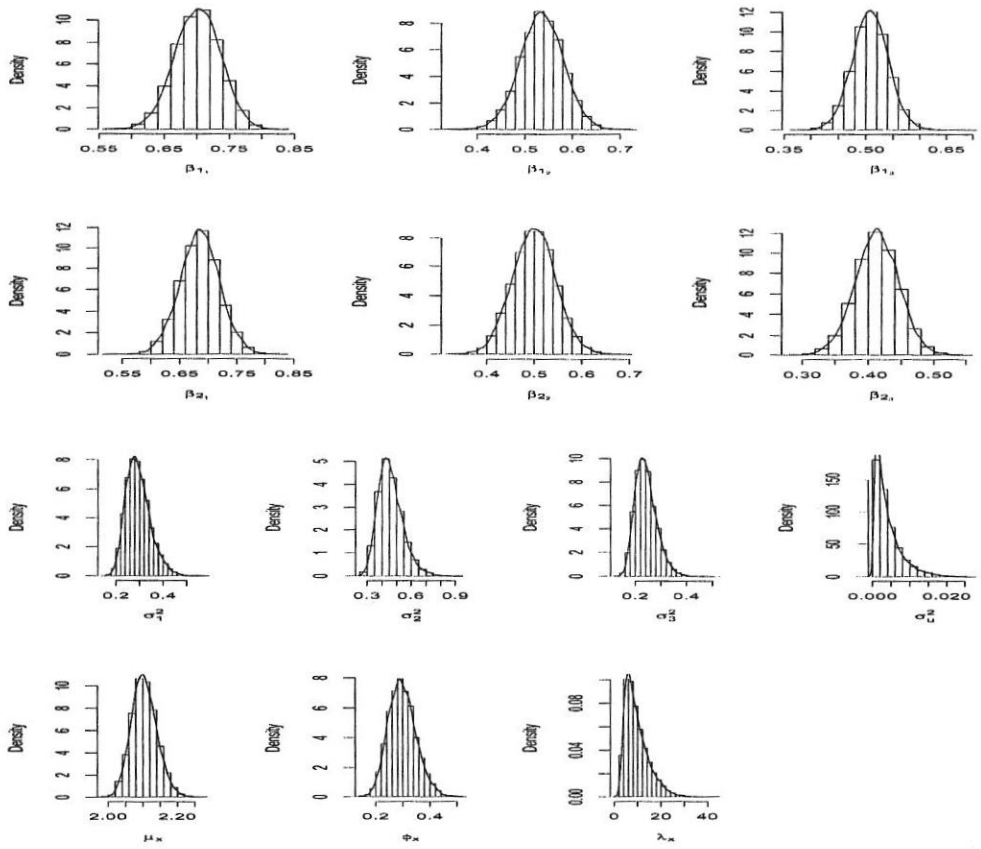

Figure 1: The multivariate skew-normal null intercept measurement error model: histogram with the approximate posterior density of β_{1_1}, β_{1_2}, β_{1_3}, β_{2_1}, β_{2_2}, β_{2_3}, σ_u^2, μ_x, ϕ_x, and λ_x

Notice that the SN-IMEM improves the corresponding normal model (N-IMEM) in all of the criterion displayed in Table 2 (the preferred model is the one with the smaller value for each criterion).

Table 2: Comparison between N-IMEM and SN-IMEM by using different Bayesian criterion.

criterion	N-IMEM	SN-IMEM
Measure L	146.3315	136.6225
DIC	-1,998	-10.17305
EAIC	9,3924	-6,2110
EBIC	49,5577	37,3013

The objective of the experiment was to compare the efficiency of the two experimental mouth rinses, A and B, with the control mouth rinse, namely, we are interested in comparing the slope parameters β_{k_2} and β_{k_3} with respect to β_{k_1}, $k = 1, 2$. One way of testing this hypothesis ($\beta_{k_1} = \beta_{k_i}$, $i = 2, 3$, $k = 1, 2$) is to consider Monte Carlo estimates based on the

generated Gibbs Samples of $\beta_{k_1} - \beta_{k_2}$ and $\beta_{k_1} - \beta_{k_3}$, $k = 1, 2$, and verify if the value zero belongs to the credible region. Another questions of interest is whether the mouth rinses continues to reduce the plaque index after three months, which corresponds to compare the slope parameters β_{1_i} and β_{2_i} $i = 1, 2, 3$. Table 3 shows the values of the mean and 95% credible interval for the questions of interest just described, as well as, the comparison of the experimental mouth rinses A and B (β_{k_2} and β_{k_3}).

Table 3: Posterior mean and 95% credible interval

$\beta_{1_1} - \beta_{1_2}$	$\beta_{1_1} - \beta_{1_3}$	$\beta_{1_2} - \beta_{1_3}$	$\beta_{2_1} - \beta_{2_2}$	$\beta_{2_1} - \beta_{2_3}$
0.1649	0.1925	0.0276	0.1830	0.2713
(0.0539 , 0.2810)	(0.0976 , 0.2861)	(-0.0816 , 0.1376)	(0.0714 , 0.2964)	(0.1770 , 0.3639)
$\beta_{2_2} - \beta_{2_3}$	$\beta_{1_1} - \beta_{2_1}$	$\beta_{1_2} - \beta_{2_2}$	$\beta_{1_3} - \beta_{2_3}$	
0.0884	0.0162	0.0343	0.0950	
(-0.0206 ,0.1952)	(-0.0767 , 0.1134)	(-0.0912 , 0.1604)	(0.0057 , 0.1875)	

After analyzing these results we conclude that both of the experimental mouth rinses are more efficient then the control mouth rinse to reduce the dental plaque after three months from the baseline and also after six months from the baseline, while there is no difference between the two experimental mouth rinses A and B. The only mouth rinse that is long lasting is the mouth rinse B.

5 Final conclusions

In this paper, we proposed the use of the skew-normal distribution to extend the model presented in Aoki et al. (2003b) to analyze the longitudinal data set. Although the model defined in Aoki et al. (2003b) takes into account the fact that the dental plaque index are measured with error, the possible dependency between measurements taken in the same subject, the restriction of the parameters of interest (which represents the percentage of dental plaque remaining after brushing) to the interval $[0, 1]$, the null intercept model (null pretest dental plaque indices imply null expected post-test values), it did not take into account the possible asymmetry in the data set. Comparing the two models, we see that the estimated values and standard deviations of the parameters of interest (slopes) are almost the same, as well as, the estimative and standard deviations for the parameters $\sigma^2_{e_i}$, $i = 1, 2, 3$. Despite the fact that the conclusions are similar for all the questions of interest under any of the proposed models, we concluded that the value λ_x is not zero, which leads to the asymmetric model. Also, the Bayesian criteria considered in Table 2 shows that the asymmetric model improves the symmetric model. In that way, considering the proposed model with the extension to the skew-normal measurement error model, we included most of the characteristics of the data set in analyze.

Acknowledgements

The authors acknowledges the partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

Aoki, R., Bolfarine, H. and Singer, J. M. (2001). Null intercept measurement error regression models. *TEST*, 10, 441-454.

Aoki, R., Achcar, J. A., Bolfarine, H. and Singer, J. M. (2003a). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. *Journal of Applied Statistics*, vol. 30, 1, 5-14.

Aoki R., Bolfarine, H., Achcar J. A. and Leão Pinto Jr D. (2003b). Bayesian Analysis of a Multivariate Null Intercept Error-in-Variables Regression Model, *Journal of Biopharmaceutical Statistics*, vol. 13, 4, 763-771.

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.

Carlin, B.P. and Louis, T.A. (2001).Bayes and Empirical Bayes Methods for Data Analysis Essays on Item Response Theory. Second edition. New York: Chapman & Hall.

Cheng, C. L. and Ness, V. (1999) - Statistical Regression with Measurement Error, first edition, Oxford University Press.

Cox, D. R. and Hinkley, D. V. (1974) - *Theoretical Statistics*, Chapman and Hall.
Fuller, A. (1987) - *Measurement Error Models*, first edition, Wiley.

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.

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

Hadgu, A. and Koch, G. (1999). Application of generalized estimating equations to a dental randomized clinical trial. *Journal of Biopharmaceutical Statistics*, 9, 161-178.

Ibrahim, J.G.; Chen, M.H. and Sinha, D. (2001) *Bayesian survival Analysis,* Springer-Verlag New York.

Lachos, V. H. Bolfarine, H. and Arellano-Valle, R. B. and Montenegro, L.C.(2006). *Likelihood based inference for multivariate skew-normal regression models.* To appear in Communication in Statistics-Special Issue.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. V. (2002). Bayesian measures of model complexity and *Journal Royal Statistical Society,* 64, 3, 583-639.