A Bayesian Skew-Normal Independent Mixture Measurement Error Model*

Themis C. Abensur Celso R. B. Cabral
José Cardoso Neto

Departamento de Estatística – Universidade Federal do Amazonas – Brazil

1 Introduction

The traditional regression model with measurement errors assumes normal distributions for the error terms and unobserved latent covariate. These assumptions are not appropriate when asymmetry, outliers, and multi-modality occur simultaneously. We propose a model that presents robustness against violations of these assumptions, assuming that the distribution of the covariate belongs to a highly flexible family of distributions, defined as a finite mixture of skew-normal independent distributions. The model can be applied in many practical situations such as comparative calibration of instruments, where the bias and precision of measurements made using some instruments are evaluated based on measurements made by a reference one. The main goals are (i) develop algorithms for Bayesian estimation of the parameters of the proposed model; (ii) investigate, through simulation, the performance of the model selection criterion DIC (Deviance Information Criterion) as a suitable method to choose between the different considered models, including the determination of the number of component mixtures and (iii) apply the proposed methodology by considering the analysis of simulated and real data sets.

2 The SNI Family

In what follows $\mathbf{0}_q$ is the zero vector of $\mathbb{R}^q$, $\mathbf{I}_q$ is the identity matrix with dimension q – we drop the index q when there is no possibility of confusion. The $q-$variate skew-normal distribution is given by the density

$$\mathrm{SN}_q(\mathbf{x}|\boldsymbol{\xi},\,\boldsymbol{\Omega},\,\boldsymbol{\lambda}) = 2\mathrm{N}_q(\mathbf{x}|\boldsymbol{\xi},\,\boldsymbol{\Omega})\Phi(\,\boldsymbol{\lambda}^\top\,\boldsymbol{\Omega}^{-1/2}(\mathbf{x}-\boldsymbol{\xi})),\quad \mathbf{x}\in\mathbb{R}^q,$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\mathrm{N}_q(\cdot|\boldsymbol{\xi},\,\boldsymbol{\Omega})$ denotes the $q-$dimensional normal density with mean vector $\boldsymbol{\xi}$ and positive definite covariance matrix $\boldsymbol{\Omega}$, $\boldsymbol{\lambda}:q\times 1$ is a

*The authors acknowledge the partial financial support from CAPES and CNPq. *Email addresses*: `tabensur@ufam.edu.br` (Themis C. Abensur), `celsoromulo@gmail.com` (Celso R. B. Cabral), `jcardoso@ufam.edu.br` (José Cardoso Neto)

vector of parameters that regulate skewness, $\boldsymbol{\lambda}^{\top}$ is the transpose of $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}^{-1/2}$ is the root of $\boldsymbol{\Omega}$, that is, a symmetric matrix that satisfies $\boldsymbol{\Omega}^{-1/2}\,\boldsymbol{\Omega}^{-1/2} = \boldsymbol{\Omega}^{-1}$. For a random vector $\mathbf{Y}$ with this distribution, we use the notation $\mathbf{Y} \sim \mathrm{SN}_q(\boldsymbol{\xi},\, \boldsymbol{\Omega},\, \boldsymbol{\lambda})$. This definition is a slight modification of the Azzalini's skew-normal distribution (Azzalini & Capitanio, 1999) – see Arellano-Valle *et al.* (2005). For computational purposes, an useful parametrization is given by $\boldsymbol{\Delta} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Omega}^{1/2}(\mathbf{I} - \boldsymbol{\delta}\boldsymbol{\delta}^{\top})\,\boldsymbol{\Omega}^{1/2}$, $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^{\top}\,\boldsymbol{\lambda})^{1/2}$.

Let $\mathbf{X} \sim \mathrm{SN}_q(\mathbf{0},\, \boldsymbol{\Omega},\, \boldsymbol{\lambda})$. An element of the SNI family is defined as the distribution of the $q-$dimensional random vector $\mathbf{Y} = \boldsymbol{\xi} + S^{-1/2}\mathbf{X}$, where S is a positive random variable – called the *scale factor* – with density $h(\cdot|\nu)$, which is independent of $\mathbf{X}$. Here ν is a scalar or vector parameter indexing the distribution of S. We use the notation $\mathbf{Y} \sim \mathrm{SNI}_q(\boldsymbol{\xi},\, \boldsymbol{\Omega},\, \boldsymbol{\lambda}, \nu)$. The following lemma is crucial in our theory:

Lemma 1. *Let* $\mathbf{Y} \sim SNI_q(\boldsymbol{\xi},\, \boldsymbol{\Omega},\, \boldsymbol{\lambda}, \nu)$, $\mathbf{a} : p \times 1$ *and* $\mathbf{B} : p \times q$. *Then*

$$\mathbf{a} + \mathbf{B}\mathbf{Y} \sim SNI_p(\mathbf{a} + \mathbf{B}\boldsymbol{\xi}, \mathbf{B}\,\boldsymbol{\Omega}\mathbf{B}^{\top}, \overline{\boldsymbol{\lambda}}, \nu),$$

where $\overline{\boldsymbol{\lambda}} = \overline{\boldsymbol{\delta}}/\sqrt{1 - (\overline{\boldsymbol{\delta}})^{\top}\overline{\boldsymbol{\delta}}}$, $\overline{\boldsymbol{\delta}} = (\mathbf{B}\,\boldsymbol{\Omega}\mathbf{B}^{\top})^{-1/2}\mathbf{B}\,\boldsymbol{\Omega}^{1/2}\boldsymbol{\delta}$.

For a proof, see Corollary 2 in Arellano-Valle *et al.* (2005). The distributions in the SNI class that will be considered in this work are: (i) *The* $q-$*variate skew-normal distribution* – when $S = 1$; (ii) *the* $q-$*variate skew-t distribution with* ν *degrees of freedom* (Branco & Dey, 2001) – when $S \sim \mathrm{Gamma}(\nu/2, \nu/2)$, $\nu > 0$; (iii) *The* $q-$*variate skew-contaminated normal distribution* – when S is discrete, taking the value τ with probability ρ and the value 1 with probability $1 - \rho$, where $\rho, \tau \in (0, 1)$ and (iv) *The* $q-$*variate skew-slash distribution* (Wang & Genton, 2006) – when S has a beta distribution with parameters $\nu > 0$ and 1.

3 The SNI Mixture Measurement Error Model

3.1 The Model

The classical *measurement error (ME) model* is defined as

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i,$$

where $\mathbf{Y}_i : r \times 1$ is a random vector of measurements made on individual i, $i = 1, \ldots, n$, $\mathbf{e}_i : r \times 1$ is a vector of independent measurement errors and x_i is the latent (unobserved) covariate for individual i. Instead, we observe $X_i = x_i + \delta_i$, where δ_i is the random error when observing X_i. $\boldsymbol{\alpha}$, $\boldsymbol{\beta} : r \times 1$ are vectors of parameters to be estimated. The following assumption is usually made:

$$\mathbf{r}_i = \left(x_i \;\; \boldsymbol{\varepsilon}_i^{\top}\right)^{\top} \overset{\text{iid}}{\sim} \mathrm{N}_{p+1}\left[\left(\xi \;\; \mathbf{0}^{\top}\right)^{\top}, \mathrm{diag}\left\{\sigma^2, \phi_1, \ldots, \phi_p\right\}\right], \tag{3.1}$$

where $p = r + 1$, $\boldsymbol{\varepsilon}_i = (\delta_i, \mathbf{e}_i^{\top})^{\top}$ and $\overset{\text{iid}}{\sim}$ denotes independent and identically distributed random vectors. Recently, Lachos *et al.* (2009a,b) have extended this model, by considering that

$$\mathbf{r}_i \overset{\text{iid}}{\sim} \mathrm{SNI}_{p+1}\left[\left(\xi \;\; \mathbf{0}^{\top}\right)^{\top}, \mathrm{diag}\left\{\sigma^2, \phi_1, \ldots, \phi_p\right\}, \left(\lambda \;\; \mathbf{0}^{\top}\right)^{\top}, \nu\right], \tag{3.2}$$

which marginally implies that

$$x_i \stackrel{\text{iid}}{\sim} \text{SNI}(\cdot|\xi, \sigma^2, \lambda, \nu), \quad \boldsymbol{\varepsilon}_i \stackrel{\text{iid}}{\sim} \text{SNI}_p(\mathbf{0},\ \mathbf{\Omega}, \mathbf{0}, \nu), \qquad i = 1, \ldots, n.$$

where $\mathbf{\Omega} = \text{diag}\left\{\phi_1, \ldots, \phi_p\right\}$. Note that the scale factor associated to subject i – and the corresponding density $h(\cdot|\nu)$ – is the same for latent variables and random errors. Thus, unless $S_i = 1$ for all i – that is, in the skew-normal case, x_i and $\boldsymbol{\varepsilon}_i$ are not independent, a typical assumption for measurement error models. But they are uncorrelated, see again Lachos *et al.* (2009a) for details.

In this context, the SNI class of models is suitable for dealing with extra skewness and discrepant observations, allowing us to model the latent covariate and errors in a more flexible way. Although this approach can be successfully used in a wide range of applications, it is clearly not sufficient when there is unobserved heterogeneity or multi-modality in the latent covariate population. To circumvent this problem, we propose to model it using a finite mixture of SNI distributions. A similar approach was adopted in some previous works, see Roy & Banerjee (2006), where the latent variable is modeled by a mixture of normal distributions, and the references herein.

Specifically, we propose to change the structure defined in (3.1) and (3.2) for

$$\mathbf{r}_i|(w_i = j) \stackrel{\text{iid}}{\sim} \text{SNI}_{p+1}\left[\begin{pmatrix}\xi_j & \mathbf{0}^\top\end{pmatrix}^\top, \text{block diag}\left\{\sigma_j^2,\ \mathbf{\Omega}\right\}, \begin{pmatrix}\lambda_j & \mathbf{0}^\top\end{pmatrix}^\top, \nu\right],$$

where w_i is a latent classification variable such that $w_i = j$ with probability p_j, $j = 1, \ldots, G$, which marginally implies that

$$x_i|w_i = j \stackrel{\text{iid}}{\sim} \text{SNI}(\xi_j, \sigma_j^2, \lambda_j, \nu), \quad \boldsymbol{\varepsilon}_i \stackrel{\text{iid}}{\sim} \text{SNI}_p(\mathbf{0},\ \mathbf{\Omega}, \mathbf{0}, \nu), \qquad i = 1, \ldots, n,$$

and it is straightforward to prove that

$$x_i \stackrel{\text{iid}}{\sim} \sum_{j=1}^{G} p_j \text{SNI}(\cdot|\xi_j, \sigma_j^2, \lambda_j, \nu), \quad \boldsymbol{\varepsilon}_i \stackrel{\text{iid}}{\sim} \text{SNI}_p(\mathbf{0},\ \mathbf{\Omega}, \mathbf{0}, \nu), \qquad i = 1, \ldots, n.$$

We call this model the *finite mixture skew-normal independent measurement error* model (FMSNI-ME). An useful application is the *comparative calibration of instruments*. In this case, measurements are made on individual i by $r+1$ instruments. X_i is an univariate measurement made by a standard – *the reference measurement* – instrument and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir})^\top$ is a vector with the remaining r measurements made by other r instruments, that is, Y_{ij} is the observation on individual i made by instrument j.

3.2 The Sample Distribution

Consider the vector of observations $\mathbf{Z}_i = (X_i, \mathbf{Y}_i^\top)^\top = \mathbf{a} + \mathbf{B}\mathbf{r}_i$, where $\mathbf{a} = \begin{pmatrix}0 & \boldsymbol{\alpha}^\top\end{pmatrix}^\top$, $\mathbf{B} = \begin{pmatrix}\mathbf{B}_1 & \mathbf{I}_{r+1}\end{pmatrix}$, with $\mathbf{B}_1 = (1 \quad \boldsymbol{\beta}^\top)^\top$. As a direct application of Lemma 1, we have that

$$\mathbf{Z}_i|w_i = j \sim \text{SNI}_{r+1}\left(\boldsymbol{\mu}_j, \mathbf{\Psi}_j, \boldsymbol{\varphi}_j, \nu\right),$$

where

$$\boldsymbol{\mu}_j = \left(\begin{array}{c} \xi_j \\ \boldsymbol{\alpha} + \boldsymbol{\beta}\xi_j \end{array} \right), \quad \boldsymbol{\Psi}_j = \sigma_j^2 \mathbf{B}_1 \mathbf{B}_1^\top + \boldsymbol{\Omega}, \quad \boldsymbol{\varphi}_j = \frac{\sigma_j \lambda_j \boldsymbol{\Psi}_j^{-1/2} \mathbf{B}_1}{\left(1 + (\lambda_j^2/\sigma_j^2)\Lambda_j\right)^{1/2}},$$

and $\Lambda_j = \sigma_j^2(1 - \mathbf{B}_1^\top \boldsymbol{\Psi}_j^{-1} \mathbf{B}_1 \sigma_j^2)$. This means that the marginal distribution of $\mathbf{Z}_i$ is a mixture of SNI distributions, with density given by

$$\pi(\mathbf{z}_i) = \sum_{j=1}^{G} p_j \mathrm{SNI}_{r+1}(\mathbf{z}_i | \boldsymbol{\mu}_j, \boldsymbol{\Psi}_j, \boldsymbol{\varphi}_j, \nu), \tag{3.3}$$

where $\mathrm{SNI}(\cdot|\cdot)$ is the SNI density with the specified parameters.

3.3 Stochastic Representation

Using the representation of the skew-normal distribution of Henze (1986) and the definition of the SNI class, we can prove that

Theorem 1. *The FMSNI-ME model admits the following stochastic representation*

$$\begin{aligned}
\mathbf{Z}_i | x_i, S_i = s_i &\sim N_p(\mathbf{a} + \mathbf{B}_1 x_i, s_i^{-1}\,\boldsymbol{\Omega}); \\
x_i | T_i = t_i, S_i = s_i, w_i = j &\sim N(\xi_j + \Delta_j t_i, s_i^{-1}\Gamma_j); \\
t_i | S_i = s_i &\sim TN(0, s_i^{-1}, (0,\infty)); \\
s_i &\sim h(\cdot|\nu); \\
P(w_i = k) &= p_k, \qquad i = 1,\ldots,n,\ k = 1,\ldots,G,
\end{aligned}$$

where $\Delta_j = \sigma_j \delta_j$, $\Gamma_j = \sigma_j^2(1-\delta_j^2) = \sigma_j^2 - \Delta_j^2$, $\delta_j = \lambda_j/(1+\lambda_j^2)^{1/2}$ *and TN denotes the truncated normal distribution on* $(0,\infty)$.

This representation is useful to obtain MCMC-type (and EM-type, for maximum likelihood estimation) algorithms for estimation.

4 Priors

We assume prior independence between the parameters and a little informative prior structure. The setup is

$(p_1,\ldots,p_k) \sim \mathrm{Dir}(\gamma_1,\ldots,\gamma_G)$, $\boldsymbol{\alpha} \sim \mathrm{N}_r(\boldsymbol{\mu}_\alpha,\ \boldsymbol{\Sigma}_\alpha)$, $\boldsymbol{\beta} \sim \mathrm{N}_r(\boldsymbol{\mu}_\beta,\ \boldsymbol{\Sigma}_\beta)$, $1/\phi_i \sim \mathrm{Gamma}(a,b)$,
$\xi_j \sim \mathrm{N}(c,d)$, $\Gamma_j^{-1}|f \sim \mathrm{Gamma}(e,f)$, $f \sim \mathrm{Gamma}(g,h)$, $\Delta_j|r,U \sim \mathrm{N}(r, \sigma_\Delta/U)$,
$r \sim \mathrm{N}_q(0,V)$, $U \sim \mathrm{Gamma}(\kappa/2, \kappa/2)$,

where γ_i, $\boldsymbol{\mu}_\alpha$, $\boldsymbol{\Sigma}_\alpha$, $\boldsymbol{\mu}_\beta$, $\boldsymbol{\Sigma}_\beta$, a, b, c, d, e, g, h, σ_Δ and κ are fixed hyperparameters. This hierarchical setup is inspired by suggestions made by Richardson & Green (1997). In addition, for the specific models a prior to ν has to be chosen according to the respective interpretation of the parameter. Our choice is (i) *For the skew-t model*: $\nu \sim \exp(\lambda)$, $\lambda \sim \text{Uniform}(\phi, \psi)$; (ii) *For the skew-contaminated normal model*: $\rho \sim \text{beta}(\rho_0, \rho_1)$ and $\tau \sim \text{beta}(\tau_0, \tau_1)$; (iii) *For the skew-slash model*: $\nu \sim \text{Gamma}(\phi_{sl}, \psi_{sl})$.

Combining this prior setup with the stochastic representation of the model given above we obtain closed full conditionals for all parameters in order to implement a Gibbs-type algorithm, except for ν. In this case, Metropolis paths are applied.

5 The DIC Criterion

We use the model selection criterion called *observed deviance information criterion* (DIC_{obs}). It is a modified version of the DIC criterion, which is not adequate for mixture models (Celeux *et al.*, 2006), because its calculation requires an estimator of the unknown $\mathbf{\Theta}$, the vector with all likelihood parameters, which may suffer from *label switching*, making DIC unstable. Letting $\mathbf{z}$ be the vector with all observations and denoting densities by $\pi(\cdot)$, it is defined as

$$\text{DIC}_{\text{obs}} = -4\text{E}\{\log[\pi(\mathbf{z}|\mathbf{\Theta})]|\mathbf{z}\} + 2\log(\text{E}[\pi(\mathbf{z}|\mathbf{\Theta})|\mathbf{z}]). \tag{5.4}$$

Using result (3.3), the first posterior expectation in this expression can be approximated by

$$\overline{D} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\log\left[\sum_{k=1}^{G}(p_k)^{(i)}\text{SNI}_{r+1}(\mathbf{z}_j|(\boldsymbol{\mu}_k)^{(i)}, (\boldsymbol{\Psi}_k)^{(i)}, (\boldsymbol{\varphi}_k)^{(i)}, \nu^{(i)})\right],$$

where $(p_k)^{(i)}$, $(\boldsymbol{\mu}_k)^{(i)}$, $(\boldsymbol{\Psi}_k)^{(i)}$, $(\boldsymbol{\varphi}_k)^{(i)}$ and $\nu^{(i)}$ are the corresponding MCMC samples, generated at the iteration i of the algorithm, $i = 1, \ldots, m$. As recommended by Celeux *et al.* (2006), the second term in the right hand of (5.4) can be approximated by $2\sum_{j=1}^{n}\log\widehat{\pi(\mathbf{z}_j)}$, where

$$\widehat{\pi(\mathbf{z}_j)} = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{G}(p_k)^{(i)}\text{SNI}_{r+1}(\mathbf{z}_j|(\boldsymbol{\mu}_k^{(i)}, (\boldsymbol{\Psi}_k)^{(i)}, (\boldsymbol{\varphi}_k)^{(i)}, \nu)$$

is the MCMC estimate of the posterior predictive density value $\text{E}[\pi(\mathbf{z}_j|\mathbf{\Theta})|\mathbf{z}]$. It is important to emphasize that this estimate is robust against possible label switching, circumventing the main drawback of using the original DIC for mixture models.

6 Application

To illustrate the usefulness of the proposed methodology, we modeled the testicular volume data set, consisting of measurements made on 42 adolescents (Chipkevitch *et al.*, 1996). Four devices are compared with a reference one. This data set was analyzed before by many authors which performed first a cubic root transformation in order to achieve normality see, for example, Galea-Rojas *et al.* (2002). Lachos *et al.* (2009b) modeled this data set considering that the distribution of the measurements of the reference device belongs to the SNI family. We can see that the histogram of the reference instrument measurements suggests multi-modality, and possibly a better fit can be achieved by using finite mixtures of SNI distributions. The results obtained were very satisfactory, implying in a better fit and shorter standard deviations for the estimates of the parameters when comparing with that obtained using the SNI distributions with only one component.

References

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

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

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

Celeux, G., Forbes, F., Robert, C. P. & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, **1**, 651–674.

Chipkevitch, E., Nishimura, R. T., Tu, D. G. S. & Galea-Rojas, M. (1996). Clinical measurement of testicular volume in adolescents: Comparison of the reability of 5 methods. *The Journal of Urology*, **156**, 2050–2053.

Galea-Rojas, M., Bolfarine, H. & de Castro, M. (2002). Local influence in comparative calibration models. *Biometrical Journal*, **44**.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.

Lachos, V. H., Garibay, V., Labra, F. V. & Aoki, R. (2009a). Robust multivariate measurement error model with skew-normal/independent distributions and Bayesian MCMC implementation. *Statistical Methodology*. doi:10.1016/j.stamet.2009.06.002.

Lachos, V. H., Labra, F. V., Bolfarine, H. & Ghosh, P. (2009b). Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics*. doi: 10.1080/02331880903236926.

Richardson, S. & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, **59**, 731–792.

Roy, S. & Banerjee, T. (2006). A flexible model for generalized linear regression with measurement error. *Annals of the Institute of Statistical Mathematics*, **58**, 153–169.

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