Robust Multivariate Measurement Error Model with Skew-Normal/ Independent Distributions

V. H. Lachos*[1], F.V. Labra* and H. Bolfarine**

* Universidade Estadual de Campinas, Campinas, Brazil

** Universidade de São Paulo, São Paulo, Brazil

Abstract

Skew-normal/independent distributions is a class of asymmetric thick-tailed distributions that includes the skew-normal distribution as a especial case. The main virtue of the members of this class of distributions is that they are easy to simulate from and they make it possible to implement the Monte Carlo EM algorithm for maximum likelihood estimation. In this paper, we take skew-normal/independent distributions (Lachos and Vilca, 2007) for the unobserved value of the covariates (latent variable) and symmetric normal/independent (Lange and Sinsheimer, 1993) distributions for the random errors providing an appealing robust alternative to the usual symmetric process in multivariate measurement errors models. Specific distributions examined include univariate and multivariate versions of the skew-normal, the skew-t, the skew-slash and the contaminated skew-normal distribution. The results and methods are applied to a real data set.

Key Words: Monte Carlo EM algorithm; skew-normal/independent distributions, Mahalanobis distance.

1 Introduction

Measurement error models (MEM) are useful for a variety of phenomena modeling in many disciplines. The MEM describes functional relationships among variables observed subject to random measurement errors. Examples include linear and nonlinear errors-in-variables regression models, factor analysis models, latent structural models, and simultaneous equations models. Such models are used, among others, in medicine, life sciences, econometrics, chemometrics and geology. Hence, a study of its properties under nonstandard assumptions, like normality for example, is very pertinent. This model corresponds to a multivariate structural linear regression model with a single predictor subject to random measurement errors (Kendall and Stuart, 1979; Fuller, 1987), which is frequently used to compare measuring instruments (Barnett, 1969, Theobald and Mallison, 1978, Shyr and Gleser, 1986, Bolfarine and Galea-Rojas, 1996 and Chipkevitch et al., 1996). If $r = 1$ then the classical linear errors-in-variables regression models follow which, under symmetrical distributions, has been widely discussed in the literature, as can be seen in texts by

[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.

Fuller (1987), Brown and Fuller (1990) and Cheng and van Ness (1999) as well as in the paper by Arellano-Valle and Bolfarine (1996). In this paper we will focus on the case $r > 1$.

On the other hand, recent statistical literature has seen an increasing interest for models providing flexibility in capturing a broad range of non-normal behavior, such as skewness, and thus, representing features of the data as adequately as possible and to reduce unrealistic assumptions. Advantages of using such general structures include easiness of interpretation, as well as estimation efficiency. In this paper we consider a classical approach of multivariate measurement errors models (MMEM), where the unobserved covariate (x) is a continuous random variable that follows skew-normal/independent distributions (SNI, Lachos and Vilca, 2007), the random errors follows symmetric normal/independent distributions and that we observe a vector of responses ($\mathbf{Y}$) of dimension r for each experimental unit. The SNI distributions is a rich class of distributions that contains as proper elements the skew-normal, the skew-t, the skew-slash, the skew-power exponential and the contaminated skew-normal distribution. This study is motivated by the fact that many data sets considered in the literature seem to present nonnormal behavior, such as asymmetry and heavy tails. This is the case with the data sets in Barnett (1969) and in Chipkevitch et al. (1996) (see also Bolfarine et al., 2002) which requires data transformation in order to be better approximated by the normal (or by a symmetric) distribution. Specifically, we first extend the normal MMEM by considering a hierarchical asymmetric version of this model, implying that the observed responses follow a skew-normal/independent distribution so that the *Skew-Normal/independent Multivariate Measurement Errors Model* (SNI-MMEM) is defined. Closed form expressions are obtained for the likelihood function which extends results in Arellano–Valle, Ozan, Bolfarine and Lachos (2005) and Lachos, Bolfarine, Vilca and Galea–Rojas (2005).

The paper is organized as follows. In Section 2, for the sake of completeness, we give a brief sketch of SNI distributions. In Section 3 we present the multivariate skew-normal/independent measurement error model (SNI-MMEM). In Section 4 we discuss the Monte Carlo EM algorithm for maximum likelihood (ML) estimation in SNI-MMEM. The observed information matrix is derived analytically in Section 5. The methodology proposed for SNI-MMEM models is illustrated in Section 6 considering a real data set and finally, some concluding remarks are presented in Section 7.

2 Skew-normal/independent distributions

A somewhat simpler departure from normality has been proposed by Azzalini (1985), by defining the univariate skew-normal distribution. An extension to the multivariate setting was proposed by Azzalini and Dalla–Valle (1996), defining the following probability density function (pdf)

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

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 $\boldsymbol{\Sigma}$, $\Phi_1(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution, and $\boldsymbol{\Sigma}^{-1/2}$ satisfies $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. When $\boldsymbol{\lambda} = \mathbf{0}$, the skew normal distribution reduces to the normal distribution ($\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$). A p-dimensional random vector $\mathbf{Y}$ with pdf as in (1) is denoted by $\mathrm{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, with marginal stochastic representation given by

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

where $|T_0|$ denotes the absolute value of T_0, $T_0 \sim \mathrm{N}_1(0, 1)$ and $\mathbf{T}_1 \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$ are independent, and " $\stackrel{d}{=}$ " means "distributed as". Reasoning as in Azzalini (1985), one can define more general models such as skew-t distributions (Sahu et al., 2003; Gupta, 2003), skew-Cauchy distributions (Arnold and Beaver, 2002), skew-slash distributions (Wang and Genton, 2006), skew-slash-t distributions (Tan and Peng, 2006) and skew-elliptical distributions (Azzalini and Capitanio, 1999; Branco and Dey, 2001; Sahu et a., 2003; Genton and Loperfido, 2005). Recently, Lachos and Vilca (2007) define the SNI distributions which combine skewness with heavy tails and study many of its properties and applications.

The asymmetrical class of SNI distributions has attracted attention, particularly due to the fact that they include distributions such as the skew-t, the skew-slash, the skew-power exponential, and the contaminated skew-normal. All these distributions have heavier tails than the skew-normal ones and can be used for robust inference in many type of models. We say that a p-dimensional random vector $\mathbf{Y}$ follows a SNI distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale matrix $\boldsymbol{\Sigma}$ (a $p \times p$ positive definite matrix) and skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$, if its pdf is given by

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))dH(u) \\ &= 2\int_0^\infty \frac{u^{p/2}}{(2\pi)^{p/2}}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{u}{2}d}\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))dH(u), \end{aligned} \tag{3}$$

where $d = (\mathbf{y} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance and U is a positive random variable with cdf $H(u; \boldsymbol{\nu})$ indexed by the parameter vector $\boldsymbol{\nu} \in \mathbb{R}^r$. For a random vector with pdf as in (3), we use the notation $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$ we refer to it as a standard SNI distribution and we denote it by $SNI_p(\boldsymbol{\lambda}; H)$. Its stochastic representation is given by

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{4}$$

where $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and U is a positive random variable with cdf H independent of $\mathbf{Z}$. Moreover, from (2) it follows that (4) can be written as

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U^{-1/2}\boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|X_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{X}_1\}, \tag{5}$$

where U, $X_0 \sim N_1(0,1)$ and $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ are all independent. Some of these distributions are described subsequently. For each element of this class, the conditional expectation $\widehat{u} = E[U|\mathbf{y}]$ and the distributional properties of the Mahalanobis distance are described, because they are extremely useful to the implementation of the EM-algorithm and in testing goodness of fit (or detecting outliers), respectively.

2.1 Examples of SNI distributions

- *The skew-t distribution, with ν degrees of freedom, $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$.* Considering $U \sim Gamma(\nu/2, \nu/2)$ with pdf of the form

$$h(u|\nu) = \frac{(\nu/2)^{\nu/2} u^{\nu/2-1}}{\Gamma(\nu/2)} \exp\left(-\frac{1}{2}\nu u\right). \tag{6}$$

Similar procedures found in Gupta (2003, Section 2) lead to the following density function:

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T_1\left(\frac{\sqrt{v+p}\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{d+p}}|0, 1, \nu+p\right), \quad \mathbf{y} \in \mathbb{R}^p, \tag{7}$$

where as usual, $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denote, respectively, the pdf and cdf of the Student-t distribution, namely $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. A particular case of the skew-t distribution is the skew-Cauchy distribution, when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case. See Gupta (2003) for further details. From Lachos and Vilca (2007), it also follows that

$$d = (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \sim pF(p, \nu).$$

$$u|\mathbf{y} \sim Gamma\left(\frac{\nu+p}{2}, \frac{\nu+d}{2}\right),$$

so that taking conditional expectation, $\widehat{u} = \mathrm{E}[U|\mathbf{y}] = \dfrac{\nu+p}{\nu+d}$.

- *The skew-slash distribution with the shape parameter $\nu > 0$, namely $SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$.* Its pdf is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = 2\nu \int_0^1 u^{\nu-1} \phi_p\left(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u}\right) \Phi_1(u^{1/2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})) du, \quad \mathbf{y} \in \mathbb{R}^p, \tag{8}$$

and U has density given by

$$h(u|\nu) = \nu u^{\nu-1} \mathbb{I}_{(0,1)}, \ \nu > 0, \tag{9}$$

where the notation $\mathbb{I}_{(A)}$ is the indicator function of the the set A. The skew-slash distribution reduces to the skew-normal distribution when $\nu \uparrow \infty$. See Wang and Genton (2006) for further details. The Mahalanobis distance has cdf

$$Pr(d \le r) = Pr(\chi^2_p \le r) - \frac{2^\nu \Gamma(p/2+\nu)}{r^\nu \Gamma(p/2)} Pr(\chi^2_{p+2\nu} \le r)$$

and the conditional distribution in this case has the form

$$u|\mathbf{y} \sim Gamma(\nu + \frac{p}{2}, \frac{d}{2})\mathbb{I}_{(0,1)},$$

so that $\widehat{u} = \left(\frac{p+2\nu}{d}\right)\frac{P_1(p/2+\nu+1, d/2)}{P_1(p/2+\nu, d/2)}$, where $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ distribution evaluated at x.

- *The contaminated skew-normal distribution,* $SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$, $0 \leq \nu \leq 1$, $0 < \gamma \leq 1$. Another SNI, arises when 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

$$h(u|\boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \;\; \boldsymbol{\nu} = (\nu, \gamma)^\top, \tag{10}$$

It follows straightforwardly that

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= 2\{\nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\gamma})\Phi_1(\gamma^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})) \\ &\quad +(1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))\}. \end{aligned} \tag{11}$$

Parameter ν can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor. The skew-contaminated normal distribution reduces to the skew-normal distribution when $\gamma = 1$. In this case, from Lachos and Vilca (2007), we have that

$$Pr(d \leq r) = \nu Pr(\chi^2_p \leq \gamma r) + (1-\nu)Pr(\chi^2_p \leq r),$$

and

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

with $p = \dfrac{u^{p/2}\exp\{-\frac{du}{2}\}}{\nu\gamma^{p/2}\exp\{-\frac{d\gamma}{2}\} + (1-\nu)\exp\{-\frac{d}{2}\}}$, so that,

$$\widehat{u} = \frac{1-\nu+\nu\gamma^{p/2+1}\exp\{(1-\gamma)d/2\}}{1-\nu+\nu\gamma^{p/2}\exp\{(1-\gamma)d/2\}}.$$

The skew–power exponential distribution is of the SNI type. However, in this case, the scale distribution $H(u;\boldsymbol{\nu})$ is not computationally attractive and it will not be dealt with in this work.

3 The model

Let n be the sample size; X_i, the observed value of the covariate in unit i; y_{ij}, the j-th observed response in unit i and x_i, the unobserved (true) covariate value for unit i, $i = 1, \ldots, n$ and $j = 1, \ldots, r$. Relating these variables we postulate as working model the equations (see, Barnett, 1969 and Shyr and Gleser, 1986),

$$X_i = x_i + u_i, \tag{12}$$

and

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i, \tag{13}$$

where $\mathbf{Y}_i = (y_{i1}, \ldots, y_{ir})^\top$, is the vector of responses for the i-th experimental unit, $\mathbf{e}_i = (e_{i1}, \ldots, e_{ir})^\top$, is a random vector of measurement errors of dimension r, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)^\top$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_r)^\top$ are parameter vectors of dimension r. Let $\boldsymbol{\epsilon}_i = (u_i, \mathbf{e}_i^\top)^\top$ and $\mathbf{Z}_i = (X_i, \mathbf{Y}_i^\top)^\top$, then the model defined by equations (4)-(5) can be written as

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b} x_i + \boldsymbol{\epsilon}_i = \mathbf{a} + \mathbf{B}\mathbf{r}_i, \tag{14}$$

where $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$ and $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top$ are $p \times 1$ vectors, with $p = r + 1$, and $\mathbf{B} = [\mathbf{b}; \mathbf{I}_p]$ a $p \times (p+1)$ matrix and $\mathbf{r}_i = (x_i, \boldsymbol{\epsilon}_i^\top)^\top$, $i = 1, \ldots, n$. Thus, using (14), the distribution of $\mathbf{Z}_i$ become specified once the distribution of $\mathbf{r}_i$ is specified, $i = 1, \ldots, n$. In this paper, we assume that

$$\mathbf{r}_i = \begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{\text{iid}}{\sim} SNI_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix}; H \right), \tag{15}$$

$i = 1, \ldots, n$, where $D(\phi_x, \boldsymbol{\phi}) = diag(\phi_x, \phi_1, \ldots, \phi_p)^\top$, with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, which will be called structural SNI-MMEM. From (4), this formulation implies that

$$\begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} | U_i = u_i \;\sim\; SN_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, u_i^{-1} D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix} \right), \tag{16}$$

$$U_i \;\sim\; h(u_i | \boldsymbol{\nu}), \tag{17}$$

$i = 1, \ldots, n$. It can be shown that conditional on U_i, $\boldsymbol{\epsilon}_i$ and x_i are independent (Azzalini and Capitanio, 1999, Proposition 6) and further,

$$\boldsymbol{\epsilon}_i | u_i \overset{\text{ind}}{\sim} N_{m+1}(\mathbf{0}, u_i^{-1}\mathrm{D}(\boldsymbol{\phi})) \;\text{ and }\; x_i | u_i \overset{\text{ind}}{\sim} SN_1(\mu_x, u_i^{-1}\phi_x, \lambda_x). \tag{18}$$

this is, conditional on U_i, $\boldsymbol{\epsilon}_i \sim SNI_p(\mathbf{0}, \mathrm{D}(\boldsymbol{\phi}), \mathbf{0}; H)$ and $x_i \sim SNI_1(\mu_x, \phi_x, \lambda_x; H)$.

The above model is considering that in the case of Barnett's (1969) data, vital capacity is not symmetrically distributed in the population. The same seems to be the case with testicular density data studied in the case of Chepkevitch et al.'s (1996) data set. On the other hand, the errors $\boldsymbol{\epsilon}_i$, are related to measurement error so that it is expected to be symmetrically distributed. The asymmetric parameter λ_x incorporates asymmetry in the latent variable x_i and consequently in the observed quantities $\mathbf{Z}_i$, $i = 1, \ldots, n$, which will be shown to have marginally multivariate SNI distributions. If $\lambda_x = 0$, then the asymmetric model reduces to the symmetric MMEM considering normal/independent (NI, Lange and Sinsheimer, 1993) distributions. Note from (2) that the regression set up defined in (14)-(15)

can be written hierarchically as

$$\mathbf{Z}_i \mid x_i, U_i = u_i \overset{\text{ind}}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, u_i^{-1}D(\boldsymbol{\phi})), \tag{19}$$

$$x_i \mid T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N_1(\mu_x + \phi_x^{1/2}\delta_x u_i^{-1/2}t_i, u_i^{-1}\phi_x(1-\delta_x^2)), \tag{20}$$

$$T_i \overset{\text{iid}}{\sim} HN_1(0,1), \tag{21}$$

$$U_i \overset{\text{iid}}{\sim} h(u_i|\boldsymbol{\nu}), \tag{22}$$

$i = 1, \ldots, n$, all independent, where $HN_1(0,1)$ denotes the standardized univariate half-normal distribution and $\delta_x = \lambda_x/(1+\lambda_x^2)^{1/2}$. As in Fernandez and Steel (1999), we assumed that $\boldsymbol{\nu}$ is known. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top$ in this type of model is based on the marginal distribution for $\mathbf{Z}_i$ (see, Bolfarine and Galea-Rojas, 1995), given in the following proposition.

Proposition 1. *Under the structural SNI-MMEM defined in (14)-(15), the marginal distribution of $\mathbf{Z}_i$ is given by*

$$f_{\mathbf{Z}_i}(\mathbf{z}_i|\boldsymbol{\theta}) = 2\int_0^\infty \phi_p(\mathbf{z}_i|\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma})\Phi_1(u_i^{1/2}\bar{\boldsymbol{\lambda}}_x^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu}))dH(u_i), \tag{23}$$

i.e., $\mathbf{Z}_i \overset{\text{iid}}{\sim} SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x; H)$, $i = 1, \ldots, n$, *where*

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x, \quad \boldsymbol{\Sigma} = \phi_x\mathbf{b}\mathbf{b}^\top + D(\boldsymbol{\phi}) \ \textit{and} \ \bar{\boldsymbol{\lambda}}_x = \frac{\lambda_x\phi_x\boldsymbol{\Sigma}^{-1/2}\mathbf{b}}{\sqrt{\phi_x + \lambda_x^2\Lambda_x}},$$

with $\Lambda_x = ({\phi_x}^{-1} + \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b})^{-1} = \phi_x/c$ *and* $c = 1 + \phi_x\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$.

Proof. See Lemmas 1 and 2 in Arellano-Valle, Bolfarine and Lachos (2005) or Proposition 5 in Lachos and Vilca (2007). □

It follows that the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}), \tag{24}$$

where $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2}log2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, with

$$K_i = K_i(d_i, A_i) = \int_0^\infty u_i^{p/2}\exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i)dH(u_i),$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\bar{\boldsymbol{\lambda}}_x$ as in Proposition 1, $d_i = (\mathbf{z}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$ and $A_i = \bar{\boldsymbol{\lambda}}_x^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu}) = A_x a_i$, with

$$A_x = \frac{\lambda_x\Lambda_x}{\sqrt{\phi_x + \lambda_x^2\Lambda_x}} \quad \text{and} \quad a_i = (\mathbf{z}_i - \boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}.$$

The result presented in Proposition 1 is important because it facilitates straightforward implementation of inferences with standard optimization routines and existing statistical softwares. The asymptotic covariance matrix of the maximum likelihood estimators can be estimated by using the Hessian matrix, which can also be computed numerically using, for instance, the *optim* routine in platform R. Here, to obtain the maximum likelihood estimator of $\boldsymbol{\theta}$ we use the EM algorithm and derived algebraically its asymptotic covariance matrix.

4 Maximum likelihood estimation

The EM algorithm (Dempster, Laird, and Rubin, 1977) is a popular iterative algorithm for ML estimation in models with incomplete data. More specifically, let $\mathbf{z}$ denote the observed data and $\mathbf{s}$ denoted the missing data. The complete data, namely $\mathbf{z}_{comp} = (\mathbf{z}, \mathbf{s})$, is $\mathbf{z}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{s})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete-data log-likelihood function and by $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ the expected complete-data log-likelihood

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{s})|\mathbf{z}, \boldsymbol{\theta}'].$$

Each iteration of the EM algorithm involves two steps, the expectation step and the maximization step:

E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ as a function of $\boldsymbol{\theta}$;

M-step: Find $\boldsymbol{\theta}^{(r+1)}$ such that $Q(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$.

Each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta})$ and the EM algorithm typically converges to a local or global maximum of the likelihood function.

Let $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$, $\mathbf{x} = (x_1, \ldots, x_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and $\mathbf{t} = (t_1, \ldots, t_n)^\top$. In the following we implement the EM algorithm for the structural SNI-MMEM by considering that $(\mathbf{x}, \mathbf{u}, \mathbf{t})$ are missing data, i.e, using triple augmentation. Thus, under the hierarchical representation (19)-(22), with $\nu_x^2 = \phi_x(1-\delta_x^2)$ and $\tau_x = \phi_x^{1/2}\delta_x$, it follows that the complete log-likelihood function associated with $(\mathbf{z}, \mathbf{x}, \mathbf{t}, \mathbf{u})$ is

$$\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{t}, \mathbf{u}) \propto -\frac{n}{2}\log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\sum_{i=1}^n u_i(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i)$$
$$-\frac{n}{2}\log(\nu_x^2) - \frac{1}{2\nu_x^2}\sum_{i=1}^n u_i(x_i - \mu_x - \tau_x u_i^{-1/2}t_i)^2. \tag{25}$$

Letting $\widehat{u}_i = E[u_i|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$, $\widehat{ut}_i = E[u_i^{1/2}t_i|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$, $\widehat{t_i^2} = E[t_i^2|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$, $\widehat{ux}_i = E[u_i x_i|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$, $\widehat{ux^2}_i = E[u_i x_i^2|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$ and $\widehat{utx}_i = E[u_i^{1/2}t_i x_i|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i]$, we obtain using known properties of conditional expectation and the moments of the truncated normal distribution (see Johnson et al., 1994, Section 10.1) that

$$\begin{aligned}
\widehat{ut}_i &= E[u_i^{1/2}A_i] = \widehat{u}_i\widehat{\mu}_{Ti} + \widehat{M}_T E[W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i], \qquad (26)\\
\widehat{t^2}_i &= E[B_i] = \widehat{u}_i\widehat{\mu}_{T_i}^2 + \widehat{M}_T^2 + \widehat{M}_T\widehat{\mu}_{T_i}E[u_i^{1/2}W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}, \mathbf{z}_i],\\
\widehat{ux}_i &= \widehat{r}_i\,\widehat{u}_i + \widehat{s}\,\widehat{ut}_i, \quad \widehat{ux^2}_i = \widehat{T}_x^2 + \widehat{r}_i^2\,\widehat{u}_i + 2\widehat{r}_i\widehat{s}\,\widehat{ut}_i + \widehat{s}^2\,\widehat{t^2}_i, \quad \text{and}\\
\widehat{utx}_i &= \widehat{r}_i\,\widehat{ut}_i + \widehat{s}\,\widehat{t^2}_i,
\end{aligned}$$

where $A_i = u_i^{1/2}\widehat{\mu}_{Ti} + W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})\widehat{M}_T$, $B_i = u_i\widehat{\mu}_{T_i}^2 + \widehat{M}_T^2 + u_i^{1/2}W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})\widehat{M}_T\widehat{\mu}_{T_i}$ with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $\widehat{M_T}^2 = [1 + \widehat{\tau}_x{}^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}_x{}^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}\widehat{\mathbf{b}}]^{-1}$, $\widehat{\mu}_{T_i} = \widehat{\tau}_x\widehat{M}_T^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}_x{}^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}(\mathbf{z}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu}_x)$, $\widehat{T_x}^2 = \widehat{\nu}_x{}^2[1 + \widehat{\nu}_x{}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}}]^{-1}$, $\widehat{r}_i = \widehat{\mu}_x + \widehat{T_x}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})(\mathbf{z}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu}_x)$ and $\widehat{s} = \widehat{\tau}_x(1 - \widehat{T_x}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}})$.

Since the conditional expectations given in (26) depends only on u_i, we need to known (and to generate) the conditional distribution $u_i|\mathbf{y}_i$, which for each element of this class of distributions can be easily derived from the results of the Proposition 1 jointly with the result of conditional distributions given in Section 2.1. Besides, for the contaminated skew-normal distribution all the quantities of the E-step (26) have closed form expression, with

$$E[W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})] = \frac{\nu\gamma^{p/2}W_{\Phi_1}(\frac{\gamma^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})\exp\{(1-\gamma)d/2\} + (1-\nu)W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T})}{1-\nu+\nu\gamma^{p/2}\exp\{(1-\gamma)d/2\}}$$

$$E[u_i^{1/2}W_{\Phi_1}(\frac{u_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})] = \frac{\nu\gamma^{(p+1)/2}W_{\Phi_1}(\frac{\gamma^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})\exp\{(1-\gamma)d/2\} + (1-\nu)W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_T})}{1-\nu+\nu\gamma^{p/2}\exp\{(1-\gamma)d/2\}}$$

For the skew-t and skew-slash, Monte-Carlo integration may be employed, which yield a so-called MC-EM algorithm.

Thus, we have the following EM type algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{u}_i$, $\widehat{t^2}_i$, $\widehat{ut}_i$, $\widehat{ux}_i$, $\widehat{ux^2}_i$, and $\widehat{utx}_i$ for $i = 1,\ldots,n$, using (26).

M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $E[\ell_c(\boldsymbol{\theta}|\mathbf{z},\mathbf{x},\mathbf{t},\mathbf{u})|\mathbf{z},\widehat{\boldsymbol{\theta}}]$ over $\boldsymbol{\theta}$, which leads to

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}} &= \overline{\mathbf{y}}_u - \overline{x}_u\widehat{\boldsymbol{\beta}}\\
\widehat{\boldsymbol{\beta}} &= \frac{\sum_{i=1}^n \widehat{ux}_i(\mathbf{y}_i - \overline{\mathbf{y}}_u)}{\sum_{i=1}^n \widehat{ux^2}_i - n\overline{\widehat{u}}\overline{x}_u^2},\\
\widehat{\phi}_1 &= \frac{1}{n}\sum_{i=1}^n(\widehat{u}_iX_i^2 - 2\widehat{ux}_iX_i + \widehat{ux^2}_i),\\
\widehat{\phi}_{j+1} &= \frac{1}{n}\sum_{i=1}^n(\widehat{u}_iy_{ij}^2 + \widehat{u}_i\alpha_j^2 + \beta_j^2\widehat{ux^2}_i - 2\widehat{u}_i\alpha_jy_{ij} - 2y_{ij}\beta_j\widehat{ux}_i + 2\alpha_j\beta_j\widehat{ux}_i), j = 1,\ldots,r,\\
\widehat{\mu}_x &= \overline{x}_u - \widehat{\tau}_x\overline{t}_u,\\
\widehat{\nu}_x^2 &= \frac{1}{n}\sum_{i=1}^n(\widehat{ux^2}_i - \widehat{\mu}_x\widehat{ux}_i) - \widehat{\tau}_x\frac{1}{n}\sum_{i=1}^n\widehat{utx}_i \text{ and}\\
\widehat{\tau}_x &= \frac{\sum_{i=1}^n(\widehat{utx}_i - \overline{x}_u\widehat{ut}_i)}{\sum_{i=1}^n(\widehat{t_i^2} - \overline{t}_u\widehat{ut}_i)},
\end{aligned}$$

where,

$$\overline{\mathbf{y}}_u = \frac{\sum_{i=1}^n \widehat{u}_i \mathbf{y}_i}{\sum_{i=1}^n \widehat{u}_i}, \overline{x}_u = \frac{\sum_{i=1}^n \widehat{ux}_i}{\sum_{i=1}^n \widehat{u}_i}, \overline{t}_u = \frac{\sum_{i=1}^n \widehat{ut}_i}{\sum_{i=1}^n \widehat{u}_i} \quad \text{and} \quad \overline{\widehat{u}} = \frac{1}{n}\sum_{i=1}^n \widehat{u}_i.$$

Note that when $u_i = 1$ the M-step equations reduce to the equations obtained in Lachos, Bolfarine, Vilca and Galea-Rojas (2005) under the skew-normal distribution and when $\lambda_x = 0$ (or $\tau_x = 0$) the M-step equations reduces to the equations obtained by Bolfarine and Galea-Rojas (1995). Moreover, when $U \sim Gamma(\nu/2, \nu/2)$ and $\lambda_x = 0$, the M-step reduces to equations obtained by Bolfarine and Galea-Rojas (1996). The shape and scale parameters of the latent variable x, can be estimated by noting that $\tau_x/\nu_x = \lambda_x$, and $\phi_x = \tau_x^2 + \nu_x^2$. Starting values are often chosen to be the corresponding estimates under a normal assumption, where the starting values for the asymmetric parameters are set to be 0 and as recommended in the literature, it is useful to run the EM-algorithm several times with different starting values. Following Arellano-Valle, Bolfarine and Lachos (2005) we also propose selecting the best fit by inspection of information criteria such as Akaike's Information Criterion (AIC, $-\ell(\widehat{\boldsymbol{\theta}})/N + P/N$), where P is the number of parameters in the model and $N = P \times n$.

5 The observed information matrix

From (24) and the notation in Proposition 1, we have after some algebraic manipulations that the log-likelihood function can be, alternatively, written as:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}), \tag{27}$$

where $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2} log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log(K_i)$, $i = 1, \ldots, n$. Thus, the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{L} = \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = -\frac{n}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \sum_{i=1}^n \frac{1}{K_i^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}}\frac{\partial K_i}{\partial\boldsymbol{\theta}^\top} + \sum_{i=1}^n \frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \tag{28}$$

where

$$\frac{\partial K_i}{\partial\boldsymbol{\theta}} = I_i^\phi(\frac{p+1}{2})\frac{\partial A_i}{\partial\boldsymbol{\theta}} - \frac{1}{2}I_i^\Phi(\frac{p+2}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}}$$

and

$$\begin{aligned}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \frac{1}{4}I_i^\Phi(\frac{p+4}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} - \frac{1}{2}I_i^\Phi(\frac{p+2}{2})\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \\ &\quad -\frac{1}{2}I_i^\phi(\frac{p+3}{2})(\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top}) - I_i^\phi(\frac{p+3}{2})A_i\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top} \\ &\quad +I_i^\Phi(\frac{p+1}{2})\frac{\partial^2 A_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},\end{aligned} \tag{29}$$

with the notation

$$I_i^{\Phi}(w) = \int_0^{\infty} u_i^w \exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i)dH(u_i),$$

$$I_i^{\phi}(w) = \int_0^{\infty} u_i^w \exp\{-\frac{1}{2}u_i d_i\}\phi_1(u_i^{1/2}A_i|0,1)dH(u_i).$$

Notice that we can also write $K_i = I_i^{\Phi}(\frac{p}{2})$. From Section 2.1 and Lema 2 in Gupta (2003), direct substitution of $H(u)$ in the integrals above yields immediately the following results for each distribution considered, namely

- *Skew-t.*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^w \nu^{\nu/2}\Gamma(w+\nu/2)}{\Gamma(\nu/2)} T_1\left(\frac{A_i}{(d_i+\nu)^{1/2}}\sqrt{w+\nu/2}|0,1,w+\nu/2\right) \quad \text{and} \\ I_i^{\phi}(w) &= \frac{2^{\nu}\nu^{\nu/2}}{\sqrt{2\pi}\Gamma(\nu/2)}\left(\frac{1}{d_i+A_i^2+\nu}\right)^{\frac{\nu+2w}{2}}\Gamma\left(\frac{\nu+2w}{2}\right). \end{aligned}$$

- *Skew-slash.*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^{w+\nu}\Gamma(w+\nu)}{d_i^{w+\nu}} P_1(w+\nu,\frac{d_i}{2})E[\Phi(S_i^{1/2}A_i)] \quad \text{and} \\ I_i^{\phi}(w) &= \frac{\nu 2^{w+\nu}\Gamma(w+\nu)}{\sqrt{2\pi}(d_i+A_i^2)^{w+\nu}} P_1(w+\nu,\frac{d_i+A_i^2}{2}), \end{aligned}$$

  where $S_i \sim Gamma(w+\nu, \frac{d_i}{2})\mathbb{I}_{(0,1)}$.

- *Contaminated skew–normal.*

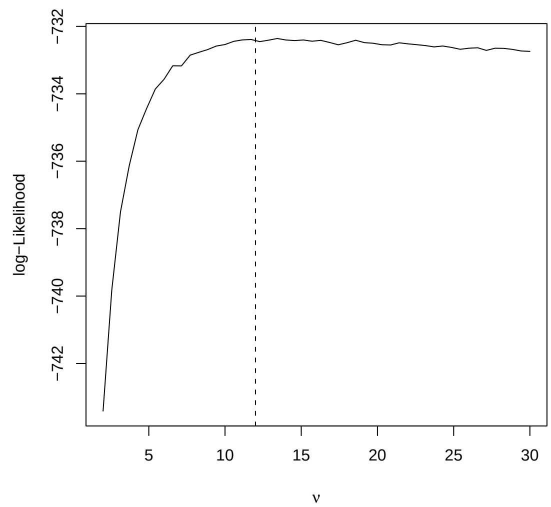
$$\begin{aligned} I_i^{\Phi}(w) &= \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1(d_i|0,\frac{1}{\gamma})\Phi(\gamma^{1/2}A_i) + (1-\nu)\phi_1(d_i|0,1)\Phi(A_i)\} \quad \text{and} \\ I_i^{\phi}(w) &= \nu\gamma^{w-1/2}\phi_1(d_i+A_i^2|0,\frac{1}{\gamma}) + (1-\nu)\phi_1(d_i+A_i^2). \end{aligned}$$

The derivatives of $\log \mathbf{\Sigma}$, d_i and A_i involved tedious but not complicated algebraic manipulations and are given in the Appendix. Asymptotic confidence intervals and test on the MLEs can be obtained using this matrix, that is, if $\mathbf{J} = -\mathbf{L}$ denotes the observed information matrix for the marginal log-likelihood $\ell(\boldsymbol{\theta})$ of the SNI-MMEM, then asymptotic confidence intervals and hypotheses tests for the parameter $\boldsymbol{\theta}$ are obtained assuming that the MLE $\boldsymbol{\theta}$ has approximately a $N_{3p+1}(\boldsymbol{\theta}, \mathbf{J}^{-1})$ distribution. In practice, $\mathbf{J}$ is usually unknown and has to be replaced by the MLE $\widehat{\mathbf{J}}$, that is, the matrix $\widehat{\mathbf{J}}$ evaluated at the MLE $\hat{\boldsymbol{\theta}}$.

6 Application

Barnett (1969) data set. In this application, the multivariate skew-normal, skew-t, skew-slash and skew-contaminated normal distributions, are applied to fit the data

Figure 1: Barnett data set. Profile Likelihood for the skew-t model.

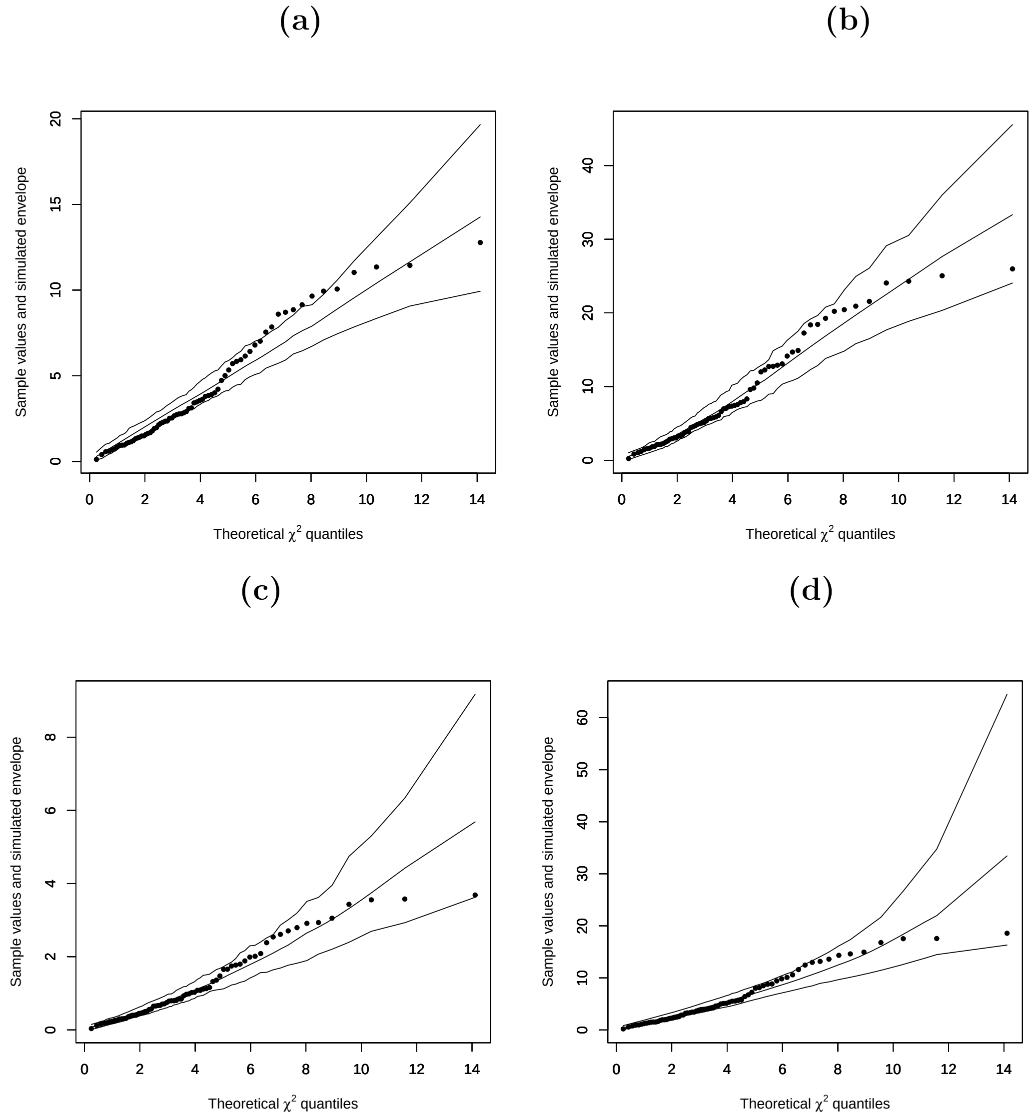


studied in Barnett (1969). We consider the measurements divided by 100 in order to achieve numerical stability. In a previous analysis of this data set (Galea-Rojas et al., 2002), a transformation was used to improve the normal fitting and Lachos, Bolfarine, Vilca e Galea-Rojas (2005) noticed left skewness and used the skew-normal distribution. We compare in the sequel skew-normal (SN), skew-t (ST), contaminated skew-normal (CSN) and skew-slash (SSL) fitting for this data set. Resulting parameter estimates for the four models are given in Table 1. As suggested by Fernandez and Steel (1999) for each model, the AIC criterion (or equivalently the log-likelihood) was used for choosing among some values of ν and γ. This strategy is illustrated in Figure 1 for the skew-t model, it is clear that, in this case, the optimum value of ν is around 12. Note that using AIC values (and log-likelihood) shown in the bottom of the Table 1 we have that they favor the SNI models. Particularly, we can see that the contaminated skew-normal fits the data better than the other three distributions. Replacing the ML estimates of $\boldsymbol{\theta}$ in the Mahalanobis distance $d_i = (\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$, we present Q-Q plots and envelopes in Figure 2 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation). It seems to us that the plots in Figure 2 provide even stronger evidence that the contaminated skew-normal distribution provides a better fit to the data set than the skew-t, skew-slash and the skew-normal distribution.

7 Final Conclusion

Paper deals with a multivariate skew–normal/independent MEM, with the skew-normal (and the normal/independent distributions) MEM as special cases. A closed form expression is obtained for the likelihood function of the observed measurements which can be maximized by using existing statistical software. An EM-type algorithm is developed by exploring statistical properties of the class considered. The

Figure 2: Barnett data set. Q-Q plots and simulated envelopes: (a) Skew-normal model (b) contaminated skew-normal model (c) skew-t model and (d) skew-slash model.



observed information matrix is derived analytically which allows direct implementation of inference on this class of models. For the Barnett (1969) data set, the skew-contaminated normal distribution seems to present a better fit. We point out that the results and methods provided in this paper is not available elsewhere in the literature and the approaches used here can be used easily extended in treating other multivariate models, for instance, we can defined of a similar form the skew-normal/independent linear mixed model which will be the subject of incoming papers.

Table 1: Results of fitting normal and skew-normal MEM to the Barnett data set. SE are the estimated asymptotic standard errors based on the observed information matrix given in Section 5.

	SN		ST		CSN		SSL	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α_1	-2.0719	1.0484	-1.9028	0.9912	-1.8924	0.9671	-1.8973	1.0061
α_2	-5.2687	1.2363	-5.1358	1.2242	-5.0278	1.2021	-5.1504	1.2278
α_3	-4.3669	1.2459	-4.2634	1.2416	-4.1993	1.2085	-4.2527	1.2465
β_1	1.0609	0.0444	1.0556	0.0434	1.0571	0.0428	1.0539	0.0435
β_2	1.1912	0.0524	1.1884	0.0537	1.1872	0.0529	1.1865	0.0533
β_3	1.1304	0.0528	1.1326	0.0543	1.1329	0.0530	1.1284	0.0539
ϕ_1	5.0399	0.9963	4.0314	0.8933	2.3271	0.4715	3.2860	0.7214
ϕ_2	1.7957	0.5801	1.3570	0.4983	0.8077	0.2639	1.1127	0.4053
ϕ_3	3.0425	0.8101	2.7341	0.7518	1.4750	0.3851	2.2039	0.5942
ϕ_4	3.9338	0.8958	3.4488	0.8600	1.8816	0.4418	2.7971	0.6786
μ_x	13.0562	1.2585	12.9702	1.3501	12.9744	0.5600	12.9651	1.3684
σ_x^2	141.8462	34.0211	125.0256	33.8647	70.1743	14.3395	99.6618	26.1853
λ_x	5.1780	2.9353	4.9176	2.9846	4.7469	1.4686	5.0885	3.2465
ν	-	-	12	-	0.67	-	3	-
γ	-	-	-	-	0.38	-	-	-
log-likelihood	-733.5978		-732.4653		-731.0912		-733.3013	
AIC	2.5924		2.5919		2.5906		2.5948	

Appendix: The observed information matrix in the multivariate skew-normal/independent measurement errors model

In this appendix the first and second derivatives $\log|\mathbf{\Sigma}|$, A_i and d_i are obtained.

• A_i

From (24), it follows that

$$\frac{\partial A_i}{\partial \boldsymbol{\gamma}} = [A_x \frac{\partial a_i}{\partial \boldsymbol{\gamma}} + a_i \frac{\partial A_x}{\partial \boldsymbol{\gamma}}],$$

$$\frac{\partial^2 A_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = [\frac{\partial A_x}{\partial \boldsymbol{\gamma}} \frac{\partial a_i}{\partial \boldsymbol{\tau}^\top} + A_x \frac{\partial^2 a_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} + \frac{\partial a_i}{\partial \boldsymbol{\gamma}} \frac{\partial A_x}{\partial \boldsymbol{\tau}^\top} + a_i \frac{\partial^2 A_x}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}]$$

with $\boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi_x, \boldsymbol{\phi}, \lambda_x$, $A_x = \lambda_x \Lambda_x / (\phi_x + \lambda_x^2 \Lambda_x)^{1/2}$, $\Lambda_x = \phi_x / c$, $a_i = \boldsymbol{X}_i^\top D^{-1}(\boldsymbol{\phi})\boldsymbol{b}$, $\boldsymbol{X}_i = \boldsymbol{Z}_i - \boldsymbol{a} - \boldsymbol{b}\mu_x$, $c = 1 + \phi_x \mathbf{b}^T D^{-1}(\boldsymbol{\phi})\mathbf{b}$, $i = 1, ..., n$. Using results in Nel (1980) related to vector derivatives it follows that,

$$\begin{aligned}
\frac{\partial A_x}{\partial \boldsymbol{\gamma}} &= 0, \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \\
\frac{\partial A_x}{\partial \boldsymbol{\beta}} &= -\frac{(2c + \lambda_x^2)}{\lambda_x^2} A_x^3 D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial A_x}{\partial \boldsymbol{\phi}} &= \frac{(2c + \lambda_x^2)}{2\lambda_x^2} A_x^3 D(\boldsymbol{b}) D^{-2}(\boldsymbol{\phi})\boldsymbol{b}, \\
\frac{\partial A_x}{\partial \phi_x} &= \frac{(2c + \lambda_x^2 - c^2)}{2\phi_x^2 \lambda_x^2} A_x^3, \\
\frac{\partial A_x}{\partial \lambda_x} &= \frac{\phi_x}{\Lambda_x^2 \lambda_x^3} A_x^3,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial a_i}{\partial \mu_x} &= -\boldsymbol{b}^\top D^{-1}(\boldsymbol{\phi})\boldsymbol{b}, \\
\frac{\partial a_i}{\partial \boldsymbol{\alpha}} &= -D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial a_i}{\partial \boldsymbol{\beta}} &= D^{-1}(\boldsymbol{\psi})\boldsymbol{W}_{2i} - \mu_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial a_i}{\partial \phi_x} &= 0, \\
\frac{\partial a_i}{\partial \boldsymbol{\phi}} &= -D(\boldsymbol{b}) D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i, \\
\frac{\partial a_i}{\partial \lambda_x} &= 0,
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -(4\frac{\phi_x}{\lambda_x^2}A_x^3 - \frac{3(2c+\lambda_x^2)^2}{\lambda_x^4}A_x^5)\boldsymbol{M}_1 - \frac{2c+\lambda_x^2}{\lambda_x^2}A_x^3 D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta} \partial \phi_x} &= -[\frac{2(c-1)}{\lambda_x^2\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{2\lambda_x^4\phi_x^2}A_x^5]D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= [2\frac{\phi_x}{\lambda_x^2}A_x^3 - \frac{3(2c+\lambda_x^2)^2}{2\lambda_x^4}A_x^5]D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\\
&\quad + \frac{2c+\lambda_x^2}{\lambda_x^2}A_x^3 \mathbb{I}_{(p)} D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta} \partial \lambda_x} &= \frac{\phi_x A_x^3}{\lambda_x^5\Lambda_x^2}(-3A_x^2(2c+\lambda_x^2) + 4\lambda_x^2\Lambda_x)D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial \phi_x \partial \phi_x} &= -\frac{\lambda_x^2+1}{\lambda_x^2\phi_x^3}A_x^3 + \frac{3(2c+\lambda_x^2-c^2)^2}{4\lambda_x^4\phi_x^4}A_x^5,\\
\frac{\partial^2 A_x}{\partial \phi_x \partial \boldsymbol{\phi}^\top} &= [\frac{(c-1)}{\lambda_x^2\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{4\lambda_x^4\phi_x^2}A_x^5]\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial \phi_x \partial \lambda_x} &= \frac{c-2}{\lambda_x^3\Lambda_x\phi_x}A_x^3 + \frac{3(2c+\lambda_x^2-c^2)}{2\lambda_x^5\Lambda_x^2\phi_x}A_x^5,\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= [-\frac{\phi_x}{\lambda_x^2}A_x^3 + \frac{3(2c+\lambda_x^2)^2}{4\lambda_x^4}A_x^5]D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\\
&\quad - \frac{2c+\lambda_x^2}{\lambda_x^2}D^2(\boldsymbol{b})D^{-3}(\boldsymbol{\phi})A_x^3,\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi} \partial \lambda_x} &= \frac{\phi_x A_x^3}{2\lambda_x^5\Lambda_x^2}[3A_x^2(2c+\lambda_x^2) - 4\lambda_x^2\Lambda_x]D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{b},\\
\frac{\partial^2 A_x}{\partial \lambda_x \partial \lambda_x} &= -\frac{3\phi_x}{\lambda_x^4\Lambda_x^2}A_x^3 + \frac{3\phi_x^2}{\lambda_x^6\Lambda_x^4}A_x^5,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 a_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} &= 0, \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \phi_x, \lambda_x \;\; \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \phi_x, \lambda_x; \quad \frac{\partial^2 a_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = 0, \;\; \boldsymbol{\gamma} = \boldsymbol{\beta}, \boldsymbol{\phi}, \;\; \boldsymbol{\tau} = \phi_x, \lambda_x,\\
\frac{\partial^2 a_i}{\partial \mu_x \partial \boldsymbol{\beta}^\top} &= -2\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi}), \quad \frac{\partial^2 a_i}{\partial \mu_x \partial \boldsymbol{\phi}^\top} = \boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}^\top} &= -D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\phi}^\top} &= \mathbb{I}_{(p)} D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -2\mu_x D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= -\mathbb{I}_{(p)} D(\boldsymbol{Z}_i - \boldsymbol{a} - 2\mu_x\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= 2D(\boldsymbol{X}_i)D(\boldsymbol{b})D^{-3}(\boldsymbol{\phi})
\end{aligned}
$$

• d_i

For $d_i = \boldsymbol{X}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i$, with $d_{i\boldsymbol{\gamma}} = \dfrac{\partial d_i}{\partial \boldsymbol{\gamma}}$ and $d_{i\boldsymbol{\gamma}\boldsymbol{\tau}^\top} = \dfrac{\partial^2 d_i}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\tau}^\top}$, it follows that

$$
\begin{aligned}
d_{i\mu_x} &= -2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i,\\
d_{i\boldsymbol{\alpha}} &= -2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i,\\
d_{i\boldsymbol{\beta}} &= -2q_i D^{-1}(\boldsymbol{\psi})\boldsymbol{W}_{2i} + 2c^{-1}\phi_x a_i q_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
d_{i\phi_x} &= -c^{-2}a_i^2,\\
d_{i\boldsymbol{\phi}} &= -D^{-2}(\boldsymbol{\phi})D(\boldsymbol{X}_i)\boldsymbol{X}_i + 2c^{-1}\phi_x a_i D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{X}_i - c^{-2}\phi_x^2 a_i^2 D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{b},\\
d_{i\lambda_x} &= 0,
\end{aligned}
$$

$$
\begin{aligned}
d_{i\mu_x\mu_x} &= 2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b},\\
d_{i\mu_x\boldsymbol{\alpha}^\top} &= 2\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
d_{i\mu_x\boldsymbol{\beta}^\top} &= -2c^{-1}\boldsymbol{A}_i,\\
d_{i\mu_x\phi_x} &= 2\frac{(c-1)}{c^2\phi_x}a_i,\\
d_{i\mu_x\boldsymbol{\phi}^\top} &= 2c^{-1}\boldsymbol{X}_i^\top \boldsymbol{\Sigma}^{-1}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}),\\
d_{i\boldsymbol{\alpha}\boldsymbol{\alpha}^\top} &= 2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
d_{i\boldsymbol{\alpha}\boldsymbol{\beta}^\top} &= 2q_i[\boldsymbol{D}^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_x \boldsymbol{M}_1] + 2c^{-1}\phi_x \boldsymbol{D}^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\boldsymbol{Y}_i - \boldsymbol{\alpha})^\top \boldsymbol{D}^{-1}(\boldsymbol{\psi}),\\
d_{i\boldsymbol{\alpha}\phi_x} &= 2c^{-2}a_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
d_{i\boldsymbol{\alpha}\boldsymbol{\phi}^\top} &= 2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}D^{-1}(\boldsymbol{\phi})[D(\boldsymbol{X}_i) - c^{-1}\phi_x a_i D(\boldsymbol{b})],
\end{aligned}
$$

$$
\begin{aligned}
d_{i\boldsymbol{\beta}\boldsymbol{\beta}^\top} &= 4\frac{\phi_x^2}{c^2}a_i[D^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi}) + D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\boldsymbol{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top D^{-1}(\boldsymbol{\psi})]\\
&\quad -2c^{-1}\phi_x \boldsymbol{D}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)(\boldsymbol{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top \boldsymbol{D}^{-1}(\boldsymbol{\psi})\\
&\quad +2\mu_x(q_i + c^{-1}\phi_x a_i)D^{-1}(\boldsymbol{\psi}) + 2\frac{\phi_x^2}{c^2}a_i^2[D^{-1}(\boldsymbol{\psi}) - 4\frac{\phi_x}{c}\boldsymbol{M}_1],\\
d_{i\boldsymbol{\beta}\phi_x} &= -2c^{-2}a_i\boldsymbol{A}_i^\top,\\
d_{i\boldsymbol{\beta}\boldsymbol{\phi}^\top} &= 2[q_i\mathbb{I}_{(p)}D(\boldsymbol{Z}_i - \boldsymbol{a} - q_i\boldsymbol{b}) + c^{-1}\phi_x\boldsymbol{A}_i^\top(\boldsymbol{Z}_i - \boldsymbol{a} - \boldsymbol{b}q_i)^\top D(\boldsymbol{b})]D^{-2}(\boldsymbol{\phi}),\\
d_{i\phi_x\phi_x} &= 2\frac{c^{-3}}{\phi_x}(c-1)a_i^2,\\
d_{i\phi_x\boldsymbol{\phi}^\top} &= (-2c^{-3}\phi_x a_i^2 D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{b} + 2c^{-2}a_i D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\boldsymbol{X}_i)^\top,\\
d_{i\boldsymbol{\phi}\boldsymbol{\phi}^\top} &= 2D^{-3}(\boldsymbol{\phi})D^2(\boldsymbol{X}_i) - 4c^{-1}\phi_x a_i D^{-3}(\boldsymbol{\phi})D(\boldsymbol{b})D(\boldsymbol{X}_i)\\
&\quad -2c^{-1}\phi_x D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{X}_i\boldsymbol{X}_i^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi})\\
&\quad +2c^{-2}\phi_x^2 D^{-2}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{X}_i\boldsymbol{X}_i^\top \boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\\
&\quad +2c^{-2}\phi_x^2 a_i^2 D^{-3}(\boldsymbol{\phi})D^2(\boldsymbol{b}) - 2c^{-3}\phi_x^3 a_i^2 D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{M}D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\\
&\quad +2c^{-2}\phi_x^2 D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b})\boldsymbol{M}\boldsymbol{X}_i\boldsymbol{X}_i^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}),
\end{aligned}
$$

• $\log|\boldsymbol{\Sigma}|$

$$\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\tau}\partial \boldsymbol{\gamma}^\top} = 0, \quad \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \lambda_x; \quad \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi_x, \boldsymbol{\phi}, \lambda_x,$$

$$\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta}\partial \phi_x} &= 2c^{-2}D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x\partial \phi_x} &= -\frac{1}{c^2\phi_x^2}(c-1)^2, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta}\partial \boldsymbol{\phi}^\top} &= -2c^{-1}\phi_x[D_1(\boldsymbol{\beta}) - c^{-1}\phi_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{b}^\top D(\boldsymbol{b})]D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= 2c^{-1}\phi_x[D^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_x \boldsymbol{M}_1], \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x\partial \boldsymbol{\phi}^\top} &= -c^{-2}\boldsymbol{b}^\top D(\boldsymbol{b})D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) - c^{-2}\phi_x^2 D(\boldsymbol{b})D^{-1}(\boldsymbol{\phi})\boldsymbol{M}D^{-1}(\boldsymbol{\phi})D(\boldsymbol{b}) + 2c^{-1}\phi_x D^2(\boldsymbol{b})D^{-3}(\boldsymbol{\phi}),
\end{aligned}$$

where $\boldsymbol{A}_i = (\boldsymbol{Y}_i - \boldsymbol{\alpha} - 2q_i\boldsymbol{\beta})^\top D^{-1}(\boldsymbol{\psi})$, $\boldsymbol{M} = D^{-1}(\boldsymbol{\phi})\boldsymbol{b}\boldsymbol{b}^\top D^{-1}(\boldsymbol{\phi})$, $\boldsymbol{M}_1 = D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi})$, $\boldsymbol{\psi} = (\phi_2, ..., \phi_p)^\top$, $q_i = \mu_x + c^{-1}\phi_x a_i$ $\boldsymbol{W}_{2i} = \boldsymbol{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x$, $\mathbb{I}_{(p)} = [\mathbf{0}, \mathbf{I}_{p-1}]$ of dimension $p - 1 \times p$.

Acknowledgements

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

References

Arellano-Valle R.B. and Bolfarine, H. (1996). Elliptical structural models. *Communications in Statistics - Theory and Methods,* 25, 2319-2341.

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

Arellano-Valle, R.B., Ozan S., Bolfarine, H. and Lachos, V.H. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis,* 98, 265-281.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* 12, 171-178.

Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distributions. *Journal Royal Statistics Society,* B, 61, 579-602.

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

Barnett, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics,* 25, 129-142.

Bolfarine, H. and Galea-Rojas, M. (1995). Maximum likelihood estimation of simultaneous pairwise linear structural ralationships. *Biometrical Journal,* 37, 673-689.

Bolfarine, H. and Galea-Rojas, M. (1996). On structural comparative calibration under a t-model. *Computational Statistics,* 11, 63-85.

Bolfarine, H., Cabral, C.R.B. and Paula, G.A. (2002). Distance tests under nonregular conditions: Applications to the comparative calibration model. *Journal of Statistics Computation and Simulation,* 72, 2, 125-140.

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

Brown, P. and Fuller, W. Eds (1990). Statistical analysis of measurement errors models and applications. *Contemporary Mathematics,* Volume 112.

Cheng, C. L. and van Ness, J. W. (1999). *Statistical regression with measurement error.* Arnold. London.

Chipkevitch, E., Nishimura, R., Tu, D. and Galea-Rojas, M. (1996). Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods. *Journal of Urology,* 156, 2050-2053.

Dempster, A.P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society,* B, 39, 1-22.

Fuller, W.A. (1987): *Measurement error models*, Wiley, New York.

Galea–Rojas, M., Bolfarine, H. and de Castro, M. (2002). Local influence in comparative calibration models. *Biometrical Journal,* 44, 59-81.

Genton, M. G., Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, **57**, 389-401.

Gupta A. K. (2003). Multivariate skew t-distributions. *Statistics,* **37(4)**, 359-363.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions,* Vol. 1. New York: Wiley.

Kendall, M. G. and Stuard, A. (1979). *The advanced theory of statistics*, 2. 4*th* ed. Hafner. New York.

Lachos, V. H., Bolfarine, H., Vilca, L. F., Galea–Rojas, M. (2005). Estimation and influence diagnostic for structural comparative calibration models under the skew-normal distribution. Technical report, RT-MAE 2005-12, IME-USP.

Lachos, V. H. and Vilca, L. F. (2007). Skew-normal/ independent distributions, with applications. Technical report, RT-IMECC 02-07, IMECC-UNICAMP.

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

Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistical Journal,* 14, 137-193.

Osorio, F.S. (2006). Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos. Ph.D. dissertation, Department of Statistics, IME-USP. São Paulo.

Sahu, S.K., Dey, D.K., and Branco, M. (2003). A new class of multivariate distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics,* 29, 129-150.

Shyr, I. and Gleser, L. (1986). Inference about comparative precision in linear structural relationships. *Journal of Statistical Planning and Inference,* 14, 339-358.

Tan, F. and Peng, H. (2006). The slash and the skew-slash Student t distributions. Submitted

Theobald, C. M. and Mallison, J. R. (1978). Comparative calibration, linear structural relationship and congeneric measurements. *Biometrics,* 34, 35-45.

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