RT-MAE 2003-10

SKEW NORMAL MEASUREMENT ERROR MODELS

by
Arellano-Valle, R.B., Ozan, S.
Bolfarine, H.
and
Lachos, V.H.

Palavras-Chave: Invariance, maximum likelihood, posterior distribution, prior distribution, structural model.
Classificação AMS: 62F05,62J12
(AMS Classification)

- Setembro de 2003 -

SKEW NORMAL MEASUREMENT ERROR MODELS

R. B. Arellano-Valle*, S. Ozan, H. Bolfarine**, V. H. Lachos****

(*) Pontificia Universidad Catolica de Chile
Casilla 306, Correo 22
Santiago-Chile

(**) Universidade de São Paulo
Caixa Postal 66281 - CEP 05315 970
São Paulo - S.P. - Brasil

Summary

In this paper we define a class of skew normal measurement error models, extending usual symmetric normal models in order to avoid data transformation. The likelihood function of the observed data is obtained, which can be maximized by using existing statistical software. Inference on the parameters of interest can be approached by using the observed information matrix, which can also be computed by using existing statistical software, such as the Ox program. Bayesian inference is also discussed for the family of asymmetric models in terms of invariance with respect to the symmetric normal distribution showing that early results obtained for the normal distribution also holds for the asymmetric family. Results of a simulation study and an analysis of a real data set analysis are provided.

Key Words: Invariance, maximum likelihood, posterior distribution, prior distribution, structural model

1. Introduction

Recent statistical literature has seen an increasing interest for models incorporating asymmetry. Advantages of using such general structures include easiness of interpretation, as well as estimation efficiency. Main references on the subject include Azzalini (1985), Azallini e Dalla Valle (1996), Branco and Dey (2001), Genton and Leporfido (2001), Arellano-Valle et al. (2002) and Arellano-Valle and Genton (2003). In this paper, we consider a linear regression model relating the variables y_i and x_i, that is,

$$y_i = \alpha + \beta x_i + e_i, \tag{1.1}$$

but we consider that x_i is not directly observed so that we observe instead

$$X_i = x_i + u_i, \tag{1.2}$$

$i = 1, \ldots, n$. The measurement error model (1.2) specifies that x_i is unavailable and the observed X_i can be seen as an unbiased estimate of the unknown x_i, $i = 1, \ldots, n$. The structure (1.2) is called additive since it is defined in terms of additive measurement errors, specified by u_i, $i = 1, \ldots, n$. The unobserved x_i can be seen as fixed parameters in which case a functional model results or a random variable, in which case a structural model results. In the functional case the x_i is an incidental parameter and it is typical to consider that

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \overset{iid}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right), \tag{1.3}$$

$i = 1, \ldots, n$, with *iid* meaning *independent and identically distributed.* In the structural case it is typically assumed that

$$x_i \overset{iid}{\sim} N_1(\mu_x, \sigma_x^2), \tag{1.4}$$

and independent of (e_i, u_i), $i = 1, \ldots, n$. It is well known that the above structural normal model is not identifiable so that additional assumptions have to be made. The most commonly adopted assumptions are that the variance σ_u^2 is known, the ratio of variances σ_e^2/σ_u^2 is known or the reliability ratio (Fuller, 1987) $k_x = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$ is known. In some situations, replications of the experiment can be used to determine such values. Maximum likelihood estimation for the structural and functional normal measurement error models is treated in detail in Fuller (1987). Some extensions for elliptical measurement error models are considered by Arellano-Valle and Bolfarine (1996) and Arellano-Valle et al., (1996). In this paper, we extend the above normal measurement error model by considering that e_i, u_i and x_i follow a skew normal distribution, which contains the normal distribution as special case. As considered in Azzalini (1985), a random variable Z follows a univariate skew normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ if the density function of Z is given by

$$f_Z(z) = \frac{2}{\sigma}\phi_1\left(\frac{z-\mu}{\sigma}\right)\Phi_1\left(\lambda\frac{z-\mu}{\sigma}\right), \tag{1.5}$$

where $\phi_1(.)$ and $\Phi_1(.)$, denote the density function and distribution function, respectively, of the standard univariate normal distribution. Note that if $\lambda = 0$ then the density of Z in (1.5) reduces to the density of the standard normal distribution. We use the notation $Z \sim SN_1(\mu, \sigma^2, \lambda)$ to denote this distribution, which will be reduced to $Z \sim SN(\lambda)$ when is assumed that $\mu = 0$ and $\sigma^2 = 1$.

Some properties of this distribution includes:

$$E[Z] = \mu + \sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}\sigma; \qquad Var[Z] = (1 - \frac{2\lambda^2}{\pi(1+\lambda^2)})\sigma^2;$$

with asymmetry and kurtosis indexes given by

$$\gamma = \frac{1}{2}(4-\pi)(\frac{E^2[X]}{Var[X]})^{3/2}; \qquad \kappa = 2(\pi - 3)(\frac{E^2[X]}{Var[X]})^2,$$

where $X = (Z-\mu)/\sigma$, implying that $-0.9953 < \gamma < 0,9953$ and $3.0000 < \kappa < 3.8692$. It follows that the even moments of $X = (Z-\mu)/\sigma$, coincide with the standard normal ones and that the odd moments are given by

$$E[X^{2k+1}] = \sqrt{\frac{2}{\pi}}\lambda(1+\lambda^2)^{-(k+\frac{1}{2})}2^{-k}[(2k+1)!]\sum_{j=0}^{k}\frac{j!(2\lambda)^{2j}}{(2j+1)!(k-j)!)}.$$

All these properties are easily obtained by using that (Henze, 1986; Azzalini, 1986) if $X \sim SN(\lambda)$, then

$$X \stackrel{d}{=} \frac{\lambda}{\sqrt{1+\lambda^2}}|X_0| + \frac{1}{\sqrt{1+\lambda^2}}X_1, \tag{1.6}$$

where X_0 and X_1 are iid $N(0,1)$ random variables.

Multivariate skew normal distributions are considered in Azzalini and Dalla Valle (1996), Branco and Dey (2001), among others. Genton et al. (2001) derive the moments of a random vector with multivariate skew normal distribution and their quadratic forms. Arellano-Valle et al. (2002) show that many of the properties of the multivariate skew normal distribution hold for a general class of skew distributions obtained from a symmetric class, defined in terms of independence conditions on signs and absolute values. From these result, Arellano-Valle and Genton (2003) introduce the class of fundamental skew distributions, giving an unified approach to obtain multivariate skew distributions starting from symmetric ones.

In this paper, we consider the skew normal measurement error model that follows by replacing the normality assumption in (1.3)-(1.4) by the assumption that the error terms e_i and u_i and the latent variable x_i have skew normal distributions. We obtain the likelihood function of the observed data $(y_1, X_1), \ldots, (y_n, X_n)$ by integrating out the variable x_i, $i = 1, \ldots, n$. The likelihood function can be maximized for obtaining maximum likelihood estimators and the matrix of second derivatives evaluated at the maximum likelihood estimators (observed information matrix, i.e., the Hessian matrix) can be used to obtain estimated standard errors of the estimates. It is interesting to note that it is not necessary to make any additional assumption to make the model identifiable as is the case with the the ordinary normal model. The special case where $\lambda_u = \lambda_e = 0$ is also studied in detail. One important and interesting characteristic of the model entertained is that under a skew normal measurement error model there is no need to make the additional assumptions mentioned above to make the estimation problem feasible.

The paper is organized as follows. Section 2 is dedicated to the derivation of the likelihood function of the observed data by integrating out the unobserved variable x. Section 3 presents a Bayesian analysis based on a special class of prior distributions yielding invariance in terms of the fact that the posterior distribution is in the class of the ordinary normal distribution, a problem studied earlier in Lindley and El-Sayad (1968) for the special case that the ratio σ_e^2/σ_u^2 is known and in Bolfarine and Cordani (1993) for the case where the reliability ratio k_x is known. In Section 4 an EM-type algorithm is developed which can overcome some difficulties detected by using direct maximization of the likelihood. Although the M-step requires a numerical maximization in each iteration, it is easily implemented. A simulation study is considered in Section 5. Finally, in Section 6, we consider an application to a data set previously analyzed in the literature using a model not incorporating measurement errors.

2. The likelihood function

The skew normal regression model with measurement errors is defined by extending the normal model defined by (1.1)-(1.3) by considering that

$$e_i \stackrel{iid}{\sim} SN_1(0, \sigma_e^2, \lambda_e), \tag{2.1}$$

(2.2) $$u_i \overset{iid}{\sim} SN_1(0, \sigma_u^2, \lambda_u),$$

$i = 1, \ldots, n$, leading, under the regression set up defined by (1.1)-(1.2), to the following functional skew normal model:

(2.3) $$y_i|x_i \overset{ind}{\sim} SN_1(\alpha + \beta x_i, \sigma_e^2, \lambda_e),$$

(2.4) $$X_i|x_i \overset{ind}{\sim} SN_1(x_i, \sigma_u^2, \lambda_u),$$

$i = 1, \ldots, n$, with *ind* meaning independent. To obtain the structural skew-normal model, we consider as an extension of (1.4) that

(2.5) $$x_i \overset{iid}{\sim} SN_1(\mu_x, \sigma_x^2, \lambda_x),$$

$i = 1, \ldots, n$. As in (1.4), λ_e, λ_u and λ_x are asymmetry parameters. It also follows that the above model is nondifferential as is considered in Bolfarine and Arellano-Valle (1998). If $\lambda_e = \lambda_u = \lambda_x = 0$, then the asymmetric model reduces to the symmetric normal model considered in (1.1)-(1.4). In the sequel we drop the subscript i in a sample unit to simplify notation. From (2.3)-(2.4), it follows that the conditional density of (y, X) given x can be written as

$$f(y, X|x) = \frac{2^2}{\sigma_e \sigma_u} \phi_1\left(\frac{y - \alpha - \beta x}{\sigma_e}\right) \phi_1\left(\frac{X - x}{\sigma_u}\right) \Phi_1\left(\lambda_e \frac{y - \alpha - \beta x}{\sigma_e}\right) \Phi_1\left(\lambda_u \frac{X - x}{\lambda_u}\right).$$

Hence, considering (2.5), the joint marginal density of (y, X) is obtained by integrating out x with respect to the above density, that is,

(2.6) $$f(y, X) = \int_{-\infty}^{\infty} f(y, X|x) f(x) dx$$

$$= \int_{-\infty}^{\infty} \frac{2^3}{\sigma_e \sigma_u \sigma_x} \phi_1\left(\frac{y - \alpha - \beta x}{\sigma_e}\right) \phi_1\left(\frac{X - x}{\sigma_u}\right) \phi\left(\frac{x - \mu_x}{\sigma_x}\right)$$
$$\Phi_1\left(\lambda_e \frac{y - \alpha - \beta x}{\sigma_e}\right) \Phi_1\left(\lambda_u \frac{X - x}{\lambda_u}\right) \Phi_1\left(\lambda_x \frac{x - \mu_x}{\sigma_x}\right) dx.$$

Making the transformation $v = x - \mu_x$, with jacobian $dx = dv$, and defining

$$\bar{e} = y - (\alpha + \beta \mu_x), \qquad \bar{u} = X - \mu_x,$$

we can write the integral (2.6) as

$$(2.7)\qquad f(y,X)=\frac{2^3}{\sigma_e\sigma_u\sigma_x}\int_{-\infty}^{\infty}\phi_1(\frac{\bar{e}-\beta v}{\sigma_e})\phi_1(\frac{\bar{u}-v}{\sigma_u})\phi_1(\frac{v}{\sigma_x})$$

$$\Phi_1(\lambda_e\frac{\bar{e}-\beta v}{\sigma_e})\Phi_1(\lambda_u\frac{\bar{u}-v}{\sigma_u})\Phi_1(\lambda_x\frac{v}{\sigma_x})dv.$$

Let

$$\mathbf{w}=(\bar{e},\bar{u},0)^T,\qquad \mathbf{b}=(\beta,1,-1)^T,\qquad \mathbf{b}_1=(\beta,1)^T$$

$$\mathbf{\Psi}=\mathrm{diag}(\sigma_e^2,\sigma_u^2,\sigma_x^2),\quad \mathbf{\Psi}_1=\mathrm{diag}(\sigma_e^2,\sigma_u^2)\quad\text{and}\quad \mathbf{\Lambda}=\mathrm{diag}(\lambda_e,\lambda_u,\lambda_x).$$

Let also $\phi_k(\cdot|\boldsymbol{\mu},\mathbf{\Sigma})$ and $\Phi_k(\cdot|\boldsymbol{\mu},\mathbf{\Sigma})$ be the density and the distribution function, respectively, of the k-dimensional normal distribution $N_k(\boldsymbol{\mu},\mathbf{\Sigma})$. Considering this notation, it follows that

$$(2.8)\qquad \phi_1(\frac{\bar{e}-\beta v}{\sigma_e})\phi_1(\frac{\bar{u}-v}{\sigma_u})\phi_1(\frac{v}{\sigma_x})=\phi_3(\mathbf{w}|\mathbf{b}v,\mathbf{\Psi})$$

and

$$(2.9)\qquad \Phi_1(\lambda_e\frac{\bar{e}-\beta v}{\sigma_e})\Phi_1(\lambda_u\frac{\bar{u}-v}{\sigma_u})\Phi_1(\lambda_x\frac{v}{\sigma_x})=\Phi_3(\mathbf{\Lambda w}-\mathbf{\Lambda b}v|\mathbf{0},\mathbf{\Psi}).$$

Further, we consider the following lemmas.

Lemma 1.

$$\phi_3(\mathbf{w}|\mathbf{b}v,\mathbf{\Psi})=\phi_2(\mathbf{z}|\boldsymbol{\mu},\mathbf{\Sigma})\phi_1\left(v|\frac{\mathbf{b}^T\mathbf{\Psi}^{-1}\mathbf{w}}{\mathbf{b}^T\mathbf{\Psi}^{-1}\mathbf{b}},\frac{1}{\mathbf{b}^T\mathbf{\Psi}^{-1}\mathbf{b}}\right),$$

where

$$\boldsymbol{\mu}=\begin{pmatrix}\alpha+\beta\mu_x\\ \mu_x\end{pmatrix},\qquad \mathbf{\Sigma}=\mathbf{\Psi}_1+\sigma_x^2\mathbf{b}_1\mathbf{b}_1^T=\begin{pmatrix}\beta^2\sigma_x^2+\sigma_e^2 & \beta\sigma_x^2\\ \beta\sigma_x^2 & \sigma_x^2+\sigma_u^2\end{pmatrix}.$$

Proof. In fact,

$$\phi_3(\mathbf{w}|\mathbf{b}v,\mathbf{\Psi})=|\mathbf{\Psi}|^{-1/2}(2\pi)^{-3/2}e^{-(1/2)Q(\mathbf{w},v)},$$

where

$$Q(\mathbf{w},v)=(\mathbf{w}-\mathbf{b}v)^T\mathbf{\Psi}^{-1}(\mathbf{w}-\mathbf{b}v)=Q_1(\mathbf{w}|v)+Q_2(\mathbf{w}),$$

with

$$Q_1(\mathbf{w}|v) = (\mathbf{b}^T\Psi^{-1}\mathbf{b})\left(v - \frac{\mathbf{b}^T\Psi^{-1}\mathbf{w}}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}\right)^2$$

and

$$Q_2(\mathbf{w}) = \mathbf{w}^T\Psi^{-1}\mathbf{w} - \frac{\mathbf{w}^T\Psi^{-1}\mathbf{b}\mathbf{b}^T\Psi^{-1}\mathbf{w}}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}$$

$$= (\mathbf{z}-\boldsymbol{\mu})^T\Psi_1^{-1}(\mathbf{z}-\boldsymbol{\mu}) - \frac{(\mathbf{z}-\boldsymbol{\mu})^T\Psi_1^{-1}\mathbf{b}_1\mathbf{b}_1^T\Psi_1^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}$$

$$= (\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu}),$$

where we use that $\Sigma^{-1} = \Psi_1^{-1} - \frac{1}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}\Psi_1^{-1}\mathbf{b}_1\mathbf{b}_1^T\Psi_1^{-1}$. Thus, the proof follows by noting that $|\Psi| = \frac{|\Sigma|}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}$.

Lemma 2. *Let* $\mathbf{V} \sim N_k(\boldsymbol{\eta}, \boldsymbol{\Omega})$. *Then*

$$E[\Phi_m(\mathbf{a}+\mathbf{A}\mathbf{V}|\boldsymbol{\gamma}, \boldsymbol{\Gamma})] = \Phi_m(\mathbf{a}|\boldsymbol{\gamma} - \mathbf{A}\boldsymbol{\eta}, \boldsymbol{\Gamma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T).$$

Proof. Notice that

$$E[\Phi_m(\mathbf{a}+\mathbf{A}\mathbf{V}|\boldsymbol{\gamma}, \boldsymbol{\Gamma})] = E[\Phi_m(\mathbf{a}|\boldsymbol{\gamma}-\mathbf{A}\mathbf{V}, \boldsymbol{\Gamma})] = E[P(\mathbf{U} \le \mathbf{a}|\mathbf{V})] = P(\mathbf{U} \le \mathbf{a}),$$

where $\mathbf{U}|\mathbf{V} = \mathbf{v} \sim N_m(\boldsymbol{\gamma} - \mathbf{A}\mathbf{v}, \boldsymbol{\Gamma})$, so that $\mathbf{U} \sim N_m(\boldsymbol{\gamma} - \mathbf{A}\boldsymbol{\eta}, \boldsymbol{\Gamma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T)$, thus concluding the proof.

We present in the sequel the marginal distribution of the observed vector $\mathbf{Z} = (y, X)^T$, which is the main result of the paper.

Theorem 1. *Under the skew-normal measurement error model defined by (2.3)-(2.5), the marginal density of* $\mathbf{Z} = (y, X)^T$ *is given by*

(2.10)

$$f_Z(\mathbf{z}|\theta, \lambda) = 2^3\phi_2(\mathbf{z}|\boldsymbol{\mu}, \Sigma)\Phi_3\left(\Lambda\Psi\mathbf{B}^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})|0, \Psi + \frac{1}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}\Lambda\mathbf{b}\mathbf{b}^T\Lambda\right),$$

where $\boldsymbol{\lambda} = (\lambda_e, \lambda_u, \lambda_x)^T$, $\boldsymbol{\theta} = (\mu_x, \alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2)^T$ *and*

$$\mathbf{B} = (\mathbf{I}_2, \mathbf{b}_1) = \begin{pmatrix} 1 & 0 & \beta \\ 0 & 1 & 1 \end{pmatrix}.$$

Proof. Replacing (2.8) and (2.9) in (2.7), we have that

$$f_Z(\mathbf{z}|\boldsymbol{\theta},\boldsymbol{\lambda}) = 2^3|\boldsymbol{\Sigma}|^{-1/2}\phi_2(\mathbf{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$$\times \int_{-\infty}^{\infty} \phi_1\left(v|\frac{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{w}}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}, \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\right)\Phi_3(\boldsymbol{\Lambda}\mathbf{w} - \boldsymbol{\Lambda}\mathbf{b}v|\mathbf{0},\boldsymbol{\Psi})dv$$

$$= 2^3|\boldsymbol{\Sigma}|^{-1/2}\phi_2(\mathbf{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})E[\Phi_3(\boldsymbol{\Lambda}\mathbf{w} - \boldsymbol{\Lambda}\mathbf{b}V|\mathbf{0},\boldsymbol{\Psi})],$$

where

$$V \sim N\left(\frac{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{w}}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}, \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\right).$$

Hence, from Lemma 2, with $k = 1$, $m = 3$, $\mathbf{a} = -\boldsymbol{\Lambda}\mathbf{b}$, $\eta = \frac{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{w}}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}$, $\boldsymbol{\Omega} = \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}$, $\gamma = \mathbf{0}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Psi}$, it follows that

$$E[\Phi_3(\boldsymbol{\Lambda}\mathbf{w} - \boldsymbol{\Lambda}\mathbf{b}V|\mathbf{0},\boldsymbol{\Psi})] = \Phi_3\left(\boldsymbol{\Lambda}\mathbf{w}|\boldsymbol{\Lambda}\mathbf{b}\frac{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{w}}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}, \boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda}\right)$$

$$= \Phi_3\left(\boldsymbol{\Lambda}\mathbf{M}\mathbf{w}|\mathbf{0}, \boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda}\right),$$

where $\mathbf{M} = \mathbf{I}_3 - \mathbf{P}$, with $\mathbf{P} = \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\mathbf{b}\mathbf{b}^T\boldsymbol{\Psi}^{-1}$. Note that, $\mathbf{P}\mathbf{b} = \mathbf{b}$, so that $\mathbf{M}\mathbf{b} = \mathbf{0}$. Furthermore, $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Psi}\mathbf{B}^T$, so that $\mathbf{B}\boldsymbol{\Psi}^{1/2}$ is a factorization of $\boldsymbol{\Sigma}$. Thus, after some algebraic manipulations, we can show that $\mathbf{M}\mathbf{w} = \boldsymbol{\Psi}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})$, concluding thus the proof.

We call attention to the fact that the density function in (2.10) is not in the class of the multivariate skew normal densities specified in Azzalini and Dalla Valle (1996), since in (2.10) the skewing function is $\Phi_3(.)$, which is of dimension 3. In their family, the skewing function is of dimension 1. However, it is in some more general family of densities, as considered in Arellano-Valle and Genton (2003).

From Theorem 1, we have that the likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{z}_1 = (y_1, X_1), \ldots, \mathbf{z}_n = (y_n, X_n)$ is given by

$$L(\boldsymbol{\theta},\boldsymbol{\lambda}|\mathbf{z}_1,\ldots,\mathbf{z}_n) = \prod_{i=1}^{n} f_Z(\mathbf{z}_i|\boldsymbol{\theta},\boldsymbol{\lambda}),$$

where $f_Z(z_i|\theta, \lambda)$ is the marginal density in (2.10) for the i-th sample unit, $i = 1, \ldots, n$. Denoting the log-likelihood function by $\ell(\theta, \lambda)$, it can be written as

$$\ell(\theta, \lambda) \propto -\frac{n}{2} log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$$

$$+\sum_{i=n}^{n} log[\Phi_3(\Lambda_*\Sigma^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu}))|0, \Omega)],$$

where

$$\Lambda_* = \Lambda\Psi\mathbf{B}^T\Sigma^{-1/2} \qquad \text{and} \qquad \Omega = \Psi + \frac{1}{\mathbf{b}^T\Psi^{-1}\mathbf{b}}\Lambda\mathbf{b}\mathbf{b}^T\Lambda,$$

so that the maximum likelihood estimators are solutions of the equations

$$\frac{\partial\ell(\theta, \lambda)}{\partial\theta} = -\frac{n}{2|\Sigma|}\frac{\partial|\Sigma|}{\partial\theta} - \frac{1}{2}\frac{\partial}{\partial\theta}\sum_{i=1}^{n}(\mathbf{z}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$$

$$+\sum_{i=1}^{n}\frac{1}{\Phi_3(\Lambda_*\Sigma^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu})|0, \Omega)}\frac{\partial}{\partial\theta}\Phi_3(\Lambda_*\Sigma^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu})|0, \Omega) = 0$$

and

$$\frac{\partial\ell(\theta, \lambda)}{\partial\lambda} = \sum_{i=n}^{n}\frac{1}{\Phi_3(\Lambda_*\Sigma^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu})|0, \Omega)}\frac{\partial}{\partial\lambda}\Phi_3(\Lambda_*\Sigma^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu})|0, \Omega) = 0.$$

We call attention to the fact that no explicit solution is available for the maximization problem and that the likelihood function has to be maximized numerically.

Some special cases may be of interest. If it is the case of comparing two measuring devices, then, since e_i and u_i are results of measurements, they may be normally distributed and x_i being true concentration level of a substance, may follow a skew distribution. Then, in such situations, for example, $\lambda_e = \lambda_u = 0$ and $\lambda_x \neq 0$, which is a special case of the above general situation and it implies that the marginal distribution of $\mathbf{Z} = (y, X)^T$ is asymmetric also. This situation is treated next.

Corollary 1. *Under the conditions of Theorem 1, with* $\lambda_e = \lambda_u = 0$*, the density function of* $\mathbf{Z} = (y, X)^T$ *is given by*

$$f_Z(\mathbf{z}|\boldsymbol{\theta}, \lambda_x) = 2\phi_2(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1\left(\gamma_x \frac{\sigma_x \mathbf{b}_1^T \boldsymbol{\Psi}_1^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\sqrt{1+\sigma_x^2 \mathbf{b}_1^T \boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}}\right), \tag{2.11}$$

where $\mathbf{b}_1 = (\beta, 1)^T$, $\boldsymbol{\Psi}_1 = \mathrm{diag}(\sigma_e^2, \sigma_u^2)$ *and*

$$\gamma_x = \frac{\lambda_x}{\sqrt{1+\lambda_x^2+\sigma_x^2 \mathbf{b}_1^T \boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}}.$$

Proof: By assumption, $\boldsymbol{\Psi} = \mathrm{diag}(\sigma_e^2, \sigma_u^2, \sigma_x^2)$ and $\boldsymbol{\Lambda} = \mathrm{diag}(0, 0, \lambda_x) = \lambda_x \mathbf{e}_3 \mathbf{e}_3^T$, where $\mathbf{e}_3 = (0,0,1)^T$. Thus,

$$\boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda} = \mathrm{diag}\left(\sigma_e^2, \sigma_u^2, \frac{\sigma_x^2 + \frac{\lambda_x^2}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\right),$$

where $\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b} = (1+\sigma_x^2 \mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1)/\sigma_x^2$, so that for the skewing factor considered by (2.10) we have that

$$\Phi_3\left(\boldsymbol{\Lambda}\boldsymbol{\Psi}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})|\mathbf{0}, \boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda}\right) =$$

$$\Phi_3\left(\left(\boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda}\right)^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Psi}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})|\mathbf{0}, \mathbf{I}_3\right),$$

where $\mathbf{B} = (\mathbf{I}_2, \mathbf{b}_1)$. Using now that $\mathbf{B}\mathbf{e}_3 = \mathbf{b}_1$, $\boldsymbol{\Psi}\mathbf{e}_3 = \sigma_x^2\mathbf{e}_3$ and

$$\mathbf{b}_1^T\boldsymbol{\Sigma}^{-1} = \mathbf{b}_1^T\left(\boldsymbol{\Psi}_1^{-1} - \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\right) = \frac{1}{1+\sigma_x^2\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1},$$

we have after some simple algebraic manipulations that

$$\boldsymbol{\Lambda}\boldsymbol{\Psi}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) = \frac{\lambda_x\sigma_x^2\,\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}(\mathbf{z}-\boldsymbol{\mu})}{1+\sigma_x^2\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}\,\mathbf{e}_3,$$

so that

$$\left(\boldsymbol{\Psi} + \frac{1}{\mathbf{b}^T\boldsymbol{\Psi}^{-1}\mathbf{b}}\boldsymbol{\Lambda}\mathbf{b}\mathbf{b}^T\boldsymbol{\Lambda}\right)^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Psi}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu}) = \gamma_x\frac{\sigma_x\,\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\sqrt{1+\sigma_x^2\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}}\,\mathbf{e}_3,$$

which concludes the proof, since $\Phi_3(a\mathbf{e}_3|\mathbf{0},\mathbf{I}_3) = \Phi_1(a)/4$.

Considering the joint density (2.11), the log-likelihood function corresponding to an observed sample $\mathbf{z}_1, \ldots, \mathbf{z}_n$ can be written as

$$(2.12) \qquad \ell(\boldsymbol{\theta}, \lambda_x) \propto -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$$

$$+\log\Phi_1\left(\gamma_x\frac{\sigma_x\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\sqrt{1+\sigma_x^2\mathbf{b}_1^T\boldsymbol{\Psi}_1^{-1}\mathbf{b}_1}}\right),$$

$i = 1, \ldots, n$. Notice that the asymmetric part gets nicely separated from the symmetric one and the likelihood function is an extension of the symmetric likelihood considered in Fuller (1987), with the assumption that one of the variances (σ_e^2 or σ_u^2) or the ratio of variances σ_e^2/σ_u^2 is known, since no assumption was made to derive the likelihoods. Although simpler the likelihood (2.12) must also be maximized numerically. We developed routines in Matlab and Ox to do this maximization. The asymptotic covariance matrix of the maximum likelihood estimators can be estimated by using the Hessian matrix, which can also be computed numerically by using the program Ox.

3. An invariance result for Bayesian inference

In this section, we consider an extension of the model considered in the previous section obtained by replacing assumptions (2.1)-(2.5) by the following assumptions:

$$(3.1) \qquad e_i \overset{ind}{\sim} SN_1(0, \sigma_e^2, \lambda_{ei}),$$

$$(3.2) \qquad u_i \overset{ind}{\sim} SN_1(0, \sigma_u^2, \lambda_{ui}),$$

$$(3.3) \qquad x_i \overset{ind}{\sim} SN_1(\mu_x, \sigma_x^2, \lambda_{xi}),$$

$i = 1, \ldots, n$, all independent. Hence, we are considering different skewing parameters for each observation. Let $\boldsymbol{\lambda}_i = (\lambda_{ei}, \lambda_{ui}, \lambda_{xi})^T$, $i = 1, \ldots, n$. Now, we consider Bayesian approach with the following prior specifications:

(i) $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n$ are independent.

(ii) The skewness parameters $\boldsymbol{\lambda}_i$, $i = 1, \ldots, n$, are i.i.d. following an asymmetric normal prior distribution.

Specifically, we consider the following prior distribution:

$$p(\boldsymbol{\theta}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n) = p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(\boldsymbol{\lambda}_i) = p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(\lambda_{ei}) p(\lambda_{ui}) p(\lambda_{xi}), \tag{3.4}$$

with

$$\lambda_{ei} \overset{iid}{\sim} SN_1(0, 1, w_e), \quad \lambda_{ui} \overset{iid}{\sim} SN_1(0, 1, w_u), \quad \lambda_{xi} \overset{iid}{\sim} SN_1(0, 1, w_x), \tag{3.5}$$

whose respective densities given by

$$p(\lambda_{ei}) = 2\phi_1(\lambda_{ei})\Phi_1(w_e\lambda_{ei}),$$
$$p(\lambda_{ui}) = 2\phi_1(\lambda_{ui})\Phi_1(w_u\lambda_{ui}),$$
$$p(\lambda_{ui}) = 2\phi_1(\lambda_{ei})\Phi_1(w_x\lambda_{xi}),$$

$i = 1, \ldots, n$. Notice that the above prior specification is considering that the λ_{ei} (as well as λ_{ui} and λ_{xi}) are exchangeable, that is, they vary with the observations but are generated by the same skew normal distribution.

Theorem 2. *Lets consider the measurement error model (1.1)-(1.2) with the assumptions (3.1)-(3.3). Then, under the the prior specification (3.4)-(3.5), it follows that the marginal posterior density for $\boldsymbol{\theta}$ is given by*

$$p(\boldsymbol{\theta}|\mathbf{z}_1, \ldots, \mathbf{z}_n) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} \phi_2(\mathbf{z}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Proof: In fact,

$$p(\boldsymbol{\theta}|\mathbf{z}_1, \ldots, \mathbf{z}_n) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} \int_{\Re^3} f_Z(\mathbf{z}_i|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) p(\boldsymbol{\lambda}_i) d\boldsymbol{\lambda}_i$$

$$= p(\boldsymbol{\theta}) \prod_{i=1}^{n} \int_{\Re^4} \frac{2^3}{\sigma_e\sigma_u\sigma_x} \phi_1(\frac{y_i - \alpha - \beta x_i}{\sigma_e}) \phi_1(\frac{X_i - x_i}{\sigma_u}) \phi(\frac{x_i - \mu_x}{\sigma_x})$$

$$\Phi_1(\lambda_{ei}\frac{y_i - \alpha - \beta x_i}{\sigma_e}) \Phi_1(\lambda_{ui}\frac{X_i - x_i}{\sigma_u}) \Phi_1(\lambda_{xi}\frac{x_i - \mu_x}{\sigma_x}) p(\boldsymbol{\lambda}_i) d\boldsymbol{\lambda}_i dx_i,$$

where the last equality follows from (2.6). Particularly, since

$$\Phi_1(\lambda_{ei}a_i)\Phi_1(w_e\lambda_{ei}) = \Phi_2(\lambda_e(a_i, w_e)^T), \quad \text{with} \quad a_i = \frac{y_i - \alpha - \beta x_i}{\sigma_e},$$

$i = 1, \ldots, n$, the integral for λ_{ei} can be written as

$$\int_{\Re} \Phi_1(\lambda_{ei}a_i)\phi_1(\lambda_{ei})\Phi_1(w_e\lambda_{ei})^T d\lambda_{ei} = \int_{\Re} \phi_1(\lambda_{ei})\Phi_2(\lambda_{ei}(a_i, w_{ei})^T)d\lambda_{ei}.$$

Hence, from Lemma 2, it follows that

$$\int_{\Re} \phi(\lambda_{ei})\Phi_2(\lambda_{ei}(a_i, w_{ei})^T)d\lambda_{ei} = \frac{1}{2},$$

$i = 1, \ldots, n$. Similar results hold with respect to λ_u and λ_x. Then, the marginal posterior density of $\boldsymbol{\theta}$ for the whole sample can be written as

$$p(\boldsymbol{\theta}|\mathbf{z}_1, \ldots, \mathbf{z}_n) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} \int_{\Re} \phi_1(\frac{y_i - \alpha - \beta x_i}{\sigma_e})\phi_1(\frac{X_i - x_i}{\sigma_u})\phi_1(\frac{x_i - \mu_x}{\sigma_x})dx_i,$$

which concludes the proof.

From Theorem 1 it follows that, under the more general skew normal model considered there, the Bayesian inference on the structural parameter θ reduces to considering the usual structural normal model, which in the special case where the ratio $k_e = \sigma_e^2/\sigma_u^2$ is known, leads to the ordinary Bayesian inference for the symmetric normal distribution considered in Lindley and El Sayad (1969), for example. The case $k_x = \sigma_x^2/\sigma_u^2$ known is treated in Bolfarine and Cordani (1993). Notice that there is no loss of generality in considering mean zero and variance 1, for the prior of λ_{ei}, λ_{ui} and λ_{xi}, as long as they are considered known. We conjecture that the above invariance result holds also in more general situations, such as when dependent skew normal priors are considered for λ_{ei}, λ_{ui} and λ_{xi}, $i = 1, \ldots, n$.

4. An EM-type algorithm

A direct maximization of the likelihood (2.12) may sometimes poses problems since it depends on terms like $\phi(w)/\Phi(w)$, which causes computational problems for w negative and moderate ($w < -3$, for example). Further, the approach seems not too robust with respect to starting values, that is, unless good starting values are used, the direct maximization approach may not converge.

Using the notation $\mathbf{Z}_i = (y_i, X_i)^T$ for the observed data and $\mathbf{r}_i = (e_i, u_i)^T$ for the random errors, the model that we are considering in this section can be specified as

$$\mathbf{Z}_i = \mathbf{a}_1 + \mathbf{b}_1 x_i + \mathbf{r}_i, \tag{4.1}$$

with the assumptions that

$$x_i \overset{iid}{\sim} SN_1(\mu_x, \sigma_x^2, \lambda_x), \quad \mathbf{r}_i \overset{iid}{\sim} N_2(0, \Psi_1), \tag{4.2}$$

$i = 1, \ldots, n$, all independent, where

$$\mathbf{a}_1 = (\alpha, 0)^T, \quad \mathbf{b}_1 = (\beta, 1)^T, \quad \Psi_1 = \mathrm{diag}(\sigma_e^2, \sigma_u^2).$$

Since $x_i = \mu_x + \sigma_x v_{xi}$, where $v_{xi} = (x_i - \mu_x)/\sigma_x$, the first assumption in (4.2) implies that $v_{xi} \overset{iid}{\sim} SN_1(\lambda_x)$, $i = 1, \ldots, n$, which jointly with (1.6) imply that $v_{xi} = \delta_x |v_{0i}| + (1 - \delta_x^2)^{1/2} v_{1i}$, so that

$$x_i = \mu_x + \sigma_x \delta_x |v_{0i}| + \sigma_x (1 - \delta_x^2)^{1/2} v_{1i}, \tag{4.3}$$

$i = 1, \ldots, n$, where v_{0i} and v_{1i} are iid $N_1(0,1)$ random variables and $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$. Moreover, the independence between x_i and (e_i, u_i), $i = 1, \ldots, n$ imply that $\mathbf{v}_i = (v_{0i}, v_{1i})^T$ and $\mathbf{r}_i = (e_i, u_i)^T$, $i = 1, \ldots, n$, are all independent. Hence, replacing (4.3) in (4.1) we have that

$$\mathbf{Z}_i = \boldsymbol{\mu} + \sigma_x \delta_x \mathbf{b}_1 t_{xi} + \mathbf{r}_{xi}, \tag{4.4}$$

where $\boldsymbol{\mu} = (\alpha + \beta\mu_x, \mu_x)^T$ is the mean vector under the usual structural normal model, and

$$t_{xi} = |v_{0i}|, \quad \mathbf{r}_{xi} = \mathbf{r}_i + \sigma_x (1 - \delta_x^2)^{1/2} \mathbf{b}_1 v_{1i},$$

which by (4.2) are such that

$$\mathbf{r}_{xi} \overset{iid}{\sim} N_2(0, \Psi_1 + \sigma_x^2 (1 - \delta_x^2) \mathbf{b}_1 \mathbf{b}_1^T), \quad t_{xi} \overset{iid}{\sim} HN_1(0,1), \tag{4.5}$$

$i = 1, \ldots, n$, all independent, where $HN_1(0,1)$ denote the standardized univariate *half-normal* distribution. Thus, the results obtained in (4.4) and (4.5) imply that the model defined by (4.1)-(4.2) can be specified as

$$\mathbf{Z}_i | t_{xi} \overset{ind}{\sim} N_2(\boldsymbol{\mu} + \mathbf{b}_x t_{xi}, \Psi_x) \quad \text{and} \quad t_{xi} \overset{iid}{\sim} HN_1(0,1), \tag{4.6}$$

$i = 1, \ldots, n$, where

$$\text{(4.7)} \qquad \mathbf{b}_x = \sigma_x \delta_x \mathbf{b}_1, \quad \mathbf{\Psi}_x = \mathbf{\Psi}_1 + \sigma_x^2(1-\delta_x^2)\mathbf{b}_1\mathbf{b}_1^T = \mathbf{\Sigma} - \mathbf{b}_x\mathbf{b}_x^T$$

and $\mathbf{\Sigma} = \mathbf{\Psi}_1 + \sigma_x^2 \mathbf{b}_1 \mathbf{b}_1^T$ is the covariance matrix under the usual structural normal model.

We call attention to the fact that the joint distribution of $\mathbf{Z}_i$ and t_{xi} that follows from (4.6) is not normal, which yields in particular the marginal density given by (2.11) for the observed vector $\mathbf{Z}_i$. In order to implement the two steeps of the EM-algorithm to obtain maximum likelihood estimates for $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2)$ and λ_x from the unobserved (complete) likelihood that follows from (4.6), where we are considering the (latent) random variables t_{xi}, $i = 1, \ldots, n$, as the missing quantities, we need some additional results which are presented next.

Proposition 3. *Under (4.6) it follows that the complete log-likelihood associated with $(\mathbf{z}_i^T, t_{xi})$, $i = 1, \ldots, n$, can be written as*

$$\text{(4.8)} \quad l_c(\boldsymbol{\theta}, \lambda_x) \propto -\frac{n}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu}) - \frac{1}{2\tau_x^2}\sum_{i=1}^{n}(t_{xi} - \eta_{xi})^2,$$

where

$$\text{(4.9)} \qquad \eta_{xi} = \frac{\mathbf{b}_x^T\mathbf{\Psi}_x^{-1}(\mathbf{z}_i - \boldsymbol{\mu})}{1 + \mathbf{b}_x^T\mathbf{\Psi}_x^{-1}\mathbf{b}_x} \quad \text{and} \quad \tau_x^2 = \frac{1}{1 + \mathbf{b}_x^T\mathbf{\Psi}_x^{-1}\mathbf{b}_x},$$

with $\mathbf{b}_x$ *and* $\mathbf{\Psi}_x$ *as defined in (4.7).*

Proof: In fact, (4.6) imply that the joint density of $\mathbf{Z}_i$ and t_{xi} is

$$f(\mathbf{z}_i, t_{xi}|\boldsymbol{\theta}, \lambda_x) = 2\phi_2(\mathbf{z}_i|\boldsymbol{\mu} + \mathbf{b}_x t_{xi}, \mathbf{\Psi}_x)\phi(t_{xi})I\{t_{xi} > 0\},$$

$i = 1, \ldots, n$ so that by independence

$$l_c(\boldsymbol{\theta}, \lambda_x) = \sum_{i=1}^{n}\log\{2\phi_2(\mathbf{z}_i|\boldsymbol{\mu} + \mathbf{b}_x t_{xi}, \mathbf{\Psi}_x)\phi(t_{xi})\},$$

where, using (4.7) we have after some simple algebraic manipulations that

$$\text{(4.10)} \qquad \phi_2(\mathbf{z}_i|\boldsymbol{\mu} + \mathbf{b}_x t_{xi}, \mathbf{\Psi}_x)\phi(t_{xi}) = \phi_2(\mathbf{z}_i|\boldsymbol{\mu}, \mathbf{\Sigma})\phi_1(t_{xi}|\eta_{xi}, \tau_x^2),$$

$i = 1, \ldots, n$, from where the proof follows.

It follows from (4.8) that to implement the E (or expectation) step is necessary to compute the following conditional moments of t_{xi} given $\mathbf{Z}_i = \mathbf{z}_i$:

$$E[t_{xi}^k|\boldsymbol{\theta}, \lambda_x, \mathbf{z}_i] = \int_{-\infty}^{\infty} t_{xi}^k f(t_{xi}|\boldsymbol{\theta}, \lambda_x, \mathbf{z}_i) dt_{xi}, \tag{4.11}$$

$k = 1, 2$, $i = 1, \ldots, n$, where $f(t_{xi}|\boldsymbol{\theta}, \lambda_x, \mathbf{z}_i)$ denotes the conditional density of t_{xi} given $\mathbf{Z}_i = \mathbf{z}_i$. In order to obtain these conditional moments, we consider first the following lemma (see Johnson et al., 1994, Section 10.1) .

Lemma 3. *Let $X \sim N(\eta, \tau^2)$. Then, for any real constant a it follows that*

$$E[X|X > a] = \eta + \frac{\phi_1\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi_1\left(\frac{a-\eta}{\tau}\right)}\tau,$$

$$E[X^2|X > a] = \eta^2 + \tau^2 + \frac{\phi_1\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi_1\left(\frac{a-\eta}{\tau}\right)}(\eta + a)\tau.$$

Proposition 4. *Lets consider $\mathbf{Z}|t_x \sim N_2(\boldsymbol{\mu} + \mathbf{b}_x t_x, \boldsymbol{\Psi}_x)$ and $t_x \sim HN_1(0, 1)$, where $\mathbf{b}_x$ and $\boldsymbol{\Psi}_x$ are given in (4.7). Then,*

$$E[t_x^k|\boldsymbol{\theta}, \lambda_x, \mathbf{z}] = E[X^k|X > 0],$$

where $X \sim N_1(\eta_x, \tau_x^2)$, with η_x and τ_x^2 given by (4.9). In particular,

$$E[t_x|\boldsymbol{\theta}, \lambda_x, \mathbf{z}] = \eta_x + \frac{\phi_1\left(\frac{\eta_x}{\tau_x}\right)}{\Phi_1\left(\frac{\eta_x}{\tau_x}\right)}\tau_x, \tag{4.12}$$

and

$$E[t_x^2|\boldsymbol{\theta}, \lambda_x, \mathbf{z}] = \eta_x^2 + \tau_x^2 + \frac{\phi_1\left(\frac{\eta_x}{\tau_x}\right)}{\Phi_1\left(\frac{\eta_x}{\tau_x}\right)}\eta_x\tau_x. \tag{4.13}$$

Proof: We note first that (4.11) can be rewritten as

$$E[t_x^k|\boldsymbol{\theta}, \lambda_x, \mathbf{z}] = \frac{1}{f_Z(\mathbf{z}|\boldsymbol{\theta}, \lambda_x)}\int_{-\infty}^{\infty} t_x^k f(\mathbf{z}, t_x|\boldsymbol{\theta}, \lambda_x) dt_x, \tag{4.14}$$

where from (2.11) and the fact that

$$\frac{\eta_x}{\tau_x} = \gamma_x \frac{\sigma_x \mathbf{b}_1^T \Psi^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\sqrt{1+\sigma_x^2 \mathbf{b}_1^T \Psi_1^{-1} \mathbf{b}_1}},$$

the marginal density of $\mathbf{Z}$ is

$$f_Z(\mathbf{z}|\boldsymbol{\theta}, \lambda_x) = 2\phi_Z(\mathbf{z}|\boldsymbol{\mu}, \Sigma)\Phi_1\left(\frac{\eta_x}{\tau_x}\right),$$

and by (4.10) the density of $(\mathbf{Z}^T, t_x)$ can be rewritten as

$$f(\mathbf{z}, t_x|\boldsymbol{\theta}, \lambda_x) = 2\phi_2(\mathbf{z}|\boldsymbol{\mu}, \Sigma)\phi_1(t_x|\eta_x, \tau_x^2)I\{t_x > 0\}.$$

Replacing now this last expression in (4.14), it follows that

$$E[t_x^k|\boldsymbol{\theta}, \lambda_x, \mathbf{z}] = \frac{1}{\Phi_1\left(\frac{\eta_x}{\tau_x}\right)}\int_0^\infty t_x^k \phi_1(t_x|\eta_x, \tau_x^2)dt_x = E[X^k|X>0],$$

where $X \sim N_1(\eta_x, \tau_x^2)$ and $\Phi_1\left(\frac{\eta_x}{\tau_x}\right) = P(X>0)$. Hence, (4.12) and (4.13) follow from Lemma 3 with $a = 0$ and η and τ^2 replaced by η_x and τ_x^2, respectively, which concludes the proof.

The EM algorithm operates as follows. Given starting values $(\boldsymbol{\theta}^{(0)}, \lambda_x^{(0)})$, compute $\widehat{t_{xi}^k} = E[t_{xi}^k|\boldsymbol{\theta}^{(0)}, \lambda_x^{(0)}, \mathbf{z}_i]$, $k = 1, 2$, $i = 1, \ldots, n$, by using (4.13) and (4.14), respectively. Replace the missing values t_{xi}^k by $\widehat{t_{xi}^k}$, $k = 1, 2$, $i = 1, \ldots, n$, in the complete log-likelihood (4.8), and maximize it with respect to $(\boldsymbol{\theta}, \lambda_x)$. This maximization step has to proceed numerically, being most easily accomplished by using Matlab, for example, and does not pose the same difficulties as is the case with a direct maximization of the observed likelihood (2.12). Further, the approach seems somewhat robust with respect to starting values. It may take more computing time but eventually will lead to the maximum of the observed likelihood.

5. A simulation study

In this section we present results of a small scale simulation study to demonstrate the usefulness of the approach developed in Section 2 in studying linear measurement error models where the distribution of (e, u, x) follows the

Table 1: Results of simulation study

n	μ_x	α	β	σ_ϵ^2	σ_u^2	$\sigma_x{}^2$	λ_x	$N.C.$
30	5.5629	0.8488	2.0267	0.9789	0.8710	0.7927	1.3778	17.5%
S.V.	0.5083	6.1461	1.0543	0.4679	0.3328	0.9292	3.7520	
50	5.5598	0.8665	2.1591	0.9846	0.9065	0.8960	1.3448	11.0%
S.V.	0.4075	5.8091	1.0015	0.4610	0.3163	0.7152	3.7707	
100	5.611	1.0477	1.9929	1.0002	0.9336	0.9274	1.2036	10%
S.V.	0.3652	3.1669	0.5477	0.2719	0.2095	0.6007	4.5251	

skew normal distribution in (1.5). For samples sizes n =30, 50 and 100, 1,000 samples were generated according to the measurement error model in (2.1)-(2.4), with $\lambda_e = \lambda_u = 0$, $\lambda_x = 3$, $\mu_x = 5$, $\alpha = 1$, $\beta = 2$ and $\sigma_e = \sigma_u = \sigma_x = 1$. For each generated sample, maximum likelihood estimators of all parameters were computed by using the EM algorithm described in Section 4. The mean values and sample (empirical) variances (S.V.) corresponding to each parameter for the 1,000 generated samples and each sample size are presented in Table 1. The mean value corresponding to the naive (least square estimator) is 0.5987, clearly indicating strong attenuation. Indeed, it seems that the attenuation factor is greater than the usual k_x, under normality. N.C. indicates percentages of samples with $\hat{\lambda}_x = \infty$.

6. An application

In this section we consider a likelihood analysis of a part of the *AIS* data set (available for download at (*http : //stat.umidp.it/SN/index.html*) considering a linear measurement error model relating SSF and $bfat$. We define the model

$$SSF_i = \alpha + \beta bfat_i + \sigma e_i,$$

$i = 1, \ldots, 202$, where $bfat_i$ is the body fat percentage of the i-th individual in the sample, SSF_i is the sum of skin folds and $e_i \overset{iid}{\sim} N(0, \sigma_e^2)$. We assume that $bfat$ is measured with error according to the equation

$$Bfat_i = bfta_i + u_i,$$

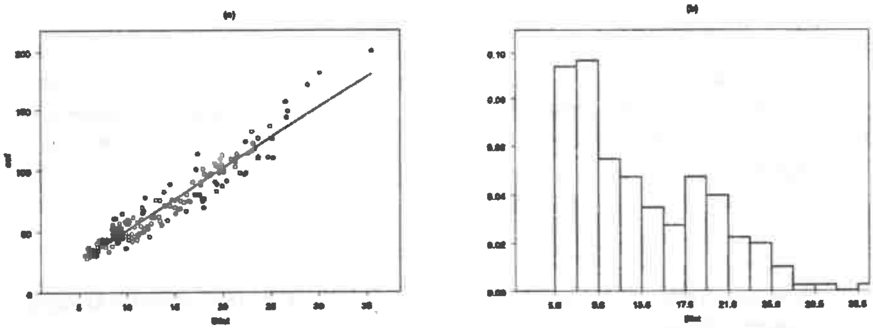
with $Bfat_i$ being an unbiased estimate of the true (unobserved) $bfat_i$, $i = 1, \ldots, n$. A simple plot of the histogram of the observed $Bfat_i$ clearly indicates its asymmetric nature so that we consider

$$bfat_i \overset{iid}{\sim} SN(\mu_x, \sigma_x^2; \lambda_x^2).$$

Table 2 reports some iterations of the algorithm illustrating the observed likelihood increase at each iteration. The likelihood (2.12) for the above data set can also be maximized directly by using Matlab, for example. However, it not simple to get convergence unless good starting values are considered. Using the estimated values obtained by using the EM as starting values for this procedure, exactly the same values were obtained.

The naive (least squares estimator) of β is $\hat{\beta}_{LS} = 5.0400$, which is about 10% smaller than the maximum likelihood estimator, thus indicating clear attenuation due to measurement error.

Figure 1: (a)Least squares line for *Bfat* vs *SSF*. (b) Histogram for *Bfat*.



7. Final conclusions

In the paper we obtain the likelihood function for the simple linear structural measurement error model with the distribution of the random quantities belonging to the family of the skew normal distributions. The likelihood is

Table 2: Convergence of the EM algorithm

iter.	μ_x	α	β	λ	$\ell(\theta)$
1	13.5074	-4.5841	5.4493	0.5769	-1167.2029
50	6.4540	-4.7712	5.4631	10.0786	-1108.5309
100	6.3153	-4.7760	5.4635	19.2492	-1105.2091
200	6.2557	-4.7712	5.4631	34.2548	-1103.7607
300	6.2367	-4.7704	5.4630	41.9238	-1103.5767
400	6.2304	-4.7767	5.4635	44.8501	-1103.5531
461	6.2282	-4.7731	5.4632	45.5745	-1103.5498
462	6.2284	-4.7751	5.4634	45.5817	-1103.5497
463	6.2284	-4.7751	5.4634	45.5820	-1103.5497

obtained by integrating out the unobserved x. We believe that this is the first attempt in working in such general distributional structure for models with measurement errors and that the approach used in the paper can be used in treating more general models which will be the subject of incoming papers. Further, as demonstrated in the simulation study and application, the maximum likelihood approach can be implemented using existing statistical software such as Ox, Matlab, and many others. We also discuss conditions under which the Bayesian inference derived under the symmetric normal model situation is robust in the more general structure of the asymmetric normal measurement error model. Finally, we want to mention that there is no difficulty in extending the approach considered in this paper to the situation where $y_i = \alpha + \mathbf{z}_i^T\beta_z + x_i\beta + e_i$, with the additional covariates $\mathbf{z}_i$, $i = 1, \ldots, n$, measured without error.

References

Arellano-Valle, R. B. and Bolfarine, H. (1996). Elliptical structural models. *Comm. Statist. Theory and Methods* **25** (10), 2319-2342.

Arellano-Valle, R. B., Bolfarine, H. and Gasco, L. (2002). Measurement error models with nonconstant covariance matrices. *Journal of Multivariate Analysis* **82**, 395-415.

Arellano-Valle, R., Bolfarine, H. and Vilca-Labra, F. (1996). Ultrastructural elliptical models. *The Canadian Journal of Statistics* **24** (2), 207-216.

Arellano-Valle, R.B., del Pino, G. and San Martin, E. (2002). Definition and probabilistic properties of skew distributions. *Statistic and Probability Letters* **58**, 111-121.

Arellano-Valle, R.B. and Genton, M. G. (2002). Fundamental skew distributions. *Institute of Statistics Mimeo Series* #2551, under review.
(http://www.stat.ncsu.edu/ mggenton/publications.html)

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

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

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

Bolfarine, H. and Arellano-Valle, R. (1999). Weak nondifferential measurement error models. *Statistics and Probability Letters*, **40**, 279-287.

Bolfarine, H. and Cordani L. (1993). Estimation of Structural linear regress model with a known ratio. *Annals of Institute of Statistical Mathematics,* **45** (3), 531-540.

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

Genton, M. G., He, L. and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistic and Probability Letters*, **51**, 319-325.

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

Johnson, N., Kotz, S. and Balakrishnan, N. (1994). Continuous univariate distributions. Volume 1. Wiley.

Lindley, D.V. and El-Sayad, G. (1968). The Bayesian estimation of a linear functional relationship. *Journal of the Royal Statistics Society,* B **30**, 190-202.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2003-01 - GARCIA,M., PAULA G. A., CYSNEIRO, F.J.A Generalized Leverage in Elliptical Nonlinear Regressio Models. 2003.28p. (RT-MAE-2003-01)

2003-02 - CORDEIRO, G.M., BARROSO, L.P. Gl third-Order Bias Corrected Estimate in Generalized Linea Models. 2003.12p. RT-MAE-2003-02)

2003-03 - GONZALES-LOPEZ, V.A., ITIRO TANAKA, N Bi-Variate Data Modeling Through Generalized Archimedea Copula. 2003.29p. RT-MAE-2003-03)

2003-04 - FERRARI, S.L.P., CRIBARI-NETO, F., Beta Regression for Modeling Rates and Proportions. 2003.19p RT-MAE-2003-04)

2003-05 - FERRARI, S.L.P., CORDEIRO G.M., URIBE-OPAZO, M.A., Improved Score Tests in Symmetric Linear Regression Models 2003.21p. RT-MAE-2003-05)

2003-06 - PEREIRA, C.A.B., STERN, J.M., WHITTLE, M.R., Genuine Bayesian Multiallelic Significance Test For The Hardy-Weinberg Equilibrium Law. 2003.13p. RT-MAE-2003-06)

2003-07 - PEREIRA, B.B., PEREIRA, C.A.B., l Likelihood Aproach to Diagnostic Test in Clinical Medicine 2003.18p. RT-MAE-2003-07)

2003-08 - GÓMEZ, H.W., SALINAS, H.S., BOLFARINE, H., Generalized Skew Normal Models: Properties and Inference 2003.14p. RT-MAE-2003-08)

2003-09 - ARELLANO-VALLE, R.B., BOLFARINE, H., IGLESIAS, P., VIVIANI,P., Systematic Risk Estimation Usin Comparative Calibration Models. 2003.17p. RT-MAE-2003-09)

The complete list of "Relatórios do Departamento d Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil