Journal of Multivariate Analysis 136 (2015) 1–11

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Estimation in skew-normal linear mixed measurement error models

Ameneh Kheradmandi, Abdolrahman Rasekh *

Department of Statistics, Shahid Chamran University, Ahvaz, Iran

ARTICLE INFO

Article history:
Received 3 July 2014
Available online 19 December 2014

AMS 2008 subject classifications:
62J05
62F10

Keywords:
Maximum likelihood
EM algorithm
Structural model
Skewness

ABSTRACT

In this paper we define a class of skew-normal linear mixed measurement error models. This class provides a useful generalization of normal linear mixed models with measurement error in fixed effects variables. It is assumed that the random effects, model errors and measurement errors follow a skew-normal distribution, extending usual symmetric normal model in order to avoid data transformation. We find the likelihood function of the observed data, which can be maximized by using standard optimization techniques. Next, an EM-type algorithm is proposed for estimating the parameters that seems to provide some advantages over a direct maximization of the likelihood. Finally, we propose results of a simulation study and an example of real data for illustration.

1. Introduction

Linear mixed effects models are the most common statistical tools for analyzing repeated measurement data and in particular, longitudinal data in biomedical, agricultural, environmental and also in economics and social sciences. In the models, independent variables are often measured with non-negligible errors (Davidian and Giltinan, [16]). Hence considerable interest has been focused on the study of the estimation of parameters in measurement error models. Main references on the subject include Armstrong [4], Fuller [19], Stefanski and Carroll [25], Cheng and Van Ness [13], Zhong et al. [29], Carroll et al. [12] and Zare et al. [27]. Let y_i be an observed continues response. We assume the following linear mixed measurement error model as:

$$\begin{aligned} y_i &= \alpha + \beta x_i + b_i z_i + e_i, \\ X_i &= x_i + u_i, \quad i = 1, \ldots, n \end{aligned} \tag{1}$$

where α and β are fixed parameters and b_i's are random effects with $b_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_b^2)$. In this model X_i is the observed value of the x_i with measurement error u_i. Furthermore, it is assumed that $(e_i, u_i)' \overset{\text{i.i.d.}}{\sim} N_2((0, 0)', \operatorname{diag}\{\sigma_e^2, \sigma_u^2\})$, with i.i.d. means independent and identically distributed. In the structural model, it is also assumed that $x_i \overset{\text{i.i.d.}}{\sim} N(\mu_x, \sigma_x^2)$, and $\{x, e, u, b\}$ are mutually independent (see, Fuller [19]). There are several proposals of estimation for mixed effects models, for example see, Davidian and Giltinan [15,16], Karcher and Wang [23], Demidenko and Stukel [17] and Vonesh et al. [26]. Zare et al. [27] studied the estimation problem for the functional mixed measurement error model. Under normality assumption, they

* Corresponding author.
E-mail addresses: amene.kheradmandi@gmail.com (A. Kheradmandi), rasekh_a@scu.ac.ir (A. Rasekh).

http://dx.doi.org/10.1016/j.jmva.2014.12.007
0047-259X/

applied the corrected score method of Nakamura [24] to obtain the estimators of the parameters. Cui et al. [14] derived the moments of estimators for the parameters in mixed effects model with error in variables.

On the other hand, the normality (symmetry) assumption is a routine but possibly restrictive assumption for different statistical models including the linear mixed measurement error models. In recent years, considerable interest has focused on the models relaxing normality assumption and incorporating asymmetry. Azzalini [6] introduces skew-normal distribution, extending usual normal model in order to avoid data transformation. The univariate skew-normal density function with location parameter μ, scale parameter σ^2 and skewness parameter λ, is defined by

$$f(x;\mu,\sigma^2,\lambda)=2\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right),\quad x,\mu,\lambda\in R,\ \sigma>0$$

where $\phi(.)$ and $\Phi(.)$ denote the probability density function and cumulative distribution function, respectively, of the normal distribution. A random variable $Z=\frac{x-\mu}{\sigma}$ follows a standard skew-normal distribution with $\mu=0$ and $\sigma^2=1$, which is denoted by $SN(\lambda)$. The skew-normal distribution has the following properties:

(i) $E(X)=\mu+\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}$.
(ii) $\mathrm{Var}(X)=(1-\frac{2\lambda}{\pi(1+\lambda^2)})\sigma^2$.
(iii) $\nu=\frac{1}{2}(4-\pi)(\frac{E^2(X)}{\mathrm{Var}(X)})^{\frac{3}{2}}$ and $\kappa=2(\pi-3)(\frac{E^2(X)}{\mathrm{Var}(X)})^2$, where ν and κ are asymmetry and kurtosis indexes, respectively.
(iv) If $\lambda=0$ then $X\sim N(\mu,\sigma^2)$.
(v) As pointed out by Henze [21], if $Z\sim SN(\lambda)$ then

$$Z\overset{d}{=}\delta|Z_0|+(1-\delta^2)^{\frac{1}{2}}Z_1. \tag{2}$$

where $Z_0\sim N(0,1)$ and $Z_1\sim N(0,1)$ are independent variables, $\delta=\frac{\lambda}{\sqrt{(1+\lambda^2)}}$ and $\overset{d}{=}$ means "distributed as".

Other properties of this distribution and its variations have been discussed by several authors including Azzalini [7], Henze [21], Azzalini and Capitanio [9], Azzalini and Dalla Valla [10], Arnold and Beaver [5] and Azzalini [8]. Arellano-Valle et al. [3] define a class of skew-normal measurement error models for a linear regression model. Skew-normal linear mixed models are introduced in Arellano-Valle et al. [1] and Bolfarine et al. [11] consider influence diagnostics for this model. In this paper, we define skew-normal linear mixed measurement error model that follows by replacing the normality assumptions by the assumptions that the random terms e_i and u_i, the random effect b_i and the independent variable x_i have the skew-normal distribution. We obtain the likelihood function of the observed data $\boldsymbol{Z}_i=(y_i,X_i),\ i=1,\ldots,n$. In addition, we consider some special cases where $\lambda_e=\lambda_u=\lambda_b=0$ or $\lambda_e=\lambda_u=\lambda_x=0$ in detail. We present an EM algorithm, which can overcome some difficulties detected by using direct maximization of the likelihood function, especially in terms of robustness with respect to starting values. The plan of the paper is as follows: In Section 2, we derive the marginal density of the observed data $\boldsymbol{Z}_i$ by integrating out the unobserved variables x_i and b_i. In Section 3, we present an EM-type algorithm and a simulation study is given in Section 4. To illustrate the usefulness of the proposed methods, an application to a real data set is reported in Section 5. Finally, Section 6 is dedicated to the concluding remarks.

2. The skew-normal linear mixed measurement error model

To consider a structural skew-normal mixed measurement error model, under the model defined by (1), we assume that:

$$\begin{aligned}&e_i\overset{\text{i.i.d.}}{\sim}SN(0,\sigma_e^2,\lambda_e),\qquad u_i\overset{\text{i.i.d.}}{\sim}SN(0,\sigma_u^2,\lambda_u),\\&x_i\overset{\text{i.i.d.}}{\sim}SN(0,\sigma_x^2,\lambda_x),\qquad b_i\overset{\text{i.i.d.}}{\sim}SN(0,\sigma_b^2,\lambda_b),\quad i=1,\ldots,n\end{aligned}\tag{3}$$

with $\{x,e,u,b\}$ are independent. Leading, under above model, we have

$$\begin{aligned}&y_i|x_i,b_i\sim SN(\alpha+\beta x_i+b_iz_i,\sigma_e^2,\lambda_e),\\&X_i|x_i,b_i\sim SN(x_i,\sigma_u^2,\lambda_u).\end{aligned}$$

In the following, we drop the subscript i in a sample unit to simplify notation. Hence, the conditional distribution of (y,X) given x and b can be computed by

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

Furthermore,

$$f(x,b)=\frac{2^2}{\sigma_x\sigma_b}\phi\left(\frac{x-\mu_x}{\sigma_x}\right)\phi\left(\frac{b}{\sigma_b}\right)\Phi\left(\lambda_x\frac{x-\mu_x}{\sigma_x}\right)\Phi\left(\lambda_b\frac{b}{\sigma_b}\right).$$

Then, the joint marginal density of (y, X) obtained by

$$f(y,X)=\iint f(y,X|x,b)f(x,b)dxdb$$
$$=\iint \frac{2^4}{\sigma_e\sigma_u\sigma_x\sigma_b}\phi\left(\frac{y-\alpha-\beta x-bz}{\sigma_e}\right)\phi\left(\frac{X-x}{\sigma_u}\right)\phi\left(\frac{x-\mu_x}{\sigma_x}\right)\phi\left(\frac{b}{\sigma_b}\right)$$
$$\times\Phi\left(\lambda_e\frac{y-\alpha-\beta x-bz}{\sigma_e}\right)\Phi\left(\lambda_u\frac{X-x}{\sigma_u}\right)\Phi\left(\lambda_x\frac{x-\mu_x}{\sigma_x}\right)\Phi\left(\lambda_b\frac{b}{\sigma_b}\right)dxdb. \quad (4)$$

Making the transformations $\nu = x-\mu_x$, $\bar{u} = X-\mu_x$ and $\bar{e} = y-\alpha-\beta\mu_x$, we have (4) as:

$$f(y,X)=\iint \frac{2^4}{\sigma_e\sigma_u\sigma_x\sigma_b}\phi\left(\frac{\bar{e}-\beta\nu-bz}{\sigma_e}\right)\phi\left(\frac{\bar{u}-\nu}{\sigma_u}\right)\phi\left(\frac{\nu}{\sigma_x}\right)\phi\left(\frac{b}{\sigma_b}\right)$$
$$\times\Phi\left(\lambda_e\frac{\bar{e}-\beta\nu-bz}{\sigma_e}\right)\Phi\left(\lambda_u\frac{\bar{u}-\nu}{\sigma_u}\right)\Phi\left(\lambda_x\frac{\nu}{\sigma_x}\right)\Phi\left(\lambda_b\frac{b}{\sigma_b}\right)d\nu db. \quad (5)$$

Let $\boldsymbol{w} = (\bar{e}, \bar{u}, 0, 0)'$, $\boldsymbol{\Gamma} = (\nu, b)'$, $\boldsymbol{c}' = \begin{pmatrix}\beta & 1 & -1 & 0\\ z & 0 & 0 & -1\end{pmatrix}$, $\boldsymbol{\Psi} = \mathrm{diag}(\sigma_e, \sigma_u, \sigma_x, \sigma_b)$, $\boldsymbol{\Psi}_1 = \mathrm{diag}(\sigma_e, \sigma_u)$, $\boldsymbol{\Psi}_2 = \mathrm{diag}(\sigma_e, \sigma_u, \sigma_x)$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_e, \lambda_u, \lambda_x, \lambda_b)$. Then it follows that

$$\phi\left(\frac{\bar{e}-\beta\nu-bz}{\sigma_e}\right)\phi\left(\frac{\bar{u}-\nu}{\sigma_u}\right)\phi\left(\frac{\nu}{\sigma_x}\right)\phi\left(\frac{b}{\sigma_b}\right)=\phi_4(\boldsymbol{w};\boldsymbol{c}\boldsymbol{\Gamma},\boldsymbol{\Psi}), \quad (6)$$

and

$$\Phi\left(\lambda_e\frac{\bar{e}-\beta\nu-bz}{\sigma_e}\right)\Phi\left(\lambda_u\frac{\bar{u}-\nu}{\sigma_u}\right)\Phi\left(\lambda_x\frac{\nu}{\sigma_x}\right)\Phi\left(\lambda_b\frac{b}{\sigma_b}\right)=\Phi_4(\boldsymbol{\Lambda}\boldsymbol{w}-\boldsymbol{\Lambda}\boldsymbol{c}\boldsymbol{\Gamma};\boldsymbol{0},\boldsymbol{\Psi}), \quad (7)$$

where ϕ_k and Φ_k are the probability density and the cumulative distribution function, respectively, of the k-dimensional normal distribution. Next, we need the following two lemmas:

Lemma 1.

$$\phi_4(\boldsymbol{w};\boldsymbol{c}\boldsymbol{\Gamma},\boldsymbol{\Psi})=\phi_2(\boldsymbol{Z};\boldsymbol{\mu},\boldsymbol{\Sigma})\phi_2(\boldsymbol{\Gamma};(\mathbf{c}'\boldsymbol{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{w},(\mathbf{c}'\boldsymbol{\Psi}^{-1}\mathbf{c})^{-1}),$$

where $\boldsymbol{Z} = (y, X)'$, $\boldsymbol{\mu} = \begin{pmatrix}\alpha+\beta\mu_x\\ \mu_x\end{pmatrix}$ *and* $\boldsymbol{\Sigma} = \begin{pmatrix}\beta^2\sigma_x^2+z^2\sigma_b^2+\sigma_e^2 & \beta\sigma_x^2\\ \beta\sigma_x^2 & \sigma_x^2+\sigma_u^2\end{pmatrix}$.

Proof. See Appendix. □

Lemma 2. *Let* $\boldsymbol{V} \sim N_k(\boldsymbol{\xi}, \boldsymbol{\Delta})$. *Then for any fixed m-dimensional vector* $\boldsymbol{a}$ *and* $m \times k$ *matrix* $\boldsymbol{A}$,

$$E\left[\Phi_m(\boldsymbol{A}+\boldsymbol{a}\boldsymbol{V};\boldsymbol{\kappa},\boldsymbol{H})\right]=\Phi_m(\boldsymbol{a};\boldsymbol{\kappa}-\boldsymbol{A}\boldsymbol{\xi},\boldsymbol{H}+\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{A}').$$

Proof. See Arellano-Valle et al. [3]. □

With the aid of above lemmas, we prove the following theorem, which is the main result of the paper.

Theorem 1. *Under the skew-normal linear mixed measurement error model given in (1) and (3), the marginal density function of* $\boldsymbol{Z} = (y, X)'$ *is given by*

$$f(\boldsymbol{Z};\boldsymbol{\theta},\boldsymbol{\lambda})=2^4\phi_2(\boldsymbol{Z};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi_4(\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{w};\boldsymbol{0},\boldsymbol{\Psi}+\boldsymbol{\Lambda}c(c'\boldsymbol{\Psi}^{-1}c)^{-1}c'\boldsymbol{\Lambda}),$$

where $\boldsymbol{\theta}' = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \sigma_b^2)$, $\boldsymbol{\lambda}' = (\lambda_e, \lambda_u, \lambda_x, \lambda_b)$, $\boldsymbol{B} = (\boldsymbol{I}_2, \mathbf{c}_1)$, $\mathbf{c}_1 = \begin{pmatrix}\beta & z\\ 1 & 0\end{pmatrix}$ *and* $\boldsymbol{M}\boldsymbol{w} = \boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu})$.

Proof. See Appendix. □

Remarks:

i. Note that the density given in Theorem 1 is in the class of fundamental skew-distributions defined by Arellano-Valle and Genton [2]. As pointed out by Arellano-Valle et al. [3], in this family of densities, the third moment is nonnull when $\boldsymbol{\lambda} \neq \boldsymbol{0}$. Hence, the skew-normal linear mixed measurement error model defined in this paper is identifiable as long as $\boldsymbol{\lambda} \neq \boldsymbol{0}$. However, parameter estimates are unstable if $\boldsymbol{\lambda}$ is close to $\boldsymbol{0}$, that is, when $\boldsymbol{Z}$ is close to being normally distributed, so that the additional assumptions on the variance have to be added to the model. This condition was also noted in simulation study and application conducted by authors.
ii. We call attention to the fact that if $\sigma_b^2 = 0$ then results obtained in Lemmas 1, 2 and Theorem 1 reduce to the results obtained in Arellano-Valle et al. [3]. Thus we have extended the skew-normal measurement error model in a nice way to the skew-normal linear mixed measurement error model.

Hence, the likelihood function of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ given the observed sample $\boldsymbol{Z}_1 = (y_1, X_1)', \ldots, \boldsymbol{Z}_n = (y_n, X_n)'$ is given by $L(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) = \prod_{i=1}^{n} f(\boldsymbol{Z}_i; \boldsymbol{\theta}, \boldsymbol{\lambda})$ and its log-likelihood function can be written as:

$$\ell(\boldsymbol{\theta}, \boldsymbol{\lambda}) \propto -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_i - \boldsymbol{\mu}) + \sum_{i=1}^{n}\log\left[\Phi_4(\boldsymbol{\Lambda}_*\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Z}_i - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Omega})\right],$$

where $\boldsymbol{\Lambda}_* = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-\frac{1}{2}}$ and $\boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda}$. The maximum likelihood estimators are given by solution of the following equation:

$$\frac{\partial\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\theta}} = -\frac{n}{2|\boldsymbol{\Sigma}|}\frac{\partial|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\theta}} - \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\theta}}\sum_{i=1}^{n}(\boldsymbol{Z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_i - \boldsymbol{\mu}) + \sum_{i=1}^{n}\left(\frac{\frac{\partial}{\partial\boldsymbol{\theta}}\Phi_4\left(\boldsymbol{\Lambda}_*\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Z}_i - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Omega}\right)}{\Phi_4\left(\boldsymbol{\Lambda}_*\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Z}_i - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Omega}\right)}\right) = \boldsymbol{0},$$

$$\frac{\partial\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = \sum_{i=1}^{n}\left(\frac{\frac{\partial}{\partial\boldsymbol{\lambda}}\Phi_4\left(\boldsymbol{\Lambda}_*\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Z}_i - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Omega}\right)}{\Phi_4\left(\boldsymbol{\Lambda}_*\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Z}_i - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Omega}\right)}\right) = \boldsymbol{0}.$$

Noting that the likelihood function has to be maximized numerically. It can be implemented by statistical softwares such as R, Matlab and OX. We are interested in some special cases such as $\lambda_e = \lambda_u = \lambda_x = 0$ and $\lambda_e = \lambda_u = \lambda_b = 0$ which are two special cases of the above general situations.

Corollary 1. *Under the conditions of Theorem 1, it follows that:*

i. *If $\lambda_e = \lambda_u = \lambda_b = 0$ the density function of $\boldsymbol{Z} = (y, X)'$ is given by*

$$f(\boldsymbol{Z}; \boldsymbol{\theta}, \lambda_x) = 2|\boldsymbol{\Psi}|^{-\frac{1}{2}}\phi_2(\boldsymbol{Z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi\left((\sigma_x^2 + f\lambda_x^2)^{-\frac{1}{2}}\lambda_x\sigma_x^2\boldsymbol{c}_1^{*'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\right) \tag{8}$$

where $\boldsymbol{c}_1^* = \binom{\beta}{1}$ and

$$f = \frac{z^2\sigma_e^{-2} + \sigma_b^{-2}}{(z^2\sigma_e^{-2} + \sigma_b^{-2})(\sigma_u^{-2} + \sigma_b^{-2}) + \beta^2\sigma_e^{-2}\sigma_b^{-2}}.$$

ii. *If $\lambda_e = \lambda_u = \lambda_x = 0$ the density function of $\boldsymbol{Z} = (y, X)'$ is given by*

$$f(\boldsymbol{Z}; \boldsymbol{\theta}, \lambda_x) = 2|\boldsymbol{\Psi}|^{-\frac{1}{2}}\phi_2(\boldsymbol{Z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\nu_b\lambda_b\sigma_b\boldsymbol{c}_2^{*'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})) \tag{9}$$

where $\boldsymbol{c}_2^* = \binom{z}{0}$ and

$$\nu_b = \left[1 + \frac{\lambda_b^2(\beta^2\sigma_u^2\sigma_x^2 + \sigma_e^2(\sigma_u^2 + \sigma_x^2))}{\beta^2\sigma_u^2\sigma_x^2 + (z^2\sigma_b^2 + \sigma_e^2)(\sigma_u^2 + \sigma_x^2)}\right]^{-\frac{1}{2}}.$$

Proof. i. Let $\boldsymbol{\Lambda} = \mathrm{diag}(0, 0, \lambda_x, 0) = \lambda_x\boldsymbol{e}_3\boldsymbol{e}_3'$ where $\boldsymbol{e}_3 = (0, 0, 1, 0)'$ and $\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda} = \mathrm{diag}(\sigma_e^2, \sigma_u^2, \sigma_x^2 + f\lambda_x^2, \sigma_b^2)$. Noting that $\boldsymbol{B}\boldsymbol{e}_3 = \boldsymbol{c}_1^*$ and $\boldsymbol{\Psi}\boldsymbol{e}_3 = \sigma_x^2\boldsymbol{e}_3$, we have that:

$$\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) = \lambda_x\sigma_x^2\boldsymbol{e}_3\boldsymbol{c}_1^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}),$$

then

$$(\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda})^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) = (\sigma_x^2 + f\lambda_x^2)^{-\frac{1}{2}}\lambda_x\sigma_x^2\boldsymbol{c}_1^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\boldsymbol{e}_3,$$

so that

$$\begin{aligned}\Phi_4(\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda}) &= \Phi_4\left((\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda})^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{I}_4\right)\\ &= \Phi_4\left((\sigma_x^2 + f\lambda_x^2)^{-\frac{1}{2}}\lambda_x\sigma_x^2\boldsymbol{c}_1^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\boldsymbol{e}_3; \boldsymbol{0}, \boldsymbol{I}_4\right),\end{aligned}$$

which concludes the proof, since $\Phi_4(a\boldsymbol{e}_3; \boldsymbol{0}, \boldsymbol{I}_4) = \frac{\Phi_1(a)}{8}$.

ii. Let $\boldsymbol{\Lambda} = \mathrm{diag}(0, 0, 0, \lambda_b) = \lambda_b\boldsymbol{e}_4\boldsymbol{e}_4'$, where $\boldsymbol{e}_4 = (0, 0, 0, 1)'$, and

$$\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda} = \mathrm{diag}(\sigma_e^2, \sigma_u^2, \sigma_x^2, \sigma_b^2\nu_b^{-2}),$$

with noting that $\boldsymbol{B}\boldsymbol{e}_4 = \boldsymbol{c}_2^*$ and $\boldsymbol{\Psi}\boldsymbol{e}_4 = \sigma_b^2\boldsymbol{e}_4$, we have that

$$\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) = \lambda_b\sigma_b^2\boldsymbol{e}_4\boldsymbol{c}_2^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}),$$

then,

$$(\boldsymbol{\Psi}+\boldsymbol{\Lambda c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda})^{-\frac{1}{2}}\boldsymbol{\Lambda\Psi B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu}) = \nu_b\lambda_b\sigma_b\boldsymbol{c}_2^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu})\boldsymbol{e}_4,$$

so that,

$$\begin{aligned}\Phi_4(\boldsymbol{\Lambda\Psi B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu});\boldsymbol{0},\boldsymbol{\Psi}+\boldsymbol{\Lambda c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda}) &= \Phi_4\left((\boldsymbol{\Psi}+\boldsymbol{\Lambda c}(\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})^{-1}\boldsymbol{c}'\boldsymbol{\Lambda})^{-\frac{1}{2}}\boldsymbol{\Lambda\Psi B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu});\boldsymbol{0},\boldsymbol{I}_4\right)\\ &= \Phi_4(\nu_b\lambda_b\sigma_b\boldsymbol{c}_2^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu})\boldsymbol{e}_4;\boldsymbol{0},\boldsymbol{I}_4),\end{aligned}$$

this concludes the proof, since $\Phi_4(a\boldsymbol{e}_4;\boldsymbol{0},\boldsymbol{I}_4)=\frac{\Phi_1(a)}{8}$. □

3. An EM-type algorithm

A direct maximization of the likelihood function (8) and (9) may sometimes poses problems since it involves terms like $\log\Phi(w)$. Further, the approach seems not too robust with starting value. As pointed out by Dempster et al. [18], the EM algorithm is a popular iterative algorithm for ML estimation in models with incomplete data. In this section we want to find an EM-type algorithm for maximization of the likelihood function (8) and (9).

3.1. EM-type algorithm when $\lambda_e=\lambda_u=\lambda_x=0$

We can write the model given in (1) as:

$$\boldsymbol{Z}_i=\boldsymbol{a}_1+\boldsymbol{c}_1^*x_i+b_i\boldsymbol{c}_2^*+\boldsymbol{r}_i, \tag{10}$$

with the assumptions that

$$\boldsymbol{r}_i \overset{\text{i.i.d.}}{\sim} N_2(\boldsymbol{0},\boldsymbol{\Psi}_1),\qquad x_i\overset{\text{i.i.d.}}{\sim} N(\mu_x,\sigma_x^2),\qquad b_i\overset{\text{i.i.d.}}{\sim} SN(0,\sigma_b^2,\lambda_b), \tag{11}$$

where $\boldsymbol{Z}_i=(y_i,X_i)'$, $\boldsymbol{a}_1=(\alpha,0)'$, $\boldsymbol{r}_i=(e_i,u_i)'$ and $\boldsymbol{\Psi}_1=\operatorname{diag}(\sigma_e^2,\sigma_u^2)$. The last assumption in (11) implies that $\alpha_{bi}=\frac{b_i}{\sigma_b}\overset{\text{i.i.d.}}{\sim}SN(\lambda_b)$, $i=1,\ldots,n$. The property given in (2) implies that $\alpha_{bi}=\delta_b|\nu_{0_i}|+(1-\delta_b^2)^{\frac{1}{2}}\nu_{1_i}$, and so

$$b_i=\sigma_b\delta_b|\nu_{0_i}|+\sigma_b(1-\delta_b^2)^{\frac{1}{2}}\nu_{1_i},\quad i=1,\ldots,n \tag{12}$$

where ν_{0_i} and ν_{1_i} are i.i.d. random variables with standard normal distributions and $\delta_b=\frac{\lambda_b}{(1+\lambda_b^2)^{\frac{1}{2}}}$. The independence between b_i and $(x_i,\boldsymbol{r}_i')$, $i=1,\ldots,n$ imply that $\boldsymbol{V}_i=(\nu_{0_i},\nu_{1_i})$ and $(x_i,\boldsymbol{r}_i')$ are all independent. Hence, replacing (12) in (10) we have that

$$\boldsymbol{Z}_i=\boldsymbol{a}_1+\sigma_b\delta_b\boldsymbol{c}_2^*t_{bi}+\boldsymbol{r}_{xb_i}, \tag{13}$$

where $t_{b_i}=|\nu_{0i}|$ and $\boldsymbol{r}_{xb_i}=\boldsymbol{c}_1^*x_i+\sigma_b(1-\delta_b^2)^{\frac{1}{2}}\boldsymbol{c}_2^*\nu_{1i}+\boldsymbol{r}_i$. Thus the assumptions given in (11), imply that

$$\boldsymbol{r}_{xb_i}\overset{\text{i.i.d.}}{\sim}N_2(\boldsymbol{c}_1^*\mu_x,\boldsymbol{\Psi}_1+\sigma_b^2(1-\delta_b^2)\boldsymbol{c}_2^*\boldsymbol{c}_2^{*'}+\sigma_x^2\boldsymbol{c}_1^*\boldsymbol{c}_1^{*'}),\qquad t_{b_i}\overset{\text{i.i.d.}}{\sim}HN_1(0,1), \tag{14}$$

where $HN_1(0,1)$ denotes the half-normal distribution. The results obtained in (13) and (14) imply that the model defined in (10) and (11) can be specified as

$$\boldsymbol{Z}_i|t_{b_i}\overset{\text{i.i.d.}}{\sim}N_2(\boldsymbol{\mu}+\boldsymbol{c}_{xb}^*t_{b_i},\boldsymbol{\Psi}_{xb}), \tag{15}$$

where $\boldsymbol{\mu}=\boldsymbol{a}_1+\boldsymbol{c}_1^*\mu_x$, $\boldsymbol{c}_{xb}^*=\sigma_b\delta_b\boldsymbol{c}_2^*$ and $\boldsymbol{\Psi}_{xb}=\boldsymbol{\Psi}_1+\sigma_b^2(1-\delta_b^2)\boldsymbol{c}_2^*\boldsymbol{c}_2^{*'}+\sigma_x^2\boldsymbol{c}_1^*\boldsymbol{c}_1^{*'}$. In order to implement the two steps of the EM algorithm, we need some lemmas, which are presented next.

Lemma 3. *The complete* log*-likelihood function of* $(\boldsymbol{Z}_i',t_{b_i})$, $i=1,\ldots,n$ *is given by*

$$\ell_c(\boldsymbol{\theta},\lambda_b)\propto-\frac{n}{2}\log\boldsymbol{\Psi}_{xb}-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Z}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_i-\boldsymbol{\mu})-\frac{1}{2\tau_{xb}^2}\sum_{i=1}^{n}(t_{b_i}-\eta_{xb_i})^2, \tag{16}$$

where

$$\eta_{xb_i}=\frac{\boldsymbol{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}(\boldsymbol{Z}_i-\boldsymbol{\mu})}{1+\boldsymbol{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}\boldsymbol{c}_{xb}^*}\quad\text{and}\quad\tau_{xb}^2=\frac{1}{1+\boldsymbol{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}\boldsymbol{c}_{xb}^*}.$$

Proof. Under the distributions of $\boldsymbol{Z}_i|t_{b_i}$ and t_{b_i} given in (14) and (15), we have

$$f(\boldsymbol{Z}_i,t_{b_i};\boldsymbol{\theta},\lambda_b)=2\phi_2(\boldsymbol{Z}_i;\boldsymbol{\mu}+\boldsymbol{c}_{xb}^*t_{b_i},\boldsymbol{\Psi}_{xb})\phi_1(t_{b_i})\boldsymbol{I}_{\left[t_{b_i}>0\right]}.$$

Also $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}_{xb}$ can be written as $\boldsymbol{\Sigma} = \boldsymbol{\Psi}_1 + \sigma_b^2 \boldsymbol{c}_2^* \boldsymbol{c}_2^{*'} + \sigma_x^2 \boldsymbol{c}_1^* \boldsymbol{c}_1^{*'}$ and $\boldsymbol{\Psi}_{xb} = \boldsymbol{\Sigma} - \boldsymbol{c}_{xb}^* \boldsymbol{c}_{xb}^{*'}$. Then, after some manipulations, we can show that

$$\phi_2(\boldsymbol{Z}_i | \boldsymbol{\mu} + \boldsymbol{c}_{xb}^* t_{b_i}, \boldsymbol{\Psi}_{xb}) \phi_1(t_{b_i}) = \phi_2(\boldsymbol{Z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \phi_1(t_{b_i}; \eta_{xb_i}, \tau_{xb}^2), \quad i = 1, \ldots, n$$

this concludes the proof.

Lemma 4. *Let $X \sim N(\eta, \tau^2)$. Then for any real constant a, it follows that*

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

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

Proof. See Johnson et al. [22]. □

Lemma 5. *Let t_{b_i} and $\boldsymbol{Z}_i | t_{b_i}$ be distributed as given in (14) and (15), respectively. It follows that $E(t_b^k | \boldsymbol{Z}) = E(X^k | X > 0)$, where $X \sim N_1(\eta_{xb}, \tau_{xb}^2)$ with η_{xb} and τ_{xb}^2 given by (16). In particular*

$$E(t_b | \boldsymbol{Z}) = \eta_{xb} + \frac{\phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)}{\Phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)} \tau_{xb}, \tag{17}$$

$$E(t_b^2 | \boldsymbol{Z}) = \eta_{xb}^2 + \tau_{xb}^2 + \frac{\phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)}{\Phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)} \eta_{xb} \tau_{xb}. \tag{18}$$

Proof. See Appendix. □

Each iteration of the EM algorithm consists of two steps, the expectation step (E-step) and the maximization step (M-step). The following steps of the EM algorithm can be formulated to obtain the ML estimate of $(\boldsymbol{\theta}, \lambda_b)$ for the likelihood of (16):

E-step:

1. Given starting values $(\boldsymbol{\theta}^{(0)}, \lambda_b^{(0)})$
2. Compute $\hat{t}_{b_i}^k = E(t_{b_i}^k | \boldsymbol{\theta}^{(0)}, \lambda_b^{(0)}, \boldsymbol{Z}_i)$, $k = 1, 2$, $i = 1, \ldots, n$, by using (17) and (18).
3. Replace the missing values $t_{b_i}^k$ by $\hat{t}_{b_i}^k$, $k = 1, 2$, $i = 1, \ldots, n$ in the complete log-likelihood (16).

M-step: This step has to proceed numerically. This maximization step does not pose the same difficulty as is the case with a direct maximization of the observed likelihood of (8).

3.2. EM-type algorithm when $\lambda_e = \lambda_u = \lambda_b = 0$

Likewise, we consider the case where $\lambda_x \neq 0$, that is, the structural mixed measurement error model in (10), with the assumptions that

$$\boldsymbol{r}_i \overset{\text{i.i.d.}}{\sim} N_2(\boldsymbol{0}, \boldsymbol{\Psi}_1), \qquad x_i \overset{\text{i.i.d.}}{\sim} SN(\mu_x, \sigma_x^2, \lambda_x), \quad \text{and} \quad b_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_b^2), \tag{19}$$

with $\{\boldsymbol{r}_i, x_i, b_i\}$ are all independent. Because of the above assumptions, it follows from property given in (2) that:

$$\boldsymbol{Z}_i = \boldsymbol{\mu} + \sigma_x \delta_x \boldsymbol{c}_1^* t_{x_i} + \boldsymbol{r}_{xb_i}, \tag{20}$$

where

$$t_{x_i} = |v_{0i}|, \qquad \boldsymbol{\mu} = \boldsymbol{a}_1 + \boldsymbol{c}_1^* \mu_x, \qquad \delta_x = \frac{\lambda_x}{(1 + \lambda_x^2)^{\frac{1}{2}}},$$

$$\boldsymbol{r}_{xb_i} = \boldsymbol{c}_1^* \sigma_x (1 - \delta_x^2)^{\frac{1}{2}} v_{1i} + b_i \boldsymbol{c}_2^* + \boldsymbol{r}_i, \tag{21}$$

which are such that

$$\boldsymbol{r}_{xb_i} \overset{\text{i.i.d.}}{\sim} N_2(\boldsymbol{0}, \boldsymbol{\Psi}_1 + \sigma_b^2 \boldsymbol{c}_2^* \boldsymbol{c}_2^{*'} + \sigma_x^2 (1 - \delta_x^2) \boldsymbol{c}_1^* \boldsymbol{c}_1^{*'}), \quad \text{and} \quad t_{x_i} \overset{\text{i.i.d.}}{\sim} HN_1(0, 1) \tag{22}$$

and they are independent, $i = 1, \ldots, n$. Therefore, (20) and (22) imply that the model defined by (10) and (19) can be written as

$$\boldsymbol{Z}_i | t_{x_i} \overset{\text{i.i.d.}}{\sim} N_2(\boldsymbol{\mu} + \boldsymbol{c}_{xb}^* t_{x_i}, \boldsymbol{\Psi}_{xb}), \tag{23}$$

Table 1
Estimated mean and standard deviation (SD) of the parameters from simulation study, with true SN(5,1,3) distribution for the x. True values of parameters are in parentheses.

		$\mu_x(5)$	$\alpha(2)$	$\beta(3)$	$\sigma_e^2(1)$	$\sigma_u^2(1)$	$\sigma_x^2(1)$	$\sigma_b^2(1)$	$\lambda_x(3)$	N.C.
$n = 30$	Mean	4.991	1.922	1.845	0.748	0.860	1.125	0.748	2.784	11.8%
	SD	0.274	3.558	0.594	0.271	0.239	0.553	0.271	2.249	
$n = 50$	Mean	5.009	1.380	1.936	0.868	0.921	1.071	0.867	3.049	10.9%
	SD	0.216	3.286	0.558	0.282	0.202	0.438	0.281	2.351	
$n = 100$	Mean	5.036	1.336	1.951	0.942	0.984	1.060	0.942	3.281	3.6%
	SD	3.042	2.193	0.440	0.361	0.294	0.421	0.361	2.265	
Normal	Mean	5.756	1.081	1.987	0.902	0.989	0.427	0.902	0	0%
	SD	0.121	0.078	0.083	0.362	0.173	0.122	0.362	0	

where $\mathbf{c}_{xb}^* = \sigma_x\delta_x\mathbf{c}_1^*$ and $\boldsymbol{\Psi}_{xb} = \boldsymbol{\Psi}_1 + \sigma_b^2\mathbf{c}_2^*\mathbf{c}_2^{*'} + \sigma_x^2(1-\delta_x^2)\mathbf{c}_1^*\mathbf{c}_1^{*'}$. Likewise, Lemma 3, because of the above results, the complete log likelihood function associated with $(\boldsymbol{Z}_i, t_{x_i})$ in the model given in (10) and (19), can be written as

$$\ell_c(\boldsymbol{\theta}, \lambda_x) \propto -\frac{n}{2}\log\boldsymbol{\Psi}_{xb} - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Z}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_i-\boldsymbol{\mu}) - \frac{1}{2\tau_{xb}^2}\sum_{i=1}^{n}(t_{x_i}-\eta_{xb_i})^2,$$

where $\eta_{xb_i} = \frac{\mathbf{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}(\boldsymbol{Z}_i-\boldsymbol{\mu})}{1+\mathbf{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}\mathbf{c}_{xb}^*}$, $\tau_{xb}^2 = \frac{1}{1+\mathbf{c}_{xb}^{*'}\boldsymbol{\Psi}_{xb}^{-1}\mathbf{c}_{xb}^*}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Psi}_{xb} + \mathbf{c}_{xb}^*\mathbf{c}_{xb}^{*'}$. We obtain from Lemma 5 that

$$\hat{t}_{x_i} = E(t_{x_i}|\boldsymbol{Z}) = \eta_{xb_i} + \frac{\phi_1\left(\frac{\eta_{xb_i}}{\tau_{xb}}\right)}{\Phi_1\left(\frac{\eta_{xb_i}}{\tau_{xb}}\right)}\tau_{xb}, \tag{24}$$

and

$$\hat{t}_{x_i}^2 = E(t_{x_i}^2|\boldsymbol{Z}) = \eta_{xb}^2 + \tau_{xb}^2 + \frac{\phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)}{\Phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)}\eta_{xb}\tau_{xb}. \tag{25}$$

Two steps of the EM algorithm in this case proceed as before, with $(\boldsymbol{\theta}^{(0)}, \lambda_x^{(0)})$, $\hat{t}_{x_i}$ and $\hat{t}_{x_i}^2$ given in (24) and (25).

4. A simulation study

To assess the performance of the proposed model and methods in studying the linear mixed measurement error models, in which the distribution of $\{x, e, u, b\}$ follows the skew-normal distribution, we conducted a simulation study. For sample sizes $n = 30$, 50 and 100, we generated 1000 data sets according to the structural mixed measurement error model in (1) and (3), with $\alpha = 1, \beta = 2, \mu_x = 5, \lambda_x = 3, \lambda_e = \lambda_u = \lambda_b = 0$ and $\sigma_e = \sigma_u = \sigma_x = \sigma_b = 1$. The skew-normal representation of the x_i with skewness parameter 3 detected a departure from normality and suggested strong evidence of skewness. For each of the 1000 generated data sets of size $n = 100$, model 1 was fitted twice under the assumptions of previous section, with the density of x_i represented by skew-normal distribution and also by the normal distribution. For each generated sample, maximum likelihood estimators of all parameters were computed by using the EM algorithm described in Section 3. The mean value and sample standard deviation corresponding to each parameter for the 1000 generated data sets and each sample size are presented in Table 1. N.C. indicates percentages of samples with $\hat{\lambda}_x = \infty$. Following other authors (e.g. Zhang and Davidian [28]) we propose to evaluate a series of fits by inspection of information criteria such as Akaike's information criterion (AIC), Schawarz's Bayesian information criterion (BIC), and the Hannan–Quinn criterion (HQ). The AIC, BIC and HQ criteria given in Table 2 indicate that the skew-normal linear mixed measurement error model presents the better fit than the normal linear mixed measurement error model, supporting the contention of a departure from normality. In addition, the advantage of estimating the x_i density may be appreciated from Fig. 1. Fig. 1(a) shows the average of density estimates over the 1000 data sets of size $n = 100$ along with the true density, the normal fit and the fit for the skew-normal. Fig. 1(a) demonstrates that the additional flexibility afforded by the skew-normal representation is sufficient to capture quite accurately the true underlying features of the x_i.

5. An application

To illustrate the usefulness of the above procedures, we consider the likelihood analysis of the part of a set of real data, which is known as the Boston Housing data set. This data set was the basis for a paper given by Harrison and Rubinfeld [20], which discussed approaches for using housing market data to estimate the willingness to pay for clean air. Zhong et al. [29] considered this data set and used the data of $n = 132$ census tracts within the 15 districts of the Boston city. Census tracts within districts are taken as repeated measurements. The pollution variable NOXSQ is assumed to have measurement errors

Table 2
Results of fitting normal and skew-normal linear mixed measurement error model to the simulated data sets.

Information criteria	Normal scenario	Skew-normal scenario
− log -likelihood	226 231.3	193 211.6
AIC	2.262383	1.932196
BIC	2.262716	1.932577
HQ	2.262484	1.932311

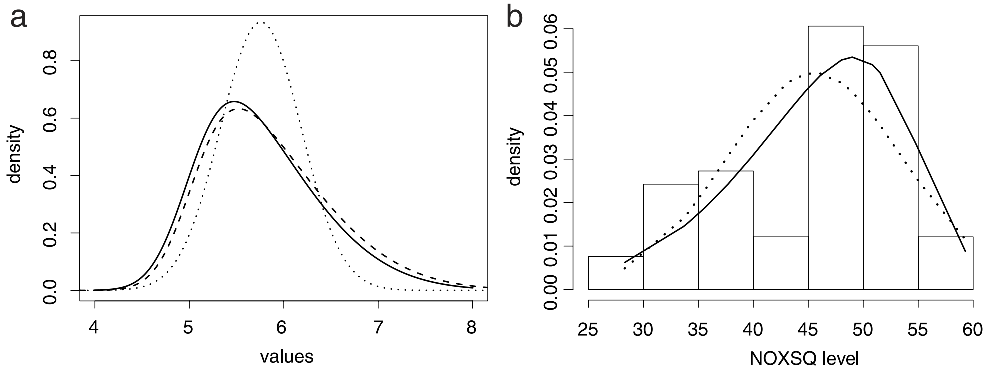

Fig. 1. (a) Simulation results based on 1000 data sets. True density (solid line) and average estimated densities: using normal (dotted line) and skew-normal (dashed–dotted). (b) Histogram for NOXSQ levels with estimated normal density (dotted) and estimated skew-normal density (solid).

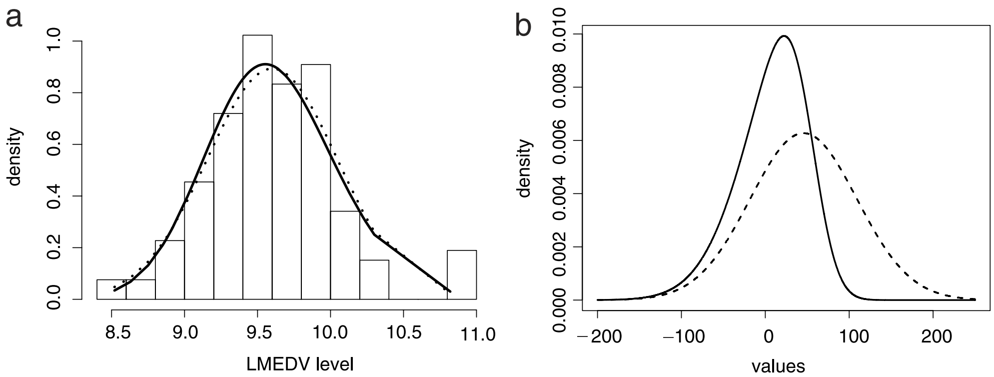

Fig. 2. (a) Histogram of LMEDV levels with estimated normal density (dotted) and estimated skew-normal density (solid). (b) Estimated marginal densities of noxsq concluded using model 3 (solid) with model1 (dotted).

according to the equation

$$NOXSQ_i = noxsq_i + u_i, \quad i = 1, \ldots, n.$$

Fig. 1(b) shows the histogram of the observed NOXSQ. Clearly, this figure indicates it is asymmetric, so that we consider $noxsq \sim SN(\mu_x, \sigma_x^2, \lambda_x)$. A simple plot of the histogram of the response variable (LMEDV) given in Fig. 2(a) indicates partially its asymmetric nature and it would be adequate to fit a skew-normal model to the data set. The asymmetric behavior of the LMEDV explains by the asymmetric behavior of the NOXSQ data and the random effect may not be normally distributed. Based on this information, three statistical models, differing in the NOXSQ data and random effect distributions are entertained. These models are:

MODEL1: A model with skew-normal distribution for NOXSQ data and symmetric normal distribution for the random effect and the errors.
MODEL2: A model with skew-normal distribution for the random effect and symmetric normal distributions for the NOXSQ data and the errors.
MODEL3: A purely Gaussian model.

Table 3 presents the results obtained by AIC, BIC and HQ criteria of three models described above. When considering MODEL 2, asymmetric not detected and the criteria are close to the ones obtained under normality (MODEL 3). The AIC, BIC and HQ criteria indicate that MODEL 1 presents the best fit, supporting the contention of a departure from normality. We present the parameter estimates obtained using the EM-type algorithm of three models described above in Table 3. SE are the estimated asymptotic standard errors based on the Hessian matrix computed numerically.

Table 3
Results of fitting models 1, 2 and 3 to the Boston city data.

Information criteria	Normal scenario	Skew-normal scenario	
Parameter	Model 1	Model 2	Model 3
α	10.2965	10.1545	10.2851
SE	0.0373	0.04147	0.0393
β	−0.0153	−0.0150	−0.0150
SE	0.0003	4.64e−05	4.59e−05
μ_x	54.3107	45.5833	45.5833
SE	0.2044	0.6973	0.6970
σ_b^2	0.0460	0.0659	0.04634
SE	0.0039	0.0363	0.0227
λ_x	−2.5242	0	0
SE	0.1532	0	0
λ_b	0	0.9105	0
SE	0	0.2940	0
− log-likelihood	316.658	351.4028	360.407
AIC	2.4318	2.7000	2.7606
BIC	2.4864	2.7546	2.8043
HQ	2.4540	2.7222	2.7784

Fig. 2(b) shows the estimated marginal densities of noxsq. It concluded using the models 1 and 3. This figure demonstrates the additional flexibility providing by the skew-normal assumption for noxsq.

6. Final conclusion

In this paper, we defined the simple linear structural mixed measurement error model with the distribution of the random quantities belonging to the family of the skew-normal distribution. An analytical expression (closed form) for the likelihood function of this model is derived by integrating out the unobserved x and the random effect b. The maximum likelihood can be implemented using standard optimization techniques and existing statistical softwares. We also developed an EM-type algorithm for evaluation of the MLE by exploring statistical properties of the considered model. A simulation study conducted to indicate that the methodology seems to work well when the normality assumption of the model does not hold. We point out that this paper is the first attempt in working on such general distributional structure for structural mixed measurement error models and that the approach considered in this paper can be extend to the situation where $y_i = \alpha + \boldsymbol{v}_i'\boldsymbol{\beta}_v + x_i\beta + b_i z_i + e_i$, with the additional fixed effect covariates $\boldsymbol{v}_i,\ i = 1, \ldots, n$.

Acknowledgments

We wish to thank the associate editor and the referees for their valuable comments and suggestions which enhanced the presentation of this paper.

Appendix

Proof of Lemma 1. We have

$$\phi_4(\boldsymbol{w}; \boldsymbol{c}\boldsymbol{\Gamma}, \boldsymbol{\Psi}) = |\boldsymbol{\Psi}|^{-\frac{1}{2}}(2\pi^{-2})e^{-\frac{1}{2}Q(\boldsymbol{w},\boldsymbol{\Gamma})},$$

with

$$Q(\boldsymbol{w}, \boldsymbol{\Gamma}) = (\boldsymbol{w} - \boldsymbol{c}\boldsymbol{\Gamma})'\boldsymbol{\Psi}^{-1}(\boldsymbol{w} - \boldsymbol{c}\boldsymbol{\Gamma}) = Q_1(\boldsymbol{w}|\boldsymbol{\Gamma}) + Q_2(\boldsymbol{w})$$

in which $Q_1(\boldsymbol{w}|\boldsymbol{\Gamma}) = (\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c})(\boldsymbol{\Gamma} - \frac{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{w}}{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}})^2$ and

$$\begin{aligned} Q_2(\boldsymbol{w}) &= (\boldsymbol{w}'\boldsymbol{\Psi}^{-1}\boldsymbol{w}) - \frac{\boldsymbol{w}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{w}}{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}} \\ &= (\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Psi}_1^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) - \frac{(\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Psi}_1^{-1}\boldsymbol{c}\boldsymbol{c}'\boldsymbol{\Psi}_1^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})}{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}} \\ &= (\boldsymbol{Z} - \boldsymbol{\mu})'\left[\boldsymbol{\Psi}_1^{-1} - \frac{\boldsymbol{\Psi}_1^{-1}\boldsymbol{c}_1\boldsymbol{c}_1{}'\boldsymbol{\Psi}_1^{-1}}{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}}\right](\boldsymbol{Z} - \boldsymbol{\mu}) \\ &= (\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}). \end{aligned}$$

Thus, the proof follows by noting that $\boldsymbol{\Sigma}^{-1} = \left[\boldsymbol{\Psi}_1^{-1} - \frac{\boldsymbol{\Psi}_1^{-1}\boldsymbol{c}_1\boldsymbol{c}_1{}'\boldsymbol{\Psi}_1^{-1}}{\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}}\right]$ and $|\boldsymbol{\Psi}| = \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{c}'\boldsymbol{\Psi}^{-1}\boldsymbol{c}|}$. $\square$

Proof of Theorem 1. Using (6) and (7), we can write the integral given in (5) as:

$$\begin{aligned} f(y, X) &= 2^4|\mathbf{\Psi}|^{-\frac{1}{2}} \iint \phi_4(\mathbf{w}; \mathbf{c}\mathbf{\Gamma}, \mathbf{\Psi})\Phi_4(\mathbf{\Lambda w} - \mathbf{\Lambda c \Gamma}; \mathbf{0}, \mathbf{\Psi}) d\nu db \\ &= 2^4|\mathbf{\Psi}|^{-\frac{1}{2}} \iint \phi_2(\mathbf{Z}; \boldsymbol{\mu}, \mathbf{\Sigma})\phi_2(\mathbf{\Gamma}; (\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{w}, (\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1})\Phi_4(\mathbf{\Lambda w} - \mathbf{\Lambda c \Gamma}; \mathbf{0}, \mathbf{\Psi}) d\mathbf{\Gamma} \\ &= 2^4|\mathbf{\Psi}|^{-\frac{1}{2}}\phi_2(\mathbf{Z}; \boldsymbol{\mu}, \mathbf{\Sigma}) E(\Phi_4(\mathbf{\Lambda w} - \mathbf{\Lambda c \Gamma}; \mathbf{0}, \mathbf{\Psi})), \end{aligned}$$

in which $\mathbf{\Gamma} \sim N_2((\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{w}, (\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1})$ and from Lemma 2, with $k = 2, m = 4, \mathbf{A} = -\mathbf{\Lambda c}, \boldsymbol{\xi} = (\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}$ $\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{w}, \boldsymbol{\kappa} = \mathbf{0}, \mathbf{a} = \mathbf{\Lambda w}, \mathbf{\Delta} = (\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}, \mathbf{V} = \mathbf{\Gamma}$, and $\mathbf{H} = \mathbf{\Psi}$, we have

$$\begin{aligned} E(\Phi_4(\mathbf{\Lambda w} - \mathbf{\Lambda c \Gamma}; \mathbf{0}, \mathbf{\Psi})) &= \Phi_4(\mathbf{\Lambda w}; \mathbf{\Lambda c}(\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{w}, \mathbf{\Psi} + \mathbf{\Lambda c}(\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\mathbf{\Lambda}) \\ &= \Phi_4(\mathbf{\Lambda M w}; \mathbf{0}, \mathbf{\Psi} + \mathbf{\Lambda c}(\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'\mathbf{\Lambda}), \end{aligned}$$

where $\mathbf{M} = \mathbf{I}_4 - \mathbf{P}$, with $\mathbf{P} = \mathbf{c}(\mathbf{c}'\mathbf{\Psi}^{-1}\mathbf{c})^{-1}\mathbf{c}'$. Note that $\mathbf{Pc} = \mathbf{c}$, $\mathbf{Mc} = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{B\Psi B}'$. We can show that $\mathbf{Mw} = \mathbf{\Psi B}'\mathbf{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu})$. □

Proof of Lemma 5. We can write $E(t_b^k|\mathbf{Z})$ and $\frac{\eta_{xb}}{\tau_{xb}}$ as

$$E(t_b^k|\mathbf{Z}) = \frac{\int t_b^k f(\mathbf{Z}, t_b) dt_b}{f(\mathbf{Z})}, \tag{A.1}$$

$$\frac{\eta_{xb}}{\tau_{xb}} = \nu_b\lambda_b\sigma_b\mathbf{c}_2^{*'}\mathbf{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \tag{A.2}$$

where from (A.2), the marginal density of $\mathbf{Z}$ and the density function of $(\mathbf{Z}', t_b)$, given in (9) and (16), respectively, can be written as

$$\begin{aligned} f(\mathbf{Z}; \boldsymbol{\theta}, \lambda_b) &= 2\phi_2(\mathbf{Z}; \boldsymbol{\mu}, \mathbf{\Sigma})\Phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right), \\ f(\mathbf{Z}, t_b) &= 2\phi_2(Z; \boldsymbol{\mu}, \mathbf{\Sigma})\phi_1(t_b; \eta_{xb}, \tau_{xb}^2)\mathbf{I}_{[t_b>0]}. \end{aligned}$$

Replacing the above functions in (A.1), it follows that

$$E(t_b^k|\mathbf{Z}) = \frac{\int_0^\infty t_b^k\phi_1(t_b; \eta_{xb}, \tau_{xb}^2) dt_b}{\Phi_1\left(\frac{\eta_{xb}}{\tau_{xb}}\right)} = E(X^k|X > 0),$$

where $X \sim N_1(\eta_{xb}, \tau_{xb}^2)$ and $\Phi_1(\frac{\eta_{xb}}{\tau_{xb}}) = P(X > 0)$. By using of Lemma 4, the proof of Lemma 5 is completed. □

References

[1] R.B. Arellano-Valle, H. Bolfarine, V.H. Lachos, Skew-normal linear mixed models, J. Data Sci. 3 (2005) 415–438.
[2] R.B. Arellano-Valle, M.G. Genton, Fundamental skew distributions, J. Multivariate Anal. 96 (2005) 93–116.
[3] R.B. Arellano-Valle, S. Ozan, H. Bolfarine, V.H. Lachos, Skew normal measurement error models, J. Multivariate Anal. 96 (2) (2005) 265–281.
[4] B. Armstrong, Measurement error in the generalized linear model, Comm. Statist. Simulation Comput. 14 (3) (1985) 529–544.
[5] B.C. Arnold, R.J. Beaver, Skewed multivariate models related to hidden truncation and/or selective reporting, TEST 11 (2002) 7–54.
[6] A. Azzalini, A class of distribution which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.
[7] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 46 (1986) 199–208.
[8] A. Azzalini, The skew normal distribution and related multivariate families, Scand. J. Statist. 32 (2005) 159–188.
[9] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3) (1999) 579–602.
[10] A. Azzalini, A. Dalla Valla, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[11] H. Bolfarine, L.C. Montenegro, V.H. Lachos, Influence diagnostics for skew-normal linear mixed models, Sankhyā 69 (4) (2007) 648–670.
[12] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models: A Modern Perspective. Vol. 105, Chapman and Hall/CRC, 2006.
[13] C.L. Cheng, J.W. Van Ness, Statistical Regression with Measurement Error. Vol. 6, Arnold, London, 1999.
[14] H. Cui, K.W. Ng, L. Zhu, Estimation in mixed effects model with errors in variables, J. Multivariate Anal. 91 (1) (2004) 53–73.
[15] M. Davidian, D.M. Giltinan, Some general estimation methods for nonlinear mixed-effects model, J. Biopharm. Statist. 3 (1) (1993) 23–55.
[16] M. Davidian, D.M. Giltinan, Nonlinear Models for Repeated Measurement Data, Chapman and Hall/CRC Press, New York, 1995.
[17] E. Demidenko, T.A. Stukel, Efficient estimation of general linear mixed effects models, J. Statist. Plann. Inference 104 (1) (2002) 197–219.
[18] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–38.
[19] W. Fuller, Measurement Error Models, Wiley, New York, 1987.
[20] D. Harrison, D.L. Rubinfeld, Hedonic housing prices and the demand for clean air, J. Environ. Econ. Manag. 5 (1) (1978) 81–102.
[21] N. Henze, A probabilistic representation of the skew-normal distribution, Scand. J. Statist. 13 (4) (1986) 271–275.
[22] N. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, Wiley, New York, 1994.
[23] P. Karcher, Y. Wang, Generalized nonparametric mixed effects models, J. Comput. Graph. Statist. 10 (4) (2001) 641–655.
[24] T. Nakamura, Corrected score function for errors-in-variablesmodels: methodology and application to generalized linear models, Biometrika 77 (1990) 127–137.
[25] L.A. Stefanski, R.J. Carroll, Deconvolution-based score tests in measurement error models, Ann. Statist. 19 (1) (1991) 249–259.

[26] E.F. Vonesh, H. Wang, L. Nie, D. Majumdar, Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models, J. Amer. Statist. Assoc. 97 (457) (2002) 271–283.
[27] K. Zare, A. Rasekh, A.A. Rasekhi, Estimation of variance components in linear mixed measurement error models, Statist. Papers 53 (4) (2012) 849–863.
[28] D. Zhang, M. Davidian, Linear mixed models with flexible distributions of random effects for longitudinal data, Biometrics 57 (2001) 795–802.
[29] X.P. Zhong, W.K. Fung, B.C. Wei, Estimation in linear models with random effects and errors-in-variables, Ann. Inst. Statist. Math. 54 (3) (2002) 595–606.