Bayesian Analysis (2007) **2**, Number 2, pp. 349–364

Bayesian inference for an extended simple regression measurement error model using skewed priors

Josemar Rodrigues* and Heleno Bolfarine†

Abstract. In this paper, we introduce a Bayesian extended regression model with two-stage priors when the covariate is positive and measured with error. Connections are made with some results in Arellano-Valle and Azzalini (2006), related to the multivariate skew-normal distributions. The usefulness of the proposed model with errors in variables, via the two-stage priors formulated by O'Hagan and Leonard (1976), is illustrated with an example abstracted from Fuller (1987, pg. 18). The main advantage of this extended Bayesian approach is the use of skewed priors, typically rare in most Bayesian applications, and to treat the true value of the explanatory variable as positive, consideration that is sometimes ignored in measurement error models. Such consideration makes naturally the model identifiable, a problem that significantly has troubled users of other approaches listed in the literature. This constraint implies also a strong asymmetry in the distribution of the response variable. Strong connections are shown with results in Copas and Li (1997) on non-random samples and with Berkson models, which are important in practical applications. Extensions of Copas and Li's results for models with vector explanatory variables are presented.

Keywords: Berkson model, non-informative prior, non-random sample, posterior distribution, pseudo-Bayes factor, regression calibration, structural error model, skew-normal distributions, Winbugs.

1 Introduction

The simple linear regression model has been subject to extensive research in the literature for over a century. In many of its real applications, mainly where the observations contain measurement errors, the explanatory variable is positive implying a strong asymmetry on the response variable which is usually ignored in the formulation of the regression models. However, we have no reports of its use in the measurement error literature. Using the unified approach of families of skew-normal distributions, introduced by Arellano-Valle and Azzalini (2006), we here formulate an extension of the simple regression model called the conditional extended simple regression model and study the impact caused by this restriction on the inference of the regression coefficients. Also, a stochastic representation of this extended distribution is obtained, which is useful for Bayesian computation purposes via Winbugs. The Bayesian results obtained in

*Universidade Federal de São Carlos-Departamento de Estatistica, Brasil, mailto:vjosemar@power.ufscar.br

†Universidade Federal de São Carlos-Departamento de Estatistica, Brasil, mailto:hbolfar@ime.usp.br

this paper when contrasted with results in Fuller (1987), show good performance for the estimation of the regression coefficient and a large degree of non-randomness in Copas and Li (1997) sense. The main advantage of the Bayesian approach is the possibility to include prior constraints on the parameters and permitting the data confirm it is true or not, or, asymptotic results, as considered in Fuller (1987) example. This kind of more complex prior consideration includes some degree of skewness and it seems to be not usually used by Bayesian statisticians, as mentioned and justified by Liseo (2004). Moreover, the likelihood function for the measurement error models with skew distributions usually is very complicated making the model introduced in this paper an interesting and attractive alternative. In our opinion, combining the degree of skewness (O'Hagan and Leonard 1976) in the prior with the degree of non-randomness in the sample (Copas and Li 1997) is a simple and natural way to solve practical problems.

The paper is organized as follows. In Section 2, we formulate the regression normal measurement error model (Fuller 1987; Carrol et al. 2006; Cheng and Ness 1999) and making use of some properties of the bivariate normal distribution, the conditional simple regression normal measurement error model is justified. In Section 3, the conditional extended regression model with errors on the real covariate value is formulated and some results are presented. The multivariate case is introduced in Section 4. Furthermore, in Section 5, an illustrative example analyzed via Winbugs is considered which shows that the approach proposed in the paper very much improves on the classical approach.

2 A simple regression model with normal measurement error

Regression with error-in-variables, also known as measurement error models, is fundamentally different from the usual simple linear regression model and the inferential problems that arise with this model are quite different and complex. The measurement error model can be thought of as a generalization of the traditional simple regression model. In this section, we consider the measurement error model relating the response variable Y_i and the explanatory variable x_i (see a comprehensive overview of this model in Fuller (1987) and, more recently, in Cheng and Ness (1999), and in Carrol et al. (2006)) as follows:

$$Y_i = \alpha + \beta x_i + e_i, \quad i = 1, \ldots, n. \tag{1}$$

Assuming that x_i is the real unobserved value of the observed variable X_i, we consider the following measurement error model:

$$X_i = x_i + u_i, \tag{2}$$

where

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N_3 \left[\begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; diag(\sigma_e^2, \sigma_u^2, \sigma_x^2) \right]$$

are independent random vectors, $i = 1, \ldots, n$ and $diag(a_1, \ldots, a_n)$ denotes a diagonal matrix with diagonal elements $a_1, \ldots, a_n$. The equations (1)-(2) define the structural measurement error model. When the unobserved value x_i is fixed, this model is known in the literature as the functional model. It is well known that the above structural model is non-identifiable, so, it is usual to make additional assumptions (such as σ_u^2 is known) to solve inferential problems.

Using well known properties of the multivariate normal distribution we can show that

$$(x_i, Y_i, X_i)^t \sim N_3(\mu; \Omega), \tag{3}$$

where,

$$\mu = \begin{pmatrix} \mu_x \\ \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \sigma_x^2 & \sigma_x^2\beta & \sigma_x^2 \\ \sigma_x^2\beta & \sigma_x^2\beta^2 + \sigma_e^2 & \sigma_x^2\beta \\ \sigma_x^2 & \sigma_x^2\beta & \sigma_x^2 + \sigma_u^2 \end{pmatrix},$$

implying that

$$(x_i, Y_i)^t | X_i \sim N_2(\mu_{X_i}, \Omega_{X_i}),$$

where

$$\mu_{X_i} = \begin{pmatrix} (1-k)\mu_x + kX_i \\ \alpha + (1-k)\beta\mu_x + k\beta X_i \end{pmatrix}, \Omega_{X_i} = \begin{pmatrix} k\sigma_u^2 & k\sigma_u^2\beta \\ k\sigma_u^2\beta & k\sigma_u^2\beta^2 + \sigma_e^2 \end{pmatrix}$$

and $k = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$.

It follows that $Y_i \mid X_i$ is normally distributed with conditional expectation and conditional variance given by

$$\begin{aligned} E(Y_i \mid X_i) &= \alpha + (1-k)\beta\mu_x + k\beta X_i, \\ V(Y_i \mid X_i) &= \sigma^2 = \sigma_e^2 + k\sigma_u^2\beta^2, \end{aligned} \tag{4}$$

$i = 1, \ldots, n$, respectively.

Hence, the conditional expectation in (4) and (1) suggest the following simple conditional regression model

$$Y_i = \alpha' + \gamma X_i + v_i$$

where

$$\begin{array}{rcl} \gamma & = & k\beta, \\ \alpha' & = & \alpha + (\beta - \gamma)\mu_x, \\ v_i & = & \overbrace{(\beta-\gamma)(X_i-\mu_x)}^{\text{global measurement error}} - \overbrace{\beta u_i}^{\text{local measurement error}} + \overbrace{e_i}^{\text{fitting}}, \\ v_i & | & X_i \sim N\left(0, \sigma^2\right), \end{array} \tag{5}$$

$i = 1, \ldots n$.

Notice that the variance of v_i can be much larger than the variance of e_i, the regression coefficient β has been attenuated by the factor k and v_i is an error variable in the usual sense. The regression coefficient γ, may be estimated by the usual least squares estimator, $\widehat{\gamma}$, which is biased for β, that is, $E(\widehat{\gamma}) = k\beta$. The ratio k is called the reliability of X_i in the social sciences and heritability in genetics. Typically, the ratio k is supposed to be known in order to avoid identifiability problems. For more details of the above model and real non-Bayesian applications we refer the reader to Fuller (1987), Carrol et al. (2006) and Cheng and Ness (1999). It is worth mentioning here that the variable ν_i was considered by DeGracie and Fuller (1972) in their studies in analysis of covariance but without the term $-\beta u_i$, which, in our opinion, seems to be incorrectly missing.

The random variable $v_i' = -\beta u_i + e_i$ in (5) is central in the study of measurement error models and differs from the error in the ordinary fixed-X regression model where X_i and v_i' are correlated. In the fixed-X regression model introduced in (5), variable v_i' is independent of X_i and $v_i = (\beta - \gamma)(X_i - \mu_x) + v_i'$ is the normal error variable for this conditional regression model. The main purpose of this paper is to introduce asymmetry in this normal error variable model by assuming the following constraint on the real value x_i:

$$x_i \geq 0 \Leftrightarrow -u_i \geq -X_i, i = 1, \ldots n, \tag{6}$$

which, as shown, makes the model (1)-(2) identifiable without additional assumptions. Formally, we do have to answer the following question: what is the distribution of v_i given that $-u_i \geq -X_i$? Using results in Arellano-Valle and Azzalini (2006) the answer will be given in the next section.

3 Extended simple regression measurement error model

In this section, we present a new measurement error model which takes into account the asymmetry of the variable error v_i by considering the restriction $x_i \geq 0$. This model will be called here the extended simple regression model with measurement errors and makes use of some connections with results in Arellano-Valle and Azzalini (2006).

We present next the main result of this section, which basically follows from result (10) in Arellano-Valle and Azzalini (2006) (and also appears in Arnold and Beaver (2000)), according to the notation used in (3).

Theorem 1: The conditional probability density function of v_i given X_i for $i = 1, \ldots n$, under (5) and the constraint $x_i \geq 0$, is given by

$$f(v_i) = \frac{1}{\sigma}\varphi\left(\frac{v_i}{\sigma}\right)\frac{\Phi\left(\frac{\sigma}{\sigma_e}\left(\frac{kX_i+(1-k)\mu_x}{k^{\frac{1}{2}}\sigma_u}\right)+\lambda\frac{v_i}{\sigma}\right)}{\Phi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}, \tag{7}$$

where $\lambda = \gamma\frac{\sigma_u k^{-\frac{1}{2}}}{\sigma_e}$, or, equivalently,

$$f(y_i) = \frac{1}{\sigma}\varphi\left(\frac{y_i-\alpha'-\gamma X_i}{\sigma}\right)\frac{\Phi\left(\frac{\sigma}{\sigma_e}\left(\frac{kX_i+(1-k)\mu_x}{k^{\frac{1}{2}}\sigma_u}\right)+\lambda\frac{y_i-\alpha'-\gamma X_i}{\sigma}\right)}{\Phi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)},$$

where $\varphi(\cdot)$ and $\Phi[\cdot]$ are the density function and the cumulative distribution function of the one-dimensional $N(0,1)$ distribution, respectively.

The class of distributions defined by (7) extends the normal class by introducing the parameter $\lambda = \gamma\frac{\sigma_u k^{-\frac{1}{2}}}{\sigma_e}$ which regulates distributional shape (or the degree of non-randomness in the sample, according to the terminology in Copas and Li (1997)); when σ_u^2 tends to zero the normal distribution with mean $\alpha + \beta X$ and variance σ_e^2 is the limiting case of (7), the classical simple regression model. It is important to note that the parameters k, σ_u, and σ_e and the sign of γ, simultaneously, regulate the asymmetry of the error variable ν_i. Moreover, this relationship avoids non-identifiability problems typically associated with measurement error models as described in Fuller (1987). Clearly, the residual (error variable) v_i has non zero (conditional) expectation since it is distributed according to the skew-normal distribution. In fact, using results in Capitanio et al. (2003) it can be shown that

$$E(v_i \mid X_i) = \frac{\varphi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}{\Phi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}\gamma k^{-\frac{1}{2}}\sigma_u \neq 0, \tag{8}$$

$$V(\nu_i \mid X_i) = \overbrace{\sigma_e^2}^{\text{fitting}} + \overbrace{\frac{\gamma^2\sigma_u^2}{k}}^{\text{measurement error}}\left[1+\overbrace{\xi_2\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}^{\text{skewness}}\right],$$

$i = 1, \ldots n$, where $\xi_2(\cdot)$ is the second derivative of $\xi_o(t) = \log[2\Phi(t)]$.

Its important to note that the sign of this residual depends on the sign of γ which in some applications like the one studied in this paper is known. A regression model

with a zero mean error variable is then easily formulated as

$$\begin{aligned} Y_i &= \alpha' + \gamma X_i^* + e_i^*, \\ X_i^* &= X_i + \xi_1\left(\frac{kX_i + (1-k)\,\mu_x}{\sigma_u k^{\frac{1}{2}}}\right) k^{-\frac{1}{2}}\sigma_u, \\ e_i^* &= v_i - \frac{\varphi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}{\Phi\left(\frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)}\gamma k^{-\frac{1}{2}}\sigma_u = v_i - \xi_1\left(\frac{kX_i + (1-k)\,\mu_x}{\sigma_u k^{\frac{1}{2}}}\right)\gamma k^{-\frac{1}{2}}\sigma_u, \end{aligned}$$

where $\xi_1[\cdot]$ is the first derivative of $\xi_o[\cdot]$.

For Bayesian inference purposes, the extended skew-normal distribution (7) is very hard to deal with directly and, hence, the next stochastic representation of a variable with density (7) will be very useful in implementing the Bayesian way in real applications, like the one to be considered in the Section 5. The proof follows directly from results in Arellano-Valle and Azzalini (2006).

Theorem 2: If the variable Y_i has probability density function given by (7), then

$$Y_i = \alpha' + \gamma\left(X_i - k^{-\frac{1}{2}}\sigma_u u_i\right) + \sigma_e u_1^*, \tag{9}$$

where

$$u_i \left| X_i, k, \mu_x, \sigma_u^2 \right. \sim N(0,1)\, 1_{\left[u_i \leq \frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right]},$$

with $u_1^* \sim N(0,1)$ and u_i and u_1^* independent random variables, $i = 1, \ldots, n$.

Hence, we have from (9) the following hierarchical representation for the extended regression model which will be used in the next section from a Bayesian perspective and will be called Berkson model:

$$Y_i \mid X_i, u_i, \theta^* \sim N\left\{\alpha' + \gamma\left(X_i - \sigma_u k^{-\frac{1}{2}} u_i\right), \sigma_e^2\right\}, \tag{10}$$

$$u_i \left| X_i, k, \mu_x, \sigma_u^2 \right. \sim N(0,1)\, 1_{\left[u_i \leq \frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}\right]},$$

and $\theta^* = \left(\alpha', \gamma, k, \mu_x, \sigma_u^2, \sigma_e^2\right)$, $i = 1, \ldots n$.

It is important to observe in (10) the intuitive selection mechanism, $u_i \leq \frac{kX_i+(1-k)\mu_x}{\sigma_u k^{\frac{1}{2}}}$, arising when assuming that the true value x_i is greater than zero. Clearly, ignoring it could cause wrong predictions if we have a large attenuation factor.

In this paper, we adopt a Bayesian strategy considering a two-stage prior on τ (a generic symbol for the parameters α, β and μ_x involved in (10)) suggested by O'Hagan and Leonard (1976), for which

$$\begin{cases} \tau \sim N\left(\tau_1, \sigma_\tau^2\right), \\ \tau_1 \sim N\left(\tau_o, \sigma_{\tau_o}^2\right), \\ \tau_1 \geq \tau_o. \end{cases} \tag{11}$$

Standard algebra shows that the marginal prior distribution for τ is then given by

$$\varphi\left(\tau; \lambda^*\right) = \frac{2}{\left(\sigma_\tau^2 + \sigma_{\tau_0}^2\right)^{\frac{1}{2}}} \varphi\left(\left(\sigma_\tau^2 + \sigma_{\tau_0}^2\right)^{-\frac{1}{2}} \left(\tau - \tau_0\right)\right) \Phi\left(\lambda^* \frac{\tau - \tau_o}{\left(\sigma_\tau^2 + \sigma_{\tau_0}^2\right)^{\frac{1}{2}}}\right), \tag{12}$$

where $\lambda^* = \frac{\sigma_{\tau_o}}{\sigma_\tau}$.

Notice that λ^* is a shape parameter which controls prior skewness on τ : the larger $\sigma_{\tau_0}^2$, the variance of τ_1, is compared to σ_τ^2, the variance of τ, the larger is the probability that τ_1 (mean of τ) assumes larger values and larger is the amount of skewness induced in the distribution of τ. Many generalizations of this result are possible (see Liseo and Loperfido 2006).

4 Linear measurement errror model with vector-valued explanatory variables

This section presents an extension of the model considered in the previous section with vector-valued explanatory variables, that is, with more than one explanatory variable. This is an extension of the results in the previous sections with the purpose to obtain Bayesian estimation of the parameters involved. The more general model we consider is written as

$$\begin{aligned} Y_i &= \alpha + x_i^t \beta + e_i, \\ X_i^t &= x_i^t + u_i^t, \; i = 1, \ldots, n, \end{aligned} \tag{13}$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^t, \beta, X_i$ and u_i are $p-$dimensional vectors, while α, Y_i and e_i are scalars. Hence, the observed vector is (X_i^t, Y_i), the unobserved vector is x_i, and the measurement error vector is (u_i^t, e_i), for $i = 1, \ldots, n$.

Assume as in the univariate case that (x_i^t, u_i^t, e_i) is jointly normally distributed and $\Sigma_{ue} = 0$. Let $\mu_x = E\left(x_i\right) = E\left(X_i\right), I_p$ be the $p \times p$ identity matrix, and the reliability matrix K be defined as

$$K = \Sigma_{xx} \left(\Sigma_{xx} + \Sigma_{uu}\right)^{-1}.$$

The extension of the reliability ratio to vector explanatory variables models is considered in Gleser (1992), but this extension can, in fact, be for models with both X and Y vector-valued allowing a connection between ordinary multivariate regression models and multivariate measurement error models (Cheng and Tsai 1996). Since a structural measurement error model is being considered, then

$$\begin{pmatrix} x_i \\ Y_i \\ X_i \end{pmatrix} \sim N_{2p+1}(\mu, \Omega)$$

where

$$\mu = \begin{pmatrix} \mu_x \\ \alpha + \mu_x^t \beta \\ \mu_x \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}\beta & \Sigma_{xx} \\ \beta^t \Sigma_{xx} & \beta^t \Sigma_{xx}\beta + \sigma_e^2 & \beta^t \Sigma_{xx} \\ \Sigma_{xx} & \Sigma_{xx}\beta & \Sigma_{xx} + \Sigma_{uu} \end{pmatrix}, \tag{14}$$

implying that

$$\begin{pmatrix} x_i \\ Y_i \end{pmatrix} | X_i \sim N_{p+1}\left[\begin{pmatrix} (I-K)\mu_x + KX_i \\ \alpha + (\beta^t - \gamma^t)\mu_x + \gamma^t X_i \end{pmatrix}, \begin{pmatrix} \Sigma_{uu}K^t & \Sigma_{uu}K^t\beta \\ \beta^t K \Sigma_{uu} & \beta^t K \Sigma_{uu}\beta + \sigma_e^2 \end{pmatrix} \right].$$

So, given X_i, we can rewrite (13) as

$$Y_i = \alpha' + \gamma^t X_i + \nu_{1i}, \tag{15}$$

where

$$\begin{aligned} \alpha' &= \alpha + (\beta^t - \gamma^t)\mu_x, \\ \gamma^t &= \beta^t K, \\ \nu_{1i} &= e_i - u_i^t \beta + \left(X_i^t - \mu_x^t\right)(\beta - \gamma), \end{aligned}$$

$i = 1, \ldots, n$. Notice that, for fixed X_i, the error variable $\nu_{1i} \sim N\left(0, \sigma^2\right)$ with

$$\sigma^2 = \sigma_e^2 + \beta^t K \Sigma_{uu} \beta. \tag{16}$$

If K is known, then Gleser (1992) showed that the ML estimators of α, γ, and σ^2 are just the ordinary least-squares estimators. The above results are also applicable when Σ_{uu} is known and K is unknown (see Cheng and Ness 1999, p. 139). Furthermore, for fixed X_i we can rewrite (15) in order to obtain a multivariate version of the the model in Copas and Li (1997) as

$$\begin{aligned} Y_i &= \alpha' + \gamma^t X_i + \nu_{1i} \\ x_i &= (I-K)\mu_x + KX_i + u_{1i}, \end{aligned} \tag{17}$$

$u_{1i} \sim N(0, \Sigma_{uu}K^t)$, $i = 1, \ldots, n$, where the second equation in (17) is called "selection equation" , where K measures the size of the attenuation in linear regression with additive measurement errors (Fuller 1987). There is another way of looking at the model in (17) which is connected with the Berkson model or the regression calibration described in detail in Chapter 4 of Carrol et al. (2006). Define $X_{blp,i} = (I - K)\,\mu_x + KX_i$, the best linear predictor of x_i given X_i. Then, from (17) we have

$$\begin{aligned} Y_i &= \alpha + \beta X_{blp,i} + \nu_{1i} \\ x_i &= X_{blp,i} + u_{1i}. \end{aligned}$$

If K and μ_x are known, we have a regression model with intercept α, slope β, error ν_{1i} uncorrelated with the regressor and u_{1i}, the classical measurement error, can be interpreted as the Berkson error. Observe that for $K = I$, we have the usual Berkson error, so K measures how much we just pretend that the measurement error u_{1i} is Berkson and not classical. Note that the slope of X_i in $X_{blp,i}$ is K , not I. In practice, $X_{blp,i}$ is not known, the parameters K and μ_x need to be estimated and this is the heart of the regression calibration method. Based on these comments the model in (17) will be called Berkson model. We have now an important question: Given the observed data (X_i^t, Y_i), how the inference procedure is modified by the componentwise restriction $x_i \geq 0$, or, $u_{1i} \leq X_i$ (X_i : fixed) ? What is the degree of non-randomness in the sample in the sense of Copas and Li (1997)? From (14), the answers to these questions are diretly obtained from Arellano-Valle and Azzalini (2006) (or G. Gonzalez-Farias and Gupta (2004)) as

$$f_{Y_i|X_i, x_i>0}\,(y_i) = \frac{1}{\sigma}\varphi\left(\frac{y_i - \alpha' - \gamma^t X_i}{\sigma}\right) G(y_i),$$

where

$$G(y_i) = \frac{\Phi_p\left\{(\Sigma_{uu}K^t)^{-\frac{1}{2}}[KX_i + (I-K)\,\mu_x] + \frac{\rho}{\sigma}\,(y_i - \alpha' - \gamma^t X_i)\,; 0, I - \rho\rho^t\right\}}{\Phi_p\left\{(\Sigma_{uu}K^t)^{-\frac{1}{2}}[KX_i + (I-K)\,\mu_x]; 0, I\right\}}, \quad (18)$$

where

$$\rho = \frac{(\Sigma_{uu}K^t)^{\frac{1}{2}}\,\beta}{\sigma},$$

and $\Phi_p\,(a, \mu, \Sigma)$ represents the p-dimensional normal distribution function with mean μ and covariance matrix Σ at the point a. Moreover, the mean, variance and the stochastic representation of (18) can be obtained from Arellano-Valle and Azzalini (2006), or, G. Gonzalez-Farias and Gupta (2004). The vector which regulates skewness, or, reflects the degree of non-randomness in the sample is given by

$$\theta = \frac{\rho}{(1 - \rho^t\rho)^{\frac{1}{2}}} = \frac{\Sigma_{uu}^{\frac{1}{2}}(K^t)^{-\frac{1}{2}}\gamma}{\sigma_e}. \quad (19)$$

For $p = 1$, the parameter θ is equal to the parameter λ in (7). Copas and Li (1997) examined the profile log-likelihood for θ and studied sensitivity of inference to small non-zero values of θ. They used Heckman's two-stage method (Heckman 1976, 1979), which is very popular in econometrics but seems to suffer from several deficiencies as pointed out by Little (1985) and others. So, as suggested by Copas and Li (1997), we carry on in the next section a Bayesian analysis of the model in (18) or (7) with the two-stage priors in (11) and (12). The special case $\theta = 0$ is the normal case or the randomness hypothesis which can be tested using (robust) Bayes factors, as illustrated in the application. To the of best of our knowledge, nobody has ever carried out a Bayesian analysis for this model (Copas and Li 1997, pg. 75).

5 An illustrative example

In this section we briefly illustrate the behavior of the Bayesian procedure for a small data set presented in Example 1.2.1, in the book by Fuller (1987, pg. 18), assuming noninformative and informative priors for the parameters involved in the conditional regression model. The data given in Table 1 are yields of corn and determinations of available soil nitrogen collected at 11 sites on Marshall soil in Iowa. The estimates of soil nitrogen contain measurement error arising from two sources. Firstly, only a small sample of soil is selected from each plot and, as a result, there is the sampling error associated with the use of a sample to represent the whole. Secondly, there is measurement error associated with the chemical analysis used to determine the level of nitrogen in the soil sample. In Fuller (1987), the measurement error variance is supposed to be known, that is, $\sigma_u^2 = 57$.

site	yield (Y)	soil nitrogen (X)	site	yield (Y)	soil nitrogen (X)
1	86	70	7	99	50
2	115	97	8	96	70
3	90	53	9	99	94
4	86	64	10	104	69
5	110	95	11	96	51
6	91	64			

Table 1: Yields of corn on Marshall soil in Iowa.

For the Bayesian analysis of our example implemented via Winbugs and to be compared with the results in Fuller (1987, pg. 19), we consider the following specific priors:

1. $\frac{1}{\sigma_u^2} \sim Gamma\left[\frac{57^2}{0.01} + 2, \frac{57^3}{0.01} + 57\right]$ (informative prior based on the prior information given in Fuller (1987), that is, an inverted gamma prior with mean equal to 57 and variance equal to 0.01.)

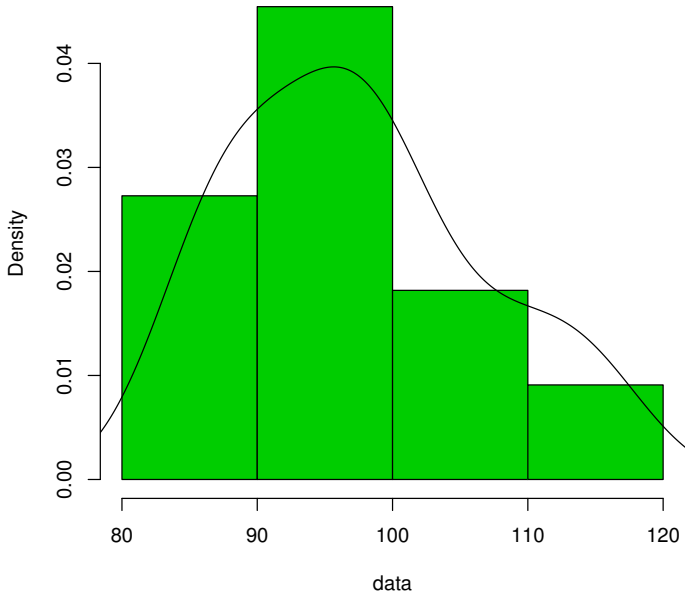

Figure 1: Adjusted check yields of corn at Iowa sites.

2. $\frac{1}{\sigma_e^2} \sim Gamma\,[0.001, 0.001]$

3. $\alpha \sim N(\alpha', 1)$ and $\alpha' \sim N(\overline{X}, 100) I_{(\alpha' \geq 0)}$ where $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ (Two-stage prior).

4. $\beta \sim N\,(\beta_o, 0.01)$ and $\beta_o \sim N\,(0, 100)\, I_{(\beta_o > 0)}$ (Two-stage prior).

5. $k \sim U(0, 1)$

6. $\mu_x \sim N\,(\mu_{xo}, 1)$ and $\mu_{xo} \sim N(\overline{X}, 100) I_{(\mu_{xo} > 0)}$ (Extended two-stage prior).

Notice that for the parameters α, β and μ_x, we use the hierarchical prior proposed by O'Hagan and Leonard (1976) in order to introduce some degree of skewness in the priors based on the constraints $\alpha' \geq 0$, $\beta_o > 0$ and $\mu_{xo} > 0$, respectively, which are natural assumptions to make. The first constraint was motivated from (5) by assuming that α is positive, which is also suggested by the nature of the problem. The second constraint also follows from assuming β to be positive, which is expressed in prior 4. The variances of these priors were chosen in order to express our degree of belief in the constraints and let the data confirm it is true or not. The constraint $\mu_{xo} > 0$ is a consequence of assuming that the real value x_i is positive. Since our approach is conditional on X_i we use this information in the formulation of our priors based on the constraints which is also suggested by the problem. The summary of the results based on (10) are reported in Table 2, in conjunction with some results reported in Fuller (1987). Note, in particular, the smaller credibility interval for the parameter β

when compared with the corresponding interval given by Fuller (1987), obtained using asymptotic arguments. Further, the posterior mean of σ_x^2 is totally inconsistent with the maximum likelihood estimator, which is expected due to large (estimated) value of k and informative prior for σ_u^2. Large estimated values should be expected in both approaches. What we indicate here is that for a given measurement error variance, we are convinced of the power of the true value (more variability) under the controlled variable rather than the classical measurement error (See also Carrol et al. 2006). It is important to remember that Fuller's results were obtained by using asymptotic results with σ_u^2 supposed to be known from other sources, that is, $\sigma_u^2 = 57$ and ignoring that the real value of nitrogen is positive. The credible interval for the skewness parameter λ, or θ, suggests a large degree of non-randomness in the observed data caused by the selection mechanism. Concluding, the model introduced in this paper can be very useful for modeling experiments with fixed target values of the explanatory variables, such as drug dosage, temperature, pressure, etc., when the unobserved true values of the explanatory variables can differ from the target values. This kind of problem is known in the literature as Berkson measurement error model, see for example, Berkson (1950), Fuller (1987, Section 1.6.4), Carrol et al. (2006) and Cheng and Ness (1999).

Parameter	Bayesian estimation					MLE estimation	
	mean	sd	2.5%	median	97.5%	Point	C.I.
α	51.58	21.6	7.20	53.84	87.79	67.5613	
β	0.66	0.32	0.14	0.62	1.36	0.4232	(0.0288, 0.8176)
k	0.54	0.23	0.12	0.52	0.96		
λ	19.66	11.34	0.70	18.14	46.01		
σ_e^2	51.41	44.24	0.02	42.68	159.8	43.29	
σ_u^2	57	0.09	56.81	57.0	57.2	57	
σ_x^2	544	26970	8.17	61.96	1613	247.85	
μ_x	69.38	10.37	48.63	69.33	89.7	70.63	

Table 2: Posterior summaries: mean, standard deviation and percentiles of posterior densities and point and interval ML estimation based on Fuller's book, (1985, p. 19) with $\sigma_u^2 = 57$.

Another aspect of interest is to decide which of the models M1: simple normal regression model with $\sigma_u^2 = 0$ and M2: Berkson error model is more appropriate for Fuller's data in Table 1. Because the Bayes factor can be extremely sensitive to the specified prior (O'Hagan 1995) and not defined for improper priors, several authors have proposed the use of robust Bayes factors. The pseudo-Bayes factor, introduced by Geisser and Eddy (1979) (see also Gelfand and Dey 1994), is one of them and is easy to implement in Winbugs. The pseudo-Bayes factor is based on the conditional predictive densities $p(y_r|y^{(r)})$, where $y^{(r)} = (y_1, y_2, \ldots, y_{r-1}, y_{r+1}, \ldots, y_n)$.

The pseudo-Bayes factor for model M1 against model M2 is defined as

$$PBF_{12} = \frac{\prod_{r=1}^{n} p_1\left(y_r|y^{(r)}\right)}{\prod_{r=1}^{n} p_2\left(y_r|y^{(r)}\right)}, \tag{20}$$

where $p_i\left(y_r|y^{(r)}\right) = \int p_i\left(y_r|\theta, y^{(r)}\right) p\left(\theta|y^{(r)}\right) d\theta, i = 1, 2$. The estimates $\widehat{p_i}\left(y_r|y^{(r)}\right)$ were computed via Winbugs and plotted against r, for r= $1, 2, \ldots, 11$.

As we can see in the Figure 2, the Berkson model (M1) performs better than the simple normal regression model with no measurement error (M2). For the model M1 we used proper noniformative priors via Winbugs.

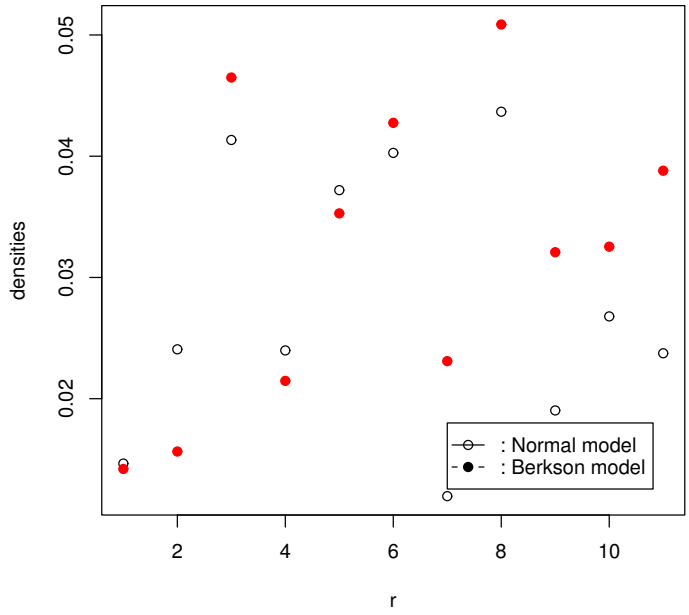

Figure 2: Conditional Predictive Densities.

Summarizing, the pseudo-Bayes factor is given by

$$PBF_{12} = 0.2080.$$

In order to check the sensitivity of the Fuller's assumption, that is, $\sigma_u^2 = 57$, we assume for model M2 a non concentrated prior on σ_u^2. The Figure 3 depicts the conditional predictive density for this model and for model M1, showing a small change with respect to the sharp prior used before. This statement confirms the pseudo-Bayes factor as a robust alternative procedure to the ordinary Bayes factor, as was mentioned before.

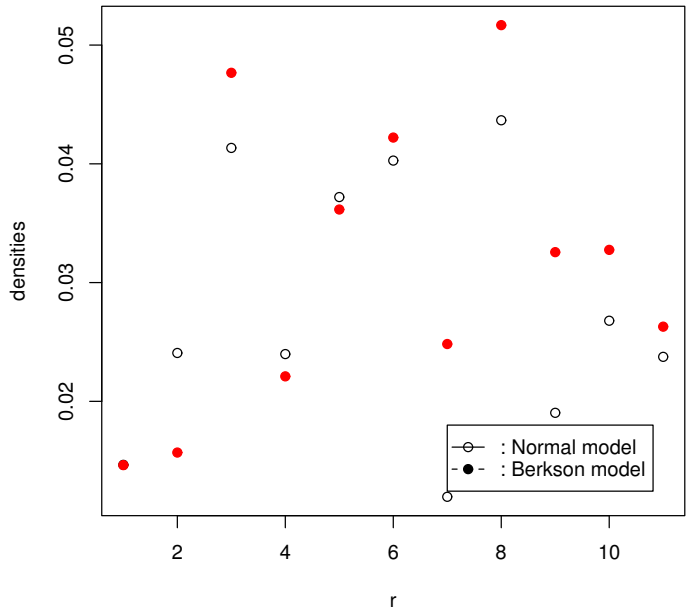

Figure 3: Conditional Predictive Densities: noniformative prior on σ_u^2.

The pseudo-Bayes factor for model M1 (the simple normal regression model) against model M2 (Berkson model with noniformative prior on σ_u^2) is

$$PBF_{12} = 0.2459.$$

6 Concluding remarks

We end this paper by calling attention to the flexibility of the proposed extended conditional regression measurement error model in avoiding the usual constraints on the regression measurement errors and the usefulness and simplicity of the Bayesian procedure with hierarchical priors via Winbugs. The extension to covariate vectors allows dealing with the situation where some of covariates are measured without error. Important contributions of this paper are the extensions of results in Copas and Li (1997) and the implementation of a Bayesian analysis with two-stage priors to the restricted regression model, which to the best our knowledge, nobody has ever carried out for this extended error model (Copas and Li 1997, pg. 75). If the reader wants to be familiarized with the extended skew-normal model, important references to be looked at are Arellano-Valle and Azzalini (2006), Arellano-Valle et al. (2006), Arnold and Beaver (2000) and Genton (2004) where detailed treatments of such models are considered. Moreover, it is important to mention that distribution (18) is a special case of the closed skew-normal distribution (CSN) introduced by G. Gonzalez-Farias and Gupta (2003) and described in detail by G. Gonzalez-Farias and Gupta (2004). From the Bayesian

view point, some details on the class of distributions considered can be found in Liseo (2004) and S.K. Sahu and Branco (2003).

Also, within the Bayesian approach it is very helpful to compare models via pseudo-Bayes factor, which is a robust alternative to the ordinary Bayes factor, which is extremely sensitive to prior specification. In a future paper we intend to introduce treatment effect to analyze the data reported by Voss (1969) for two soil types studied by DeGracie and Fuller (1972) where the variance of the measurement error is considered as known. It is also worth undertaking a study of Berkson measurement error models reported in Carrol et al. (2006) and Cheng and Ness (1999) using the approach considered in this paper.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). "On the unification of families of skew-normal distributions." *Scandinavian Journal of Statistics*, 33: 561–574.

Arellano-Valle, R. B., Branco, M., and Genton, M. G. (2006). "A unified view on skewed distributions arising from selections." *The Canadian Journal of Statistics*, 34: 581–601.

Arnold, B. C. and Beaver, R. J. (2000). "Hidden truncation models." *Sankhya*, 62: 22–35.

Berkson, J. (1950). "Are there two regressions?" *Journal of the American Statistical Association*, 45: 164–180.

Capitanio, A., Azzalini, A., and Stanghellini, E. (2003). "Graphical models for skew-normal variate." *Scandinavian Journal of Statistics*, 30: 129–144.

Carrol, R. J., Stefanski, L. A., and Ruppert, D. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. London: Chapman & Hall.

Cheng, C.-L. and Ness, J. W. V. (1999). *Statistical Regression with Measurement Error, Kendall's Library on Statistics*. London: Arnold.

Cheng, C.-L. and Tsai, C.-L. (1996). "Scores tests in measurement error models: heteroscedasticity, autocorrelation an transformation." *Institute of Statistical Science, Academia Sinica*, (Technical Report RT-MAE-2004-15).

Copas, J. B. and Li, H. G. (1997). "Inference for non-random samples (with discussion)." *Journal Royal Statistical Society, B*, 59: 55–95.

DeGracie, J. S. and Fuller, W. A. (1972). "Estimation of the slope and analysis of covariance when the concomitant variable is measured with error." *Journal of the American Statistical Association*, 67: 930–937.

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

G. Gonzalez-Farias, J. D.-M. and Gupta, A. (2003). "Additive properties of skew-normal random vectors." *Journal of Statistical Planning and Inference, 126*, 2: 521–534.

— (2004). "The closed skew-normal distribution." In *Skew-Elliptical Distributions and Their Application: A Journey Beyond Normality*, 25–42. CRC, Boca Raton, FL: Chapman & Hall.

Geisser, S. and Eddy, W. (1979). "A predictive approach to model selection." *Journal of the American Statistical Association*, 74: 153–160.

Gelfand, A. and Dey, D. (1994). "Bayesian model choice: asymptotic and exact calculations." *Journal Royal Statistical Society, B*, 56: 501–519.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Application: A Journey Beyond Normality*. Boca Raton: Chappman & Hall/CRC.

Gleser, L. (1992). "The importance of assessing measurement reliability in multivariate regression." *Journal of the American Statistical Association*, 87: 696–707.

Heckman, J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models." *Annals of Economical and Social Measurement*, 5: 475–492.

— (1979). "Sample selection bias as a specification error." *Econometrica*, 47: 153–161.

Liseo, B. (2004). "Skew-Elliptical distributions in Bayesian Inference." In *Skew-Elliptical Distributions and Their Application: A Journey Beyond Normality*, 153–190. CRC, Boca Raton,FL: Chapman & Hall.

Liseo, B. and Loperfido, N. (2006). "A note on reference priors for the scalar skew-normal distribution." *Journal Statistical Planning Inference*, 136: 373–389.

Little, R. (1985). "A note about models for selectivity bias." *Econometrica*, 53: 1469–1474.

O'Hagan, A. (1995). "Fractional Bayes factors for model comparison." *Journal Royall Statistical Society*, 57: 99–138.

O'Hagan, A. and Leonard, T. (1976). "Bayes estimates subject to uncertainty about parameter constraints." *Biometrika*, 63: 201–202.

S.K. Sahu, D. D. and Branco, M. (2003). "A new class of Multivariate Skew Distributions with Applications to Bayesian Regression Models." *Canadian Journal of Statistics*, 31: 129–150.

Voss, E. R. (1969). "Response by Corn to NPK Fertilization on Marshall and Manona Soils as Influenced by Management and Meteorological Factors." In *Ph.D thesis*. Iowa State University, Ames, Iowa: Unpublished.

Acknowledgments

We are very grateful to the referees and Editor for their helpful and insightful comments. Partial financial support from CNPq-Brasil is also acknowledged.