Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 618–639
DOI: 10.1214/16-BJPS326

Bayesian analysis of flexible measurement error models

Luz Marina Rondon[a,b] **and Heleno Bolfarine**[a]

[a]*Universidade de São Paulo*
[b]*Universidad Nacional de Colombia*

Abstract. This paper proposes the Bayesian inference for flexible measurement error models, in which their systematic components include explanatory variable vectors with and without measurement errors, as well as nonlinear effects that are approximated by using *B*-splines. The model investigated is the structural version, as the error-prone variables follow scale mixtures of normal distributions such as Student-*t*, slash, contaminated normal, Laplace and symmetric hyperbolic distributions. To draw samples of the posterior distribution of the model parameters, an MCMC algorithm is proposed. The performance of this algorithm is assessed through simulations. In addition, the function `fmem()` of the R package **BayesGESM** is presented, which provides an easy way to apply the methodology presented in this paper. The proposed methodology is applied to a real data set, which shows that ignoring measurement errors (i.e., analyze the data by using the traditional methodology) can lead to wrong conclusions.

1 Introduction

Regression models under the assumption of independent and normally distributed errors are very useful statistical tools for data analysis. However, in practice, there are data sets in which the presence of explanatory variables measured with error can substantially affect the good performance of the parameter estimators by introducing asymptotic bias and by producing low coverage rates for confidence intervals (see Fuller, 1987; Cheng and Van Ness, 1999). Therefore, the regression models where the measurement errors are taken into account, termed measurement error models or error-in-variables models, have been studied by many authors as, for example, Arellano-Valle, Bolfarine and Labra (1996), Carroll, Roeder and Wasserman (1999), Kulathinal, Kuulasmaa and Gasbarra (2002), Li, Palta and Shao (2004), Patriota, Bolfarine and de Castro (2009) and Cao, Lin and Zhu (2012) under the frequentist approach; and Kelly (2007), Carroll et al. (2006), Chapter 9 and de Castro, Bolfarine and Galea (2013) under the Bayesian approach. However, most of the proposals presented in these papers are based on the assumption of a random term that exhibits a normal distribution, which is known to be vulnerable in the presence of extreme or outlying observations (see Maronna, Martin

Key words and phrases. Bayesian analysis, measurement error models, semi-parametric models, MCMC algorithm, *B*-splines, scale mixtures of normal distributions.
Received March 2015; accepted June 2016.

and Yohai, 2006). Nonetheless, most of these papers do not allow the simultaneous presence of vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects whose functional form is unknown.

Thus, in this paper, a measurement error model which admits vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects approximated by using B-splines is introduced. The model investigated is the structural version, as the error-prone variables follow scale mixtures of normal distributions such as Student-t, slash, contaminated normal, Laplace and symmetric hyperbolic distributions. In addition, the model considered assumes an error term whose distribution also belongs to the class of scale mixtures of normal distributions. Some of these distributions present heavier tails than the normal ones. Accordingly, the regression models based on them seem to be reasonable choice for robust inference. Lemonte and Patriota (2011) present a general formulation of models based on elliptical distributions. In the parametric case (that is, when there are not nonparametric effects), the models presented by those authors are more general than the model addressed in this paper. An ultrastructural measurement error model with an error term distributed according to the Student-t distribution has been studied by using the frequentist approach (see, for instance, Arellano-Valle, Bolfarine and Labra, 1996). Similarly, Cao, Lin and Zhu (2012) proposed a structural heteroscedastic measurement error model where the random terms follow scale mixtures of normal distributions. However, given the complexity of the model investigated in this paper, the Bayesian approach is considered, which has to be based on the Markov chain Monte Carlo (MCMC) methodology for parameter estimation (see Gamerman and Lopes, 2006).

The rest of this paper is organized as follows: Section 2 describes the random component of model investigated, that is, the class of multivariate scale mixtures of normal distributions. Section 3 formulates the flexible measurement error models that allow explanatory variables with and without measurement error, as well as the presence of a nonlinear effects which are approximated by using B-splines. Section 4 is dedicated to the specification of the prior distributions and the Gibbs sampler to draw samples from the posterior distribution of the interest parameters. Model selection is also discussed in that section. Section 5 presents a simulation study where the performance of the proposed MCMC algorithm is illustrated. That section also presents the function `fmem()` of the R package **BayesGESM**, which has been developed to provide a easy way to apply the statistical methodology presented in this paper. In Section 6, the proposed methodology is applied to a real data set. It reveals that the onset of measurement errors in one of the covariates is a better model than the model which does not take into account measurement errors (i.e., *naive* approach). Section 7 presents some concluding remarks.

2 The multivariate scale mixture of normal distributions

The class of multivariate scale mixture of normal distributions provides a rich set of multivariate symmetric distributions, some of which have heavier/lighter tails than the normal ones as well as distributions with different levels of kurtosis. In fact, the distributions with heavier tails than the normal ones can be used to obtain robust inferences in datasets with outlying observations where statistical inference based on the normal distribution is known to be vulnerable. In the univariate case, this class of distributions includes Student-t, slash (see Rogers and Tukey, 1972), contaminated normal, Laplace (see Box and Tiao, 1973) and symmetric hyperbolic (see Barndorff-Nielsen, 1977) distributions. Moreover, following Andrews and Mallows (1974), a continuous random vector $\mathbf{Y} = (Y_1, \ldots, Y_r)^{\mathrm{T}}$ follows a multivariate scale mixtures of normal distribution, which is denoted by $\mathcal{SMN}_r$, if it can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{\frac{1}{2}}(U)\mathbf{Z},$$

where $\boldsymbol{\mu} \in \mathbb{R}^r$ is the location parameter, $\mathbf{Z} \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ as an $(r \times r)$ positive definite matrix, $\kappa(\cdot)$ is a positive function and U is a positive random variable independent of $\mathbf{Z}$ and having cumulative distribution function (c.d.f.) denoted by $H(u; \boldsymbol{\eta})$, in which $\boldsymbol{\eta}$ is a parameter or parameter vector indexing the distribution of U. Thus, the probability density function (p.d.f.) of $\mathbf{Y}$ has an $\mathcal{SMN}_r$ representation if it can be expressed as

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) = \int_0^\infty \phi_r\big(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}\big)\, dH(u; \boldsymbol{\eta}), \tag{1}$$

where $\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\{-\frac{1}{2}\delta^2\} \times (2\pi)^{-r/2}|\boldsymbol{\Sigma}|^{-r/2}$ and $\delta^2 = (\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$. If $\mathbf{Y}$ has a p.d.f. as described in (1), then it is denoted by $\mathbf{Y} \sim \mathcal{SMN}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H, \kappa)$. In hierarchical form, the distribution of $\mathbf{Y}$ can be specified as $\mathbf{Y}|U = u \sim \mathcal{N}_r(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})$ and $U \sim \mathcal{H}(u; \boldsymbol{\eta})$. Below are examples of distributions that belong to the $\mathcal{SMN}_r$ family:

- *Multivariate Student-t*. In this case, $U \sim \mathcal{G}amma(\eta/2, \eta/2)$, $\eta > 0$, and $\kappa(u) = 1/u$. Thus, according to (1) the density of the random vector $\mathbf{Y}$ is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta) = \frac{\Gamma(\frac{\eta+r}{2})}{\Gamma(\frac{\eta}{2})|\boldsymbol{\Sigma}|^{1/2}(\pi\eta)^{r/2}}\left[1 + \frac{\delta^2}{\eta}\right]^{-(\frac{\eta+r}{2})}.$$

  Then, $\mathbf{Y} \sim t_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$.
- *Multivariate slash.* Let $\kappa(u) = 1/u$ and $U \sim \mathcal{B}eta(\eta, 1)$, $\eta > 0$. Then, the p.d.f. of $\mathbf{Y}$ becomes

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta) = \frac{\eta}{(2\pi)^{r/2}|\boldsymbol{\Sigma}|^{1/2}} \int_0^1 u^{\eta+\frac{r}{2}-1} \exp\big[-u\delta^2/2\big]\, du.$$

  Hence, $\mathbf{Y} \sim \mathcal{S}l_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$.

- *Multivariate contaminated normal.* In this case, $\kappa(u) = 1/u$ and U is a discrete random variable which takes value η_2 with probability η_1 and value 1 with probability $(1 - \eta_1)$. Thus, the p.d.f. of U is given by $h(u|\boldsymbol{\eta} = (\eta_1, \eta_2)^{\mathrm{T}}) = \eta_1 \mathbb{I}_{(u=\eta_2)} + (1 - \eta_1)\mathbb{I}_{(u=1)}$. According to (1), the density of the random vector $\mathbf{Y}$ is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) = \eta_1 \phi_r\big(\mathbf{y}|\boldsymbol{\mu}, \eta_2^{-1}\boldsymbol{\Sigma}\big) + (1 - \eta_1)\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

  where $0 < \eta_1 < 1$ and $0 < \eta_2 < 1$. So, $\mathbf{Y} \sim \mathcal{CN}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$.
- *Multivariate Laplace.* Here $U \sim \mathcal{E}xp(1/8)$ and $\kappa(u) = u$, thus

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\mathbf{K}_a(\sqrt{\delta^2/4})(\delta^2)^{-\frac{r+2}{4}}}{2^{r+1}\pi^{r/2}|\boldsymbol{\Sigma}|^{1/2}},$$

  where $a = -\frac{r}{2} + 1$ and $\mathbf{K}_a(\eta) = \frac{1}{2}\int_0^\infty x^{a-1}\exp(-\frac{1}{2}\eta(x + x^{-1}))\partial x$ is the modified Bessel function type three of order a. Therefore, $\mathbf{Y} \sim \mathcal{L}aplace_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- *Multivariate symmetric hyperbolic.* In this case, $\kappa(u) = u$ and U follows an generalized inverse Gaussian distribution, that is, $U \sim \mathcal{GIG}(1, 1, \eta^2)$ (see Jörgensen, 1982). The p.d.f. of a random variable $U \sim \mathcal{GIG}(a, b, c)$ is given by

$$h(u|a, b, c) = \frac{(c/b)^{\frac{a}{2}}}{2\mathbf{K}_a(\sqrt{bc})}u^{a-1}\exp\left[-\frac{1}{2}\big(bu^{-1} + cu\big)\right].$$

  Therefore, according to (1) the p.d.f. of $\mathbf{Y}$ reduces to

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta) = \frac{\mathbf{K}_a(\eta\sqrt{\delta^2 + 1})\eta^{r/2}(\delta^2 + 1)^{-\frac{r}{4}+\frac{1}{2}}}{2^{r/2}\pi^{r/2}|\boldsymbol{\Sigma}|^{1/2}\mathbf{K}_1(\eta)},$$

  where $a = -\frac{r}{2} + 1$. Hence, $\mathbf{Y} \sim \mathcal{SH}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$, $\eta > 0$.

Figure 1 presents the probability density functions of some standard (i.e., $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$) $\mathcal{SMN}_2$ distributions.

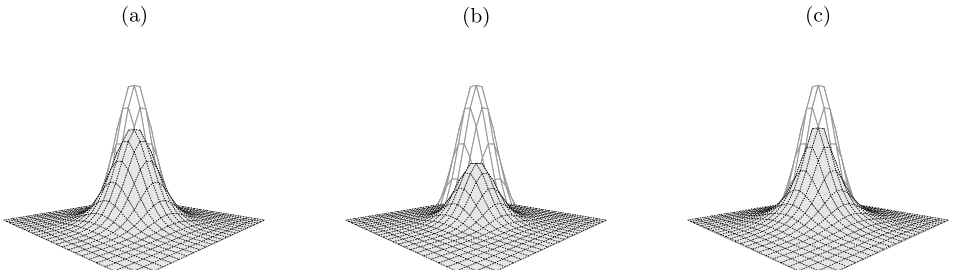

Figure 1 *The probability density functions of some standard $\mathcal{SMN}_2$ distributions*: *Slash* ($\eta = 2$) (a), *Contaminated normal* ($\boldsymbol{\eta} = (0.6, 0.2)^{\mathrm{T}}$) (b) *and Symmetric hyperbolic* ($\eta = 1$) (c) *when compared with standard bivariate normal distribution*.

3 Flexible measurement error models

Initially, the semi-parametric measurement error model is defined to connect the response variable to the explanatory variables, where the s unspecified nonlinear effects are also included and which are approximated by using B-splines. More specifically, the following stochastic mechanism is assumed to have generated the data set with n observations:

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{m}_i^{\mathrm{T}}\boldsymbol{\rho} + \sum_{j=1}^{s} f_j(v_{ij}) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $(\mathbf{x}_i^{\mathrm{T}}, \mathbf{m}_i^{\mathrm{T}}, v_{i1}, \ldots, v_{is})^{\mathrm{T}}$ represents the explanatory variables associated with sample individual i; $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_q)^{\mathrm{T}}$ are vectors of unknown parameters to be estimated; $f_j(\cdot)$ $(j = 1, \ldots, s)$ are unknown, smooth and continuous functions approximated by using B-splines and ε_i is the model random error. Furthermore, the explanatory variable vector $\mathbf{m}_i$ is not observed directly (see Cheng and Van Ness, 1999) but only through a (additive) random mechanism yielding an "estimate" of it, denoted as $\mathbf{M}_i$, represented by the equation

$$\mathbf{M}_i = \mathbf{m}_i + \boldsymbol{\xi}_i, \qquad i = 1, \ldots, n.$$

In addition, it is assumed that

$$\begin{pmatrix} \varepsilon_i \\ \boldsymbol{\xi}_i \\ \mathbf{m}_i \end{pmatrix} \sim \mathcal{SMN}_{2q+1}\left[\begin{pmatrix} 0 \\ \mathbf{0} \\ \boldsymbol{\mu}_m \end{pmatrix}; \begin{pmatrix} \sigma_y^2 & 0 & 0 \\ 0 & \sigma_\xi^2 \mathbf{I}_q & \mathbf{0} \\ 0 & \mathbf{0} & \boldsymbol{\Sigma}_m \end{pmatrix} \right], \qquad i = 1, \ldots, n,$$

are independent random vectors where $\boldsymbol{\mu}_m$, σ_y^2, σ_ξ^2 and $\boldsymbol{\Sigma}_m$ are unknown parameters to be estimated. This model is the flexible semi-parametric measurement error model. It is assumed that the ratio $\omega = \sigma_y^2/\sigma_\xi^2$ is known, an assumption typically made to make the model identifiable. Because a Bayesian approach is adopted, specification of proper prior distributions for the model parameters is an alternative way of making the model identifiable.

The nonparametric components for model (2) are approximated by using B-splines (see De Boor, 1978). For instance, suppose that we are estimating the function $f(v)$ in the interval $[a, b]$, then $a = t_0 < t_1 < \cdots < t_k < t_{k+1} = b$ is a partition of this interval, where t_i $(i = 1, \ldots, k)$ are the internal knots. In this setup, the values $f_j(v_{ij})$ may be approximated by $\mathbf{b}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j$, where $\mathbf{b}_{ij} = (b_{1j}(v_{ij}), \ldots, b_{K_j j}(v_{ij}))^{\mathrm{T}}$ is the vector of basis functions, $\boldsymbol{\alpha}_j \in \mathbb{R}^{K_j}$ is the B-spline coefficient vector for variable v_j, and $K_j = M_j + k_j$, with M_j and k_j as the degree and the number of internal knots of the spline, respectively. Selection of knots is generally an important aspect of spline smoothing. The number of internal knots is assumed to be $k_j = [n^{1/5}]$ according to the proposal by He, Fung and Zhu (2005), where n is the sample size and $[x]$ is the integer part of x. More specifically,

the internal knots are selected as $\{q(v, 1/(k_j+1)), \ldots, q(v, k_j/(k_j+1))\}$, where $q(x, p)$ is the quantile of order $0 < p < 1$ of x. Under the frequentist approach, the penalty term of $\boldsymbol{\alpha}_j$ is given by $\frac{1}{2\tau^2_{\alpha_j}}\boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{\alpha}_j$, where $\tau^2_{\alpha_j}$ is the smoothing parameter. In fact, this approximation of $f_j(v_j)$ can be considered to be an *P*-spline (see Eilers and Marx, 1996) with a penalty term of order 0. Therefore, under the Bayesian approach, this type of spline induces a well-known proper priori distribution for $\boldsymbol{\alpha}_j$ (i.e., multivariate normal distribution).

4 Bayesian inference

Under the Bayesian approach the inference about the interest parameters is based on their posterior distribution. In this section, the prior distributions for the model parameters is described and the Markov chain Monte Carlo (MCMC) (see Gamerman and Lopes, 2006) algorithm is proposed to draw samples from the posterior distribution. Additionally, the model comparison is presented.

4.1 Prior distributions

One important step in a Bayesian approach is the specification of the prior distributions for the model parameters. It is assumed a priori that the parameters $\boldsymbol{\beta}$, $\boldsymbol{\rho}$, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$, $\boldsymbol{\mu}_m^{\mathrm{T}}$, $\boldsymbol{\Sigma}_m$ and σ_y^2 are independent and have the following distributions

$$\begin{aligned}
&\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\beta}_0, \mathbf{S}_\beta), \qquad \boldsymbol{\rho} \sim \mathcal{N}_q(\boldsymbol{\rho}_0, \mathbf{S}_\rho),\\
&\boldsymbol{\alpha}_j \sim \mathcal{N}_{K_j}\big(\boldsymbol{\alpha}_{j0}, \tau^2_{\alpha_j}\mathbf{I}_{K_j}\big), \qquad \boldsymbol{\mu}_m \sim \mathcal{N}_q(\boldsymbol{\mu}_{m_o}, \boldsymbol{\Sigma}_{\mu_o}),\\
&\boldsymbol{\Sigma}_m^{-1} \sim \mathcal{W}ishart(q, \boldsymbol{\Omega}_m), \qquad \sigma_y^2 \sim \mathcal{IG}\left(\frac{a}{2}, \frac{b}{2}\right), \qquad \tau^2_{\alpha_j} \sim \mathcal{IG}(a_{\tau_{\alpha_j}}, b_{\tau_{\alpha_j}}),
\end{aligned}$$

where the hyperparameters $\boldsymbol{\beta}_0$, $\boldsymbol{\rho}_0$, $\boldsymbol{\alpha}_{j0}$, $\boldsymbol{\mu}_{m_o}$, $\mathbf{S}_\beta > 0$, $\mathbf{S}_\rho > 0$, $\boldsymbol{\Sigma}_{\mu_o} > 0$, $\boldsymbol{\Omega}_m > 0$, $a > 0$, $b > 0$, $a_{\tau_{\alpha_j}} > 0$ and $b_{\tau_{\alpha_j}} > 0$ $(j = 1, \ldots, s)$ are assumed known, $\mathbf{I}_n$ corresponds to the identity matrix of order n, and the p.d.f. of a random variable $X \sim \mathcal{IG}(a, b)$ is given by

$$p(x|a, b) \propto x^{-a-1}\exp\left(-\frac{b}{x}\right).$$

4.2 MCMC algorithm

The algorithm described in the sequel uses the fact that the distribution of the response vector follows a multivariate scale mixture of normal distribution. It is a data augmented algorithm (see Tanner and Wong, 1987) using u_i and $\mathbf{m}_i$, $i = 1, \ldots, n$, as latent (unobserved) variables. Therefore, according to the specifications above, the augmented likelihood function for the parameter vector

$(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\rho}^{\mathrm{T}}, \boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_s^{\mathrm{T}}, \boldsymbol{\mu}_m^{\mathrm{T}}, \boldsymbol{\Sigma}_m, \sigma_y^2)$ can be expressed as

$$
\begin{aligned}
&L(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_s^{\mathrm{T}}, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \sigma_y^2 | \mathbf{y}, \mathbf{X}, \mathbf{M}, \mathbf{v}, \mathbf{u}, \mathbf{m}) \\
&\quad \propto \prod_{i=1}^{n} (\sigma_y^2)^{-\frac{1}{2}} (\sigma_\xi^2)^{-\frac{q}{2}} \kappa(u_i)^{-\frac{(q+2)}{2}} |\boldsymbol{\Sigma}_m|^{-\frac{1}{2}} \\
&\qquad \times \exp\Bigg[-\frac{(\mathbf{M}_i - \mathbf{m}_i)^{\mathrm{T}} (\mathbf{M}_i - \mathbf{m}_i)}{2\sigma_\xi^2 \kappa(u_i)} - \frac{(y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} - \mathbf{m}_i^{\mathrm{T}} \boldsymbol{\rho} - \sum_{j=1}^{s} \mathbf{b}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_j)^2}{2\sigma_y^2 \kappa(u_i)} \\
&\qquad - \frac{(\mathbf{m}_i - \boldsymbol{\mu}_m)^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} (\mathbf{m}_i - \boldsymbol{\mu}_m)}{2\kappa(u_i)} \Bigg].
\end{aligned}
$$

To apply the Gibbs sampling, the likelihood and the prior distributions described above are combined to obtain the complete conditional posterior distributions for each parameter in model (2). Then, the algorithm encompasses the following steps:

1. Start the algorithm with an initial value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \boldsymbol{\rho}^{(0)}, \boldsymbol{\mu}_m^{(0)}, \boldsymbol{\Sigma}_m^{(0)}, \boldsymbol{\alpha}_1^{(0)}, \ldots, \boldsymbol{\alpha}_s^{(0)}, \sigma_y^{2(0)})$;
2. Calculate the quantity $S_i^{(l)}$, $i = 1, \ldots, n$, where $S_i^{(l)}$ is S_i evaluated at $\boldsymbol{\theta}^{(l)}$, and

$$
\begin{aligned}
S_i = &\frac{(y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} - \mathbf{m}_i^{\mathrm{T}} \boldsymbol{\rho} - \sum_{j=1}^{s} \mathbf{b}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_j)^2}{\sigma_y^2} + \frac{(\mathbf{M}_i - \mathbf{m}_i)^{\mathrm{T}} (\mathbf{M}_i - \mathbf{m}_i)}{\omega \sigma_y^2} \\
&+ (\mathbf{m}_i - \boldsymbol{\mu}_m)^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} (\mathbf{m}_i - \boldsymbol{\mu}_m).
\end{aligned}
$$

3. Generate $u_i^{(l+1)} \sim p(u_i | S_i^{(l)})$, $i = 1, \ldots, n$, independent, according to the distribution of the random component $(\varepsilon, \boldsymbol{\xi}^{\mathrm{T}}, \mathbf{m}^{\mathrm{T}})^{\mathrm{T}}$:
   (a) Normal distribution: $P(u_i = 1 | S_i^{(l)}) = 1$.
   (b) Student-t distribution:

$$
p(u_i | S_i^{(l)}) \propto u_i^{\frac{q+\eta}{2}} \exp\Bigg[-\frac{u_i}{2} (S_i^{(l)} + \eta) \Bigg].
$$

   Thus, $u_i | S_i^{(l)} \sim \mathcal{G}amma(\frac{\eta+q+2}{2}, \frac{S_i^{(l)}+\eta}{2})$.
   (c) Slash distribution:

$$
p(u_i | S_i^{(l)}) \propto u_i^{\eta+\frac{q}{2}} \exp\Bigg[-\frac{u_i}{2} S_i^{(l)} \Bigg] I_{(0,1)}(u_i).
$$

   Therefore, $u_i | S_i^{(l)} \sim \mathcal{T}run\mathcal{G}amma(\eta + 1 + \frac{q}{2}, \frac{S_i^{(l)}}{2}; (0, 1))$. that is, $u_i | S_i^{(l)}$ follows a truncated gamma distribution (see Nadarajah and Kotz, 2006).
   (d) Contaminated normal distribution:

$$
p(u_i | S_i^{(l)}) = p_\eta \mathbb{I}_{(u_i = \eta_2)} + (1 - p_\eta) \mathbb{I}_{(u_i = 1)},
$$

where

$$p_\eta \propto \eta_1 \eta_2^{\frac{q}{2}+1} \exp\left\{-\frac{\eta_2 S_i^{(l)}}{2}\right\} \quad \text{and} \quad (1-p_\eta) \propto (1-\eta_1)\exp\left\{-\frac{S_i^{(l)}}{2}\right\}.$$

(e) Symmetric hyperbolic distribution:

$$p\big(u_i|S_i^{(l)}\big) \propto u_i^{-\frac{q}{2}-1} \exp\left\{-\frac{1}{2}\left[\frac{S_i^{(l)}+1}{u_i} + \eta^2 u_i\right]\right\},$$

that is, $u_i|S_i^{(l)} \sim \mathcal{GIG}(-\frac{q}{2}, S_i^{(l)}+1, \eta^2)$.

(f) Laplace distribution:

$$p\big(u_i|S_i^{(l)}\big) \propto u_i^{-\frac{q}{2}-1} \exp\left\{-\frac{1}{2}\left[\frac{S_i^{(l)}}{u_i} + \frac{u_i}{4}\right]\right\},$$

consequently, $u_i|S_i^{(l)} \sim \mathcal{GIG}(-\frac{q}{2}, S_i^{(l)}, \frac{1}{4})$.

4. Calculate the matrix $\mathbf{L}_u^{(l+1)} = \operatorname{diag}\{u_1^{(l+1)}, \ldots, u_n^{(l+1)}\}$.
5. Generate $\tilde{\boldsymbol{\beta}}^{(l+1)} \sim \mathcal{N}_{p+q}(\boldsymbol{\mu}_{\tilde{\beta}}, \boldsymbol{\Sigma}_{\tilde{\beta}})$, in which

$$\boldsymbol{\Sigma}_{\tilde{\beta}} = \left[\begin{pmatrix} \mathbf{S}_\beta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_\rho^{-1} \end{pmatrix} + \frac{1}{\sigma_y^{2(l)}} \overline{\mathbf{X}}^{\mathrm{T}(l)} \big[\mathbf{L}_u^{(l+1)}\big]^{-1} \overline{\mathbf{X}}^{(l)}\right]^{-1} \quad \text{and}$$

$$\boldsymbol{\mu}_{\tilde{\beta}} = \boldsymbol{\Sigma}_{\tilde{\beta}}\left[\begin{pmatrix} \mathbf{S}_\beta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_\rho^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\rho}_0 \end{pmatrix} + \frac{1}{\sigma_y^{2(l)}} \overline{\mathbf{X}}^{\mathrm{T}(l)} \big[\mathbf{L}_u^{(l+1)}\big]^{-1} \left(\mathbf{y} - \sum_{j=1}^{s} \mathbf{B}_j \boldsymbol{\alpha}_j^{(l)}\right)\right],$$

where $\overline{\mathbf{X}}^{(l)} = [\mathbf{X}, \mathbf{m}^{(l)}]$, $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\rho}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{B}_j = (\mathbf{b}_{1j}, \ldots, \mathbf{b}_{nj})^{\mathrm{T}}$.

6. Generate the ith row of $\mathbf{m}^{(l+1)}$, which is denoted by $\mathbf{m}_i^{(l+1)}$, with $\mathbf{m}_i^{(l+1)} \sim \mathcal{N}_q(\boldsymbol{\mu}_{m_i}, \boldsymbol{\Sigma}_{m_i})$, where

$$\boldsymbol{\Sigma}_{m_i} = \left[\frac{(\boldsymbol{\Sigma}_m^{(l)})^{-1}}{\kappa(u_i^{(l+1)})} + \frac{\mathbf{I}_q}{\omega\sigma_y^{2(l)}\kappa(u_i^{(l+1)})} + \frac{\boldsymbol{\rho}^{(l+1)}\boldsymbol{\rho}^{\mathrm{T}(l+1)}}{\sigma_y^{2(l)}\kappa(u_i^{(l+1)})}\right]^{-1} \quad \text{and}$$

$$\boldsymbol{\mu}_{m_i} = \boldsymbol{\Sigma}_{m_i}\left[\frac{(\boldsymbol{\Sigma}_m^{(l)})^{-1}\boldsymbol{\mu}_m^{(l)}}{\kappa(u_i^{(l+1)})} + \frac{\mathbf{M}_i}{\omega\sigma_y^{2(l)}\kappa(u_i^{(l+1)})} + \frac{\boldsymbol{\rho}^{(l+1)}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(l+1)} - \sum_{j=1}^{s}\mathbf{b}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j^{(l)})}{\sigma_y^{2(l)}\kappa(u_i^{(l+1)})}\right].$$

7. Generate $\boldsymbol{\mu}_m^{(l+1)} \sim \mathcal{N}_q(\boldsymbol{\mu}_{\mu_m}, \boldsymbol{\Sigma}_{\mu_m})$, where

$$\boldsymbol{\Sigma}_{\mu_m} = \left[(\boldsymbol{\Sigma}_m^{(l)})^{-1}\left(\sum_{r=1}^{n}\frac{1}{\kappa(u_i^{(l+1)})}\right) + \boldsymbol{\Sigma}_{\mu_0}^{-1}\right]^{-1} \quad \text{and}$$

$$\boldsymbol{\mu}_{\mu_m} = \boldsymbol{\Sigma}_{\mu_m}^{(l+1)}\left[(\boldsymbol{\Sigma}_m^{(l)})^{-1}\left(\sum_{r=1}^{n}\frac{\mathbf{m}_i^{(l+1)}}{\kappa(u_i^{(l+1)})}\right) + \boldsymbol{\Sigma}_{\mu_0}^{-1}\boldsymbol{\mu}_{m_0}\right].$$

8. Generate $(\boldsymbol{\Sigma}_m^{(l+1)})^{-1} \sim \mathcal{W}ishart(q+n, \boldsymbol{\Omega}_m^*)$, where

$$\boldsymbol{\Omega}_m^* = \left[\boldsymbol{\Omega}_m^{-1} + \sum_{r=1}^{n}\frac{(\mathbf{m}_i^{(l+1)} - \boldsymbol{\mu}_m^{(l+1)})(\mathbf{m}_i^{(l+1)} - \boldsymbol{\mu}_m^{(l+1)})^{\mathrm{T}}}{\kappa(u_i^{(l+1)})}\right]^{-1}.$$

9. Generate $\tau_{\alpha_j}^{2(l+1)} \sim \mathcal{IG}(\frac{K_j}{2} + a_{\tau_{\alpha_j}}, \frac{2b_{\tau_{\alpha_j}} + (\boldsymbol{\alpha}_j^{(l)} - \boldsymbol{\alpha}_{j0})^{\mathrm{T}}(\boldsymbol{\alpha}_j^{(l)} - \boldsymbol{\alpha}_{j0})}{2})$, $j = 1, \ldots, s$.
10. Generate $\boldsymbol{\alpha}_j^{(l+1)} \sim \mathcal{N}_{K_j}(\boldsymbol{\mu}_{\alpha_j}, \boldsymbol{\Sigma}_{\alpha_j})$, $j = 1, \ldots, s$, where

$$\boldsymbol{\Sigma}_{\alpha_j} = \left[\frac{1}{\tau_{\alpha_j}^{2(l+1)}}\mathbf{I}_{K_j} + \frac{1}{\sigma_y^{2(l)}}\mathbf{B}_j^{\mathrm{T}}[\mathbf{L}_u^{(l+1)}]^{-1}\mathbf{B}_j\right]^{-1} \quad \text{and}$$

$$\boldsymbol{\mu}_{\alpha_j} = \boldsymbol{\Sigma}_{\alpha_j}\left[\frac{1}{\tau_{\alpha_j}^{2(l+1)}}\boldsymbol{\alpha}_{j0} + \frac{1}{\sigma_y^{2(l)}}\mathbf{B}_j^{\mathrm{T}}[\mathbf{L}_u^{(l+1)}]^{-1} \times \left(\mathbf{y} - \overline{\mathbf{X}}^{(l+1)}\tilde{\boldsymbol{\beta}}^{(l+1)} - \sum_{0<i<j}\mathbf{B}_i\boldsymbol{\alpha}_i^{(l+1)} - \sum_{j<i\leq s}\mathbf{B}_i\boldsymbol{\alpha}_i^{(l)}\right)\right].$$

11. Generate

$$\sigma_y^{2(l+1)} \sim \mathcal{IG}\left(\frac{n(1+q)+a}{2}, \frac{1}{2}\left[\sum_{r=1}^{n} b_i^{(l+1)} + b\right]\right),$$

where $b_i = \frac{(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{m}_i^{\mathrm{T}}\boldsymbol{\rho} - \sum_{j=1}^{s}\mathbf{b}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j)^2}{\kappa(u_i)} + \frac{(\mathbf{M}_i - \mathbf{m}_i)^{\mathrm{T}}(\mathbf{M}_i - \mathbf{m}_i)}{\omega\kappa(u_i)}$, and $b_i^{(l+1)}$ is b_i evaluated at $\boldsymbol{\beta}^{(l+1)}$, $\boldsymbol{\rho}^{(l+1)}$, $\mathbf{m}_i^{(l+1)}$, $\boldsymbol{\alpha}_1^{(l+1)}, \ldots, \boldsymbol{\alpha}_s^{(l+1)}$ and $u_i^{(l+1)}$.
12. Repeat 2–11 until convergence is reached.

Therefore, following the MCMC algorithm, it is possible to sample from the marginal posterior distributions for parameters $\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s, \boldsymbol{\mu}_m$ and σ_y^2. Moreover, given the generated sample of size R, the following summary statistics can be computed:

$$\bar{\boldsymbol{\beta}} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\beta}^{(r)}, \qquad \bar{\boldsymbol{\rho}} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\rho}^{(r)}, \qquad \bar{\boldsymbol{\alpha}}_j = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\alpha}_j^{(r)},$$

$$\bar{\boldsymbol{\mu}}_m = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\mu}_m^{(r)} \quad \text{and} \quad \bar{\sigma}_y^2 = \frac{1}{R}\sum_{r=1}^{R}\sigma_y^{2(r)}.$$

4.3 Unknown extra parameter $\boldsymbol{\eta}$

In the previous MCMC algorithm, the parameter $\boldsymbol{\eta}$ of the $\mathcal{SMN}_r$ distributions is assumed to be known. However, if it is unknown, it is possible introduce a new

step in the algorithm, denoted by step 11.b, to deal with this parameter. Next, the step 11.b under some $\mathcal{SMN}_r$ distributions is provided.

11.b Generate $\boldsymbol{\eta}^{(l+1)} \sim p(\eta|\boldsymbol{\theta}^{(l)})$ according to the distribution of the random component $(\varepsilon, \boldsymbol{\xi}^{\mathrm{T}}, \mathbf{m}^{\mathrm{T}})^{\mathrm{T}}$.

(a) Slash distribution

$$p(\eta|\boldsymbol{\theta}) \propto \eta^{n+a_\eta-1} \exp\left[-\eta\left(b_\eta - \sum_{i=1}^{n} \log u_i\right)\right].$$

Thus, $\eta|\boldsymbol{\theta} \sim \mathcal{G}amma(n + a_\eta, b_\eta + \sum_{i=1}^{n} \log u_i)$. In this case, the prior distribution for η is $\eta \sim \mathcal{G}amma(a_\eta, b_\eta)$, where the hyperparameters $a_\eta > 0$ and $b_\eta > 0$ are assumed to be known.

(b) Contaminated normal distribution: Here, $\pi(\eta_1, \eta_2) = \pi(\eta_1)\pi(\eta_2)$, in which $\eta_1 \sim \mathcal{B}eta(a_{\eta_1}, b_{\eta_1})$ and $\eta_2 \sim \mathcal{T}run\mathcal{G}amma(\frac{a_{\eta_2}}{2}, \frac{b_{\eta_2}}{2}; (0, 1))$ are the prior distributions for $\boldsymbol{\eta}$, and

$$p(\eta_1|\boldsymbol{\theta}) \propto \eta_1^{a_{\eta_1}+g_\eta-1}(1-\eta_1)^{b_{\eta_1}+n-g_\eta-1} \quad \text{and}$$

$$p(\eta_2|\boldsymbol{\theta}) \propto \eta_2^{\frac{g_\eta+a_{\eta_2}}{2}-1} \exp\left[-\frac{\eta_2}{2}\left(\sum_{i:u_i\in g(\eta)} S_i + b_{\eta_2}\right)\right] I_{(0,1)}(\eta_2),$$

where $g_\eta = \sum_{i=1}^{n} U_i$ and

$$U_i = \begin{cases} 1, & \text{if } u_i = \eta_2, \\ 0, & \text{if } u_i = 1. \end{cases}$$

Then, $\eta_1|\boldsymbol{\theta} \sim \mathcal{B}eta(a_{\eta_1} + g_\eta, b_{\eta_1} + n - g_\eta)$ and $\eta_2|\boldsymbol{\theta} \sim \mathcal{T}run\mathcal{G}amma(\frac{g_\eta+a_{\eta_2}}{2}, \frac{\sum_{i:u_i\in g(\eta)} S_i + b_{\eta_2}}{2}; (0, 1))$, with the hyperparameters $a_{\eta_1} > 0$, $b_{\eta_1} > 0$, $a_{\eta_2} > 0$ and $b_{\eta_2} > 0$ are assumed to be known.

(c) Under Student-t and symmetric hyperbolic distributions a Metropolis–Hastings step is required.

4.4 Model comparison

In the literature, there are many methodologies for comparing the goodness-of-fit (penalized by the model complexity) of competitive models in order to select the one that best fits the data. In this paper, the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) and the conditional predictive ordinate (CPO) studied by Gelfand, Dey and Chang (1992), are considered. These measures can be estimated using a single sample drawn from the posterior distribution. The DIC is a generalization of the AIC (Akaike Information Criterion) based on the posterior mean of the deviance, and it is calculated as follows

$$\widehat{\mathrm{DIC}} = 2\overline{\mathrm{D}} - \mathrm{D}(\overline{\boldsymbol{\theta}}),$$

where $\overline{\mathrm{D}} = R^{-1}\sum_{r=1}^{R}\mathrm{D}(\boldsymbol{\theta}^{(r)})$ and $\overline{\boldsymbol{\theta}} = R^{-1}\sum_{r=1}^{R}\boldsymbol{\theta}^{(r)}$, with $\mathrm{D}(\boldsymbol{\theta}) = -2\sum_{i=1}^{n}\log f(\mathbf{y}_i^*|\boldsymbol{\theta})$ and $\boldsymbol{\theta}^{(r)}$ as the rth element of the posterior sample of $\boldsymbol{\theta}$, $r = 1, \ldots, R$, in which $f(\mathbf{y}_i^*|\boldsymbol{\theta}) \equiv f(\mathbf{y}_i^*|\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*, \eta)$, $\mathbf{y}_i^* = (y_i, \mathbf{M}_i^{\mathrm{T}})^{\mathrm{T}}$,

$$\boldsymbol{\mu}_i^* = \begin{pmatrix} \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \sum_{j=1}^{s}\mathbf{b}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j + \boldsymbol{\mu}_m^{\mathrm{T}}\boldsymbol{\rho} \\ \boldsymbol{\mu}_m \end{pmatrix} \quad \text{and}$$

$$\boldsymbol{\Sigma}_i^* = \begin{pmatrix} \sigma_y^2 + \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\Sigma}_m\boldsymbol{\rho} & \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\Sigma}_m \\ \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\Sigma}_m & \omega\sigma_y^2\mathbf{I}_q + \boldsymbol{\Sigma}_m \end{pmatrix}.$$

Given a set of candidate models, the model yielding the smallest value of the DIC could be considered as the one that best fits the data. The CPO is a measure based on the cross validation criterion. This measure is the predictive density of one observation conditional on the rest of the data. An estimate of CPO_i for a posterior sample of size R is given by

$$\widehat{\mathrm{CPO}}_i = \left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{f(\mathbf{y}_i^*|\boldsymbol{\theta}^r)}\right]^{-1}.$$

Therefore, a statistic that summarizes the CPO_i values is the log-marginal pseudo likelihood, which is expressed by $\mathrm{LMPL} = \sum_{i=1}^{n}\log(\widehat{\mathrm{CPO}}_i)$, where the best fit is obtained by the model with the larger value of this measure.

5 Simulation study

This section presents a simulation study aimed at evaluating the performance of the algorithm derived under the model in (2). A sample of size $n = 500$ of (y, M) is generated from the following mechanism

$$\begin{cases} y_i = \beta_1 + \beta_2 \mathrm{x}_i + \rho \mathrm{m}_i + \dfrac{1}{2}\sin(2\pi v_i) + \varepsilon_i, \\ \mathrm{M}_i = \mathrm{m}_i + \xi_i, \qquad i = 1, \ldots, n, \end{cases} \tag{3}$$

where $(\varepsilon_1, \xi_1, \mathrm{m}_1)^{\mathrm{T}}, \ldots, (\varepsilon_n, \xi_n, \mathrm{m}_n)^{\mathrm{T}}$ are independent and identically distributed random vectors with a $\mathcal{SMN}_3$ distribution; x_i is generated according to the $\mathcal{U}(-1, 1)$; $v_i \sim \mathcal{U}(0, 1)$; $\beta_1 = \beta_2 = 1$, $\rho = 0.5$, $\sigma_y^2 = 1$, $\mu_m = 1$, $\sigma_m^2 = 0.5$ and $\omega = 1$.

For the error term $(\varepsilon_i, \xi_i, \mathrm{m}_i)^{\mathrm{T}} \sim \mathcal{SMN}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ the following distributions are considered: (i) Normal; (ii) Student-t for $\eta = 3, 5, 8$ and 12; (iii) Slash for $\eta = 2, 4, 7$ and 11; (iv) Symmetric hyperbolic for $\eta = 0.8, 1.0, 1.2$ and 1.4; (v) Laplace; and (vi) Contaminated normal for $\boldsymbol{\eta} = (0.4, 0.2)^{\mathrm{T}}$, $(0.5, 0.2)^{\mathrm{T}}$, $(0.55, 0.2)^{\mathrm{T}}$ and $(0.6, 0.2)^{\mathrm{T}}$, where $\boldsymbol{\mu} = (0, 0, \mu_m)^{\mathrm{T}}$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_y^2, \sigma_\xi^2, \sigma_m^2)$. Next, a sample is taken from the conditional posterior distribution for model (3), in which $f(v) =$

$\frac{1}{2}\sin(2\pi v)$ is approximated by using cubic B-splines with $[n^{1/5}]$ internal knots. Prior distributions for model parameters are as considered in the previous section with the following hyperparameters values: $\beta_0 = \rho_0 = \mu_{m_o} = 0$, $S_\beta = S_\rho = \Sigma_{\mu_o} = \Omega_m = 10^3$, $\boldsymbol{\alpha}_0 = \mathbf{0}_K$, and $a_{\tau_\alpha} = b_{\tau_\alpha} = a = b = 0.001$. We consider scenarios where the parameter $\boldsymbol{\eta}$ is assumed to be known and unknown. The MCMC procedure with 55,000 iterations was implemented, which includes a *burn-in* of 5000 and thinning of 10 so that a sample of size $R = 5000$ was retained. This procedure was repeated 100 times, where the values of $\mathbf{x}$ and $\mathbf{v}$ are kept fixed. The following summary statistics were considered:

$$\mathrm{M}(\theta_j) = \frac{1}{100}\sum_{r=1}^{100}\bar{\theta}_j^{(r)}, \qquad \mathrm{D}(\theta_j) = \left\{\frac{1}{99}\sum_{r=1}^{100}[\bar{\theta}_j^{(r)} - \mathrm{M}(\theta_j)]^2\right\}^{1/2},$$

where $\theta_1 = \beta_1$, $\theta_2 = \beta_2$, $\theta_3 = \rho$, $\theta_4 = \mu_m$, $\theta_5 = \sigma_m^2$, $\theta_6 = \sigma_y^2$ and $\bar{\theta}_j^{(r)}$ are the posteriori mean of θ_j in the replicate j, $j = 1, \ldots, 100$. For the nonparametric component, the following summary statistic was considered:

$$\hat{f}(v) = \frac{1}{100}\sum_{j=1}^{100}\mathbf{B}_1\bar{\boldsymbol{\alpha}}_1^{(j)},$$

where $\bar{\boldsymbol{\alpha}}_1^{(j)}$ is a posterior mean of $\boldsymbol{\alpha}_1$ in the replicate j.

Tables 1 and 2 present the values of $\mathrm{M}(\cdot)$ and $\mathrm{D}(\cdot)$ for each one of parameters of model (3) in the simulation scenarios considered. It can be seen that estimates are close to the true parameter values. An exception seems to be parameter σ_m^2 for which to get a closer estimate larger sample sizes seem to be required. Moreover, in all cases the values of $\mathrm{D}(\cdot)$ increases as the tails of the error term distribution becomes heavier. In general, we observe that the values of $\mathrm{D}(\cdot)$ are higher when the extra parameter $\boldsymbol{\eta}$ is considered to be unknown.

The focus is now on studying the performance of the estimates of the function $f(v)$. With such in view, Figures 2 and 3 present the true function (full line) and their estimates (dotted lines) under different simulation scenarios. It can be concluded that the estimates of the nonparametric function present behaviour similar to the true one irrespective to the distribution of the error term $(\varepsilon_i, \xi_i, \mathrm{m}_i)^{\mathrm{T}}$. The estimates of the parameters and the nonparametric function are very close to the true values, even when the extra parameter is assumed to be unknown.

A second simulation study was performed. The data were simulated using the same setup than that of the first simulation study (that is, number of replicates, parameter values, size of posterior sample, thinning, etc.), but in all cases the contaminated normal distribution ($\mathcal{CN}(0.6, 0.2)$) was considered for the error term. Some $\mathcal{SMN}$ models were fitted to the data and the values of the M and D statistics were calculated. The results are summarized in the Table 3.

Table 1 *Summary statistics* M(·) *and* D(·) *when the parameter* $\boldsymbol{\eta}$ *is considered to be known*

	M						D					
Distribution	β_1	β_2	ρ	μ_m	σ_m^2	σ_y^2	β_1	β_2	ρ	μ_m	σ_m^2	σ_y^2
Normal	0.98	1.03	0.53	0.99	0.51	1.02	0.118	0.082	0.108	0.050	0.102	0.073
$t(3)$	0.93	0.96	0.54	1.01	0.65	1.08	0.148	0.107	0.132	0.058	0.179	0.118
$t(5)$	0.94	1.00	0.56	0.99	0.47	1.07	0.202	0.089	0.206	0.056	0.176	0.123
$t(8)$	0.88	1.01	0.62	1.02	0.41	1.06	0.271	0.088	0.260	0.049	0.179	0.112
$t(12)$	0.93	0.97	0.57	0.99	0.47	1.02	0.195	0.097	0.179	0.049	0.164	0.108
$\mathcal{S}l(2)$	0.90	0.98	0.59	1.02	0.39	1.06	0.205	0.098	0.201	0.061	0.156	0.108
$\mathcal{S}l(4)$	0.92	1.04	0.60	1.00	0.44	1.02	0.288	0.089	0.288	0.054	0.148	0.080
$\mathcal{S}l(7)$	0.90	1.00	0.60	1.00	0.38	1.03	0.271	0.082	0.271	0.057	0.151	0.097
$\mathcal{S}l(11)$	0.91	1.00	0.60	0.99	0.39	1.03	0.243	0.081	0.246	0.048	0.155	0.097
$\mathcal{SH}(0.8)$	0.87	0.95	0.61	0.98	0.40	1.06	0.339	0.123	0.295	0.078	0.204	0.124
$\mathcal{SH}(1.0)$	0.89	0.98	0.60	1.03	0.43	1.05	0.263	0.112	0.243	0.081	0.182	0.122
$\mathcal{SH}(1.2)$	0.92	1.04	0.57	0.99	0.47	1.04	0.200	0.096	0.191	0.057	0.149	0.099
$\mathcal{SH}(1.4)$	0.91	0.99	0.60	1.00	0.44	1.05	0.260	0.086	0.243	0.058	0.183	0.112
$\mathcal{CN}(0.4, 0.2)$	0.95	1.03	0.55	0.94	0.69	1.16	0.175	0.112	0.151	0.073	0.159	0.108
$\mathcal{CN}(0.5, 0.2)$	0.94	0.95	0.55	1.05	0.50	1.04	0.177	0.122	0.146	0.084	0.166	0.107
$\mathcal{CN}(0.55, 0.2)$	0.95	0.96	0.54	1.06	0.53	1.03	0.179	0.129	0.141	0.086	0.158	0.101
$\mathcal{CN}(0.6, 0.2)$	0.95	0.95	0.54	1.06	0.55	1.03	0.179	0.134	0.136	0.089	0.149	0.090
Laplace	0.96	0.98	0.54	1.04	0.57	1.05	0.184	0.160	0.150	0.133	0.196	0.118

Table 2 *Summary statistics* M(·) *and* D(·) *when the parameter* $\boldsymbol{\eta}$ *is considered to be unknown*

	M							D						
Distribution	β_1	β_2	ρ	μ_m	σ_m^2	σ_y^2	η	β_1	β_2	ρ	μ_m	σ_m^2	σ_y^2	η
Normal	0.98	1.03	0.53	0.99	0.51	1.02		0.118	0.082	0.108	0.050	0.102	0.073	
$t(3)$	0.92	0.98	0.56	1.02	0.67	1.08	3.35	0.189	0.085	0.174	0.057	0.192	0.140	0.422
$t(5)$	0.88	0.99	0.60	1.00	0.47	1.07	6.65	0.276	0.106	0.253	0.055	0.189	0.134	1.549
$t(8)$	0.87	1.01	0.63	1.01	0.40	1.06	11.85	0.284	0.081	0.278	0.048	0.194	0.134	4.758
$t(12)$	0.93	0.98	0.57	1.00	0.49	1.02	18.90	0.223	0.082	0.209	0.055	0.157	0.108	9.357
$\mathcal{S}l(2)$	0.89	0.99	0.60	1.03	0.40	1.06	2.17	0.255	0.096	0.235	0.069	0.161	0.124	0.314
$\mathcal{S}l(4)$	0.85	1.05	0.65	1.02	0.38	1.02	6.69	0.276	0.099	0.255	0.060	0.182	0.138	2.398
$\mathcal{S}l(7)$	0.92	1.01	0.59	1.00	0.37	1.03	9.11	0.214	0.078	0.194	0.054	0.140	0.104	2.564
$\mathcal{S}l(11)$	0.88	1.03	0.62	1.00	0.35	1.03	10.03	0.270	0.090	0.258	0.054	0.146	0.104	2.110
$\mathcal{SH}(0.8)$	0.91	0.96	0.60	0.99	1.32	1.06	1.77	0.286	0.151	0.313	0.087	0.918	2.302	0.688
$\mathcal{SH}(1.0)$	0.86	0.99	0.62	1.03	1.09	1.05	1.82	0.369	0.123	0.353	0.076	0.903	1.807	0.808
$\mathcal{SH}(1.2)$	0.89	1.01	0.62	0.98	1.13	1.04	2.22	0.274	0.098	0.273	0.063	1.015	1.608	1.000
$\mathcal{SH}(1.4)$	0.90	0.99	0.60	1.01	0.95	1.05	2.47	0.287	0.084	0.268	0.050	0.673	1.337	0.949
$\mathcal{CN}(0.4, 0.2)$	0.91	0.93	0.58	1.04	0.53	1.16	0.45, 0.25	0.232	0.108	0.209	0.069	0.197	0.212	0.088, 0.034
$\mathcal{CN}(0.5, 0.2)$	0.90	0.95	0.61	1.05	0.53	1.04	0.52, 0.26	0.328	0.128	0.313	0.076	0.236	0.318	0.098, 0.044
$\mathcal{CN}(0.55, 0.2)$	0.95	0.95	0.57	1.06	0.63	1.03	0.56, 0.27	0.197	0.118	0.168	0.075	0.267	0.370	0.109, 0.060
$\mathcal{CN}(0.6, 0.2)$	0.93	0.95	0.58	1.04	0.69	1.03	0.60, 0.27	0.267	0.124	0.235	0.073	0.289	0.388	0.111, 0.056

In all cases, the values of M are close to the true parameter values. However, according to the values of D, when the error distribution have heavier tails than those of the normal, the variability of the estimates around the true values is lower

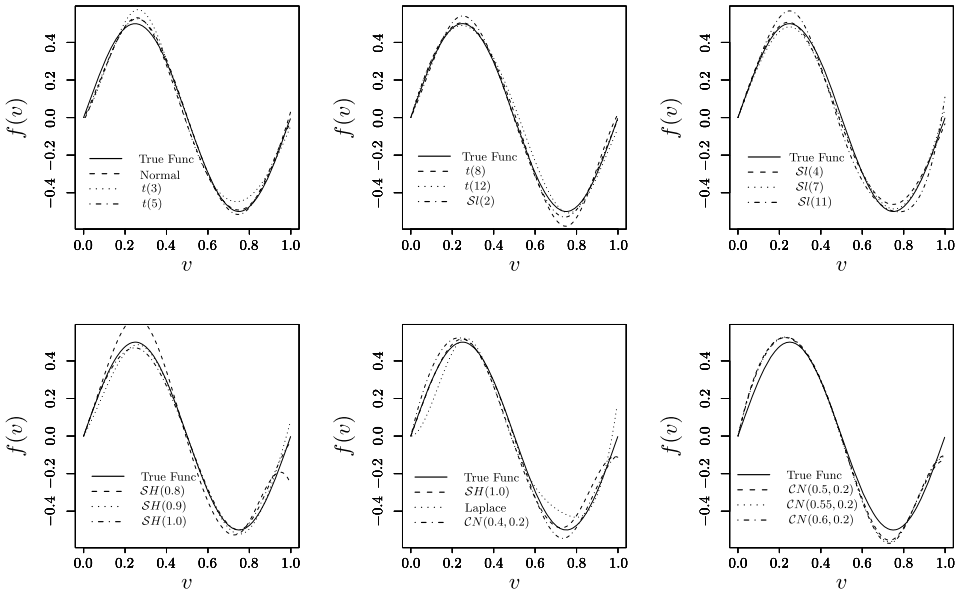

Figure 2 *True function* $f(v)$ (*full line*) *and estimates* (*dotted line*) *when the parameter* $\boldsymbol{\eta}$ *is considered known.*

than when the error distribution is normal. That is, under the presence of outlying observations, the models based on the $\mathcal{SMN}$ distributions (i.e., those with heavier tails than those of the normal) outperform the usual measurement error model.

5.1 Computational implementation

A function in the R (`www.R-project.org`) package **BayesGESM** has been developed to provide an easy way to apply the statistical methodology presented in this paper. The package **BayesGESM** (Rondon and Bolfarine, 2014) may be freely downloaded from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=BayesGESM`. In fact, the function `fmem()` supports the flexible measurement error models described in Section 3 under the following prior distributions

$$\begin{aligned}
\boldsymbol{\beta} &\sim \mathcal{N}_p(\mathbf{0}, 10^5\mathbf{I}_p), \qquad \boldsymbol{\rho} \sim \mathcal{N}_q(\mathbf{0}, 10^5\mathbf{I}_q),\\
\boldsymbol{\alpha} &\sim \mathcal{N}_K(\mathbf{0}, \tau_\alpha^2\mathbf{I}_K), \qquad \boldsymbol{\mu}_m \sim \mathcal{N}_q(\mathbf{0}, 10^5\mathbf{I}_q),\\
\boldsymbol{\Sigma}_m^{-1} &\sim \mathcal{W}ishart(q, 10^5\mathbf{I}_q),\\
\sigma_y^2 &\sim \mathcal{IG}\left(\frac{10^{-5}}{2}, \frac{10^{-5}}{2}\right) \quad \text{and}\\
\tau_\alpha^2 &\sim \mathcal{IG}(10^{-5}, 10^{-5}).
\end{aligned}$$

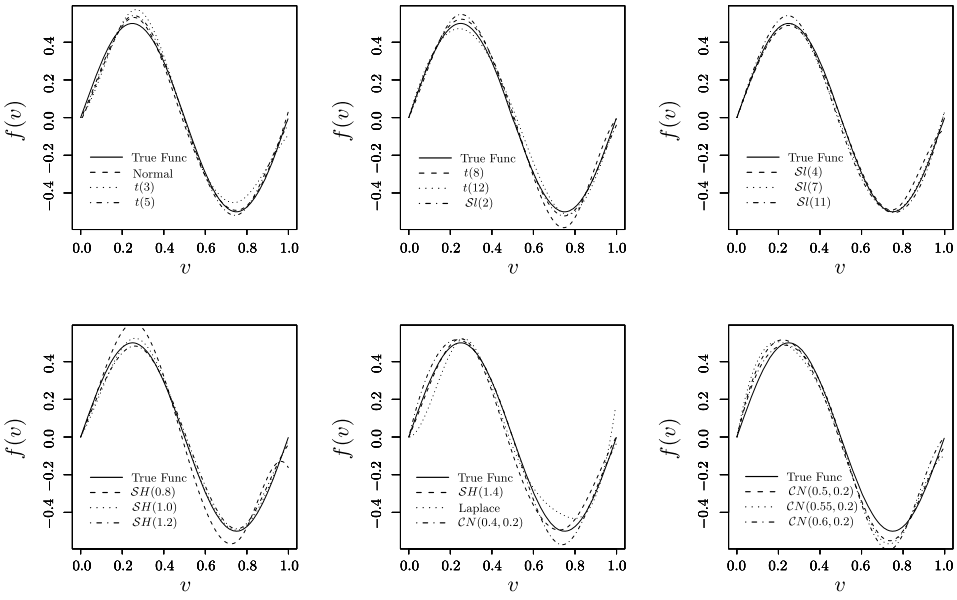

Figure 3 *True function* $f(v)$ *c*(*dotted line*) *and estimated ones* (*dotted lines*) *when the parameter* $\boldsymbol{\eta}$ *is considered known.*

Table 3 *Summary statistics* M(·) *and* D(·) *for second simulation study*

	M			D		
Distribution	β_1	β_2	ρ	β_1	β_2	ρ
Normal	0.9390	0.9175	0.5580	0.2461	0.1502	0.1940
$t(3)$	0.9499	0.9436	0.5476	0.1793	0.1335	0.1384
$\mathcal{S}l(2)$	0.9687	0.9504	0.5409	0.1745	0.1319	0.1354
$\mathcal{SH}(0.8)$	0.9479	0.9474	0.5411	0.1822	0.1402	0.1345
Laplace	0.9710	0.9339	0.5228	0.1616	0.1468	0.1314
$\mathcal{CN}(0.6, 0.2)$	1.0027	0.9447	0.4900	0.1561	0.1309	0.1308

The arguments of the function `fmem()` are

```
fmem(formula, data, family, eta, omeg=1, burn.in,
  post.sam.s, thin=1).
```

The systematic component of the model must be specified in the argument `formula`. This argument comprises of three parts, namely: (i) observed response variable; (ii) covariates with measurement error; and (iii) covariates without measurement error. The first two parts are separated by the symbol "∼" and the second and third parts are separated by the symbol "|". In addition, the nonparametric component can be specified in the third part of `formula` by using the function

`bsp()`. For example, a model with observed response variable `y`, covariates `x1` and `x2` with measurement error, and covariates `w1` and `w2` without measurement error, where the latter has a nonlinear effect whose functional form is unknown, must be specified as

```
fmem(formula= y ~ x1 + x2 | w1 + bsp(w2), ...).
```

Moreover, the distribution of the multivariate error term and its extra parameter must be specified in the arguments `family` and `eta`, respectively. The supported distributions include normal, Student-t, Laplace, symmetric hyperbolic, slash, and contaminated normal. The absence of argument `eta` in the call to the function `fmem()` indicates that a model with a unknown extra parameter $\boldsymbol{\eta}$ is required. The arguments `burn.in` and `post.sam.s` are used to specify the number of burn-in iterations and the required sample size for the posterior distribution. Additionally, the option `thin` allows for specification of the thinning interval to be used in the simulation to obtain the required sample size for the posterior distribution.

The function `fmem()` provides a matrix with the simulated chains for all interest parameters, in which each column represents the marginal posterior sample of each involved parameter. In addition, the function `fmem()` calculates goodness-of-fit measures such as DIC and LMPL as well as some diagnostic measures such as residuals and global influence measures based in Kullback–Leibler divergence and the χ^2-distance proposed by Weiss and Cook (1992) and Peng and Dey (1995). Examples and additional information on the `fmem()` function can be found in `http://cran.r-project.org/web/packages/BayesGESM/BayesGESM.pdf`.

6 Application

In this section, the data set BOSTON is used to illustrate the proposed methodology. This data set, which consists of 506 individuals and was discussed by Belsley, Kuh and Welsch (2005) and Harrison and Rubinfeld (1978), relates the impact of the air pollution and other explanatory variables on the price of owner-occupied homes in Boston. The interest variable is the logarithm of the median value of owner-occupied homes (`log(medv)`), and it is related with 14 explanatory variables, six of them are defined from census track and the remaining variables are defined for clusters. To this illustration, the following variables are considered:

- `nox`: nitrogen oxides concentration (parts per 10 million);
- `crim`: per capita crime rate by town;
- `rm`: average number of rooms;
- `lstat`: lower status of the population (percent);
- `dis`: weighted mean of distances to five Boston employment centres.

The dispersion graphs of variable `log(medv)` against explanatory variables `lstat` and `dis` suggest a nonlinear relation between them, so that these relations are described by using nonparametric functions. The explanatory variable `nox` is assumed to be measured with error because this type of variables are prone to errors introduced by the measurement tools. Therefore, the following model is used to analyze the data set:

$$\begin{cases} \texttt{log(medv)}_i = \beta_0 + \beta_1 \texttt{crim}_i + \beta_2 \texttt{rm}_i + \rho_1 \texttt{nox}_i + f_1(\texttt{lstat}_i) \\ \qquad\qquad + f_2(\texttt{dis}_i) + \varepsilon_i \\ \texttt{Nox}_i = \texttt{nox}_i + \xi_i, \qquad i = 1, \ldots, 506, \end{cases}$$

where $(\varepsilon_1, \xi_1, \texttt{nox}_1)^{\mathrm{T}}, \ldots, (\varepsilon_n, \xi_n, \texttt{nox}_n)^{\mathrm{T}}$ are independent and identically distributed random vectors with $\mathcal{SMN}_3$ distribution. Several values of ω were considered, but its value was fixed at 4 because in that case the best value of the DIC statistic is obtained. Furthermore, the following independent prior distributions are specified:

$$\begin{aligned} \beta_0 &\sim \mathcal{N}(0, 10^3), \qquad \beta_j \sim \mathcal{N}(0, 10^3), \\ \rho_1 &\sim \mathcal{N}(0, 10^3), \qquad \boldsymbol{\alpha}_j \sim \mathcal{N}_{K_j}(\mathbf{0}, \tau^2_{\alpha_j} \mathbf{I}_{K_j}), \\ \mu_{\texttt{nox}} &\sim \mathcal{N}(0, 10^3), \qquad \tau^2_{\alpha_j} \sim \mathcal{GI}(0.001, 0.001) \quad \text{and} \\ \sigma^2_y &\sim \mathcal{GI}(0.0005, 0.0005), \qquad j = 1, 2. \end{aligned}$$

60,000 iterations of the MCMC algorithm described in Section 4 are simulated, which include a burn-in period of 10,000 and thinning of 10 iterations, so that an approximately independent sample of size $R = 5000$ is obtained. For the error term $(\varepsilon_i, \xi_i, \texttt{nox}_i)^{\mathrm{T}}$, normal, Student-$t$, slash, symmetric hyperbolic, Laplace and contaminated normal distributions were considered. In all cases, the extra parameters are assumed to be unknown. Model comparison was implemented by using DIC and LMPL criteria. According to the DIC statistic the best model is that with the lowest value on this criterion. Similarly, according to the LMPL statistic the best model is that with the highest value on this criterion.

Table 3 presents the values of these criteria for each considered model. The model where the error term follows contaminated normal distribution presents the smallest DIC and the highest LMPL values. Thus, it may be considered as the best model to describe the data set. In `fmem()` the chosen model can be fitted via

```
fit <- fmem(log(medv) ~ nox | crim + rm + bsp(lstat)
     + bsp(dis), data=Boston,
    family="ContNormal", burn.in=10,000,
    post.sam.s=5000, omeg=4, thin=10)
summary(fit)
```

Table 4 *DIC and LMPL values for the considered models*

Distribution	DIC	LMPL
Normal	−987.34	489.03
Student-*t*	−1011.78	501.83
$\mathcal{SI}$	−1006.99	498.27
$\mathcal{SH}$	−1010.62	501.39
$\mathcal{CN}$	**−1017.56**	**504.63**
Laplace	−972.60	480.89

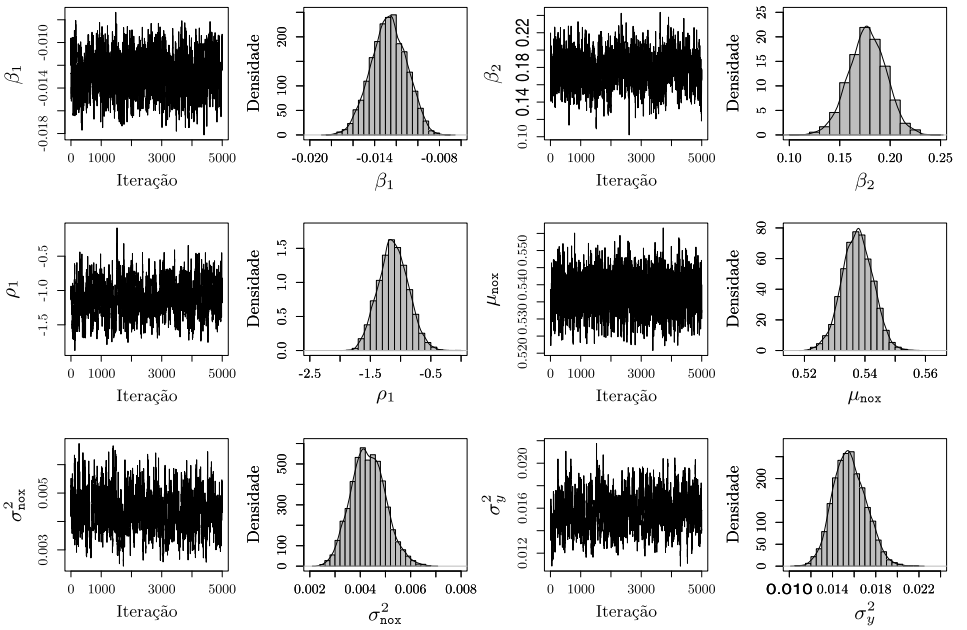

Figure 4 *Chains and marginal posterior densities plots for parameters* β_1, β_2, ρ_1 , $\mu_{\texttt{nox}}$, $\sigma^2_{\texttt{nox}}$ *and* σ^2_y *under the contaminated normal model.*

Figure 4 reveals the behaviour of the chains as well as the approximation of the marginal posterior densities for β_1, β_2, ρ_1, $\mu_{\texttt{nox}}$, $\sigma^2_{\texttt{nox}}$ and σ^2_y corresponding to the contaminated normal model. These graphs suggest that the convergence was reached and that the marginal posterior densities are approximately symmetric for parameters β_1, β_2, ρ_1 and $\mu_{\texttt{nox}}$. Figure 5 depicts the behavior of the nonparametric functions $f_1(\texttt{lstat})$ and $f_2(\texttt{dis})$, which are approximated by using *B*-splines under the contaminated normal model. A nonlinear behavior is revealed not only by the data set but also by the estimated functions. Analysis of residu-

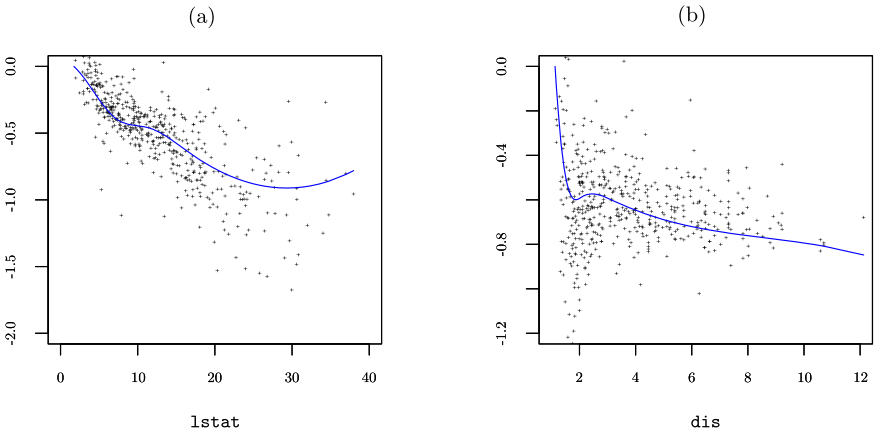

Figure 5 *Graphs of* $\hat{f}_1(\texttt{lstat})$ (a) *and* $\hat{f}_2(\texttt{dis})$ (b) *under the contaminated normal model.*

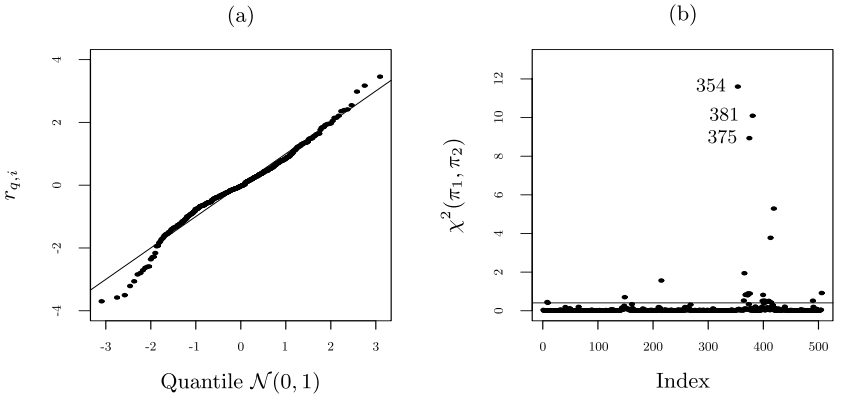

Figure 6 *Graphs of the quantile residuals* (a) *and the influence measure* (b) *for the normal model fitted to the Boston data.*

als and influence measures (omitted here) indicate that the contaminated normal model provides a suitable fit to the data set.

Figures 6 and 7 present some diagnostic plots for the normal and contaminated normal models fitted to the Boston data. These graphs indicate that the contaminated normal model fit the data better than the normal one, especially in the left tail of the distribution. Furthermore, the graphs of the influence measures suggest that the contaminated normal model is also less sensitive than the normal one. Therefore, the contaminated normal model is chosen as the best model to describe the Boston data.

To investigate how measurement errors in covariate `nox` affects modelling results, we further compare two methods of estimation. One is based on the flexible measurement error model (FMEM) taking into account measurement errors

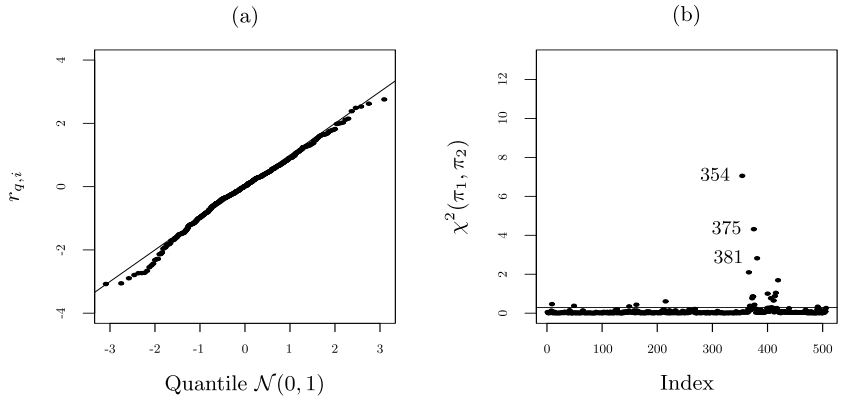

Figure 7 *Graphs of the quantile residuals* (a) *and the influence measure* (b) *for the contaminated normal model fitted to the Boston data.*

Table 5 *Behavior of parameter estimates for different values of* ω *for* $\mathcal{CN}(0.3, 0.5)$ *model*

Model	Measure	β_0	β_1	β_2	ρ_1	σ_y^2	DIC	LMPL
With measurement	Mean	3.6988	−0.0127	0.1773	−1.1192	0.0156	−1017.565	504.631
error*	Standard D.	0.2213	0.0016	0.0185	0.2441	0.0015		
Without measurement	Mean	3.4241	−0.0127	0.1774	−0.6132	0.0180	−1017.301	504.528
error**	Standard D.	0.2046	0.0016	0.0184	0.1492	0.0014		

*$\omega = 4$, i.e., $\sigma_\xi^2 = \sigma_y^2/4$; **$\omega = 10{,}000$, i.e., $\sigma_\xi^2 = \sigma_y^2/10{,}000 \approx 0$.

through the equation $\texttt{Nox}_i = \texttt{nox}_i + \xi_i$. The other, the *naive* method, uses the raw (observed) $\texttt{Nox}_i$ to directly replace the true (unobserved) $\texttt{nox}_i$. It can be seen from Table 5 that there are important differences in the estimates of ρ_1. The naive model may produce unreasonable estimates and substantially underestimate the covariate Nox effect. Furthermore, the model that best fits the data is the FMEM as it has the smaller DIC value and the highest LMPL value.

7 Conclusions

A regression model based on scale mixture of normal distributions and covariates measured with errors, is introduced. Parameter estimation is conducted by using the Bayesian approach, which, given the model complexity, is implemented by using Monte Carlo Markov chains. The nonparametric part of the model is approximated via *B*-splines. Simulation studies and a real data application reveal that this approach can produce satisfactory estimates and can be useful in practice.

References

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 99–102. MR0359122

Arellano-Valle, R. B., Bolfarine, H. and Labra, V. (1996). Ultrastructural elliptical models. *Canadian Journal of Statistics* **24**, 207–216. MR1406176

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society. Series A, Mathematical, Physical and Engineering Sciences* **353**, 401–419.

Belsley, D. A., Kuh, E. and Welsch, R. E. (2005). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley. MR0576408

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.

Cao, C. Z., Lin, J. G. and Zhu, X. X. (2012). On estimation of a heteroscedastic measurement error model under heavy-tailed distributions. *Computational Statistics and Data Analysis* **56**, 438–448. MR2842351

Carroll, R. J., Roeder, K. and Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics* **55**, 44–54.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. Boca Raton: Chapman and Hall. MR2243417

Cheng, C. L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. London: Arnold. MR1719513

De Boor, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences.* New York: Springer. MR0507062

de Castro, M., Bolfarine, H. and Galea, M. (2013). Bayesian inference in measurement error models for replicated data. *Environmetrics* **24**, 22–30. MR3042271

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121. MR1435485

Fuller, M. A. (1987). *Measurement Error Models*. New York: Wiley. MR0898653

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*, 2nd ed. Boca Raton, FL: Chapman and Hall. MR2260716

Gelfand, A., Dey, D. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics* **4**, 147–167. MR1380275

Harrison, D. Jr. and Rubinfeld, D. L (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management* **5**, 81–102.

He, X., Fung, W. K. and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association* **100**, 1176–1184. MR2236433

Jörgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. *Lecture Notes in Statistics* **9**. New York: Springer. MR0648107

Kelly, B. C. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal* **665**, 1489–1506.

Kulathinal, S. B., Kuulasmaa, K. and Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement errors vary betwwen the observations. *Statistics in Medicine* **21**, 1089–1101.

Lemonte, A. J. and Patriota, A. G. (2011). Multivariate elliptical models with general parameterization. *Statistical Methodology* **8**, 389–400. MR2800359

Li, L., Palta, M. and Shao, J. (2004). A measurement error model with a Poisson distributed surrogate. *Statistics in Medicine* **23**, 2527–2536.

Maronna, R. A., Martin, D. R. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. New York: Wiley. MR2238141

Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae* **92**, 97–111. MR2265333

Patriota, A. G., Bolfarine, H. and de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology* **6**, 408–423. MR2751083

Peng, F. and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics* **23**, 199–213.

Rondon, L. M. and Bolfarine, H. (2014). BayesGESM: Bayesian Analysis of Generalized Elliptical Semiparametric Models. R package version 1.1. http://CRAN.R-project.org/package=BayesGESM.

Rogers, W. H. and Tukey, J. W. (1972). Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica* **26**, 211–226. MR0383605

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Var der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Methodological)* **64**, 583–640. MR1979380

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550. MR0898357

Weiss, R. and Cook, R. (1992). A grafical case statistic for assessing posterior influence. *Biometrika* **79**, 51–55.

Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo
Brazil
and
Departamento de Estadística
Universidad Nacional de Colombia
Bogotá
Colombia
E-mail: lumarp@gmail.com

Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo
Brazil
E-mail: hbolfar@ime.usp.br