

Bayesian Analysis of Censored Linear Regression Models with Scale Mixtures of Normal Distributions

Aldo M. Garay^a and Heleno Bolfarine^a

and Victor H. Lachos^{b1} and Celso R. B. Cabral^c

^a Department of Statistics, São Paulo University, Brazil

^b Department of Statistics, Campinas State University, Brazil

^c Department of Statistics, Federal University of Amazonas, Brazil

Abstract

As is the case of many studies, the data collected are limited and an exact value is recorded only if it falls within an interval range. Hence, the responses can be either left, interval or right censored. Linear (and nonlinear) regression models are routinely used to analyze these types of data and are based on normality assumptions for the errors terms. However, those analyses might not provide robust inference when the normality assumptions are questionable. In this article, we develop a Bayesian framework for censored linear regression models by replacing the Gaussian assumptions for the random errors with scale mixtures of normal (SMN) distributions. The SMN is an attractive class of symmetric heavy-tailed densities that includes the normal, Student-t, Pearson type VII, slash and the contaminated normal distributions, as special cases. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference. A new hierarchical prior distribution is suggested for the degrees of freedom parameter in the Student-t distribution. The likelihood function is utilized to compute not only some Bayesian model selection measures but also to develop Bayesian case-deletion influence diagnostics based on the q-divergence measures. The proposed Bayesian methods are implemented in the R package `BayesCR`. The newly developed procedures are illustrated with an application and simulated data.

Keywords: Bayesian modeling, Censored regression models, MCMC, Scale mixtures of normal distributions.

¹*Address for correspondence:* Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: hlachos@ime.unicamp.br.

1 Introduction

Regression models with normal observational errors are usually applied to model symmetrical data. However, it is well-known that several phenomena are not always in agreement with the assumptions of the normal model, yielding data with heavier tails or skewed or multimodal distributions. A good alternative is to consider errors with a more flexible class of distributions, such as, the Student-t distribution. For instance, Fernández and Steel (1999) discuss some inferential procedures in regression models with Student-t distribution for the errors. Pinheiro et al. (2001) proposed a multivariate Student-t linear mixed model (t-LME) and demonstrated its robustness against outliers through extensive simulations. Lin and Lee (2006) and Lin and Lee (2007) developed some additional tools for t-LME from likelihood-based and Bayesian perspectives. More recently, Ibacache-Pulgar and Paula (2011), proposed local influence measures in the Student-t partially linear regression model. Other existing methods for robust estimation are based on the class of scale mixtures of normal (SMN) distributions. This class of distributions is symmetric and thick-tailed and includes as special cases many symmetric distributions, such as, the normal, Pearson type VII, Student-t, slash and the contaminated normal. For an overview and applications, see for instance, Andrews and Mallows (1974), Lange and Sinsheimer (1993) and Rosa et al. (2003).

In this work, we are interested in fitting regression models when the responses are possibly censored. Censoring occurs in several practical situations, for reasons such as limitations of measuring equipment or from experimental design. Roughly speaking, a censored observation contains only partial information about an event of interest. For example, the needle of a scale that does not provide a reading over 200 kg will show 200 kg for all the objects that weigh more than the limit. Another interesting example is the following, extracted from Breen (1996): on a school examination, the pass mark is 40%. A certificate, containing the status of the student (passed or not passed) is given to all of them, but only the students who meet the pass mark have reported their scores. Suppose that we want to study the relation between the scores and some other explanatory variables, like social class, gender and parental education. In this case, the scores are the responses and are *left-censored* because, if y_i denotes the score of student i and he or she did not meet the limit, we only know that $y_i \leq 39$. The case of censored responses with normal observational errors, denoted by N-CR, has been studied extensively

in the literature, see for example, Nelson (1977), Stapleton and Young (1984), Chib (1992), Thompson and Nelson (2003), Park et al. (2007) and Vaida and Liu (2009), to mention a few. However, to the best of our knowledge, a little attention has been given to the issue of observational errors having distribution with heavier tails than the normal. In censored linear regression models with Student-t distribution, the works by Arellano-Valle et al. (2012) and Matos et al. (2013) can be mentioned, but they use a classic approach. Besides, they only deal with the issue of estimation.

From a Bayesian point of view, conventional methods and analytic approximations have become unattractive for computation of sophisticated models within a Bayesian framework. The computer-intensive sampling methods, such as Markov chain Monte Carlo (MCMC) methods, (Smith and Roberts, 1993; Tierney, 1994), provide an efficient alternative for handling complicated Bayesian calculations. In particular, the Gibbs-type algorithm (Gelfand et al., 1990) approach enables simulating posterior samples from a set of iteratively updated conditional distributions. Choy and Smith (1997) make use of normal scale mixtures representation for the Student-t, symmetric stable and exponential power (EP) densities in Bayesian hierarchical models. In the context of mixed models, Rosa et al. (2003) changed the model with normal effects and Student-t random errors for a distribution belonging to the class of SMN distributions and Frühwirth-Schnatter et al. (2004) analyzed properties of the Gibbs-type algorithm suggested by Lenk and DeSarbo (2000), where the random effects are mixtures of distributions that belong to exponential families, and proposed alternative algorithms based on blocking and collapsing, to improve mixing properties.

In this paper, we propose a robust parametric model of the censored linear regression models based on the SMN distributions, denoted by SMN-CR, from a Bayesian perspective using a Bayesian framework to carry out posterior inference. In addition, we suggest an efficient Gibbs-type algorithm for posterior Bayesian inference and discuss some Bayesian diagnostic measures based on the q -divergence, as proposed by Peng and Dey (1995) and Lachos et al. (2013), to detect influential observations, which are an essential part of the analysis when using this kind of model, showing the drawbacks of the normal one and justifying the usefulness of more flexible class of the SMN distributions. These Bayesian diagnostic measures can be easily implemented directly from the MCMC output.

The rest of the paper is organized as follows. In Section 2, after briefly outlining some basic notations and conventions, we introduce the SMN class of distributions.

Then, the SMN censored linear regression model is presented and defined. In Section 3, we present a Gibbs-type algorithm for Bayesian estimation, specifying priors distributions for the parameters of interest. The model selection and influence diagnostics issue is considered in Section 4. The proposed method is illustrated in Section 5, by considering the analysis of simulated data sets. An application considering a data set of housewife wages is presented in Section 6. Section 7 concludes with a short discussion of issues raised by our study and some possible directions for the future research

2 Linear regression models with censored responses

2.1 Basic Notations and Conventions

Vectors of \mathbb{R}^p are $p \times 1$ matrices and are represented by boldface letters when $p > 1$. In general, we denote a random vector by an upper case letter and its realization by the corresponding lower case letter – like \mathbf{Y} and \mathbf{y} , for example. The transpose of a matrix \mathbf{A} is denoted by \mathbf{A}^\top . The identity matrix of order p is denoted by \mathbf{I}_p . The null vector of order p is denoted by $\mathbf{0}_p$. $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Omega})$ denotes a random vector with p -variate normal distribution with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Omega}$ – we drop the index p in the univariate case; $N_p(\cdot | \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the corresponding density; $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution.

Let \mathbf{Y} be a random vector and $\mathcal{B} \subset \mathbb{R}$ a Borel set. The conditional distribution of $\mathbf{Y} | (\mathbf{Y} \in \mathcal{B})$ is called *distribution of \mathbf{Y} truncated to \mathcal{B}* . Assuming that \mathbf{Y} has density $g(\cdot)$ and that $P(\mathbf{Y} \in \mathcal{B}) > 0$, the density of $\mathbf{Y} | (\mathbf{Y} \in \mathcal{B})$ is given by

$$f(\mathbf{y}) = \frac{1}{P(\mathbf{Y} \in \mathcal{B})} g(\mathbf{y}) \mathbb{I}_{\mathcal{B}}(\mathbf{y}), \quad (1)$$

where $\mathbb{I}_{\mathcal{B}}(\cdot)$ denotes the indicator function, that is, $\mathbb{I}_{\mathcal{B}}(\mathbf{y}) = 1$ if $\mathbf{y} \in \mathcal{B}$ and $\mathbb{I}_{\mathcal{B}}(\mathbf{y}) = 0$ otherwise.

For two arbitrary random vectors \mathbf{X} and \mathbf{Y} , $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ means that \mathbf{X} has the same distribution as \mathbf{Y} and $\mathbf{X} \perp \mathbf{Y}$ means that they are independent. If \mathbf{Y}_i , $i = 1, \dots, n$ are random vectors, then $\mathbf{Y}_i \stackrel{\text{iid.}}{\sim} P$ means that all \mathbf{Y}_i 's are independent and have the same distribution P . $\mathbf{Y}_i \stackrel{\text{ind.}}{\sim} P_i$ means that the \mathbf{Y}_i 's are independent, and each \mathbf{Y}_i follows a P_i distribution.

2.2 The scale mixtures of normal (SMN) distributions

In this section we define the censored regression model. Before this, let us recall that a random variable Y is said to follow a SMN distribution with location parameter μ and scale parameter $\sigma^2 > 0$, that we denote by $Y \sim \text{SMN}(\mu, \sigma^2, \boldsymbol{\eta})$, if it has the following stochastic representation:

$$Y \stackrel{d}{=} \mu + U^{-1/2}Z; \quad Z \sim N(0, \sigma^2); \quad Z \perp U, \quad (2)$$

where U is a positive random variable with cdf $H(\cdot|\boldsymbol{\eta})$ and $\boldsymbol{\eta}$ is a scalar or vector parameter indexing the distribution of U . When $\mu = 0$ and $\sigma^2 = 1$ we call it *the standard SMN distribution*. In this case, we denote the distribution function by $F_{SMN}(\cdot)$.

The random variable U can be discrete or continuous and the form of a SMN distribution is determined by the distribution of U . Thus, for the standard cases, we have:

- *Pearson type VII distribution:* In this case we consider $U \sim \text{Gamma}(\boldsymbol{\eta}/2, \delta/2)$, with $\boldsymbol{\eta} > 0$ and $\delta > 0$, where $\text{Gamma}(a, b)$ denotes the gamma distribution with mean a/b . The density of the random variable Y , defined in (2), takes the form

$$f_{PVII}(y|\boldsymbol{\eta}, \delta) = \frac{1}{B(\boldsymbol{\eta}/2, 1/2)\sqrt{\delta}} \left(1 + \frac{y^2}{\delta}\right)^{-\frac{\boldsymbol{\eta}+1}{2}}, \quad y \in \mathbb{R},$$

where $\boldsymbol{\eta}$ and δ are shape parameters and $B(a, b)$ represents the beta function. When $\delta = \boldsymbol{\eta}$ we have the Student-t distribution with degrees of freedom $\boldsymbol{\eta}$. The Student-t distribution tends to the normal case when $\boldsymbol{\eta} \rightarrow \infty$. Also, we have the Cauchy distribution when $\delta = \boldsymbol{\eta} = 1$.

- *Slash distribution:* In this case we have $U \sim \text{Beta}(\boldsymbol{\eta}, 1)$, with positive shape parameter $\boldsymbol{\eta}$. The density of the random variable Y , defined in (2), is given by

$$f_{sl}(y|\boldsymbol{\eta}) = \boldsymbol{\eta} \int_0^1 u^{\boldsymbol{\eta}-1} \phi(y\sqrt{u}) du, \quad y \in \mathbb{R}.$$

- *Contaminated normal distribution:* Here U is a discrete random variable taking one of two states 1 or γ . In this case the probability function of U is given by

$$h(u|\boldsymbol{\eta}, \gamma) = \boldsymbol{\eta} \mathbb{I}_{\{\gamma\}}(u) + (1 - \boldsymbol{\eta}) \mathbb{I}_{\{1\}}(u) \quad \boldsymbol{\eta}, \gamma \in (0, 1).$$

It follows immediately that the density of the random variable Y , defined in (2), is given by

$$f_{CN}(y|\boldsymbol{\eta}, \gamma) = \boldsymbol{\eta}\phi(y|0, \gamma^{-\frac{1}{2}}) + (1 - \boldsymbol{\eta})\phi(y).$$

From relation (2), it is straightforward to obtain the respective densities for general location and scale parameters μ and σ^2 , respectively. In this case, we use the notations $Y \sim \text{PVII}(\mu, \sigma^2; \boldsymbol{\eta}, \delta)$, $Y \sim \text{SL}(\mu, \sigma^2; \boldsymbol{\eta})$ and $Y \sim \text{SCN}(\mu, \sigma^2; \boldsymbol{\eta})$ to say that Y has the Pearson type VII distribution and so on. In the case of the Student-t distribution we write $Y \sim \text{t}_{\boldsymbol{\eta}}(\mu, \sigma^2)$. The respective pdf's are denoted by $f_{PVII}(\cdot|\mu, \sigma^2, \boldsymbol{\eta}, \delta)$, $f_{sl}(\cdot|\mu, \sigma^2, \boldsymbol{\eta})$, etc.

Now, consider a random variable Y with a truncated SMN distribution on the interval $[a, b]$, for $b > a$, that is, $Y \stackrel{d}{=} X|(X \in [a, b])$, where $X \sim \text{SMN}(\mu, \sigma^2, \boldsymbol{\eta})$. Here, $-\infty \leq a < b \leq \infty$ and the notation $[a, b]$ means that the interval can be open, closed or half-open. In this case, according to (1), Y has density

$$f_{TSMN}(y|\mu, \sigma^2, \boldsymbol{\eta}; [a, b]) = f_{SMN}(y|\mu, \sigma^2, \boldsymbol{\eta}) \left[F_{SMN}\left(\frac{b-\mu}{\sigma}\right) - F_{SMN}\left(\frac{a-\mu}{\sigma}\right) \right]^{-1} \mathbb{I}_{[a,b]}(y),$$

where F_{SMN} is the cdf of the SMN distribution. We use the notation $Y \sim \text{TSMN}(\mu, \sigma^2, \boldsymbol{\eta}; [a, b])$.

2.3 The SMN censored linear regression model

The SMN linear regression model is a natural extension of the normal linear regression model and is defined as

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \text{SMN}(0, \sigma^2, \boldsymbol{\eta}), \quad (3)$$

$i = 1, \dots, n$, where Y_i is a response, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of explanatory variable values for subject i and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of regression parameters. Estimation and diagnostic analysis for this model have been widely discussed in the literature, see for example, Cysneiros and Paula (2005), Villegas et al. (2012) and Zhu et al. (2013).

In this work we are interested in the situation in which the vector of responses is not fully observed for all subjects. Assuming left-censoring, the observations are of the form

$$Y_{\text{obs}_i} = \begin{cases} \kappa_i & \text{if } Y_i \leq \kappa_i; \\ Y_i & \text{if } Y_i > \kappa_i, \end{cases} \quad (4)$$

$i = 1, \dots, n$, for some threshold point κ_i . The compound of (3) and (4) is called the *SMN censored linear regression model* or, hereafter, the *SMN-CR model*. Supposing that there are $1 \leq m \leq n$ censored values of the characteristic of interest, we can partition the observed sample $\mathbf{y}_{\text{obs}} = \{y_{\text{obs}_1}, \dots, y_{\text{obs}_n}\}$ in two subsamples of m censored – y_{obs_i} , $i = 1, \dots, m$, say – and $n - m$ uncensored values (or one sample with n censored values, when $n = m$). Consider the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\eta})^\top$. The log-likelihood function of the SMN-CR model is given by

$$\pi(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}) = \sum_{i=1}^m \log \left[F_{SMN} \left(\frac{\kappa_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right] + \sum_{i=m+1}^n \log [f_{SMN}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta})]. \quad (5)$$

For simplicity, we will assume that the data are left-censored with the SMN-CR model being thus defined. The extensions of these results to arbitrary censoring can be easily presented.

3 Bayesian Inference for the SMN-CR model

In what follows, for any random vectors \mathbf{X} and \mathbf{Y} , we use $\pi(\mathbf{x})$ to denote the density of \mathbf{X} and $\pi(\mathbf{x}|\mathbf{y})$ to denote the conditional density of $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ which, although being an abuse of notation, greatly simplifies the exposition.

3.1 Prior distributions

In the Bayesian context, distributional prior specifications are needed for posterior inference. We assume that $\boldsymbol{\beta} \sim N_p(\mathbf{b}_0, \mathbf{S}_\beta)$, where \mathbf{b}_0 is a $p \times 1$ fixed hyperparameters vector and \mathbf{S}_β is a $p \times p$ known positive definite matrix. We suppose that the common scale parameter σ^2 follows an inverted gamma distribution, that is, $\sigma^{-2} \sim \text{Gamma}(a/2, b/2)$, where $a > 0$ and $b > 0$ are known.

Each specific model in the SMN distributions has a scale factor parameter with specific interpretation and obviously deserves a different treatment for prior choice. For instance, there are several suggestions for the estimation of the unknown degree of freedom of the Student-t distribution. In this respect, see excellent discussions in Geweke (1993), Fonseca et al. (2008) and Cabral et al. (2012). Based in the simulation study given in Subsection 5.2, we will use the suggestion given by Cabral et al. (2012), that is, $\boldsymbol{\eta} \sim \text{Exp}(\lambda)$ and $\lambda \sim U(c, d)$ where $\text{Exp}(\lambda)$ denotes the

exponential distribution with mean $1/\lambda$ ($\lambda > 0$) and $U(c, d)$ denotes the uniform distribution on the interval (c, d) . For the slash distribution, we consider the same suggestion given by Cabral et al. (2012) for the degree of freedom of the Student-t distribution, $\eta \sim \text{Exp}(\lambda)$ and $\lambda \sim U(e, f)$, where e and f are known and positive. For the contaminated normal distribution, a simple prior setup can be considered, such as $\eta \sim \text{Beta}(\eta_0, \eta_1)$ and $\gamma \sim \text{Beta}(\gamma_0, \gamma_1)$, where $\eta_0, \eta_1, \gamma_0, \gamma_1$ are known and positive. As we show below, the prior choice of η can be highly influential on the posterior distribution tail thickness.

Supposing that the elements of θ are a priori independent, the complete prior distribution is:

$$\pi(\theta) = \pi(\beta) \pi(\sigma^2) \pi(\eta). \quad (6)$$

3.2 MCMC estimation

In the Bayesian framework, estimators are obtained as characteristics associated to the posterior distribution, like expectations, modes, etc. Due to its complex form, it is clear that it is prohibitive to approximate its moments using techniques like numerical integration. Nowadays, it is well-known that an efficient way to approximate these integrals is through the generation of samples from the posterior distribution via an MCMC-type algorithm (Gamerman and Lopes, 2006). In our case, this algorithm can be easily developed using a data augmentation scheme, which is shown next. It suffices to obtain the full conditional distribution of each parameter in the model, that is, the distribution of each one given the remaining data (including the observed data). Then, we draw samples from these (full conditional) distributions.

Let Y_i as given in (3). Then

$$Y_i | U_i = u_i \sim N(\mathbf{x}_i^\top \beta, u_i^{-1} \sigma^2),$$

$$U_i \sim H(\cdot | \eta), \quad i = 1, \dots, n.$$

Corresponding to the m censored values, there are m unobserved values of the characteristic of interest. Let us group them into a vector, $\mathbf{y}_L = (y_1, \dots, y_m)^\top$, say. Obviously, the order of the elements of \mathbf{y}_L is irrelevant and is established here only for convenience. Thus, y_i is a realization of the latent unobservable variable $Y_i \sim \text{SMN}(\mathbf{x}_i^\top \beta, \sigma^2, \eta)$, $i = 1, \dots, m$. The key to the development of our MCMC-type

algorithm is to consider the augmented data $\{\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}\}$, where $\mathbf{u} = (u_1, \dots, u_n)^\top$, that is, we treat the problem as if \mathbf{y}_L and \mathbf{u} were in fact observed.

Compared to the uncensored case, there is no additional difficulty to obtain the full conditionals. The trick is very simple: it is enough to observe that, given the value $Y_i = y_i$, then $Y_{\text{obs}_i} = y_i$. The algorithm is as follows.

Step 1. Sample y_i , $i = 1, \dots, m$ independently from $\pi(y_i|y_{\text{obs}_i}, u_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta})$, which is truncated normal distribution

$$\text{TN}(\mathbf{x}_i^\top \boldsymbol{\beta}, u_i^{-1} \sigma^2; [-\infty, \kappa_i]).$$

Thus, the new $\mathbf{y}_{\text{obs}} = (y_1, \dots, y_m, y_{m+1}, \dots, y_n)$ is the sample generated for the m censored cases and the values observed y_i , $i = m + 1, \dots, n$ for the uncensored cases.

Step 2. Sample u_i , $i = 1, \dots, n$ independently from $\pi(u_i|y_{\text{obs}_i}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta})$, which is

(a) for the student-t case,

$$\text{Gamma}\left(\frac{\boldsymbol{\eta} + 1}{2}, \frac{(y_{\text{obs}_i} - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2} + \frac{\boldsymbol{\eta}}{2}\right);$$

(b) for the slash case,

$$\text{TGamma}\left(\boldsymbol{\eta} + \frac{1}{2}, \frac{(y_{\text{obs}_i} - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2} + \frac{\boldsymbol{\eta}}{2}; [0, 1]\right),$$

a truncated gamma distribution;

(c) for the contaminated normal case, a discrete distribution taking values γ with probability $\frac{p_1^*}{p_1^* + p_2^*}$ and 1, with probability $\frac{p_2^*}{p_1^* + p_2^*}$, where

$$p_1^* = \boldsymbol{\eta} \gamma^{(1/2)} \exp\left(-\frac{\gamma}{2} \left(\frac{y_{\text{obs}_i} - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right)^2\right) \text{ and}$$

$$p_2^* = (1 - \boldsymbol{\eta}) \exp\left(-\frac{1}{2} \left(\frac{y_{\text{obs}_i} - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right)^2\right).$$

Step 3. Sample $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta}|\mathbf{y}_{\text{obs}}, \mathbf{u}, \sigma^2, \boldsymbol{\eta})$, which is

$$N_p\left(\mathbf{A}_\beta \left[\mathbf{S}_\beta^{-1} \mathbf{b}_0 + \sigma^{-2} \left(\sum_{i=1}^n u_i y_{\text{obs}_i} \mathbf{x}_i \right) \right], \mathbf{A}_\beta\right),$$

where $\mathbf{A}_\beta = (\mathbf{S}_\beta^{-1} + \sigma^{-2} \sum_{i=1}^n u_i \mathbf{x}_i \mathbf{x}_i^\top)^{-1}$.

Step 4. Sample σ^{-2} from $\pi(\sigma^{-2}|\mathbf{y}_{\text{obs}}, \mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\eta}, \lambda)$, which is

$$\text{Gamma}\left(\frac{n+a}{2}, \frac{b + \sum_{i=1}^n u_i (y_{\text{obs}_i} - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{2}\right).$$

Step 5. Sampling $\boldsymbol{\eta}$ must be carried out considering the choice for the scale factor distribution, and each case deserves a specific treatment. For example,

(a) for the Student-t case,

- i. Sample λ from $\pi(\lambda|\boldsymbol{\eta})$, which is $\text{TGamma}(2, \boldsymbol{\eta}; [c, d])$.
- ii. Using a Metropolis-Hastings path, sample $\boldsymbol{\eta}$ from the marginal conditional distribution

$$\begin{aligned} \pi(\boldsymbol{\eta}|\mathbf{y}_{\text{obs}}, \boldsymbol{\beta}, \sigma^2, \lambda) \propto \\ \exp(-\lambda\boldsymbol{\eta}) \left[\sum_{i=1}^m \log \text{T}_{\boldsymbol{\eta}}\left(\frac{\kappa_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) + \sum_{i=m+1}^n \log \text{t}_{\boldsymbol{\eta}}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2) \right] \end{aligned} \quad (7)$$

The proposals are obtained in the following way: given an observation $\boldsymbol{\eta}^{(j-1)}$ obtained at stage $j-1$, generate a candidate $\boldsymbol{\eta}^*$ from the lognormal distribution

$$\text{LN}(\log \boldsymbol{\eta}^{(j-1)}, \delta_{\boldsymbol{\eta}}^2).$$

The new observation $\boldsymbol{\eta}^*$ is accepted with probability

$$\min\left\{\frac{\pi(\boldsymbol{\eta}^*|\dots)\boldsymbol{\eta}^*}{\pi(\boldsymbol{\eta}^{(j-1)}|\dots)\boldsymbol{\eta}^{(j-1)}}, 1\right\},$$

where $\pi(\boldsymbol{\eta}^*|\dots)$ denotes (7) evaluated using the current values of $\boldsymbol{\beta}$, σ^2 , λ and \mathbf{y}_{obs} . In this case, the samples are obtained from partially marginalized distributions, integrating out the latent variables u_1, \dots, u_n . This method, known as the *collapsing* principle, is usually more efficient than sampling from the full conditional distribution, which often yields slow convergence and highly correlated draws. See Liu (1994) for a detailed discussion;

(b) for the slash case,

- i. Sample λ from $\pi(\lambda|\boldsymbol{\eta})$, which is $\text{TGamma}(2, \boldsymbol{\eta}; [e, f])$.
- ii. Sample $\boldsymbol{\eta}$ from $\pi(\boldsymbol{\eta}|\mathbf{u})$ which is equal to

$$\text{Gamma}\left(n+1, \lambda - \sum_{i=1}^n \log(u_i)\right);$$

(c) for the contaminated normal case,

- i. Sample $\boldsymbol{\eta}$ from $\pi(\boldsymbol{\eta}|\mathbf{u}, \gamma)$, which is Beta $(\boldsymbol{\eta}_0 + m_\gamma; \boldsymbol{\eta}_1 + n - m_\gamma)$, where $m_\gamma = (n - \sum_{i=1}^n S_i) / (1 - \gamma)$ and

$$S_i = \begin{cases} 1 & \text{if } u_i = \gamma \\ 0 & \text{if } u_i \neq \gamma. \end{cases}$$

- ii. Consider the marginal conditional distribution of $\gamma|\mathbf{y}_{\text{obs}}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta}$

$$\pi(\gamma|\mathbf{y}_{\text{obs}}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta}) \propto \gamma^{\boldsymbol{\eta}_0-1} (1 - \gamma)^{\boldsymbol{\eta}_1-1} \times \left[\sum_{i=1}^m \log F_{CN} \left(\frac{\kappa_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}, \boldsymbol{\eta}, \gamma \right) + \sum_{i=m+1}^n \log f_{CN}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2) \right],$$

where F_{CN} is the cdf of the contaminated normal distribution. Given the parameterization $\gamma_r = \gamma/1 - \gamma$ we sample γ_r from its distribution, given by

$$\pi(\gamma_r|\mathbf{y}_{\text{obs}}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta}) = \frac{1}{(1 + \gamma_r)^2} \pi\left(\frac{\gamma_r}{1 + \gamma_r}|\mathbf{y}_{\text{obs}}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta}\right),$$

using a Metropolis-Hastings path with a lognormal proposal density: given an observation γ_r^{j-1} obtained at stage $j - 1$, generate a candidate γ_r^* from the distribution

$$\text{LN}(\log \gamma_r^{(j-1)}, \delta_{\gamma_r}^2).$$

The new observation γ_r^* is accepted with probability

$$\min \left\{ \frac{\pi(\gamma_r^*|\dots)\gamma_r^*}{\pi(\gamma_r^{(j-1)}|\dots)\gamma_r^{(j-1)}}, 1 \right\}.$$

In this case, the samples are obtained from partially marginalized distributions, integrating out the latent variables u_1, \dots, u_n .

For the Metropolis-Hastings paths the proposal scales were adjusted in order to have an acceptance rate in the interval (0.15, 0.30)

4 Bayesian model selection and influence diagnostics

4.1 Model comparison criteria

There are several propositions for Bayesian model choice criteria, which are useful to compare competing models fitting the same data set. For a review, see Ando (2010).

One of the most used in applied works is derived from the conditional predictive ordinate (CPO) statistic, which is based on the cross validation criterion to compare the models. Let $\mathbf{Z} = \{z_1, \dots, z_n\}$ be an observed sample from $\pi(\cdot|\boldsymbol{\theta})$. For the i -th observation, the CPO_i is written as

$$\text{CPO}_i = \int \pi(z_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{Z}_{(-i)})d\boldsymbol{\theta} = \left(\int \frac{\pi(\boldsymbol{\theta}|\mathbf{Z})}{\pi(z_i|\boldsymbol{\theta})}d\boldsymbol{\theta} \right)^{-1}, \quad (8)$$

where $\mathbf{Z}_{(-i)}$ is the sample without the i -th observation. For the proposed model, the CPO_i does not have a closed form. However, it is easy to see, from (8), that a Monte Carlo approximation can be obtained by using a MCMC sample $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q\}$ (after burn-in) from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{Z})$. It is given by Dey et al. (1997)

$$\widehat{\text{CPO}}_i = \left(\frac{1}{Q} \sum_{q=1}^Q \frac{1}{\pi(z_i|\boldsymbol{\theta}_q)} \right)^{-1}.$$

A summary statistic of the CPO_i 's, is the so-called *log-marginal pseudo likelihood for the model*, defined by

$$\text{LPML} = \sum_{i=1}^n \log(\widehat{\text{CPO}}_i).$$

Larger values of LMPL indicate better fit. Moreover using LMPL, we can estimate the *pseudo-Bayes factor* (PBF), widely used for comparing two competing models, say, M_1 and M_2 , and defined by

$$\text{PBF}(M_1, M_2) = \exp(\text{LPML}_1 - \text{LPML}_2),$$

where LMPL_i is the log-marginal pseudo likelihood for the model M_i , $i = 1, 2$.

The deviance information criterion (DIC) was proposed to work simultaneously as a measure of fit and as a measure of complexity. The *deviance* is defined as

$$D(\boldsymbol{\theta}) = -2 \log \pi(\mathbf{Z}|\boldsymbol{\theta}).$$

In connection with a measure of model complexity, the criterion considers a measure of the *effective number of parameters in the model*. It is defined by

$$\rho_D = \bar{D}(\boldsymbol{\theta}) - D(\tilde{\boldsymbol{\theta}}),$$

where the first term is the posterior expectation of the deviance, given by

$$\bar{D}(\boldsymbol{\theta}) = -2\text{E}[\log \pi(\mathbf{Z}|\boldsymbol{\theta})|\mathbf{Z}],$$

and the second term is the deviance evaluated at some estimate $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. The posterior mean $E[\boldsymbol{\theta}|\mathbf{Z}]$ is a natural choice for $\tilde{\boldsymbol{\theta}}$. Other alternatives are the posterior mode or median. Finally, we define the DIC by

$$\text{DIC} = \bar{D}(\boldsymbol{\theta}) + \rho_D = 2\bar{D}(\boldsymbol{\theta}) - D(\tilde{\boldsymbol{\theta}}).$$

Again, we can see that the computation of the integral $\bar{D}(\boldsymbol{\theta})$ is a complex numerical problem, and a good solution can be obtained using a MCMC sample $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q\}$ from the posterior distribution. Thus, we can obtain an approximation of the DIC by first computing the sample posterior mean of the deviations

$$\bar{D} = -2 \frac{1}{Q} \sum_{i=1}^Q \log \pi(\mathbf{Z}|\boldsymbol{\theta}_i)$$

and, after this, we compute

$$\widehat{\text{DIC}} = 2\bar{D} - D(\tilde{\boldsymbol{\theta}}).$$

Others criteria that can be used to compare models fitting the same data set are: the expected Akaike information criterion (EAIC), see Brooks (2002), and the expected Bayesian information criterion (EBIC), see Carlin and Louis (2001). These criteria can be estimated by:

$$\widehat{EAIC} = \bar{D} + 2\vartheta \quad \text{and} \quad \widehat{EBIC} = \bar{D} + \vartheta \log(n),$$

where ϑ is the number of model parameters.

Note that, for all these criteria, the evaluation of the likelihood function $\pi(\mathbf{Z}|\boldsymbol{\theta})$ is a key aspect. In our case, it is available in a closed form and is given by (5).

To determine model adequacy, we use a discrepancy measure based on the posterior predictive distribution. If the observed value is extreme relative to the reference distribution (the posterior predictive distribution), there is some concern with respect to assessment of model-fit to the data. Define y_i to be the observed data. Then the discrepancy measure between model and data is computed as a summary statistic (Gelman et al., 2004) using model parameters and data defined as

$$T(y, \boldsymbol{\theta}) = -2 \sum_{i=1}^n \log [\pi(y_i | \boldsymbol{\theta})], \quad (9)$$

The Bayesian p-value/posterior predictive p-value (Gelman et al., 2004) p_B is defined as the number of times $T(y_{pr}, \boldsymbol{\theta})$ exceeds $T(y, \boldsymbol{\theta})$ out of L simulated draws, *i.e.*, $p_B = \Pr(T(y_{pr}, \boldsymbol{\theta}) \geq T(y, \boldsymbol{\theta}) | Y = y)$, where y_{pr} denotes a simulated draw from

the posterior predictive distribution. A very large (> 0.95) or very small p-value (< 0.05) signals model misspecification, *i.e.*, the observed pattern would be unlikely to be seen in replications of the data under the true model (Gelman et al., 2004)

4.2 Influential observations

In this section we consider some Bayesian diagnostic measures of influence. Our focus is on case deletion methods, which detect observations that have a global influence in the inferential process.

Computation of divergence measures between posterior distributions with and without a given subset of the data is a useful way of quantifying influence. The q -divergence measure between two densities $\pi_1(\cdot)$ and $\pi_2(\cdot)$ for $\boldsymbol{\theta}$ (Csiszár, 1967) is defined by

$$d_q(\pi_1, \pi_2) = \int q \left(\frac{\pi_1(\boldsymbol{\theta})}{\pi_2(\boldsymbol{\theta})} \right) \pi_2(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (10)$$

where q is a convex function such that $q(1) = 0$. Some specific divergence measures are obtained by considering different options for $q(\cdot)$. For example, the *Kullback-Leibler divergence* is obtained when $q(z) = -\log(z)$; the *J-distance divergence* (a symmetric version of Kullback-Leibler divergence) is obtained when $q(z) = (z - 1)\log(z)$ and the *L₁-distance divergence* is obtained when $q(z) = |z - 1|$.

Let $\mathbf{y} = \{y_1, \dots, y_n\}$ be the sample and I a subset of $\{1, \dots, n\}$. Let us define $\mathbf{y}_I = \{y_i; i \in I\}$ and denote its complement set by \mathbf{y}_{-I} . The q -influence of \mathbf{y}_I on the posterior distribution of $\boldsymbol{\theta}$ is obtained by considering $\pi_1(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta}|\mathbf{y}_{(-I)})$ and $\pi_2(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{y})$ in (10). This influence measure can be written as

$$d_q(I) = \mathbb{E} \left[q \left(\frac{\pi_1(\boldsymbol{\theta}|\mathbf{y}_{(-I)})}{\pi_2(\boldsymbol{\theta}|\mathbf{y})} \right) | \mathbf{y} \right]. \quad (11)$$

It is important to note that all these measures can be approximated by using the MCMC posterior samples. Observe that they do not determine when a specific set of observations is influential or not. A way to circumvent this drawback is to establish a threshold point to help each a decision. In this direction, a proposition was made by Peng and Dey (1995) and Vidal and Castro (2010), which is given next.

Suppose that we toss a coin one time with probability $p \in [0, 1]$ of heads. If $x = 1$ means “heads” and $x = 0$ otherwise, the associated probability function is $\pi_1(x|p) = p^x(1-p)^{1-x}$, with $x = 0, 1$. If the coin is unbiased, we have $\pi_2(x|p) = 0.5$, $x = 0, 1$. From (10), the q -divergence between a (possibly) biased and an unbiased

coin is given by

$$d_q^*(p) = \frac{q(2p) + q(2(1-p))}{2}.$$

Note that $d_q^*(p)$ increases as p moves away from 0.5, is symmetric around $p = 0.5$ and achieves its minimum value at $p = 0.5$, which is the point where $\pi_1(\cdot) = \pi_2(\cdot)$ (in this case, we also have $d_q^*(0.5) = q(1) = 0$). Regarding the L_1 distance divergence measure, if we consider $p \geq 0.75$ as a strong bias, then we can say that the observation i is influential when $d_{L_1}(\{i\}) \geq 0.25$, since $d_{L_1}^*(0.75) = 0.25$. Similarly, for the Kullback-Leibler and J -distance divergences, we have $d_{KL}^*(0.75) \approx 0.143841$ and $d_J^*(0.75) \approx 0.274653$, respectively. Thus, if we use the Kullback-Leibler divergence, we can consider an influential observation when $d_{KL}(\{i\}) > 0.14$ and, using the J -distance, an observation with $d_J(\{i\}) > 0.27$ can be considered as influential.

5 Simulation study

In order to study the performance of our proposed model and algorithm, we present two simulation studies. The computational procedures were implemented using the R software (R Development Core Team, 2013). The first part of this simulation study shows the consequences on the parameter inference when the normality assumption is inappropriate. The goal of the second part is to compare the estimates of the degrees of freedom η , for the t-CR model, when we consider some prior distributions $\pi(\eta)$ proposed in the literature.

We consider artificial data generated from the linear left censored regression model, the setting given in (4), where $i = 1, \dots, 300$, $\varepsilon_i \sim t_\eta(0, \sigma^2)$, with $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (1, 2)$, $\sigma^2 = 3$, $\eta = 4$ and $\mathbf{x}_i^\top = (1, x_{i2})$. We considered, x_{i2} , $i = 1, \dots, 300$ as a random sample from a uniform distribution on the interval $(1, 3)$. These values were fixed throughout the simulations.

5.1 Study I

The main focus of this simulation study is to investigate the consequences on parameter inference when the normality assumption is inappropriate for different levels of censoring. To do this, we chose several censoring proportion settings (5%, 10%, 15%, 20%, 25%, 30% and 50%) and we considered the prior specification in Subsection 3.1, with $\mathbf{b}_0^\top = \mathbf{0}_2$, $\mathbf{S}_\beta = 10^3 \mathbf{I}_2$, $a = 2$, $b = 0.02$, $c = 0.02$ and $d = 0.5$.

For each level of censoring, we simulated 400 data sets and for each data set, we fitted the t-CR and the N-CR models and recorded the MCMC estimates of the parameters. Table 1 presents the summary statistics for β and σ^2 for the 7 censoring patterns. In this table, MC Mean and MC Sd denote the average values of the estimates and of the posterior standard deviations, respectively, computed across all samples. MC coverage is the percentage of times, computed across all samples, the 95% highest posterior density intervals contain the true value of the parameter.

From Table 1, we observe that the t-CR model presents better performance at all levels of censoring. Figure 1 shows that there is a strong increase of the bias (the deviations of the parameter estimates from the true value) when using the normal assumption for the errors. Clearly, the Student-t assumption yields more accurate estimates. Besides this, we observe in Figure 2 that estimates provided by fitting the t-CR model are more stable, especially when the censoring proportion increases. Therefore, this study shows evidence that the model with heavier tails than the normal one produces more accurate estimates in the context of censored data, since all the measures strongly favor the t-CR model, detecting clear departures from normality.

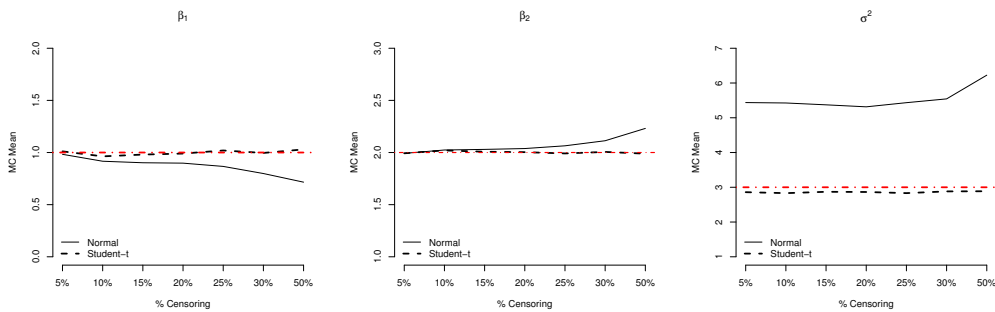


Figure 1: MC mean of parameters β_1, β_2 and σ^2 for the N-CR and t-CR models with different settings of censoring proportions in comparison with the true value of parameters (red line)

5.2 Study II

To assess the performance of the estimates of the degrees of freedom η , here we perform a sensitivity analysis by considering different prior densities for the degrees

Censoring	Parameters	N-CR			t-CR		
		MC Mean	MC Sd	MC Cover	MC Mean	MC Sd	MC Cover
5%	β_1	0.9827	(0.5049)	95.70%	1.0111	(0.4125)	97.97%
	β_2	1.9918	(0.2426)	96.71%	1.9925	(0.2032)	97.72%
	σ^2	5.4371	(0.9013)	95.70%	2.8588	(0.3834)	98.27%
10%	β_1	0.9163	(0.5064)	96.96%	0.9631	(0.4429)	97.47%
	β_2	2.0241	(0.2358)	97.22%	2.0201	(0.2111)	98.48%
	σ^2	5.4253	(1.1646)	94.19%	2.8310	(0.3701)	96.71%
15%	β_1	0.9018	(0.5155)	96.20%	0.9803	(0.4489)	98.23%
	β_2	2.0300	(0.2543)	94.94%	2.0075	(0.2159)	96.71%
	σ^2	5.3721	(2.5729)	94.34%	2.8697	(0.3858)	95.80%
20%	β_1	0.8980	(0.5596)	94.94%	0.9900	(0.4826)	97.72%
	β_2	2.0381	(0.2713)	94.18%	2.0034	(0.2330)	98.23%
	σ^2	5.3144	(0.9679)	93.39%	2.8643	(0.3803)	95.65%
25%	β_1	0.8667	(0.5214)	95.19%	1.0195	(0.4549)	96.20%
	β_2	2.0643	(0.2549)	95.44%	1.9912	(0.2217)	97.47%
	σ^2	5.4340	(1.4972)	95.95%	2.8315	(0.3910)	96.12%
30%	β_1	0.7985	(0.5297)	95.19%	0.9960	(0.4663)	97.72%
	β_2	2.1130	(0.2616)	94.43%	2.0056	(0.2288)	96.96%
	σ^2	5.5431	(1.2213)	95.70%	2.8813	(0.3988)	97.42%
50%	β_1	0.7156	(0.5925)	94.18%	1.0298	(0.4949)	95.95%
	β_2	2.2311	(0.3167)	92.91%	1.9890	(0.2551)	96.46%
	σ^2	6.2252	(1.5874)	94.43%	2.8859	(0.5018)	96.20%

Table 1: Summarized Monte Carlo results based on 400 simulated t-CR samples. MC mean, MC Sd (in parentheses) and MC Coverage (MC Cover) from fitting t-CR and N-CR models with different settings of censoring proportions.

of freedom η in the t-CR model. Thus, we consider the linear left censored regression model with 8% of censoring level and the degrees of freedom are fixed at different values, say $\eta = 4, 6, 8, 10, 12, 14$ or 20 . Following Fonseca et al. (2008) and Branco et al. (2012), we consider the following priors for η

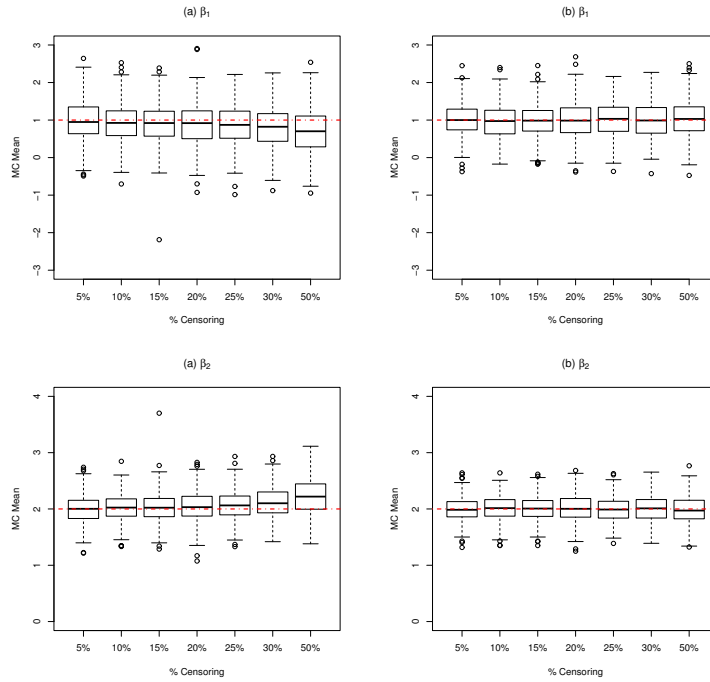


Figure 2: Box plots of MC mean of parameters β_1 and β_2 , from fitting (a) N-CR model and (b) t-CR model with different settings of censoring proportions. The red line represents the true value of parameters.

1. $\eta \sim \text{Exp}(\lambda_i)\mathbb{I}_{[2, \infty]}(\eta)$, $i = 1, 2, 3$, that is, a truncated exponential distribution on $[2, \infty]$, where $\lambda_1 = 0.01$, $\lambda_2 = 0.33$ and $\lambda_3 = 1.5$. This proposition was made originally by Geweke (1993);
2. $\eta \sim U(2, 100)$;
3. $\pi(\eta) = \left(\frac{\eta}{\eta+3}\right)^{1/2} \left\{ \psi_1\left(\frac{\eta}{2}\right) - \psi_1\left(\frac{\eta+1}{2}\right) - \frac{2(\eta+3)}{\eta(\eta+1)^2} \right\}^{1/2}$, where $\psi_1(x) = \frac{d^2}{dx^2} \log \Gamma(x)$ is the trigamma function. This is a Jeffreys prior proposed by Fonseca et al. (2008);
4. The hierarchical prior suggested by Cabral et al. (2012), that is, $\eta \sim \text{Exp}(\lambda)$ with $\lambda \sim U(0.02, 0.5)$. The choice of $c = 0.02$ and $d = 0.5$ ensures that the expected value of η is in the range $[2, 50]$.

We generated $R = 400$ replicates for each combination of parameter values and priors for η . To evaluate the results we considered the relative mean bias (RELMB) and the root relative mean squared error (RRELMSE), see Branco et al. (2012),

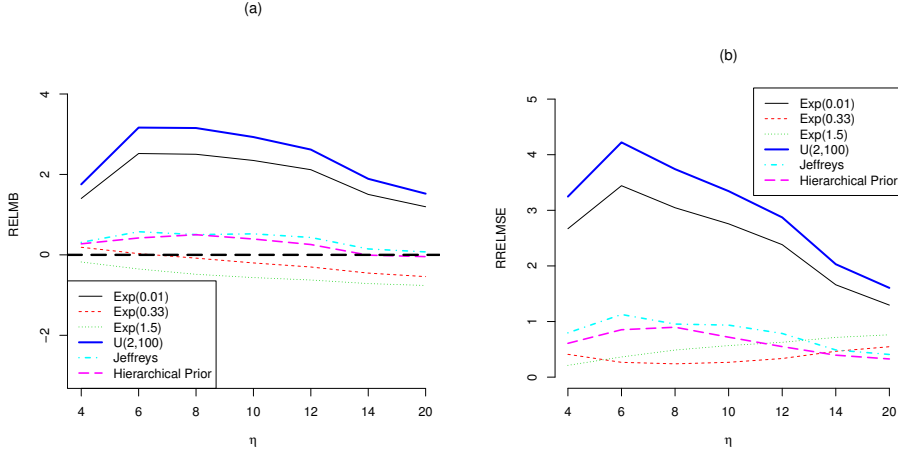


Figure 3: (a) Relative mean bias (RELMB) and (b) Root relative mean squared error (RRELMSE) of the simulation results for η considering different priors.

defined by

$$\text{RELMB} = \frac{1}{R} \sum_{i=1}^R \frac{(\hat{\eta}_i - \eta)}{\eta} \quad \text{and} \quad \text{RRELMSE} = \left[\frac{1}{R} \sum_{i=1}^R \frac{(\hat{\eta}_i - \eta)^2}{\eta^2} \right]^{1/2},$$

where $\hat{\eta}_i$ is the estimate of η computed from sample i , $i = 1, \dots, R$.

Table 2 and Figure 3 report the RELMB and the RRELMSE, respectively. These results show that the hierarchical prior is superior to the other priors in terms of those measures for all values of η . Moreover, it is interesting to note that when we change the values of the hyperparameters of the exponential prior, $\eta \sim \text{Exp}(\lambda_i) \mathbb{I}_{[2, \infty]}(\eta)$, the values of RELMB and RRELMSE also change significantly, consequently the estimates of η are quite sensitive to the choice of hyperparameters when we consider the Exponential prior. Despite the Jeffreys prior has the same decreasing behavior as the hierarchical prior, we can observe that the latter has lower values of RELMB and RRELMSE for all values of the parameter η .

6 Data analysis

We consider the wage rate data set of Mroz (1987), where a measure of the wage of 753 married white women, with ages between 30 and 60 years old in 1975, is evaluated. Of 753 women considered in this study, 428 worked at some point during that year. Thus, the variables are:

		RELMB						
Priors	η							
	4	6	8	10	12	14	20	
Exp(0.01)	1.401	2.519	2.501	2.346	2.117	1.503	1.196	
Exp(0.33)	0.188	0.030	-0.081	-0.200	-0.301	-0.453	-0.540	
Exp(1.50)	-0.176	-0.351	-0.483	-0.564	-0.625	-0.713	-0.761	
U(2, 100)	1.753	3.164	3.153	2.929	2.617	1.890	1.521	
Jeffreys	0.311	0.575	0.504	0.522	0.433	0.147	0.073	
Hierarchical	0.274	0.420	0.501	0.393	0.257	-0.007	-0.043	

		RRELMSE						
Priors	η							
	4	6	8	10	12	14	20	
Exp(0.01)	2.669	3.442	3.046	2.756	2.382	1.660	1.295	
Exp(0.33)	0.410	0.266	0.239	0.266	0.334	0.465	0.547	
Exp(1.50)	0.212	0.362	0.487	0.567	0.626	0.714	0.761	
U(2, 100)	3.246	4.221	3.740	3.344	2.872	2.030	1.604	
Jeffreys	0.796	1.126	0.954	0.935	0.784	0.488	0.407	
Hierarchical	0.607	0.852	0.897	0.718	0.548	0.395	0.326	

Table 2: Relative mean bias (RELMB) and root relative mean squared error (RRELMSE) for different values of η considering different priors.

- y_i : *wage rates*, defined as the average hourly earnings. If the wage rates are set equal to zero, these wives did not work in 1975. Therefore, these observations are considered left censored at zero.
- x_{i2} : wife's age.
- x_{i3} : years of schooling.
- x_{i4} : the number of children younger than six years old in the household.
- x_{i5} : the number of children between six and nineteen years old,

where, each of the vectors of explanatory variable values is given by $\mathbf{x}_i^\top = (1, x_{2i}, x_{3i}, x_{4i}, x_{5i})$ for $i = 1, 2, \dots, 753$. This data set was analyzed by Arellano-Valle et al. (2012) using the t-CR model, from a frequentist point of view.

Parameters	Models							
	N-CR			t-CR				
	Mean	Sd	HPD (95%)	Mean	Sd	HPD (95%)		
β_1	-2.7695	1.7542	(-6.247; 0.579)	-1.1945	1.4236	(-4.100; 1.469)		
β_2	-0.1056	0.0278	(-0.157;-0.049)	*	-0.110	0.0229	(-0.154;-0.066)	*
β_3	0.7324	0.0847	(0.570; 0.899)	*	0.6534	0.0719	(0.516; 0.801)	*
β_4	-3.0521	0.4513	(-3.970;-2.210)	*	-3.1649	0.3885	(-3.914;-2.392)	*
β_5	-0.2197	0.1557	(-0.519; 0.082)		-0.2905	0.1310	(-0.546;-0.034)	*
σ^2	21.3451	1.6156	(18.110;24.426)	*	11.5843	1.0460	(9.539;13.625)	*
η	—	—	—		5.2789	0.6709	(4.543; 6.581)	*

Parameters	SL-CR			CN-CR				
	Mean	Sd	HPD (95%)	Mean	Sd	HPD (95%)		
	β_1	-1.1931	1.4000	(-3.909; 1.613)	-1.6361	1.4675		(-4.388; 1.276)
β_2	-0.1093	0.0223	(-0.151;-0.063)	*	-0.1057	0.0233	(-0.150;-0.059)	*
β_3	0.6494	0.0710	(0.518; 0.791)	*	0.6651	0.0722	(0.518; 0.797)	*
β_4	-3.1325	0.3905	(-3.912;-2.390)	*	-3.0721	0.3949	(-3.866;-2.295)	*
β_5	-0.2959	0.1272	(-0.532;-0.036)	*	-0.2796	0.1323	(-0.517;-0.009)	*
σ^2	6.9515	0.8672	(5.374; 8.723)	*	12.7946	1.9148	(9.797;16.931)	*
η	1.4379	0.2094	(1.059; 1.848)	*	0.0989	0.0568	(0.015; 0.205)	*
γ	—	—	—		0.1761	0.0892	(0.113; 0.237)	*

Table 3: Wage rate data. Posterior mean, standard deviation (SD), and HPD (95%) interval for the parameters of the SMN-CR models.

Here, we revisit this data set in order to evaluate the performance of the proposed Bayesian methods. As in Arellano-Valle et al. (2012), we consider the SMN-CR model, the setting given in (3) and (4), to describe the *wage rates*. In our analysis we assume the normal (*N-CR*), Student-t (*t-CR*), slash (*SL-CR*) and contaminated normal (*CN-CR*) distributions from the SMN class for comparative purposes.

6.1 Estimation

In the estimation process, we consider the prior densities discussed in the previous Subsection 3.1 and for η of the *t-CR* model, we assume the hierarchical prior. We generated four parallel independent MCMC runs of size 50,000 with widely dispersed initial values for each parameter for all the four subclasses, where the first 10,000 iterations were discarded as burn-in samples. To eliminate potential problems due to autocorrelation, we considered a spacing of size 20. The convergence of the MCMC chains was monitored using trace plots, autocorrelation plots, and Gelman-Rubin

\widehat{R} diagnostics. Sensitivity analysis on the routine use of the inverse-gamma prior on the variance components reveals that the results are fairly robust under different prior choices. Table 3 reports the posterior means (Mean), standard deviations (SD) and HPD (95%) interval of the model parameters after fitting the different SMN-CR models. We can notice from this table that the slope estimates are similar among the four fitted models, however the standard deviations of the t-CR, SL-CR and CN-CR normal models are smaller than the ones of the N-CR model, indicating that the three models with longer-than-normal tails seem to produce more accurate Bayesian estimates. We can see that the HPD (95%) interval of the intercept β_1 contains zero, i.e., the intercept is considered nonsignificant at the 5% level for the four fitted models. Note that the posterior estimates of $\beta_2, \beta_3, \beta_4$ and β_5 are significant for all the SMN-CR models with heavy tails. This is in contrast to the N-CR model, where the 95% confidence interval of β_5 (*number of children between six and nineteen years old*) includes zero.

Model	LPML	DIC	EAIC	EBIC	p_B
N-CR	-1489.68	8901.99	2967.38	2976.63	0.609
t-CR	-1447.26	8656.71	2887.62	2901.49	0.332
SL-CR	-1443.63	8631.85	2878.64	2892.51	0.433
CN-CR	-1475.16	8655.73	2878.74	2892.61	0.762

Table 4: Wage rate data. Comparison between the SMN-CR models.

Table 4 compares the four sub-classes of SMN-CR models using the model selection criteria discussed in Section 4.1. Notice that all the 3 members of the SMN-CR models (with heavy tails) perform significantly better than the N-CR model, with the SL-CR outperforming all the rest. For the t-CR and SL-CR models, as η (the degrees of freedom) $\rightarrow \infty$, it approaches the N-CR model as a limiting case. For both these models, the estimated value of η is small, indicating the lack of adequacy of the normal assumption for the wage rate data. In Table 4, we also report the posterior values of p_B for the model parameters from the four fitted SMN-CR models. Note that these values indicate no overall lack of fit.

6.2 Bayesian case influence diagnostics

Considering the sample of the posterior distributions of the parameters of the four models, the q -divergence measures, described in Section 4.2, were computed. The cases #74, #185, #349, #394 and #408 were identified as influential under the N-CR model, because they exceed the specified thresholds. However, for the t-CR and SL-CR models we do not find highly influential cases. Figure 4 to Figure 6 depict the index plot of KL , J and L_1 distance and as expected, the effect of influential observations on the Bayesian estimates of the parameters are attenuated when heavy-tailed distributions are considered. Figure 7 depicts the history of the chains and the approximate posterior marginal densities of the parameters β_2 and η for the SL-CR model.

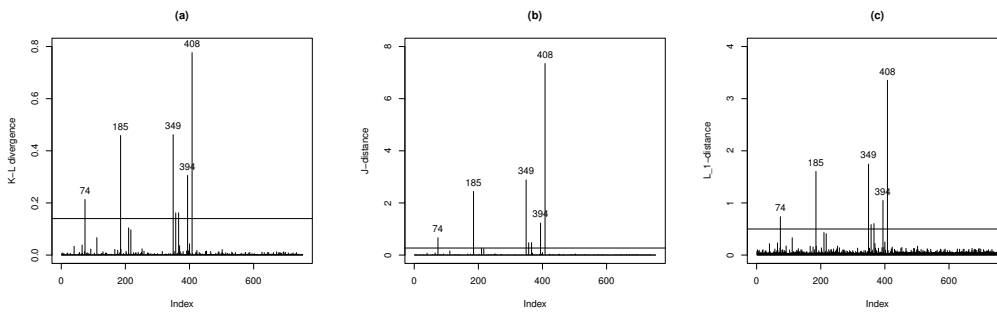


Figure 4: Influence diagnostics for the N-CR model: (a) K-L divergence, (b) J-distance, (c) L_1 -distance

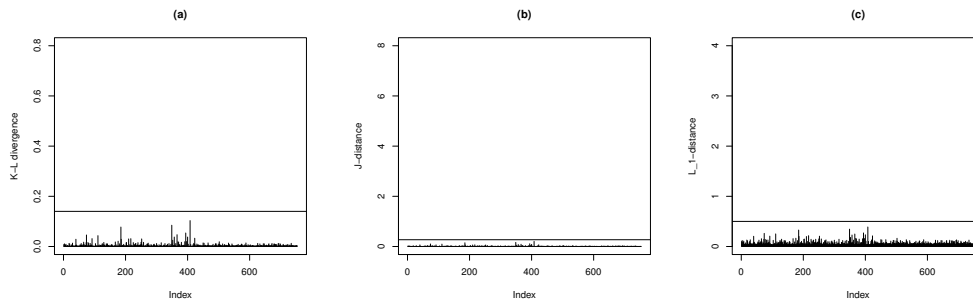


Figure 5: Influence diagnostics for the t-CR model: (a) K-L divergence, (b) J-distance, (c) L_1 -distance

In order to reveal the impact of these five observations on the parameter estimates we refitted the parameters of SL-CR model, first removing one by one and later all

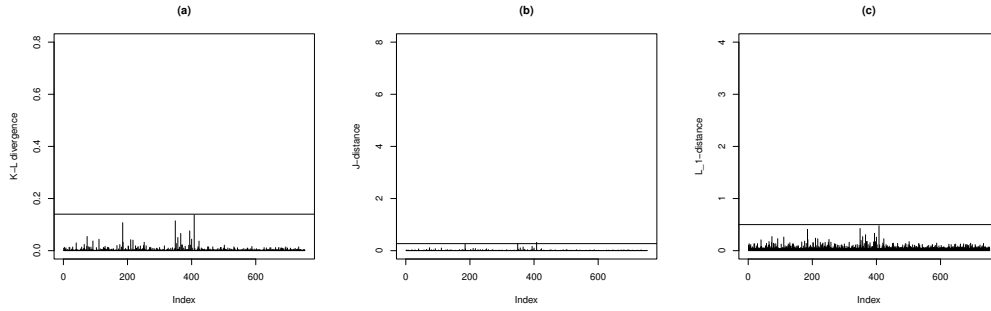


Figure 6: Influence diagnostics for the SL-CR model: (a) K-L divergence, (b) J-distance, (c) L_1 -distance

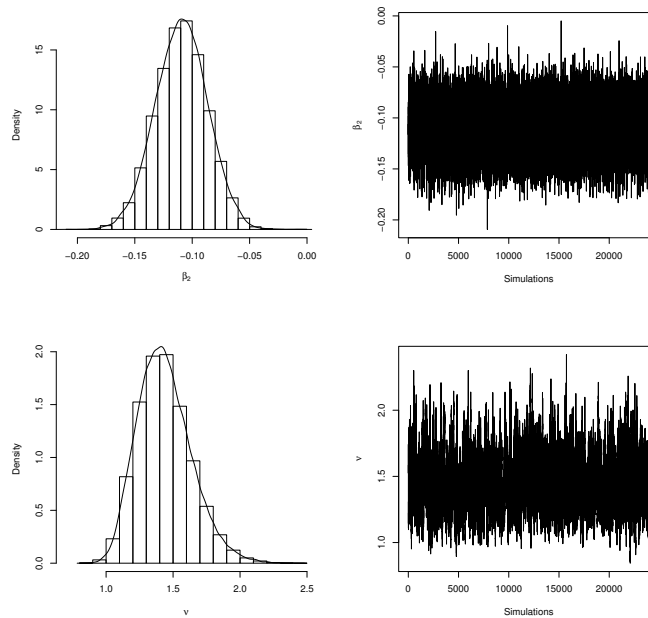


Figure 7: History of the chains and the approximate posterior marginal densities of β_2 and η considering the SL-CR model

five influential points. In Table 5 we show the relative changes (in percentage) of each parameter estimate, which is defined by $\mathbf{RC}_{\theta_j} = \left| (\hat{\theta}_j - \hat{\theta}_{j(I)}) / \hat{\theta}_j \right| \times 100$, where $\hat{\theta}_{j(I)}$ denotes the estimate of θ_j after the set I of observations was removed. Note that the parameter β_1 is heavily impacted by these observations as compared to the other covariates. The parameter significance and sign of the coefficients remained the same.

Set{ I }	β_1	β_2	β_3	β_4	β_5	σ^2
All observations	-	-	-	-	-	-
All - {#408}	[139]	[51]*	[46]*	[56]*	[42]*	[65]*
All - {#394}	[140]	[51]*	[46]*	[56]*	[43]*	[65]*
All - {#349}	[140]	[51]*	[46]*	[57]*	[42]*	[65]*
All - {#185}	[136]	[51]*	[46]*	[56]*	[43]*	[65]*
All - {#74}	[140]	[50]*	[46]*	[56]*	[41]*	[65]*
All - {#74, #185, #349, #394, #408}	[130]	[51]*	[44]*	[56]*	[40]*	[61]*

Table 5: Wage rate data. Relative changes [-**RC**-in %] for all the parameters in the SL-CR model. * indicates parameter significance.

7 Conclusions

In this paper, we considered use of the class of SMN distributions as a replacement for the conventional choice of normality distribution for the censored linear models where computational issues and outlier identification are concerned. Thus, this paper generalizes the works by Barros et al. (2010), Arellano-Valle et al. (2012) and Massuia et al. (2012) from a Bayesian perspective.

In order to explore the statistical properties of the proposed models, an efficient MCMC- type algorithm, in the sense of Liu and Rubin (1994), has been coded and implemented in the R package **BayesCR** (Garay and Lachos, 2013) which is available for download at CRAN repository. Two simulation studies were performed. The first simulation study revealed gain in efficiency and accuracy for parameter estimates of the t-CR model, where typical assumptions of normality are questionable. In the second simulation study, we showed that the hierarchical prior (Cabral et al., 2012) of the η for the t-CR model is superior to the others priors specifications previously proposed in the literature, in terms of relative mean bias (RELMB) and the root relative mean squared error (RRELMSE).

We also applied our method to the wage rate data set of Mroz (1987), in order to illustrate how the procedure developed can be used to evaluate model assumptions, identify outliers and obtain robust parameter estimates. As expected, our proposed SMN-CR model showed considerable flexibility to accommodate outliers.

While our SMN-CR model showed considerable flexibility to accommodate outliers, the robustness aspects could be seriously affected by the presence of skewness. In this setup, a natural extension would be to incorporate skewness and heavy-tailedness simultaneously, using scale mixtures of skew-normal (**SMSN**) distribu-

tions, as proposed in Lachos et al. (2010). Other extensions of the current work include considering autoregressive models to take into account for the time dependence, as proposed in Park et al. (2007).

Acknowledgement

The research of V. H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and Grant 2011/17400-6 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). A. M. Garay was supported by Grant 161119/2012-3 from CNPq. Heleno Bolfarine was supported by CNPq and C.R. B. Cabral was supported by CNPq (via the Universal and CT-Amazônia projects) and CAPES (via Project PROCAD 2007).

References

- Ando, T., 2010. Bayesian Model Selection and Statistical Modeling. Chapman & Hal/CRC.
- Andrews, D. F., Mallows, C. L., 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B.* 36, 99–102.
- Arellano-Valle, R. B., Castro, L. M., González-Farías, G., Muñoz-Gajardo, K. A., 2012. Student-t censored regression model: properties and inference. *Statistical Methods & Applications* 21, 453–473.
- Barros, M., Galea, M., González, M., Leiva, V., 2010. Influence diagnostics in the tobit censored response model. *Statistical Methods & Applications* 19, 716–723.
- Branco, M. D., Genton, M. C., Liseo, B., 2012. Objective Bayesian analysis of skew-t distributions. *Scandinavian Journal of Statistics* 40, 63–85.
- Breen, R., 1996. *Regression Models: Censored, Sample Selected, or Truncated Data.* Sage Publications.
- Brooks, S. P., 2002. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002) 64 (4), 616–618.

- Cabral, C. R. B., Lachos, V. H., Madruga, M. R., 2012. Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference* 142, 181–200.
- Carlin, B. P., Louis, T. A., 2001. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Edition. Chapman & Hall/CRC, Boca Raton.
- Chib, S., 1992. Bayes inference in the Tobit censored regression model. *Journal of Econometrics* 51, 79–99.
- Choy, S. T. B., Smith, A. F. M., 1997. Hierarchical models with scale mixtures of normal distributions. *Test* 6, 205–221.
- Csiszár, I., 1967. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2, 299–318.
- Cysneiros, F. J. A., Paula, G. A., 2005. Restricted methods in symmetrical linear regression models. *Computational Statistics & Data Analysis* 49, 689–708.
- Dey, D. K., Chen, M. H., Chang, H., 1997. Bayesian approach for the nonlinear random effects models. *Biometrics* 53, 1239–1252.
- Fernández, C., Steel, M. J. F., 1999. Multivariate Student-t regression models: Pitfalls and inference. *Biometrika* 86, 153.
- Fonseca, T. C. O., Ferreira, M. A. R., Migon, H. S., 2008. Objective Bayesian analysis for the Student-t regression model. *Biometrika* 95, 325–333.
- Frühwirth-Schnatter, S., Tüchler, R., Otter, T., 2004. Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics* 22 (1), 2–15.
- Gamerman, D., Lopes, H. F., 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd Edition. Chapman & Hall/CRC.
- Garay, A. M., Lachos, V. H., 2013. BayesCR: Bayesian analysis of censored linear regression models with scale mixtures of normal (SMN) distributions. R package v. 1.0.
URL <http://CRAN.R-project.org/package=BayesCR>

- Gelfand, A. E., Hills, S. E., Racine-Poon, A., Smith, A. F. M., 1990. Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association* 85, 972–985.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. *Bayesian data analysis*. Chapman & Hall/CRC.
- Geweke, J., 1993. Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics* 8, S19–S40.
- Ibacache-Pulgar, G., Paula, G. A., 2011. Local influence for student-t partially linear models. *Computational Statistics & Data Analysis* 55, 1462–1478.
- Lachos, V. H., Castro, L. M., Dey, D. K., 2013. Bayesian inference in nonlinear mixed-effects models using normal independent distributions. *Computational Statistics & Data Analysis* 64, 237–252.
- Lachos, V. H., Ghosh, P., Arellano-Valle, R. B., 2010. Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* 20, 303–322.
- Lange, K. L., Sinsheimer, J. S., 1993. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* 2, 175–198.
- Lenk, P., DeSarbo, W., 2000. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* 65, 93–119.
- Lin, T. I., Lee, J. C., 2006. A robust approach to t linear mixed models applied to multiple sclerosis data. *Statistics in Medicine* 25, 1397–1412.
- Lin, T. I., Lee, J. C., 2007. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* 137, 484–495.
- Liu, C., Rubin, D. B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* 80, 267–278.
- Liu, J. S., 1994. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* 89, 958–966.

- Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H., 2012. Influence diagnostics for Student-t censored linear regression models. Technical Report No. 9–23.
- Matos, L. A., Prates, M. O., Chen, M. H., Lachos, V. H., 2013. Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica* 23, 1323–1345.
- Mroz, T. A., 1987. The sensitivity of an empirical model of married women’s hours of work to economic and statistical assumptions. *Econometrica* 55, 765–799.
- Nelson, F. D., 1977. Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics* 6, 309–327.
- Park, J. W., Genton, M. G., Ghosh, S. K., 2007. Censored time series analysis with autoregressive moving average models. *Canadian Journal of Statistics* 35, 151–168.
- Peng, F., Dey, D. K., 1995. Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics* 23, 199–213.
- Pinheiro, J. C., Liu, C. H., Wu, Y. N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics* 10, 249–276.
- R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL <http://www.R-project.org>
- Rosa, G. J. M., Padovani, C. R., Gianola, D., 2003. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal* 45, 573–590.
- Smith, A. F. M., Roberts, G. O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B.* 55, 3–24.
- Stapleton, D. C., Young, J. D., 1984. Censored normal regression with measurement error on the dependent variable. *Econometrica*, 737–760.

- Thompson, M. L., Nelson, K. P., 2003. Linear regression with Type I interval and left-censored response data. *Environmental and Ecological Statistics* 10, 221–230.
- Tierney, L., 1994. Markow chain for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701–1762.
- Vaida, F., Liu, L., 2009. Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics* 18, 797–817.
- Vidal, I., Castro, L. M., 2010. Influential observations in the independent Student-t measurement error model with weak nondifferential error. *Chilean Journal of Statistics* 1, 17–34.
- Villegas, C., Paula, G. A., Cysneiros, F. J. A., Galea, M., 2012. Influence diagnostics in generalized symmetric linear models. *Computational Statistics & Data Analysis* 59, 161–170.
- Zhu, X. X., Zhu, B., Cao, C. Z., 2013. Diagnostics for a linear model with first-order autoregressive symmetrical errors. *Communications in Statistics - Theory and Methods* 42, 2335–2350.