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

Monique B. Massuia, Aldo M. Garay, Victor H. Lachos*

Departamento de Estatística, Universidade Estatual de Campinas, Campinas, São Paulo, SP-Brazil

Celso R. B. Cabral

Departamento de Estatística, Universidade Federal do 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 the normality assumption for the errors terms. However, those analyses might not provide robust inference when the normality assumption (or symmetry) is questionable. In this article, we develop a Bayesian framework for censored linear regression models by replacing the Gaussian assumption for the random errors with the asymmetric class of scale mixtures of skew-normal (SMSN) distributions. The SMSN is an attractive class of asymmetrical heavy-tailed densities that includes the skew-normal, skew-t, skew-slash, the skew-contaminated normal and the entire family of scale mixtures of normal distributions as special cases. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference. 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`, proposed by the authors. The newly developed procedures are illustrated with applications using real and simulated data.

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

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. A good alternative is to consider a more flexible distribution for the errors, such as the Student-t. This is done in Fernández and Steel (1999), where some inferential procedures are discussed. Ibacache-Pulgar and Paula (2011), propose local influence measures in the Student-t partially linear

*Corresponding author: Department of Statistics, Campinas State University, Brazil. Rua Sérgio Buarque de Holanda, 651 Cidade Universitária, "Zeferino Vaz", Campinas, São Paulo. Brazil

regression model. Other existing methods for robust estimation are based on the class of scale mixtures of normal (SMN) distributions presented by Andrews and Mallows (1974). These distributions have heavier tails than the normal one, and thus they seem to be a reasonable choice for robust inference and includes as special cases many symmetric distributions, such as the normal, Pearson type VII, Student-t, slash and contaminated normal. For an interesting review, including applications in mixed models, see Meza et al. (2012). An another important wide class of distributions is the scale mixtures of skew–normal distributions (SMSN), presented by Branco and Dey (2001). This class of distributions deals with heavy tails and skewness simultaneously, and contains the entire family of SMN distributions as special case.

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 a 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. Arellano-Valle et al. (2012) and Massuia et al. (2015) proposed extensions of the N-CR model by considering that the error term follows a Student-t distribution. Symmetric extensions of the N-CR model can be obtained by assuming that the distribution of the perturbations belongs to the scale mixture of normal (SMN) distributions family as in Garay et al. (2015b). These papers provide extensions of the normal censored model for statistical modeling of censored datasets involving observed variables with heavier tails than the normal distribution. The work of Massuia et al. (2015) examines the performance of the model through case-deletion and local influence techniques.

Here we suggest to use a flexible class of SMSN distributions, extending the previous cited works of Arellano-Valle et al. (2012), Massuia et al. (2015) and Garay et al. (2015b), where the error component distributions is assumed to follow a SMSN distributions. It is important to notice that the skew–normal and skewed versions of some other classical symmetric distributions are SMSN members: the skew–t (ST), the skew–slash (SSL) and the skew contaminated normal (SCN), for example. These distributions have heavier tails than the skew–normal (and the normal) one, and thus they seem to be a more reasonable choice for robust inference. In this paper, we propose a robust parametric approach of the censored linear regression models based on the SMSN distributions, denoted by SMSN-CR, from a Bayesian perspective.

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 the more flexible class of the SMSN 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 SMSN class of distributions. The SMSN censored linear regression model model is presented in Section 3. We present a Gibbs-type algorithm for Bayesian estimation, specifying priors distributions for the parameters of interest in Section 4. The model selection and influence diagnostics issue is considered in Section 5. The proposed method is illustrated in Section 6, by the analysis of a data set of housewife wages, and in Section 7, by considering the analysis of simulated data sets. Section 8 concludes with a short discussion of issues raised by our study and some possible directions for the future research.

2. Preliminaries

2.1. Notations and definitions

In this paper $X \sim \mathrm{N}(\mu, \sigma^2)$ denotes a random variable X with normal distribution with mean μ and variance σ^2 and $\phi\left(\,\cdot\,;\mu,\sigma^2\right)$ denotes its probability density function (pdf). $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the pdf and the cumulative distribution function (cdf) of the standard normal distribution. $X \sim Gamma(a, b)$ denotes a random variable with Gamma distribution with mean a/b and variance a/b^2, with $a > 0$ and $b > 0$. We use the traditional convention denoting a random variable (or a random vector) by an upper case letter and its realization by the corresponding lower case. Random vectors and matrices are denoted by boldface letters. $\mathbf{X}^\top$ is the transpose of $\mathbf{X}$. $X \perp Y$ indicates that the random variables X and Y are independent. For the 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.

2.2. Scale mixtures of skew-normal (SMSN) distributions

In order to define the linear regression model with censored responses under the SMSN class, we start with the definition of this family of distributions, its hierarchical formulation, and then we will introduce some further properties. This class of distributions was proposed by Branco and Dey (2001) and contains the entire family of SMN distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993), and skewed versions of classic symmetric distributions such as the skew-Student-t and the skew-slash, among others. Before we define the SMSN class, we present the fundamental concept of skew-normal (SN) distribution, given in Azzalini (1985).

Definition 1. *A random variable Z has a skew-normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by $Z \sim SN(\mu, \sigma^2, \lambda)$, if its pdf is given by*

$$f(z) = 2\phi(z; \mu, \sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right). \tag{1}$$

The relation between the SMSN class and the SN distribution is given in the next definition.

Definition 2. *We say that a random variable Y has a SMSN distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by SMSN$(\mu, \sigma^2, \lambda; H)$, if it has the following stochastic representation:*

$$Y = \mu + \kappa(U)^{1/2}Z, \quad U \perp Z, \tag{2}$$

where $Z \sim SN(0, \sigma^2, \lambda)$, $\kappa(\cdot)$ is a positive function and U is a positive random variable with cdf $H(\,\cdot\,; \boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$.

The random variable U is known as *the scale factor* and its cdf $H(\,\cdot\,; \boldsymbol{\nu})$ is called the *mixing distribution function*. Note that, when $\lambda = 0$, the SMSN family reduces to the symmetric class of SMN distributions.

Now we present a hierarchical representation for $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda; H)$ which is very convenient to derive some mathematical properties and also to make Bayesian inference through the implementation of a Gibbs-type algorithm. It was used by Basso et al. (2010) in the context of mixtures of SMSN distributions and by Cancho et al. (2011) in the context of non-linear regression models, among others. This representation is given by

$$Y = \mu + \Delta T + \kappa(U)^{1/2}\tau^{1/2}T_1, \tag{3}$$

where

$$\Delta = \sigma\delta, \quad \tau = (1-\delta^2)\sigma^2, \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}},$$
$$T = \kappa(U)^{1/2}|T_0|, \quad T_0 \perp T_1,$$

where T_0 and T_1 are standard normal random variables and $|\cdot|$ denotes absolute value.

The scale factor U can be discrete or continuous and the form of the SMSN distribution is determined by its distribution. In this paper we take into account three members of SMSN class: SN, St and SSL distributions, although there are another examples of distributions which belong to the SMSN family, as the skew-contaminated normal, the skew-Cauchy, the skew-Pearson VII and the corresponding symmetric versions.

Using the representation in Equation (2), we observe that

$$Y|U = u \sim SN(\mu, \kappa(u)\sigma^2, \lambda).$$

Thus, integrating out U from the joint density of Y and U will lead to the following marginal density of Y

$$f(y) = 2\int_0^\infty \phi(y; \mu, \kappa(u)\sigma^2)\Phi\left(\frac{\lambda(y-\mu)}{\sigma\kappa(u)^{1/2}}\right)dH(u). \tag{4}$$

Also, considering the stochastic representation given in Equation (3), we have that

$$\begin{aligned} Y|T = t, U = u &\sim N(\mu + \Delta t\ ,\ \kappa(u)\tau), \\ T|U = u &\sim TN(0, \kappa(u)\ ;\ (0, \infty)), \end{aligned}$$

where $\text{TN}(\mu, \sigma^2 \; ; \; \mathbb{A})$ denotes the density of a normal distribution with mean μ and variance σ^2 truncated in the set $\mathbb{A}$. Therefore, another way to write the pdf of Y is

$$\begin{aligned} f(y) &= \int\int f(y|t,u) f(t|u) f(u) \; dt \; du \qquad (5) \\ &= 2\int_0^\infty \int_0^\infty \phi(y \; ; \; \mu + \Delta t \; , \; \kappa(u)\tau) \\ &\quad \times \phi(t \; ; \; 0, \kappa(u)) \; dt \; dH(u). \end{aligned}$$

The following Lemmas are useful to compute the model comparison criteria, which will be seen in Section 5. Some new notation is now in order: $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $m-$variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\Phi_m(\cdot \; ; \; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the corresponding cdf. $T_m\left(\cdot; \; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu\right)$ represents the cdf of the $m-$variate Student-t distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. The special notation $T\left(\cdot; \; \nu\right)$ is used for the univariate case with mean zero and scale 1.

Lemma 1. *Let $Y \sim SMSN(\mu, \sigma^2, \lambda; H)$. Then, the cdf of Y can be written in the following ways:*

$$\begin{aligned} F(y) &= 2\int_0^\infty \int_0^\infty \phi(t) \qquad (6) \\ &\quad \times \Phi\Big(\frac{y-\mu}{\sigma\kappa(u)^{1/2}}\sqrt{1+\lambda^2} - \lambda t\Big) \; dt \; dH(u) \quad \textit{and} \\ F(y) &= \int_0^\infty 2\Phi_2\Big(\boldsymbol{y}(u)^*; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big) dH(u), \qquad (7) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{y}(u)^* &= (\kappa(u)^{-1/2}y, 0)^\top, \quad \boldsymbol{\mu}^* = (\mu, 0)^\top, \qquad (8) \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma^2 & -\delta\sigma \\ -\delta\sigma & 1 \end{pmatrix} \quad \textit{and} \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}. \end{aligned}$$

Proof. See Appendix Appendix A. □

Lemma 2. *Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $U \sim Gamma\left(\alpha, \beta\right)$ be independent. Then, for any fixed vector $\boldsymbol{w} \in \mathbb{R}^p$,*

$$E\left[\boldsymbol{\Phi}_p\left(\sqrt{U}\boldsymbol{w} \; ; \; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right] = T_p\left(\sqrt{\frac{\alpha}{\beta}}\boldsymbol{w} \; ; \; \boldsymbol{\mu}, \boldsymbol{\Sigma}, 2\alpha\right).$$

Proof. See Lemma 1 of Prates et al. (2014). □

Following Basso et al. (2010), in this paper we consider $\kappa(u) = 1/u$ in the stochastic representation (3), since this choice leads to interesting mathematical properties. Besides, our study focus on some particular cases of SMSN distributions. For each specific SMSN distribution, we compute its cdf, which is useful for evaluating the likelihood in SMSN-CR models, and $k_m = E[U^{-m/2}]$, useful for the implementation of the Gibbs-type algorithm for posterior inference about the parameters in these models. We consider the following distributions:

- *The skew-normal distribution:* in this case we consider, in Definition 2, $P\left(U = 1\right) = 1$, which implies $k_m = 1$. The density of Y is defined in (1) and, using Equation (7) of Lemma 1, the cdf of Y is given by

$$F(y) = 2\Phi_2\Big(\boldsymbol{y}^*; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big), \qquad (9)$$

where $\mathbf{y}^* = (y,0)^\top$ and $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are like in (8).

- *The skew–t distribution:* this case arises when we consider $U \sim Gamma(\nu/2, \nu/2)$, leading to

$$k_m = (\nu/2)^{(m/2)} \frac{\Gamma((\nu-m)/2)}{\Gamma(\nu/2)}.$$

If Y has this distribution, we use the notation $Y \sim ST\left(\mu, \sigma^2, \lambda; \nu\right)$. The density of Y is

$$\begin{aligned} f(y|\mu,\sigma^2,\lambda;\nu) = & \frac{2\,\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{d(y)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ & \times T\left(\lambda\, d(y)\sqrt{\frac{\nu+1}{\nu+d(y)^2}};\ \nu+1\right), \\ & y \in \mathbb{R}, \end{aligned} \tag{10}$$

where $d(y) = (y-\mu)/\sigma$. A particular case of the skew-t distribution is the skew–Cauchy distribution, when $\nu = 1$. Also, when $\nu \to \infty$, we get the skew-normal distribution as the limiting case. Using Equation (7) of Lemma 1, we obtain the following expression for the cdf of Y

$$F(y) = 2\ T_2\left(\boldsymbol{y}^*\ ;\ \boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \nu\right), \tag{11}$$

where $\mathbf{y}^* = (y,0)^\top$ and $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are like in (8). Results (10) and (11) are proven in Appendix Appendix B.

- *The skew–slash distribution*: in this case $U \sim Beta(\nu, 1)$, with density $h(u|\nu) = \nu u^{\nu-1}$, $0 < u < 1$, with $\nu > 0$, so

$$k_m = \frac{\nu}{\nu - m/2}, \quad \nu > m/2.$$

For a random variable Y with skew-slash distribution, we use the notation $Y \sim SSL(\mu, \sigma^2, \lambda; \nu)$. The density of Y is given by

$$\begin{aligned} f(y|\mu,\sigma^2,\lambda;\nu) = 2\nu & \int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2) \\ & \times \Phi\left(u^{1/2}\frac{\lambda(y-\mu)}{\sigma}\right) du, \quad y \in \mathbb{R}. \end{aligned}$$

The cdf of the skew-slash distribution does not have a closed form expression. However, using Equation (7) of Lemma 1, we can write it in terms of the following integral, which can be approximated by numerical methods,

$$F(y) = \int_0^\infty 2\nu\Phi_2\Big(\boldsymbol{y}(u)^*; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big) u^{\nu-1}\, du,$$

where $\mathbf{y}(u)^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are like in (8).

3. The SMSN censored linear regression model

The linear regression model under SMSN distributions is defined as

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{12}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression parameters. For subject i, Y_i is a response variable and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of explanatory variables. We assume that

$$\varepsilon_i \sim SMSN\left(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H\right), \quad i = 1, \ldots, n, \tag{13}$$

are independent random variables. The value of the location parameter of ε_i is chosen in order to obtain $E[\varepsilon_i] = 0$, as in the normal model, see Lemma 1 in Basso et al. (2010). Thus, if the moments exist, we have

$$\begin{aligned} Y_i &\sim SMSN(\mathbf{x}_i^\top\boldsymbol{\beta} + b\Delta, \sigma^2, \lambda; H), \\ E[Y_i] &= \mathbf{x}_i^\top\boldsymbol{\beta}, \text{ and } Var[Y_i] = k_2\sigma^2 - b^2\Delta^2, \ \ i = 1, \ldots, n, \end{aligned}$$

where $b = -\sqrt{\frac{2}{\pi}}k_1$.

Estimation and diagnostic analysis for linear and nonlinear models under SMSN distributions have been widely discussed in the literature from a Bayesian and frequentist perspective see, for example, Cancho et al. (2010), Zeller et al. (2011), Garay et al. (2011) and Labra et al. (2012), among others. In this work we are interested in the situation in which the response variable can not be fully observed for all subjects, i.e. when Y_i in model defined in (12)–(13) is censored. Assuming left-censoring, Y_i is a latent variable, and we observe the variable V_i defined as

$$V_i = \begin{cases} c_i & \text{if} \quad Y_i \le c_i; \\ Y_i & \text{if} \quad Y_i > c_i, \end{cases} \tag{14}$$

for some known threshold point c_i, $i = 1, 2, \ldots, n$. The censored linear regression model using SMSN distributions, hereafter SMSN-CR model, is defined by combining (12)– (14). In this work we restrict ourselves to the cases where ϵ_i, in (13), is skew-Student-t (the St-CR model) or skew-slash (the SSL-CR model).

Consider the vector of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu\right)^\top$ and the vector of observations $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ of $\mathbf{V} = (V_1, V_2, \ldots, V_n)$. The log-likelihood function is given by

$$\begin{aligned} \ell(\boldsymbol{v}|\boldsymbol{\theta}) = &\sum_{i=1}^{n} \log\left[F_{SMSN}\left(v_i|\boldsymbol{\theta}; H\right)\right] \ \mathbb{1}_{\{c_i\}}(v_i) \\ &+ \sum_{i=1}^{n} \log\left[f_{SMSN}(v_i|\boldsymbol{\theta}; H)\right] \ \mathbb{1}_{(c_i,\infty)}(v_i), \end{aligned} \tag{15}$$

where $\mathbb{1}_{\mathbb{A}}(\cdot)$ is the usual indicator function of the set $\mathbb{A}$, that is, $\mathbb{1}_{\mathbb{A}}(x) = 1$ if $x \in \mathbb{A}$ and $\mathbb{1}_{\mathbb{A}}(x) = 0$ if $x \notin \mathbb{A}$.

In our theoretical development, we will use a left censoring pattern. Because the response Y_i is defined over the real line, extensions to right censored data are immediate. The right censored problem can be represented by a left censored problem by transforming the response Y_i and censoring level c_i to $-Y_i$ and $-c_i$, respectively.

4. Bayesian Inference for the SMSN-CR model

In this section we develop the Gibbs sampling algorithm to carry out Bayesian inference for the SMSN-CR model. To do so, as stated in Section 3, the stochastic representation given in (3) plays a key role. Following Cancho et al. (2011), we consider a reparameterization of the SMSN class of distributions based on the representation mentioned before in order to simplify the mathematical development of the algorithm.

Let $\boldsymbol{\omega} = (\boldsymbol{\beta}^\top, \Delta, \tau, \nu)^\top$ be the vector of parameters in focus, which has a one-to-one correspondence to the original vector of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu\right)^\top$, since

$$\Delta = \sigma\sqrt{\frac{\lambda}{\lambda^2+1}} \quad \text{and} \quad \tau = \frac{\sigma^2}{\lambda^2+1}.$$

Thus, we can obtain σ^2 and λ from Δ and τ using $\sigma^2 = \tau + \Delta^2$ and $\lambda = \Delta/\sqrt{\tau}$. Therefore, the Gibbs sampler can be used to draw samples from the posterior distribution of $\boldsymbol{\omega}$ or $\boldsymbol{\theta}$, indistinctly.

4.1. Prior distributions

In the Bayesian context, distributional prior specifications are needed for posterior inference. Following Cancho et al. (2011), we assume that $\boldsymbol{\beta} \sim \mathrm{N}_p\left(\boldsymbol{\mu}_0, \Sigma_0\right)$, where $\boldsymbol{\mu}_0$ and Σ_0 (positive definite) are known. Also, using the reparameterization mentioned in the beginning of this section, we assume that $\Delta \sim \mathrm{N}(\mu_\Delta, \sigma^2_\Delta)$ and $\tau \sim IGamma(a_\tau, b_\tau)$, the inverse gamma distribution, where μ_Δ, σ^2_Δ, a_τ and b_τ are known. These choices are made to ensure conjugacy.

Regarding ν, the parameter that indexes the mixing distribution $H(\cdot;\nu)$, we use the suggestion given in Cabral et al. (2012), i.e., $\nu \sim \mathrm{Texp}\left(\gamma\,;\,\mathbb{A}\right)$ and $\gamma \sim \mathrm{Unif}\left(a, b\right)$, where a and b are known hyperparameters. $\mathrm{Texp}\left(\gamma\,;\,\mathbb{A}\right)$ denotes the exponential distribution with rate parameter $\gamma > 0$ truncated on the interval $\mathbb{A}$, and $\mathrm{Unif}\left(a, b\right)$ denotes the uniform distribution on the interval (a, b). In order to guarantee the existence of the first two moments, we set $\mathbb{A} = (2, \infty)$ and $\mathbb{A} = (1, \infty)$ for the St-CR model and for the SSL-CR model, respectively.

We also assume independence between the parameters, thus the joint prior distribution of the parameter vector $\boldsymbol{\omega}$ is

$$\pi(\boldsymbol{\omega}) = \pi\left(\boldsymbol{\beta}\right)\pi\left(\Delta\right)\pi\left(\tau\right)\pi\left(\nu|\,\lambda\right).$$

Note that, although our prior assumption of independence may not be realistic for some sets of parameters, it leads to posterior distributions with good mathematical properties, as conjugacy, leading to an easy implementation of the Gibbs sampler. Moreover, if this assumption is not true, it will be corrected by the posterior distribution and will not undermine the inference process.

4.2. 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 developed using a data augmentation scheme, that consists in assume that the latent variables in the model, given by the vector of censored responses $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ and by the vectors $\boldsymbol{U} = (U_1, U_2, \ldots, U_n)^\top$ and $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)$ – see representation (3) – could be observed, and then obtain the full conditional distribution for each parameter in the model and for each latent variable, defined as the conditional distribution of one variable given values of all the remaining (the observed data included). Then, we draw samples from these (full conditional) distributions.

If we consider the augmented data, the stochastic representation of a random variable with SMSN distribution is given by

$$\begin{aligned} Y_i|U_i = u_i, T_i = t_i &\sim \mathrm{N}(\mathbf{x}_i^\top \boldsymbol{\beta} + \Delta t_i, u_i^{-1}\tau), \\ T_i|U_i = u_i &\sim \mathrm{TN}(b, u_i^{-1}\ ;\ (b, \infty)), \\ U_i &\sim H(\cdot|\nu), \end{aligned}$$

for $i = 1, 2, \ldots, n$.

The algorithm is as follows.

Step 1. For $i = 1, 2, \ldots, n$; if $v_i = c_i$ sample y_i (independently) from $\pi(y_i \mid v_i, t_i, u_i, \boldsymbol{\beta}, \Delta, \tau)$, which is a truncated normal distribution

$$\mathrm{TN}(\mathbf{x}_i^\top \boldsymbol{\beta} + \Delta t_i, u_i^{-1}\tau\ ;\ (-\infty, c_i]).$$

Otherwise $y_i = v_i$.

Step 2. For $i = 1, 2, \ldots, n$, sample t_i independently from $\pi(t_i \mid v_i, y_i, u_i, \boldsymbol{\beta}, \Delta, \tau)$, which is

$$\mathrm{TN}(\mu_{t_i}, \sigma^2_{t_i}\ ;\ [b, \infty)),$$

with $\mu_{t_i} = \frac{\Delta}{\Delta^2+\tau}\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} + \frac{b\tau}{\Delta}\right)$ and $\sigma^2_{t_i} = \frac{\tau}{u_i(\Delta^2+\tau)}$.

Step 3. Sample $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta} \mid \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{u}, \Delta, \tau, \boldsymbol{\nu})$, which is $\mathrm{N}_p\left(\boldsymbol{\mu}^*, \Sigma^*\right)$ with

$$\boldsymbol{\mu}^* = \Sigma^* \left(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \frac{\boldsymbol{X}^{*\top}\boldsymbol{y}^*}{\tau} - \frac{\Delta \boldsymbol{X}^{*\top}\boldsymbol{t}^*}{\tau}\right) \quad \text{and}$$

$$\Sigma^* = \left(\frac{\boldsymbol{X}^{*\top}\boldsymbol{X}^*}{\tau} + \Sigma_0^{-1}\right)^{-1},$$

where $\boldsymbol{t}^* = (t_1^*, \ldots, t_n^*)^\top, \boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top, \boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)^\top$ and, for $i = 1, 2, \ldots, n$, $t_i^* = \sqrt{u_i}t_i$, $y_i^* = \sqrt{u_i}y_i$ and $\boldsymbol{x}_i^* = (\sqrt{u_i}x_{i1}, \ldots, \sqrt{u_i}x_{ip})^\top$.

Step 4. Sample Δ from $\pi(\Delta \mid \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{u}, \boldsymbol{\beta}, \tau, \boldsymbol{\nu})$, which is $\mathrm{N}\left(\mu^*_\Delta, \sigma^{2*}_\Delta\right)$, with

$$\mu^*_\Delta = \sigma^{2*}_\Delta \left(\frac{\mu_\Delta}{\sigma^2_\Delta} + \frac{1}{\tau}\sum_{i=1}^n u_i t_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\right) \quad \text{and}$$

$$\sigma^{2*}_\Delta = \left(\frac{1}{\tau}\sum_{i=1}^n u_i t_i^2 + \frac{1}{\sigma^2_\Delta}\right)^{-1}.$$

Step 5. Sample τ from $\pi(\tau \mid \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{u}, \boldsymbol{\beta}, \Delta, \boldsymbol{\nu})$, which is a inverse gamma distribution

$$IGamma\left(a_\tau + \frac{n}{2}\ ,\ b_\tau + \frac{1}{2}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \Delta t_i)^2\right).$$

Step 6. Sample u_i, $i = 1, \ldots, n$, independently from $\pi(u_i \mid v_i, y_i, t_i, \boldsymbol{\beta}, \Delta, \tau, \boldsymbol{\nu})$, which is

1. for the skew-t case,
$$\text{Gamma}\left(\frac{\nu}{2} + 1, \frac{\nu + A_i}{2}\right),$$
where $A_i = \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \Delta t_i\right)^2 / \tau + (t_i - b)^2$;
2. for the skew-slash case,
$$\text{TGamma}\left(\nu + 1, \frac{A_i}{2}\ ;\ (0, 1)\right),$$
a truncated gamma distribution on $(0, 1)$;
3. for the skew-normal case, set $u_i = 1$ for $i = 1, 2, \ldots, n$.

Step 7. For skew-slash or skew-t, we need to sample ν from its full conditional distribution.

1. for the skew-t case,
   (a) Sample γ from $\pi(\gamma|\nu)$, which is $\text{TGamma}(2, \nu; (a, b))$.
   (b) Using a Metropolis-Hastings path, sample ν from its full conditional distribution

$$\begin{aligned} &\pi(\nu \mid \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{u}, \boldsymbol{\beta}, \Delta, \tau, \lambda) \propto \\ &\left(\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\right)^n \exp\left\{-\nu\left(\frac{1}{2}\sum_{i=1}^{n} u_i + \gamma\right)\right\} \\ &\times \prod_{i=1}^{n} u_i^{\frac{\nu}{2}-1} \mathbb{1}_{(2,\infty)}(\nu). \end{aligned} \tag{16}$$

We use the following artificial Gaussian state space model proposed by Abanto-Valle et al. (2015): given an observation $\nu^{(j-1)}$ obtained at stage $j-1$, generate a candidate ν^* from $q(\cdot|\nu^{(j-1)}, \cdots)$, which is the density of the truncated normal distribution

$$\text{TN}\left(\nu^{(j-1)} - \frac{c_{\nu^{(j-1)}}}{d_{\nu^{(j-1)}}}, \sqrt{-\frac{1}{d_{\nu^{(j-1)}}}}, (2, \infty)\right),$$

where

$$\begin{aligned} c_{\nu^{(j-1)}} &= \left.\frac{\partial \log \pi(\nu|\cdots)}{\partial \nu}\right|_{\nu=\nu^{(j-1)}} \quad \text{and} \\ d_{\nu^{(j-1)}} &= \left.\frac{\partial^2 \log \pi(\nu|\cdots)}{\partial^2 \nu}\right|_{\nu=\nu^{(j-1)}}. \end{aligned}$$

The new observation ν^* is accepted with probability

$$\min\left\{\frac{\pi(\nu^*|\cdots)\ q\left(\nu^*|\nu^{(j-1)}, \cdots\right)}{\pi(\nu^{(j-1)}|\cdots)\ q\left(\nu^{(j-1)}|\nu^*, \cdots\right)}, 1\right\},$$

where $\pi(\nu^*|\cdots)$ denotes (16) evaluated using the current values of γ and $\boldsymbol{u}$.

2. for the skew-slash case,
   (a) Sample γ from $\pi(\gamma|\nu)$, which is $\text{TGamma}(2, \nu; (a, b))$.
   (b) Sample ν from $\pi(\nu \mid \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{u}, \boldsymbol{\beta}, \Delta, \tau, \lambda)$ which is equal to

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

5. Bayesian model selection and influence diagnostics

5.1. 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 $\mathbf{z} = \{z_1, \ldots, z_n\}$ be an observed random sample from $\pi(\cdot|\boldsymbol{\theta})$. For the i-th observation, $i = 1, 2, \ldots, n$ the CPO_i is written as

$$\text{CPO}_i = \int \pi(z_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{z}_{(-i)})d\boldsymbol{\theta} = \left(\int \frac{\pi(\boldsymbol{\theta}|\mathbf{z})}{\pi(z_i|\boldsymbol{\theta})}d\boldsymbol{\theta}\right)^{-1}, \tag{17}$$

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 (17), that a Monte Carlo approximation can be obtained by using a MCMC sample $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_Q\}$ from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{z})$ (after burn-in and thinning). 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 (LPML) for the model*, defined by

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

Larger values of LMPL indicate better fit.

The deviance information criterion (DIC), proposed by Spiegelhalter et al. (2002), measures at the same time goodness of fit and model complexity. The *deviance* is defined as

$$\text{D}(\boldsymbol{\theta}) = -2\log\left(\prod_{i=1}^{n}\pi(z_i|\boldsymbol{\theta})\right).$$

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

$$\rho_{\text{DIC}} = \overline{\text{D}}(\boldsymbol{\theta}) - \text{D}(\tilde{\boldsymbol{\theta}}),$$

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

$$\overline{\text{D}}(\boldsymbol{\theta}) = -2\sum_{i=1}^{n}\text{E}\left[\log\pi(z_i|\boldsymbol{\theta})|\mathbf{z}\right],$$

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

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

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

$$\widehat{\overline{\text{D}}}(\boldsymbol{\theta}) = -\frac{2}{Q}\sum_{j=1}^{Q} \log\left(\prod_{i=1}^{n} \pi(z_i|\boldsymbol{\theta}_j)\right)$$

and, after this, we compute

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

More recently, Watanabe (2010) introduced another criterion for model selection that takes into account goodness of fit and complexity, the *Watanabe-Akaike information criterion* (WAIC), and proved that it is asymptotically equivalent to the Bayes cross-validation loss. First, let us define the *log pointwise predictive density*, given by

$$p(\mathbf{z}) = \sum_{i=1}^{n} \log \int \pi(z_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{z})\, d\boldsymbol{\theta}.$$

Basically, WAIC is $p(\mathbf{z})$ plus a correction for the effective number of parameters to adjust for overfitting. There are two different approachs to calculate this correction and both can be viewed as approximations to cross-validation, as discussed in Gelman et al. (2014). The first of them is similar to the one used on ρ_{DIC} and is given by

$$\rho_{\text{WAIC}_1} = 2\, p(\mathbf{z}) + \overline{\text{D}}(\boldsymbol{\theta}).$$

The other one is defined by

$$\rho_{\text{WAIC}_2} = \sum_{i=1}^{n} Var\left[\log \pi(z_i|\boldsymbol{\theta})|\boldsymbol{z}\right].$$

Finally, the two versions of the WAIC criterion are given by

$$\text{WAIC}_k = 2\,\rho_{\text{WAIC}_k} - 2\,p(\mathbf{z}) \qquad k = 1, 2. \tag{18}$$

It is important to notice that in Watanabe's original definition, the WAIC criterion was defined only as $-p(\mathbf{z})/n$ plus a corretion. Here, following the suggestion made by Gelman et al. (2014), we multiplied this term by -2 so as to be on deviance scale.

Again, computation of both versions of WAIC involves calculation of integrals which usually raise numerical problems. Thus, one can approximate the value of WAIC using a MCMC sample, as it was done in the DIC criterion case. First, the approximation of $p(\mathbf{z})$ is given by

$$\widehat{p(\mathbf{z})} = \sum_{i=1}^{n} \log \left(\frac{1}{Q} \sum_{j=1}^{Q} \pi(z_i|\boldsymbol{\theta}_j) \right)$$

and then, considering the approximation of $\overline{\mathrm{D}}(\boldsymbol{\theta})$ given before, the approximation of first version of WAIC criteria is given by

$$\widehat{\mathrm{WAIC}}_1 = 2\, \widehat{p(\mathbf{z})} + 2\, \widehat{\overline{\mathrm{D}}}(\boldsymbol{\theta}).$$

The approximation of the second version of WAIC, $\widehat{\mathrm{WAIC}}_2$, can be calculated if we consider the sample variance $V_{j=1}^{Q}(x) = \frac{1}{Q-1}\sum_{j=1}^{Q}(x_j - \bar{x})^2$ as an estimate of the variance, where $\bar{x} = \frac{1}{Q}\sum_{j=1}^{Q} x_j$, and use the MCMC sample $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_Q\}$ to approximate the value of $\pi(z_i|\boldsymbol{\theta})$, $i = 1, 2, \ldots, n$, that is

$$\widehat{\pi(z_i|\boldsymbol{\theta})} = \frac{1}{Q} \sum_{j=1}^{Q} \pi(z_i|\boldsymbol{\theta}_j).$$

Comparing the two versions of the WAIC criterion, Gelman et al. (2014) points out that $\widehat{\mathrm{WAIC}}_2$ is more recommended for pratical use than $\widehat{\mathrm{WAIC}}_1$, since its series expansion has closer resemblance to the series expansion for leave-one-out cross-validation and gives results that are closer to this method.

We also use the expected Akaike information criterion (EAIC), see Brooks (2002), and the expected Bayesian information criterion (EBIC), see Carlin and Louis (2001), to compare models. These criteria are defined by

$$EAIC = \overline{\mathrm{D}}(\boldsymbol{\theta}) + 2\vartheta \quad \text{and} \quad EBIC = \overline{\mathrm{D}}(\boldsymbol{\theta}) + \vartheta \log(n),$$

where ϑ is the number of parameters in the model. Replacing $\overline{\mathrm{D}}(\boldsymbol{\theta})$ by $\widehat{\overline{\mathrm{D}}}(\boldsymbol{\theta})$, one can obtain an estimate of these criteria.

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 given by (15).

To evaluate model adequacy, we use a discrepancy measure based on the posterior predictive distribution. One can use any pre-fixed statistic to measure if its observed value is extreme relative to the reference distribution (the posterior predictive distribution). If this is the case, there is some concern with respect to the assessment of model fit to the data. Define y_i to be the observed data. Gelman et al. (2004) use a function of the log-likelihood as a summary statistic, given by

$$T(y, \boldsymbol{\theta}) = -2 \sum_{i=1}^{n} \log \left[\pi(y_i \mid \boldsymbol{\theta}) \right]. \tag{19}$$

The Bayesian p-value/posterior predictive p-value, proposed by Rubin (1984), is defined to be

$$p_B = \Pr(T(y_{pr}, \boldsymbol{\theta}) \geq T(y, \boldsymbol{\theta}) | Y = y),$$

where y_{pr} denotes a simulated draw from the posterior predictive distribution. It is the number of times $T(y_{pr}, \boldsymbol{\theta})$ exceeds $T(y, \boldsymbol{\theta})$ out of L simulated draws. According to Gelman et al. (2004, pp. 180), a model is suspect if a discrepancy is of practical importance and its p-value is close to 0 or 1. An extreme p-value implies that the model cannot be expected to capture this aspect of the data. A very small or very large p-value (< 0.05 or > 0.95, say) signals model misspecification, *i.e.*, the observed pattern would be unlikely to be seen in replications of the data under the true model.

5.2. 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* $\boldsymbol{\theta}$ (Csiszar, 1967) is defined by

$$d_q(\pi_1, \pi_2) = \int q\left(\frac{\pi_1(\boldsymbol{\theta})}{\pi_2(\boldsymbol{\theta})}\right) \pi_2(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{20}$$

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 L_1*-distance divergence* is obtained when $q(z) = |z-1|$.

Let $\mathbf{y} = \{y_1, \ldots, y_n\}$ be the sample and I a subset of $\{1, \ldots, n\}$. 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 (20). This influence measure can be written as

$$d_q(I) = \mathrm{E}\left[q\left(\frac{\pi_1(\boldsymbol{\theta}|\mathbf{y}_{(-I)})}{\pi_2(\boldsymbol{\theta}|\mathbf{y})}\right) |\mathbf{y}\right]. \tag{21}$$

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 (20), 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 $p \geq 0.80$ as a strong bias, then we can say that the observation i is influential when $d_{L_1}(\{i\}) \geq 0.60$, since $d_{L_1}^*(0.80) = 0.60$. Similarly, for the Kullback-Leibler and $J-$distance divergences, we have $d_{KL}^*(0.80) \approx 0.2231436$ and $d_J^*(0.80) \approx 0.4158883$, respectively. Thus, if we use the Kullback-Leibler divergence, we can consider an influential observation when $d_{KL}(\{i\}) > 0.22$ and, using the J-distance, an observation with $d_J(\{i\}) > 0.41$ can be considered as influential.

6. Data analysis

In order to study the performance of our proposed model and algorithm, we analyze a real data set. The computational procedures of this section, and of the next section, were implemented using the `R` software (R Development Core Team, 2015), through the package `BayesCR` (Garay et al., 2015c). We consider the wage rate data set described in 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:

- 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_{i1}: wife's age;
- x_{i2}: years of schooling;
- x_{i3}: the number of children younger than six years old in the household;
- x_{i4}: the number of children between six and nineteen years old.

Each of the vectors of explanatory variable values is given by $\mathbf{x}_i^\top = (1, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ for $i = 1, 2, \ldots, 753$. This data set was analyzed by Arellano-Valle et al. (2012) using a censored regression model with Student-t responses and, more recently, Garay et al. (2015a) and Garay et al. (2015b) using a censored regression model with SMN responses from a Bayesian and a frequentist point of view, respectively. Here, we revisit this data set in order to evaluate the performance of the proposed Bayesian methods considering the class SMSN-CR.

6.1. Estimation

In the estimation process, we consider the prior densities discussed in Subsection 4.1. We generated two parallel independent MCMC runs of size 400,000 with widely dispersed initial values for each parameter, considering a burn-in of 100,000 iterations and a thin of 30.

The convergence of the MCMC chains was monitored using trace plots, autocorrelation plots, and Gelman-Rubin $\widehat{R}$ diagnostics. Table 1 reports the posterior means (Mean), standard deviations (SD), highest posterior density (HPD) credible intervals (95%) and Gelman-Rubin statistic ($\hat{R}$) of the parameters after fitting the different SMSN-CR models. One can notice that for all the models, except the SSL-CR, the HPD interval for β_1 contains the value 0. Also, notice that the small value of the estimate of ν for the St-CR and SSL-CR models indicates a lack of adequacy of the skew-normal (or normal) assumption. Table 2 compares the fit of the three asymmetric models and the N-CR model using the model selection criteria discussed in Section 5.1. Note that the models with heavy tails (St-CR and SSL-CR) perform significantly better than the N-CR and SN-CR models. Moreover, it seems that adding the extra skewness parameter also improves data fitting, as can be seen comparing the symmetric model N-CR with its asymmetric version, SN-CR. In the end, the SSL-CR model outperforms all the rest. In

Table 2 one can also find the values of p_B, the Bayesian p-value calculated with the posterior sample of the parameters. These values indicate no lack of fit at all.

Model	Parameters	Mean	SD	HPD (95%)	$\hat{R}$
N-CR	β_1	-2.752	1.748	(-6.133; 0.665)	1.000003
	β_2	-0.106	0.028	(-0.161;-0.051)	1.000007
	β_3	0.731	0.084	(0.569; 0.896)	0.999999
	β_4	-3.056	0.448	(-3.923;-2.188)	1.000000
	β_5	-0.215	0.153	(-0.521; 0.077)	1.000003
	σ^2	21.325	1.5999	(18.222;24.483)	1.000010
SN-CR	β_1	-1.034	1.632	(-4.178;2.206)	1.000004
	β_2	-0.120	0.026	(-0.170;-0.070)	0.999999
	β_3	0.675	0.081	(0.519; 0.836)	0.999999
	β_4	-3.243	0.442	(-4.112;-2.389)	1.000005
	β_5	-0.259	0.146	(-0.542;0.030)	1.000001
	σ^2	33.708	3.270	(27.143; 39.833)	1.000229
	λ	1.803	0.380	(1.159; 2.576)	1.000663
St-CR	β_1	-3.058	1.516	(-5.856; 0.083)	1.000025
	β_2	-0.088	0.024	(-0.133;-0.040)	1.000011
	β_3	0.673	0.068	(0.540; 0.806)	1.000012
	β_4	-2.809	0.387	(-3.569;-2.065)	1.000011
	β_5	-0.267	0.128	(-0.510;-0.011)	1.000007
	σ^2	22.562	4.495	(13.774;31.283)	0.999999
	λ	-1.422	0.377	(-2.141;-0.656)	1.000060
	ν	4.877	0.255	(4.656; 5.369)	1.006467
SSL-CR	β_1	-4.127	1.485	(-7.097; -1.349)	1.000003
	β_2	-0.079	0.023	(-0.124; -0.036)	1.000013
	β_3	0.669	0.065	(0.542; 0.796)	1.000006
	β_4	-2.688	0.366	(-3.406; -1.979)	0.999998
	β_5	-0.265	0.122	(-0.505; -0.030)	1.000003
	σ^2	13.424	2.369	(8.938; 18.123)	1.000026
	λ	-1.940	0.397	(-2.728; -1.183)	1.000036
	ν	1.063	0.064	(1.001; 1.191)	1.000144

Table 1: Wage rate data. Posterior mean, standard deviation (SD), HPD (95%) interval and Gelman and Rubin potential scale reduction statistic ($\hat{R}$) for the parameters in the SMSN-CR models.

	Model			
Criterion	N-CR	SN-CR	St-CR	SSL-CR
LPML	-1489.290	-1479.075	-1441.834	**-1432.518**
DIC	2975.017	2955.640	2881.913	**2863.778**
EAIC	2975.381	2955.402	2884.199	**2864.841**
EBIC	3003.126	2987.770	2921.192	**2901.834**
WAIC_1	2978.080	2958.067	2883.431	**2864.796**
WAIC_2	2978.651	2958.144	2883.766	**2865.119**
p_B	0.3693	0.6098	0.5293	0.5425

Table 2: Wage rate data. Comparison between the SMSN-CR models.

6.2. Bayesian case influence diagnostics

Considering the samples of the posterior distributions of the parameters of the four models, q-divergence measures, described in Section 5.2, were computed (using p=0.80). The cases #185, #349, #394 and #408 were identified as influential under the N-CR model by the K-L divergence, because they exceed the specified thresholds. When SN-CR model was fitted, observation #394 was not considered influential anymore, although it was still close to the threshold. However, no observation was influential for the St-CR and SL-CR models. Figure 1 depicts the index plot of the K-L divergence and, as expected, the effect of influential observations on the Bayesian estimates of the parameters are attenuated when heavy-tailed and/or asymmetric distributions are considered.

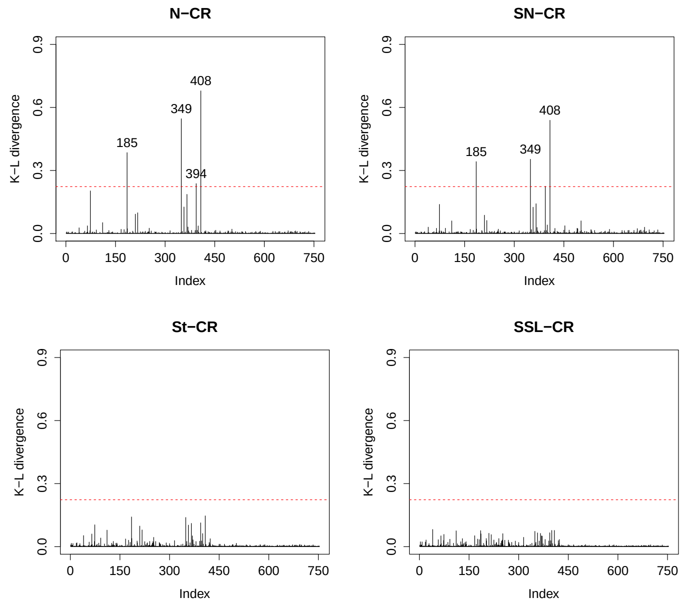

Figure 1: Wage rate data. K-L divergence for N-CR, SN-CR, St-CR and SSL-CR models

In order to reveal the impact of these four observations on the parameter estimates we refitted the N-CR and SSL-CR models (the ones with the worst and best values for model selection criteria in Table 2, respectively), first removing one by one and then all four influential points. In Table 3 we show the relative changes (in percentage) of each parameter estimate, which is defined by

$$\mathbf{RC}_{\hat{\theta}_{j(I)}} = \left|(\hat{\theta}_j - \hat{\theta}_{j(I)})/\hat{\theta}_j\right| \times 100,$$

where $\hat{\theta}_{j(I)}$ denotes the Bayesian estimate of θ_j after the set I of observations was removed. In this Table, $^{(*)}$ indicates parameters that were not significant in the original fitting and that became significant when influential observations were removed. Note that the intercept β_1 is heavily impacted by these observations when compared to the other regression coefficients. Note that all the relatives changes are smaller under the SSL-CR model than under the N-CR model, showing that SSL-CR is more robust, as expected. Besides, the parameter significance was unaltered under the SSL-CR model fit, while β_5, which was not considered significant under the original fit of the N-CR model, became significant when this model was adjusted without observation #185, as well as when all influential observations were removed – this fact shows one more time the robustness of the SSL-CR model when compared to the N-CR model.

7. Simulation study

In order to study the performance of our proposed models and algorithm, we present two simulation studies. The computational procedures were implemented using the `R` software (R Development Core

	N-CR					
Set{I}	β_1	β_2	β_3	β_4	β_5	σ^2
All - {#185}	2.43	3.63	0.86	3.30	3.56$^{(*)}$	6.68
All - {#349}	22.59	10.19	1.01	0.44	19.85	6.35
All - {#394}	8.46	3.14	4.09	2.07	5.65	4.32
All - {#408}	7.16	0.36	0.80	3.48	19.47	7.73
All - {#185, #349, #394, #408}	33.15	1.02	7.36	0.46	35.53$^{(*)}$	25.64
	SSL-CR					
Set{I}	β_1	β_2	β_3	β_4	β_5	σ^2
All - {#185}	2.46	2.27	0.22	1.11	1.91	3.73
All - {#349}	0.37	0.06	0.19	0.62	0.59	3.09
All - {#394}	1.16	1.52	0.49	0.32	2.64	3.03
All - {#408}	2.06	1.42	0.14	0.39	1.65	4.03
All - {#185, #349, #394, #408}	7.79	6.37	0.20	3.23	3.57	16.46

Table 3: Wage rate data. Relative changes (in %) for all parameters in N-CR and SSL-CR models.

Team, 2015). 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 performance of the three assymetric models, SN-CR, St-CR and SSL-CR, when some observations are perturbed, generating outliers.

7.1. 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 so we generated a left-censored variable with Normal Inverse Gaussian distribution (Barndorff-Nielsen, 1997) with shape parameter 5, skewness parameter 4.9 and scale parameter 2. The vector with location parameters is given by $\boldsymbol{X}^\top\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector $(-10;\ 2)^\top$ and $\boldsymbol{X}$ is a 200×2 matrix with a all-ones first column and a second column generated from a Uniform distribution on the interval $(0, 4)$.

We have chosen several censoring proportion settings (10%, 25%, 40% and 50%) and the prior specification has been fixed as in Subsection 4.1, with $\boldsymbol{\mu}_0 = \mathbf{0}_2$, $\boldsymbol{\Sigma}_0 = 100\mathbf{I}_2$, $\mu_\Delta = 0$, $\sigma^2_\Delta = 100$,$a_\tau = 2.1$, $b_\tau = 3$, $c = 0.02$, $d = 0.49$, $e = 0.02$ and $f = 0.9$. Here, $\mathbf{0}_2$ denotes a vector with length 2 with all components equal to zero and $\mathbf{I}_2$ denotes the identity matrix with dimension 2.

For each level of censoring, we simulated 150 data sets and, for each data set, we fitted the N-CR, SN-CR, St-CR and SSL-CR models and recorded the MCMC estimates of the parameters. Then, we computed the estimated bias and the estimated mean squared error (MSE) for the estimates of the regression coefficient $\boldsymbol{\beta}$ in each model. We ran 60,000 iterations of the Gibbs sampler, burned-in the first 18,000 and used a thin of 3, so each final chain of MCMC observations has size 14000. For the parameter β_j, $j = 1, 2$, we define the estimated Bias and MSE as

$$\text{Bias} = \frac{1}{150}\sum_{i=1}^{150}(\hat{\beta}_j^{(i)} - \beta_j) \qquad , \qquad \text{MSE} = \frac{1}{150}\sum_{i=1}^{150}(\hat{\beta}_j^{(i)} - \beta_j)^2$$

where $\hat{\beta}_j^{(i)}$ is the Bayesian estimate of β_j for the $i-$th simulated data set, for $j = 1, 2$ and $i = 1, 2, \ldots, 150$.

Figure 2 presents Bias and MSE for all the four models fitted and the four censoring patterns. Figure 3 summarizes via box-plot all the 150 pontual estimates for β_1 and β_2, comparing them with the real values of these parameters, for all the models fitted and censoring patterns.

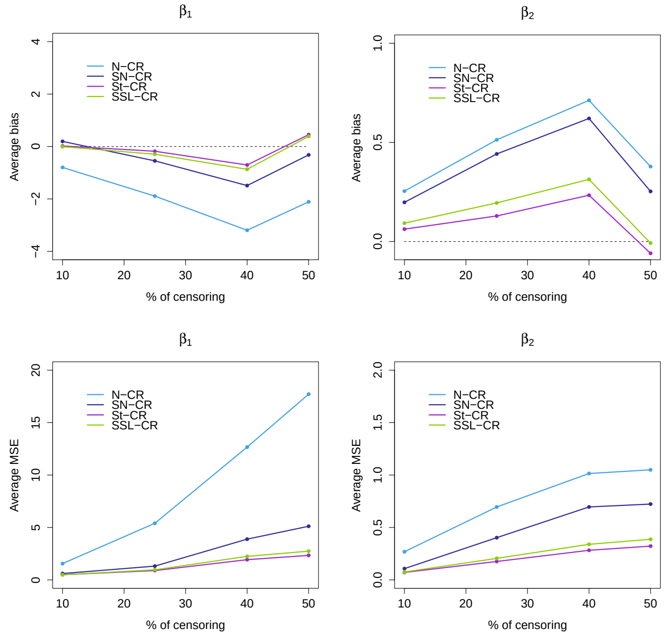

Figure 2: Bias and MSE of parameters β_1 and β_2 for N-CR, SN-CR, St-CR and SSL-CR models with different settings of censoring proportions

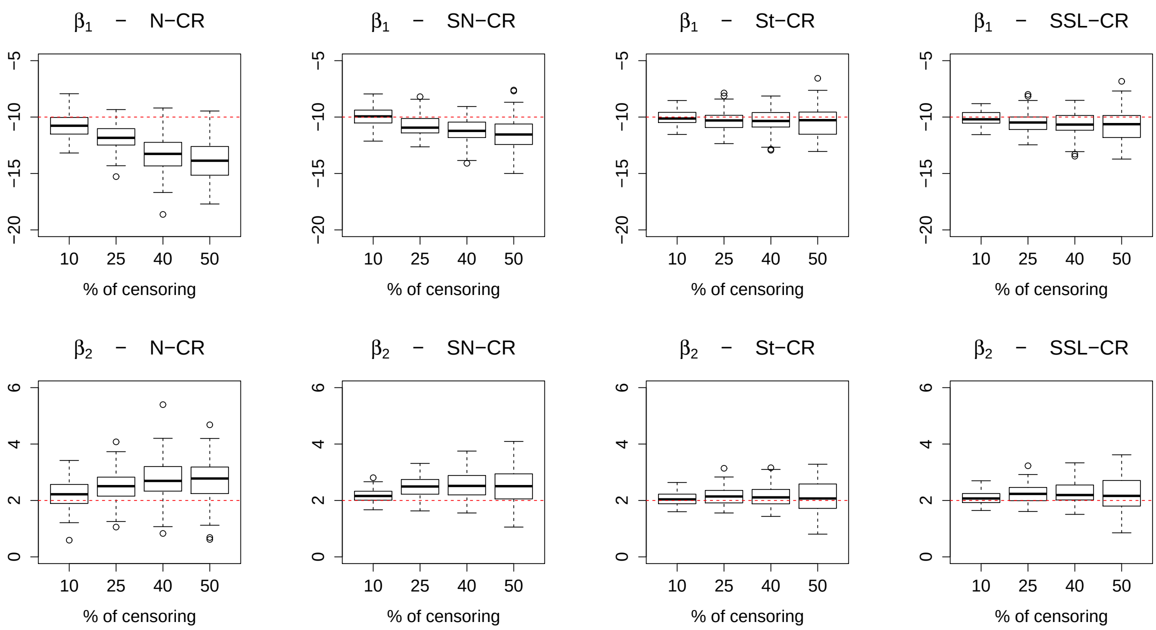

Figure 3: Box-plot for the 150 point estimates (posterior mean via MCMC samples) of β_1 and β_2, for N-CR, SN-CR, St-CR and SSL-CR models with different settings of censoring proportions, in comparison with the true value of parameters (red line).

From Figures 2 and 3 we can observe that the St-CR model presents better performance at all levels of censoring and it is no much different from the SSL-CR model. Comparing the symmetric and the asymmetric versions of Normal model, it is clear that the fitting performance is improved a lot when we add the skewness parameter for all the censoring levels, especially for the β_1 estimation. This also happens when we compare normal and skew-normal models with the heavy-tailed ones, showing that the kurtosis parameter plays an important role on the estimation process. On the other hand, all the models seems to loose performance when the censoring levels increases.

7.2. Study II

The goal of this study is to compare the performance of the parameter estimates for the SN-CR, St-CR and SSL-CR models in the presence of outliers on the response variable.

We performed a simulation study based on the SN-CR model. Specifically, we considered $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (10, 15)$, $\boldsymbol{x}_i^\top = (1, x_i)$, and errors ε_i with skew-normal distribution with scale parameter $\sigma^2 = 2$ and shape parameter $\lambda = -4$, $i = 1, \ldots, n$. The values x_i, $i = 1, \ldots, n$, were generated independently from a uniform distribution on the interval $(1, 3)$. A sample of size $n = 100$ was generated from this model with 10% of censoring. We perturbed observations #3 ($y_3 = 43.22178$), #66 ($y_{66} = 51.17056$) and #92 ($y_{92} = 31.82169$), which were randomly chosen among the non-censored observations, by increasing them Λ% of their original value, for $\Lambda = 10, 20, 30, \ldots, 150$. It means that, if y denotes the original observation, the perturbed observation y^* is given by:

$$y^* = \left(1 + \frac{\Lambda}{100}\right) y.$$

For each one of the 15 patterns of perturbation, we fitted the SN-CR, St-CR and SSL-CR models and computed the relative change in $\boldsymbol{\beta}$ estimates (comparing with the fit of the non-perturbed data). Looking at the graph, of Figure 4, of the relative change for β_1 (the intercept) one can see that for perturbations smaller than 100%, the relative changes are not so significant (smaller than 5%) and there is no precise pattern. But, when Λ becomes larger than 100, we see that the relative change for St-CR and SSL-CR models seems to stabilize near 5%, while for the SN-CR model it presents an increasing pattern, reaching 10% when $\Lambda = 150$. For β_2, one can see that for small ($\Lambda \in \{10, 20, 30\}$) perturbations the three models behave in a very similar way but, when Λ increases, the SN-CR model loses performance as it is less robust than the St-CR and SSL-CR models to deal with outliers.

We also recorded the LPML, DIC, WAIC, EAIC and EBIC criteria. Figure 5 shows the results for the LPML and WAIC criteria. We observe that the SN-CR fit is as good as the the other ones for small values of Λ, what is expected as the data were generated from a skew-normal distribution. However, as perturbation increases, the St-CR and SSL-CR models are preferred. In Appendix Appendix C one can find the graphs for DIC, EAIC and EBIC criteria, with similar results.

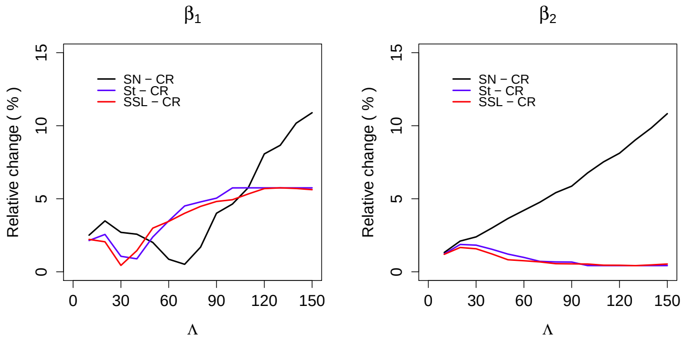

Figure 4: RC (in %) for β_1 and β_2 for the SN-CR, St-CR and SSL-CR models with different levels of perturbation Λ.

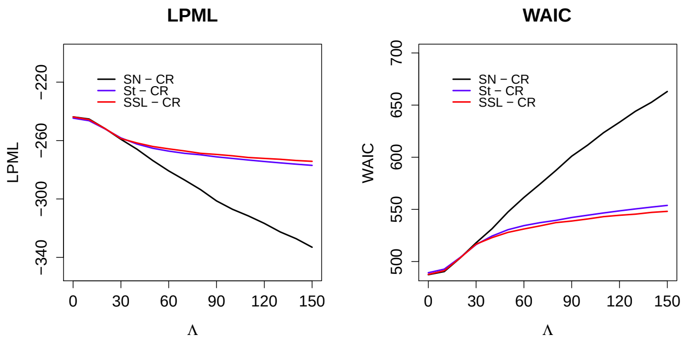

Figure 5: Model selection criteria for the SN-CR, St-CR and SSL-CR models with different levels of perturbation Λ.

8. Conclusions

In this paper, we proposed the class of SMSN distributions as a replacement for the conventional choice of normal distribution for censored linear models where computational issues and outlier identification are concerned. It generalizes the works of Barros et al. (2010), Arellano-Valle et al. (2012) and Massuia et al. (2015), using a Bayesian approach.

In order to explore the statistical properties of the proposed models, an efficient Gibbs-type algorithm, in the sense of Liu and Rubin (1994), has been coded and implemented using the `R` package `BayesCR` (Garay et al., 2015c) which is available for download at the CRAN repository. Two simulation studies were performed. The first simulation study revealed gain in efficiency and accuracy for parameter estimates (especially for the β_1 esimation) for all the censoring levels when we add the skewness parameter and the typical assumptions of normality is questionable. In the second simulation study, we showed that the performnce of the parameter estimates for the St-CR and SSL-CR models are better that under the N-CR and SN-CR models when the perturbation increases.

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 SMSN-CR model showed considerable flexibility to accommodate outliers.

Acknowledgment

Monique B. Massuia, Víctor H. Lachos and Aldo M. Garay would like to acknowledge the support of grants 2012/18702-9, 2013/21468-0 and 2014/02938-9, São Paulo Research Foundation (FAPESP), respectively. Celso R. B. Cabral was supported by CNPq (via BPPesq, Universal Project and grant 167731/2013-0), and FAPEAM (via Universal Amazonas Project).

Appendix A. Proof of Lemma 1

Proof. Let $Y \sim SMSN(\mu, \sigma^2, \lambda\ ;\ H)$. Using the pdf of Y given in Equation (5), we have that:

$$
\begin{aligned}
F(y) &= \int_{-\infty}^{y}\int_{0}^{\infty}\int_{0}^{\infty} 2\phi(z\ ;\ \mu+\Delta t, \kappa(u)\tau)\phi(t\ ;\ 0, \kappa(u))dt\, dH(u)\, dz \\
&= \int_{0}^{\infty}\int_{0}^{\infty} 2\left[\int_{-\infty}^{y} \phi(z\ ;\ \mu+\Delta\kappa(u)^{1/2}x, \kappa(u)\tau)\, dz\right]\phi(x)\, dx\, dH(u) \qquad \text{(A.1)} \\
&= 2\int_{0}^{\infty}\int_{0}^{\infty} \Phi\left(\frac{y-\mu-\Delta\kappa(u)^{1/2}x}{\kappa(u)^{1/2}\tau^{1/2}}\right)\phi(x)\, dx\, dH(u) \\
&= 2\int_{0}^{\infty}\int_{0}^{\infty} \Phi\left(\frac{y-\mu}{\kappa(u)^{1/2}\sigma\sqrt{1-\delta^2}} - \frac{\delta}{\sqrt{1-\delta^2}}x\right)\phi(x)\, dx\, dH(u) \qquad \text{(A.2)} \\
&= 2\int_{0}^{\infty}\int_{0}^{\infty} \Phi\left(\frac{(y-\mu)\sqrt{1+\lambda^2}}{\kappa(u)^{1/2}\sigma} - \lambda x\right)\phi(x)\, dx\, dH(u). \qquad \text{(A.3)}
\end{aligned}
$$

Equation (A.1) is obtained using the transformation $x = t/\sqrt{\kappa(u)}$. Equations (A.2) and (A.3) are consequence of considering the relations $\Delta = \sigma\delta$, $\tau = \sigma^2(1-\delta^2)$ and $\delta = \lambda/\sqrt{1+\lambda^2}$, and we have obtained Equation (6).

To obtain Equation (7), we use the following result: let $\mathbf{Z} = (X, W)^\top$ be a random vector with bivariate normal distribution with $E[X] = E[W] = 0$, $Var[X] = Var[W] = 1$ and correlation coefficient ρ. Then, the cdf of $\mathbf{Z}$ can be written as

$$
F_{\mathbf{Z}}(x, w) = \int_{-\infty}^{x} \phi(s)\Phi\left(\frac{w-\rho s}{\sqrt{1-\rho^2}}\right) ds. \qquad \text{(A.4)}
$$

A proof of this result can be found in Parrish and Bargmann (1981). Using expression (4), we have that $Y \sim SMSN(\mu, \sigma^2, \lambda\ ;\ H)$ has cdf

$$
F(y) = 2\int_{0}^{\infty}\int_{-\infty}^{y} \phi(z; \mu, \kappa(u)\sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma\kappa(u)^{1/2}}\right) dz\, dH(u).
$$

Using the transformation $s = (z-\mu)/(\sigma\kappa(u)^{1/2})$, we obtain

$$
F(y) = 2\int_{0}^{\infty}\int_{-\infty}^{\frac{y-\mu}{\sigma\kappa(u)^{1/2}}} \phi(s)\Phi(\lambda s) ds\, dH(u).
$$

If we make $\lambda = -\rho/(\sqrt{1-\rho^2})$ in (A.4), which implies $\rho = -\delta$, we have that

$$
\begin{aligned}
F(y) &= 2\int_{0}^{\infty} F_{\mathbf{Z}}\left(\frac{y-\mu}{\sigma\kappa(u)^{1/2}}, 0\right) dH(u) \\
&= 2\int_{0}^{\infty} P\left(\sigma X + \mu \leq \frac{y}{\kappa(u)^{1/2}}, W \leq 0\right) dH(u) \\
&= \int_{0}^{\infty} 2\Phi_2\Big(\boldsymbol{y}(u)^*; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big) dH(u). \qquad \text{(A.5)}
\end{aligned}
$$

Observe that the random vector $(\sigma X + \mu, W)^\top$ has a bivariate normal distribution with mean vector $\boldsymbol{\mu}^*$ and covariance matrix $\boldsymbol{\Sigma}$, implying Equation (A.5).

□

Appendix B. Development of the pdf and cdf of the St distribution

In this Appendix, we derive the pdf and cdf of the St distribution. Using the general definition of pdf of the SMSN family given in Equation (4), the density of the St distribution is given by

$$\begin{aligned} f(y) &= 2\int_0^\infty \phi(y\ ;\ \mu, u^{-1}\sigma^2)\ \Phi\left(\frac{\lambda(y-\mu)}{u^{-1/2}\sigma}\right) \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} u^{\nu/2-1} \exp\left\{-\frac{\nu}{2}u\right\}\ du \\ &= \frac{\sqrt{2}\ (\nu/2)^{\nu/2}}{\sqrt{\pi}\sigma\Gamma(\nu/2)} \int_0^\infty u^{\frac{\nu-1}{2}} \exp\left\{-u\left(\frac{\nu}{2}+\frac{d(y)^2}{2}\right)\right\} \Phi\left(\lambda\ d(y)\ \sqrt{u}\right) du \\ &= \frac{\sqrt{2}(\nu/2)^{\nu/2}}{\sqrt{\pi}\sigma\Gamma(\nu/2)} \Gamma\left(\frac{\nu+1}{2}\right) \left(\frac{\nu+d(y)^2}{2}\right)^{-\frac{\nu+1}{2}} E_X\left[\Phi\left(\lambda\ d(y)\ \sqrt{X}\ ;\ 0,1\right)\right], \end{aligned}$$

where $d(y) = \frac{y-\mu}{\sigma}$ and $X \sim Gamma\left(\frac{\nu+1}{2}, \frac{\nu+d(y)^2}{2}\right)$. Using Lemma 2, we have that:

$$f(y) = \frac{2\ \Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi\nu}\sigma} \left(1+\frac{d(y)^2}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\lambda\ d(y)\ \sqrt{\frac{\nu+1}{\nu+d(y)^2}}\ ;\ \nu+1\right).$$

Using Equation (7) of Lemma 1, the cdf of the Skew-t distribution becomes

$$F(y) = 2E_U\left[\Phi_2\left(\boldsymbol{y}(U)^*\ ;\ \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\right)\right],$$

where $U \sim Gamma(\nu/2, \nu/2)$, $\boldsymbol{y}(u)^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are defined in Lemma (1). Thus, by Lemma 2, we have that

$$F(y) = 2\ T_2\left(\begin{pmatrix} y \\ 0 \end{pmatrix}\ ;\ \boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \nu\right).$$

Appendix C. Complementary results of the simulation study II: Performance of model selection criteria

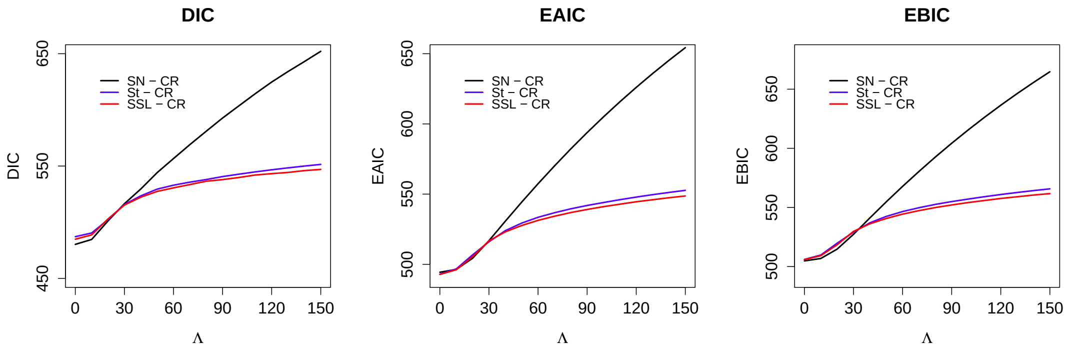

Figure C.6: DIC, EAIC and EBIC criteria for simulation study II

References

Abanto-Valle, C., Lachos, V., Dey, D., 2015. Bayesian estimation of a skew-student-t stochastic volatility model. Methodology and Computing in Applied Probability 17, 721–738.

Ando, T., 2010. Bayesian Model Selection and Statistical Modeling. Chapman & Hall/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.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

Barndorff-Nielsen, O. E., 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. Scandinavian Journal of Statistics 24, 1–13.

Barros, M., Galea, M., González, M., Leiva, V., 2010. Influence diagnostics in the tobit censored response model. Statistical Methods & Applications 19, 716–723.

Basso, R. . M., Lachos, V. H., Cabral, C. R., Ghosh, P., 2010. Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis 54, 2926–2941.

Branco, M. D., Dey, D. K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99–113.

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, 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.

Cancho, V. G., Dey, D. K., Lachos, V. H., Andrade, M. G., 2011. Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. Computational Statistics & Data Analysis 55, 588–602.

Cancho, V. G., Lachos, V. H., Ortega, E. M. M., 2010. A nonlinear regression model with skew-normal errors. Statistical Papers 51, 547–558.

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.

Csiszar, I., 1967. Information-type measures of difference of probability distributions and indirect observations. Studia Scientiarum Mathematicarum Hungarica 2, 299–318.

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.

Gamerman, D., Lopes, H. F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edition. Chapman & Hall/CRC.

Garay, A. M., Bolfarine, H., Lachos, V. H., Cabral, C. R. B., 2015a. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. Journal of Applied Statistics, DOI 10.1080/02664763.2015.1048671.

Garay, A. M., Lachos, V. H., Abanto-Valle, C. A., 2011. Nonlinear regression models based on scale mixtures of skew-normal distributions. Journal of the Korean Statistical Society 40, 115–124.

Garay, A. M., Lachos, V. H., Bolfarine, H., Cabral, C. R. B., 2015b. Linear censored regression models with scale mixtures of normal distributions. Statistical Papers, DOI 10.1007/s00362–015–0696–9.

Garay, A. M., Massuia, M. B., Lachos, V. H., 2015c. BayesCR: Bayesian analysis of censored linear regression models with scale mixtures of normal (SMSN) distributions. R package v. 2.0.
URL `http://CRAN.R-project.org/package=BayesCR`

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman & Hall/CRC.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24, 997–1016.

Ibacache-Pulgar, G., Paula, G. A., 2011. Local influence for Student-t partially linear models. Computational Statistics & Data Analysis 55, 1462–1478.

Labra, F. V., Garay, A. M., Lachos, V. H., Ortega, E. M. M., 2012. Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. Journal of Statistical Planning and Inference 142, 2149–2165.

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.

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.

Liu, C., Rubin, D. B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. Biometrika 80, 267–278.

Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H., 2015. Influence diagnostics for Student-t censored linear regression models. Statistics 49 (5), 1074–1094.

Meza, C., Osorio, F., De la Cruz, R., 2012. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. Statistics and Computing 22, 121–139.

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.

Parrish, R., Bargmann, R., 1981. A method for evaluation of cumulative probabilities. of bivariate distributions using the Pearson family. Statistical Distributions in Scientific Work 5, 241–257.

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

Prates, M. O., Costa, D. R., Lachos, V. H., 2014. Generalized linear mixed models for correlated binary data with t-link. Statistics and Computing 24, 1111–1123.

R Development Core Team, 2015. 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

Rubin, D. B., 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics 12, 1151–1172.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. A Bayesian measure of model complexity and fit (with discussion). Journal of Royal Statistical Society 64, 583–639.

Stapleton, D. C., Young, J. D., 1984. Censored normal regression with measurement error on the dependent variable. Econometrica 52, 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.

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.

Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. The Journal of Machine Learning Research 11, 3571–3594.

Zeller, C. B., Lachos, V. H., Vilca-Labra, F. E., 2011. Local influence analysis for regression models with scale mixtures of skew-normal distributions. Journal of Applied Statistics 38, 343–368.