Likelihood Based Inference for Censored Linear Regression Models with Scale Mixtures of Skew-Normal Distributions

Thalita do Bem Mattos[a], Aldo M. Garay[a,b], Victor H. Lachos*[a]

[a]*Departamento de Estatística, Universidade Estatual de Campinas, Campinas, São Paulo, Brazil*
[b]*Departamento de Estatística, Universidade Federal de Pernambuco, Brazil*

Abstract

In many studies the data collected are subject to some upper and lower detection limits. Hence, the responses are either left or right censored. A complication arises when these continuous measures present heavy tails and asymmetrical behavior, simultaneously. For such data structures, we propose a robust censored linear model based on the 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, skew-contaminated normal and the entire family of scale mixtures of normal (SMN) distributions as special cases. We propose a fast estimation procedure to obtain the maximum likelihood (ML) estimates of the parameters, using a stochastic approximation of the EM (SAEM) algorithm. This approach allows us to estimate the parameters of interest easily and quickly, obtaining as a byproduct the standard errors, predictions of unobservable values of the response and the log-likelihood function. The proposed methods are illustrated through a real data application and several simulation studies.

Keywords: Censored regression models; Heavy tails; SAEM algorithm; Scale mixtures of skew-normal distributions.

1. Introduction

Linear and non linear regression models with normal observational errors are usually applied to model symmetrical data. However, several phenomena are not always in agreement with the assumptions of the normal model. To deal with this problem, some proposals have been made in the literature to replace the normality assumption with more flexible classes of distributions. For instance, Fernández and Steel (1999) discuss some inferential procedures in regression models with Student-t distribution for the errors. Ibacache-Pulgar and Paula (2011), propose 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 introduced by Andrews and Mallows (1974). These distributions have heavier tails than the normal one, so they seem to be a reasonable choice for robust

*Address for correspondence: Víctor Hugo Lachos Dávila, Department of Statistics, Campinas State University, Brazil. Rua Sérgio Buarque de Holanda, 651 Cidade Universitária, "Zeferino Vaz", Campinas, São Paulo. Brazil

Email addresses: `thalitadobem@gmail.com` (Thalita do Bem Mattos[a]), `agaray@de.ufpe.br` (Aldo M. Garay[a,b]), `hlachos@ime.unicamp.br` (Victor H. Lachos*[a])

inference. They include as special cases many symmetric distributions, such as the normal (N), Pearson type VII (P-VII), Student-t (T), slash (SL) and contaminated normal (CN). Although these models are attractive, there is a need to check the distributional assumptions of the model errors because these can present skewness and heavy tail behavior, simultaneously. To overcome the problem of atypical data in an asymmetrical context, Branco and Dey (2001) proposed the class of scale mixtures of skew-normal (SMSN) distributions. This class of distributions contains the entire family of SMN distributions, and skewed versions of classic asymmetric distributions such as the skew-normal (SN), skew-t (ST), skew-slash (SSL) and skew contaminated normal (SCN) distributions.

In general, censored regression (CR) models are based on the development of the so called Tobit model, which is constructed in terms of the normal assumption (Tobin, 1958). However, many models do not fit the assumption of normality. Thus, in recent years several authors have studied CR models for statistical modeling of censored datasets involving observed variables with heavier tails than the normal distribution. For instance, Arellano-Valle et al. (2012) and Massuia et al. (2014) proposed an extension of the CR model with normal errors (N-CR) to Student-t (T-CR) errors. Garay et al. (2015b) proposed a robust CR model where the observational errors follow a SMN distribution (SMN-CR model). More recently, Massuia et al. (2015) developed a Bayesian framework for CR models by assuming that the random errors follow a SMSN distribution. In this work, we suggest an attractive ML estimation procedure for CR models considering the SMSN class of distributions, extending the works by Arellano-Valle et al. (2012), Massuia et al. (2014), Garay et al. (2015b) and supplementing the work by Massuia et al. (2015) from a likelihood-based perspective.

A typical algorithm for ML estimation in models involving the class of SMSN distributions is the EM algorithm and its variants. See, for instance, Basso et al. (2010), Lachos et al. (2010) and Garay et al. (2011). However, in some cases EM-type algorithms are not appropriate due to the computational difficulty in the E-step, which involves the computation of expected quantities that cannot be obtained analytically and must be calculated using stochastic simulation. To deal with this problem, Delyon et al. (1999) proposed a stochastic approximation version of the EM algorithm, the so-called SAEM algorithm. This algorithm consists of replacing the E-step by a stochastic approximation obtained using simulated data, while the M-step remains unchanged. Jank (2006) showed that the computational effort of SAEM is much smaller and reaches convergence in just a fraction of the simulation size when compared to Monte Carlo EM (MCEM). This is due the memory effect contained in the SAEM method, in which the previous simulations are considered in the computation of the posterior ones. In this paper, we develop a full likelihood approach for SMSN-CR models, including the implementation of the SAEM algorithm for ML estimation with the likelihood function, predictions of unobservable values of the response and the asymptotic standard errors as a byproduct.

The rest of the paper is organized as follows. In Section 2, we describe the family of SMSN distributions, including an outline of the SAEM algorithms. The SMSN-CR model and the ML estimation procedure based in the SAEM algorithm are described in Section 3. In Section 4, we discuss how to obtain the approximated standard errors. To examine the performance of our proposed methods, we present various simulation studies in Section 5. In Section 6 the proposed method is illustrated by the analysis of a real dataset. Section 7 concludes with a short discussion of issues raised by our study and some possible directions for a future research.

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

2.1. Preliminaries

In order to define the SMSN-CR model, we first make some remarks related to the SMSN class of distributions. This class of distributions was proposed by Branco and Dey (2001) and is a group of skew-thick-tailed distributions that are useful for robust inference and that contain as proper elements the SN, ST, SSL, SCN distributions and the entire family of SMN distributions proposed by Andrews and Mallows (1974) (see also, Lange and Sinsheimer, 1993). Thus, in the following we present some definitions where we explain first the fundamental concept of the SN distribution proposed by Azzalini (1985), and its relation with the SMSN class of distributions.

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 probability density function (pdf) is given by*

$$f_{SN}(z|\mu, \sigma^2, \lambda) = 2\phi(z|\mu, \sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right), \qquad z \in \mathbb{R}, \tag{1}$$

where $\phi\left(\,\cdot\,|\mu, \sigma^2\right)$ denotes the density of the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard univariate normal distribution.

Definition 2. *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, U is a positive random variable with cdf $H(\,\cdot\,|\boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$. $U \perp Z$ represents that the random variables U and Z are independent.

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

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

and integrating out U from the joint density of Y and U leads to the following marginal density of Y:

$$f_{SMSN}(y|\mu, \sigma^2, \lambda; H) = 2\int_0^\infty \phi(y|\mu, \kappa(u)\sigma^2)\Phi\left(\frac{\lambda(y-\mu)}{\sigma\kappa(u)^{1/2}}\right) dH(u). \tag{3}$$

Another important class of distribution, which will be useful for implementing the SAEM algorithm, is the truncated SMSN distributions, given by the following definition:

Definition 3. *Let $W \sim$ SMSN$(\mu, \sigma^2, \lambda; H)$ and $\mathbb{P}\left(a < W < b\right) > 0$, with $a < b$. A random variable Y has a truncated SMSN distribution in the interval $\lfloor a, b \rfloor$, denoted by $Y \sim$ TSMSN$(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$, if it has the same distribution as $W|W \in \lfloor a, b \rfloor$. Here $\lfloor a, b \rfloor$ means that each extreme of the interval can be either open or closed.*

Thus, the pdf of the random variable $Y \sim \text{TSMSN}(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$ is

$$f_{TSMSN}(y \mid \mu, \sigma^2, \lambda \,;\, H, \lfloor a, b \rfloor) = \frac{f_{SMSN}(y \mid \mu, \sigma^2, \lambda \,;\, H)}{F_{SMSN}\,(b \mid \mu, \sigma^2, \lambda \,;\, H) - F_{SMSN}\,(a \mid \mu, \sigma^2, \lambda \,;\, H)} \mathbb{1}_{\lfloor a,b \rfloor}(y),$$

where $\mathbb{1}_{\mathbb{A}}(\cdot)$ denotes the indicator function of the set $\mathbb{A}$, i.e., $\mathbb{1}_{\mathbb{A}}(y) = 1$ if $y \in \mathbb{A}$ and $\mathbb{1}_{\mathbb{A}}(y) = 0$ otherwise. $f_{SMSN}(\cdot \mid \mu, \sigma^2, \lambda; H)$ and $F_{SMSN}(\cdot \mid \mu, \sigma^2, \lambda; H)$ represent the pdf and cdf of the SMSN distribution, respectively.

The following lemmas show a convenient stochastic representation of a SMSN random variable as well as its cdf. These lemmas will be useful to implement the proposed SAEM algorithm.

Lemma 1. *The random variable $Y \sim$ SMSN$(\mu, \sigma^2, \lambda; H)$, has a stochastic representation given by*

$$Y = \mu + \Delta T + \kappa(U)^{1/2} \tau^{1/2} T_1, \tag{4}$$

where $\Delta = \sigma\delta$, $\tau = (1 - \delta^2)\sigma^2$, $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$, $T = \kappa(U)^{1/2}|T_0|$, T_0 and T_1 are independent standard normal random variables and $|\cdot|$ denotes absolute value.

Proof. See Basso et al. (2010). □

The representation given in Lemma 1 is very appropriate to derive some mathematical properties and can be used to simulate pseudo-realizations of Y. It is important to stress that this representation was used by Basso et al. (2010) in the context of finite mixtures of SMSN distributions and by Cancho et al. (2011), Garay et al. (2011) and Labra et al. (2012) in the context of non-linear regression models for complete data. For instance, from Equation (2), we have the following hierarchical representation:

$$\begin{aligned} Y|T = t, U = u &\sim N(\mu + \Delta t\ ,\ \kappa(u)\tau), \\ T|U = u &\sim TN(0, \kappa(u)\ ;\ \lfloor 0, \infty \rfloor), \\ U &\sim H(\ \cdot\ ; \boldsymbol{\nu}), \end{aligned} \tag{5}$$

where $\text{TN}(\mu, \sigma^2\ ;\ \lfloor a, b \rfloor)$ denotes a normal distribution with mean μ and variance σ^2 truncated in the interval $\lfloor a, b \rfloor$.

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

$$F_{SMSN}(y \mid \mu, \sigma^2, \lambda\ ;\ H) = \int_0^\infty 2\Phi_2\Big(\boldsymbol{y}(u)^* | \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big) dH(u), \tag{6}$$

where

$$\boldsymbol{y}(u)^* = (\kappa(u)^{-1/2} y, 0)^\top, \quad \boldsymbol{\mu}^* = (\mu, 0)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & -\delta\sigma \\ -\delta\sigma & 1 \end{pmatrix} \tag{7}$$

and $\Phi_m(\cdot \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denotes the cdf of the $m-$variate normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$.

Proof. See Appendix 1 in Massuia et al. (2015). □

2.2. Particular cases of SMSN distributions.

Although we can deal with any $\kappa(\cdot)$ function, we restrict our attention to the case where $\kappa(u) = 1/u$, since it leads to good mathematical properties. Moreover, the scale factor U can be discrete or continuous and the form of the SMSN distribution is determined by its distribution. We take into account four members of SMSN class: skew–normal, skew–t, skew–slash and skew contaminated normal distributions. For each specific SMSN distribution described below, we compute its cdf, which is useful to evaluate the likelihood function related to CR models.

- *The skew–t distribution.* Denoted by $Y \sim ST\left(\mu, \sigma^2, \lambda; \nu\right)$, this case arises when we consider $U \sim Gamma(\nu/2, \nu/2)$ in Definition 2. Thus, the density of Y takes the form

$$f_{ST}(y|\mu, \sigma^2, \lambda; \nu) = \frac{2\,\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1 + \frac{d(y)^2}{\nu}\right)^{-\frac{\nu+1}{2}} T_1\left(\lambda\, d(y)\sqrt{\frac{\nu+1}{\nu+d(y)^2}} \mid 0, 1, \nu+1\right), \quad y \in \mathbb{R}, \tag{8}$$

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 Lemma 2, we obtain the following expression for the cdf of Y:

$$F_{ST}(y \mid \mu, \sigma^2, \lambda\,; \nu) = 2\,T_2\left(\boldsymbol{y(u)}^* \mid \boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \nu\right), \tag{9}$$

where $\boldsymbol{y(u)}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (7) and $T_m\left(\cdot\mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \nu\right)$ represents the cdf of the m-variate Student-t distribution with location vector $\boldsymbol{\mu}_0$, scale matrix $\boldsymbol{\Sigma}_0$ and ν degrees of freedom. The proof of these results are given in Massuia et al. (2015).

- *The skew–slash distribution.* Denoted by $Y \sim SSL(\mu, \sigma^2, \lambda; \nu)$, in this case we consider $U \sim Beta(\nu, 1)$ with $\nu > 0$ in Definition 2. The density of Y is given by

$$f_{SSL}(y|\mu, \sigma^2, \lambda; \nu) = 2\nu \int_0^1 u^{\nu-1}\phi(y|\mu, u^{-1}\sigma^2)\Phi(u^{1/2}A(y))du, \quad y \in \mathbb{R}, \tag{10}$$

where $A(y) = \lambda(y-\mu)/\sigma$. The cdf of the skew-slash distribution does not have a closed form expression. However, using Lemma 2, we can write it in terms of the following integral, which can be approximated by numerical methods:

$$F_{SSL}(y|\mu, \sigma^2, \lambda; \nu) = \int_0^\infty 2\nu\Phi_2\Big(\boldsymbol{y}(u)^*|\boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big)u^{\nu-1}\,du, \tag{11}$$

where $\boldsymbol{y}(u)^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (7).

- *The skew-contaminated normal distribution.* Denoted by $Y \sim SCN(\mu, \sigma^2, \lambda; (\nu, \gamma))$, here U is a discrete random variable taking one of two states of $\boldsymbol{\nu} = (\nu, \gamma)^\top$. In this case the pdf of U is given by

$$U = \begin{cases} \gamma & \text{with probability } \nu; \\ 1 & \text{with probability } 1-\nu. \end{cases}$$

It follows immediately that

$$f_{SCN}(y|\mu,\sigma^2,\lambda;\boldsymbol{\nu}) \quad = \quad 2\{\nu\phi(y|\mu,\gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A(y)) + (1-\nu)\phi(y|\mu,\sigma^2)\Phi(A(y))\} \tag{12}$$

and

$$F_{SCN}(y|\mu,\sigma^2,\lambda;\boldsymbol{\nu}) \quad = \quad 2\{\nu\Phi_2\Big(\gamma^{1/2}\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big) + (1-\nu)\Phi_2\Big(\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big)\}, \tag{13}$$

where $A(y) = \lambda(y-\mu)/\sigma$.

- *The skew-normal distribution.* This distribution is obtained when $U = 1$ (a degenerated random variable) in Definition 2. The density of Y was defined in (1) and clearly, from Lemma 2, the cdf of Y is given by

$$F(y) = 2\Phi_2\Big(\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big), \tag{14}$$

where $\boldsymbol{y}^* = (y,0)^\top$ and $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (7).

In Table 1, we present the expected values $k_m = E[U^{-m/2}]$ for all the SMSN distributions discussed above, which are useful to define the SMSN-CR model.

Table 1: $k_m = E[U^{-m/2}]$ for different SMSN models.

Model	k_m
SN	1
ST	$\left(\frac{\nu}{2}\right)^{m/2}\frac{\Gamma\left(\frac{\nu-m}{2}\right)}{\Gamma(\nu/2)}$
SSL	$\frac{2\nu}{\nu-m/2}$
SCN	$\frac{\nu}{\gamma^{m/2}}+1-\nu$

2.3. Algorithms for ML estimation

In models with non-observed or incomplete data, the EM algorithm is a very popular iterative optimization strategy commonly used. This algorithm has many attractive features such as numerical stability and simplicity of implementation, and its memory requirements are quite reasonable (Couvreur, 1996). Letting $\mathbf{y}_{comp} = (\mathbf{y}^m, \mathbf{y}^o)$ the complete data vector, where $\mathbf{y}^m$ represents the missing data and $\mathbf{y}^o$ the observed data respectively and $\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp})$ the complete data log-likelihood function, then the EM-algorithm proceeds in two steps:

- **E-step:** Let $\widehat{\boldsymbol{\theta}}^{(j)}$ be the current $j-$th step estimate of $\boldsymbol{\theta}$. By using the property of conditional expectation, we can compute the $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ function by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)}) = E\left[\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp})|\mathbf{y}^o,\widehat{\boldsymbol{\theta}}^{(j)}\right]. \tag{15}$$

- **M-step:** Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ with respect to $\boldsymbol{\theta}$, obtaining $\widehat{\boldsymbol{\theta}}^{(j+1)}$.

As mentioned by Meza et al. (2012), each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y}^o)$ and the EM sequence $\boldsymbol{\theta}^{(j)}$ converges to a stationary point of the observed likelihood under mild regularity conditions (for more details see Wu (1983) and Vaida (2005)).

For cases in which the E-step has no analytic form, Wei and Tanner (1990) proposed the Monte Carlo EM (MCEM) algorithm, in which the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the missing data. In order to reduce the number of required simulations compared to the MCEM algorithm, Delyon et al. (1999) proposed the stochastic approximation version of the EM algorithm, the so-called SAEM algorithm, which consists of replacing the E-step by a stochastic approximation, obtained using simulated data, while the M-step is unchanged. The SAEM algorithm consists, at each iteration, of successively simulating the missing data with the conditional distribution, and updating the unknown parameters of the model. Thus, the $j-$th iteration of SAEM algorithm consists of the following steps:

- **S-step:** Draw the missing data $\mathbf{y}^{m(j)}$ with the conditional distribution $p(\mathbf{y}^m|\mathbf{y}^o, \widehat{\boldsymbol{\theta}}^{(j-1)})$.
- **AE-step:** Update $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ according to

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)}) \approx Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j-1)}) + \gamma_j \left[\frac{1}{m} \sum_{\ell=1}^{m} \ell_{comp}(\theta|\mathbf{y}^o, \mathbf{y}^{m(j)}) - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j-1)}) \right]. \tag{16}$$

- **M-step:** Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ with respect to $\boldsymbol{\theta}$ obtaining $\widehat{\boldsymbol{\theta}}^{(j+1)}$,

where γ_j is a decreasing sequence of positive numbers such that

$$\sum_{j=1}^{\infty} \gamma_j = \infty \quad \text{and} \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty, \tag{17}$$

as presented by Kuhn and Lavielle (2004).

Thus, the SAEM algorithm performs a Monte Carlo E-step, like MCEM, but with a small and fixed Monte Carlo sample sizes ($m \leq 20$). This is possible because unlike the traditional EM algorithm and its variants, the SAEM algorithm uses not only the current simulation of the missing data at the $j-$iteration, denoted by $\mathbf{y}^m$, but also some or all previous simulations, where this *'memory'* property is set by the smoothing parameter γ_j.

Note, in Equation (16), that sequence γ_j has a strong impact on the speed of convergence of the algorithm. Thus, if the smoothing parameter γ_j is equal to 1 for all j, the SAEM algorithm will have *'no memory'*, and will be equivalent to the MCEM algorithm. The SAEM with no memory will converge quickly (convergence in distribution) to a solution neighborhood, but the algorithm with memory will converge slowly (almost sure convergence) to the ML solution. As suggested by Galarza et al. (2015), we use the following choice of the smoothing parameter:

$$\gamma_j = \begin{cases} 1, & \text{for} \quad 1 \leq j \leq cS, \\ \frac{1}{j-cS}, & \text{for} \quad cS+1 \leq j \leq S, \end{cases}$$

where S is the maximum number of iterations, and c a cutoff point ($0 \leq c \leq 1$) that determines the percentage of initial iterations with no memory. For example, if $c = 0$, the algorithm will have memory for all iterations, and hence will converge slowly to the ML estimates. If $c = 1$, the algorithm will have no memory, and so will converge quickly to a solution neighborhood. For the first case, S would need to be large in order to achieve the ML estimates. For the second, the algorithm will yield a Markov Chain where, after applying a *burn-in* and *thinning*, the mean of the chain observations can be a reasonable estimate. A number c between 0 and 1 ($0 < c < 1$) will assure an initial convergence in distribution to a solution neighborhood for the first cS iterations and an almost sure convergence for the rest of the iterations. Hence, this combination will lead to a fast algorithm with good estimates. To implement SAEM, the user must fix several constants matching the number of total iterations S and the cutoff point c that defines the start of the smoothing step of the SAEM algorithm. However, those parameters will vary depending of the model and the data. To determine those constants, a graphical approach is recommended to monitor the convergence of the estimates for all the parameters, and if possible, to monitor the difference (relative difference) between two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_o)$, given by:

$$||\ell(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}^o) - \ell(\boldsymbol{\theta}^{(j)}|\mathbf{y}^o)|| \quad \text{or} \quad ||\ell(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}^o)/\ell(\boldsymbol{\theta}^{(j)}|\mathbf{y}^o) - 1||,$$

respectively.

3. The SMSN censored linear regression model

3.1. Model specification

The SMSN-CR model that we are going to discuss is defined by:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{18}$$

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

In this work, we assume that

$$\varepsilon_i \sim SMSN\left(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H\right), \quad i = 1, \ldots, n, \tag{19}$$

are independent random variables. The value of the location parameter, $-\sqrt{\frac{2}{\pi}}k_1\Delta$, of ε_i is chosen in order to obtain $E[\varepsilon_i] = 0$, as in the normal model. For more details, see Lemma 1 in Basso et al. (2010). Thus, when the moments exist, we have

$$Y_i \sim SMSN\left(\mathbf{x}_i^\top \boldsymbol{\beta} - \sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H\right),$$

where $E[Y_i] = \mathbf{x}_i^\top \boldsymbol{\beta}$ and $Var[Y_i] = k_2\sigma^2 - \dfrac{2k_1^2\Delta^2}{\pi}$, for $i = 1, \ldots, n$. The values of k_1 and k_2 are given in Table 1, for particular cases of SMSN distributions.

In this work we are interested in the situation in which the response variable is not fully observed for all subjects. Thus, for the $i-$th subject and assuming left-censoring, Y_i is a latent variable and the observe data (V_i, ρ_i) is of the form

$$V_i = \begin{cases} c_i & \text{if} \quad \rho_i = 1 \quad (\text{i.e. } Y_i \leq c_i); \\ Y_i & \text{if} \quad \rho_i = 0 \quad (\text{i.e. } Y_i > c_i), \end{cases} \tag{20}$$

for some known threshold point c_i, $i = 1, 2, \ldots, n$.

The SMSN-CR model is defined by combining (18)–(20). The log-likelihood function of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu\right)^\top$ given the observed data $(\boldsymbol{v}, \boldsymbol{\rho})$, is

$$\ell(\boldsymbol{\theta}|\boldsymbol{v}, \boldsymbol{\rho}) = \log\left\{\prod_{i=1}^{n}\left[F_{SMSN}\left(\frac{v_i - \mathbf{x}_i^\top\boldsymbol{\beta}}{\sigma}|\boldsymbol{\theta}; H\right)\right]^{\rho_i}\left[f_{SMSN}(v_i|\boldsymbol{\theta}; H)\right]^{1-\rho_i}\right\}, \tag{21}$$

where $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is the observed sample of $\mathbf{V} = (V_1, V_2, \ldots, V_n)$. Thus, $\rho_i = 1$, with 0 indicating whether the $i-$th observation is censored, i.e. $Y_i \leq c_i$, or not respectively. $f_{SMSN}(\cdot|\boldsymbol{\theta}; H)$ and $F_{SMSN}(\cdot|\boldsymbol{\theta}; H)$ represent the pdf and cdf of the SMSN class, respectively.
For simplicity, we will assume the data are left censored, and develop the SAEM algorithm for ML estimation. Extensions, to right censored data are immediate.

3.2. ML estimation via the SAEM algorithm

In this section we consider the ML estimation of the parameters in the SMSN-CR models introduced in Section 2.3. In particular, we show how to implement the SAEM algorithm for the particular cases of the SMSN class, that is, the SN, ST, SSL and SCN distributions.

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

$$\Delta = \sigma\frac{\lambda}{\sqrt{\lambda^2 + 1}} = \sigma\delta \quad \text{and} \quad \tau = \left(1 - \delta^2\right)\sigma^2 = \frac{\sigma^2}{\lambda^2 + 1},$$

we can obtain σ^2 and λ from Δ and τ considering

$$\sigma^2 = \tau + \Delta^2 \quad \text{and} \quad \lambda = \Delta/\sqrt{\tau}. \tag{22}$$

We observe that a useful straightforward result, used by Basso et al. (2010) and Massuia et al. (2015), is that the conditional distribution of T_i given y_i and u_i is $TN(\mu_{Ti} - \sqrt{\frac{2}{\pi}}k_1, u_i^{-1}M_T^2; \lfloor -\sqrt{\frac{2}{\pi}}k_1, \infty \rfloor)$, with

$$\mu_{T_i} = \frac{\Delta}{\Delta^2 + \tau}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}) \quad \text{and} \quad M_T^2 = \frac{\tau}{\Delta^2 + \tau}. \tag{23}$$

In order to implement the SAEM algorithm, we consider a data augmentation scheme that consists of assuming that the latent variables (*missing data*) in the model, given by the vector of censored responses $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^\top$, the vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)^\top$ - see representation (4) - can be observed. Thus, considering the observed data $(\mathbf{V}, \boldsymbol{\rho})$ and the latent variables $(\mathbf{Y}, \mathbf{t}, \mathbf{u})$, we define the complete data by $\mathbf{Y}_{comp} = (\mathbf{V}^\top, \boldsymbol{\rho}^\top, \mathbf{Y}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$. Then, it is easy to derive the complete data log-likelihood, defined by $\ell_{comp}(\boldsymbol{\omega}|\mathbf{Y}_{comp})$, using the representation (5) as:

$$\ell_{comp}(\boldsymbol{\omega}|\mathbf{Y}_{comp}) \propto cte - \frac{n}{2}\log\tau - \frac{1}{2\tau}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta} - \Delta t_i)^2 + \sum_{i=1}^{n}\log h(u_i|\boldsymbol{\nu}), \tag{24}$$

where cte is a constant that is independent of $\boldsymbol{\omega}$ and $h(\cdot|\boldsymbol{\nu})$ is the pdf of the random variable U. In what follows the superscript (j) indicates the estimate of the related parameter at stage j of the algorithm. Thus, we have:

- **E-step:** Given the current estimate $\boldsymbol{\omega}^{(j)} = (\beta^{(j)\top}, \Delta^{(j)}, \tau^{(j)}, \nu^{(j)})^\top$ at the $j-$th iteration, we obtain the conditional expectation of the complete data log-likelihood function (Q-function), which is given by

$$\begin{aligned} Q(\boldsymbol{\omega}|\boldsymbol{\omega}^{(j)}) &= E\left[\ell_{comp}(\boldsymbol{\omega}|\mathbf{Y}_{comp})|\mathbf{V}, \boldsymbol{\rho}, \boldsymbol{\omega}^{(j)}\right] \\ &= cte - \frac{n}{2}\log(\tau) - \frac{1}{2\tau}\sum_{i=1}^{n}\Big[\mathcal{E}_{02i}(\boldsymbol{\omega}^{(j)}) - 2\mathcal{E}_{01i}(\boldsymbol{\omega}^{(j)})\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i^\top\boldsymbol{\beta})^2 \\ &- 2\Delta\mathcal{E}_{11i}(\boldsymbol{\omega}^{(j)}) + 2\Delta\mathcal{E}_{10i}(\boldsymbol{\omega}^{(j)})\mathbf{x}_i^\top\boldsymbol{\beta} + \Delta^2\mathcal{E}_{20i}(\boldsymbol{\omega}^{(j)})\Big] + \mathrm{E}\left[\log\{h(U_i|\boldsymbol{\nu})\}|V_i, \rho_i, \boldsymbol{\omega}^{(j)}\right]. \end{aligned}$$

  Observe that the expression of the Q-function is completely determined by the knowledge of the following expectations:

$$\mathcal{E}_{rsi}(\boldsymbol{\omega}^{(j)}) = \mathrm{E}[U_i T_i^r Y_i^s|V_i, \rho_i, \boldsymbol{\omega}^{(j)}] \quad \text{for} \quad r, s = 0, 1, 2,$$

  as well as

$$\mathrm{E}[\log\{h(U_i|\boldsymbol{\nu})\}|V_i, \rho_i].$$

  As presented by Basso et al. (2010), considering known properties of conditional expectation and Equation (23), we obtain

$$\mathcal{E}_{10i}(\boldsymbol{\omega}^{(j)}) = \mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})\mu_{T_i}^{(j)} + M_T^{(j)}\psi_i^{(j)}, \tag{25}$$

$$\mathcal{E}_{20i}(\boldsymbol{\omega}^{(j)}) = \mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})\mu_{T_i}^{2(j)} + M_T^{2(j)} + M_T^{(j)}\mu_{T_i}^{(j)}\psi_i^{(j)}, \tag{26}$$

  where

$$\psi_i^{(j)} = \mathrm{E}\left[U_i W_\Phi\left(\frac{U_i\mu_{T_i}^{(j)}}{M_T^{(j)}}\right)|V_i, \rho_i, \boldsymbol{\omega}^{(j)}\right] \quad \text{and} \quad W_\Phi(a) = \frac{\phi(a)}{\Phi(a)} \text{ for } a \in \mathbb{R}.$$

  Thus, at each step, to compute $\mathcal{E}_{rsi}(\boldsymbol{\omega}^{(j)})$ we need to obtain the conditional expectations $\mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})$ and $\psi_i^{(j)}$ for the different SMSN distributions considering two different situations:

  a) *For an uncensored observation i:*
  In this case we have that $\rho_i = 0$, thus $V_i = Y_i \sim SMSN\left(\mathbf{x}_i^\top\boldsymbol{\beta} - \sqrt{\frac{2}{\pi}}k_1\Delta, \tau + \Delta^2, \lambda; H\right)$ and, therefore,

$$\mathcal{E}_{rsi}(\boldsymbol{\omega}^{(j)}) = y_i^s\mathcal{E}_{r0i}(\boldsymbol{\omega}^{(j)}), \tag{27}$$

  where $\mathcal{E}_{r0i}(\boldsymbol{\omega}^{(j)})$ can be obtained using equations (25)-(26) and the results given by Basso et al. (2010). Thus, for example,

* **For the skew-t case**

$$\begin{aligned}
\mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)}) &= \frac{2^2\nu^{\nu^{(j)}/2}\Gamma\left(\frac{\nu^{(j)}+3}{2}\right)(\nu^{(j)}+\mathrm{d}^{(j)}(y_i))^{-\frac{\nu^{(j)}+3}{2}}}{f_{ST}(y_i)\Gamma\left(\frac{\nu^{(j)}}{2}\right)\sqrt{\pi}(\tau^{(j)}+\Delta^{2(j)})^{1/2}} \\
&\quad \times T\left(\sqrt{\frac{\nu^{(j)}+3}{\nu^{(j)}+\mathrm{d}^{(j)}(y_i)}}A_i^{*(j)};\nu^{(j)}+3\right), \\
\psi_i^{(j)} &= \frac{2\Gamma(\frac{\nu^{(j)}+2}{2})\nu^{\nu^{(j)}/2}(\nu^{(j)}+\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)})^{-\frac{\nu^{(j)}+2}{2}}}{f_{ST}(y_i)\Gamma\left(\frac{\nu^{(j)}}{2}\right)\pi(\tau^{(j)}+\Delta^{2(j)})^{1/2}},
\end{aligned}$$

as defined in (8), $f_{ST}(\cdot)$ represents the pdf of skew-t distribution and $T(\cdot;\nu)$ is the cdf of the standard Student-t distribution.

* **For the skew-slash case**

$$\begin{aligned}
\mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)}) &= \frac{\nu^{(j)}2^{\nu^{(j)}+2}\Gamma\left(\frac{2\nu^{(j)}+3}{2}\right)P_1\left(\frac{2\nu^{(j)}+3}{2},\frac{\mathrm{d}^{(j)}(y_i)}{2}\right)\mathrm{d}^{(j)}(y_i)^{-\frac{2\nu^{(j)}+3}{2}}}{f_{SSL}(y_i)\sqrt{\pi}(\tau^{(j)}+\Delta^{2(j)})^{1/2}} \\
&\quad \times E\left[\Phi(S_i^{(j)1/2}A_i^{*(j)})\right], \\
\psi_i^{(j)} &= \frac{\nu^{(j)}2^{\nu^{(j)}+1}\Gamma\left(\frac{2\nu+2}{2}\right)}{f_{SSL}(y_i)\pi(\tau^{(j)}+\Delta^{2(j)})^{1/2}}\left(\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)}\right)^{-\frac{2\nu^{(j)}+2}{2}} \\
&\quad \times P_1\left(\frac{2\nu^{(j)}+2}{2},\frac{\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)}}{2}\right),
\end{aligned}$$

where $S_i^{(j)} \sim Gamma(\frac{2\nu^{(j)}+3}{2},\frac{\mathrm{d}^{(j)}(y_i)}{2})\mathbb{I}_{(0,1)}$ is a truncated gamma distribution on $(0,1)$, with the parameter values in parentheses before truncation and $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ evaluated at x . As defined in (10), $f_{SSL}(\cdot)$ represents a density of skew-slash distribution.

* **For the skew contaminated normal case**

$$\begin{aligned}
\mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)}) &= \frac{2}{f_{SCN}(y_i)}\left\{\nu^{(j)}\gamma^{(j)}\phi\left(y_i;\mu^{*(j)},\gamma^{-1(j)}(\tau^{(j)}+\Delta^{2(j)})\right)\Phi\left(\gamma^{1/2}A_i^{*(j)}\right)\right. \\
&\quad \left.+(1-\nu^{(j)})\phi\left(y_i;\mu^{*(j)},\tau^{(j)}+\Delta^{2(j)}\right)\Phi\left(A_i^{*(j)}\right)\right\} \\
\psi_i^{(j)} &= \frac{2}{f_{SCN}(y_i)}\left\{\nu^{(j)}\gamma^{(j)}\phi\left(y_i;\mu^{*(j)},\gamma^{-1(j)}(\tau^{(j)}+\Delta^{2(j)})\right)\Phi\left(\gamma^{1/2}A_i^{*(j)}\right)\right. \\
&\quad \left.+(1-\nu^{(j)})\phi\left(y_i;\mu^{*(j)},\tau^{(j)}+\Delta^{2(j)}\right)\phi\left(A_i^{*(j)}\right)\right\},
\end{aligned}$$

where $f_{SCN}(\cdot)$ represents the pdf of the skew contaminated normal distribution, as defined in (12).

In all cases described before, $\mu^{*(j)} = \mathbf{x}_i^\top\boldsymbol{\beta}^{(j)} - \sqrt{\frac{2}{\pi}}k_1\Delta^{(j)}$, $A^{*(j)} = \frac{\mu_{T_i}^{(j)}}{M_T^{(j)}}$ and $\mathrm{d}^{(j)}(y_i) = \frac{(y_i-\mu^{*(j)})}{\sqrt{\tau^{(j)}+\Delta^{2(j)}}}$ represents the Mahalanobis distance. Thus, in each step, the conditional expectations $\mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})$ and $\psi_i^{(j)}$ can be easily obtained.

For the skew–t and skew contaminated normal distributions we have computationally attractive expressions that can be easily implemented. However, this is not the case for the skew–slash one, where Monte Carlo integration can be employed, as suggested by Basso et al. (2010) and Lachos et al. (2010).

b) *For a censored observation* i:

In this case, we have that $\rho_i = 1$, i.e. $Y_i \leq c_i$, therefore

$$\mathcal{E}_{rsi}\left(\boldsymbol{\omega}^{(j)}\right) = \mathrm{E}[U_i T_i^r Y_i^s | V_i, Y_i \leq c_i, \boldsymbol{\omega}^{(j)}] \text{ with } r, s = 0, 1, 2. \tag{28}$$

As this conditional expectation does not have closed form, we need to introduce two intermediate steps in order to replace the E-step by a stochastic approximation using simulated data. Thus, the iteration j consists of the following steps:

* **S-step (Sampling)**
  Let $\mathbf{Y}^{(c)} = \left(Y_1^{(c)}, Y_2^{(c)}, \ldots, Y_{n^c}^{(c)}\right)$ the vector of n^c censored cases, where $Y_i^{(c)}$ is generated from TSMSN $\left(\mathbf{x}_i^\top \boldsymbol{\beta} - \sqrt{\frac{2}{\pi}} k_1 \Delta, \tau + \Delta^2, \lambda; H, \lfloor -\infty, c_i \rfloor\right)$ for $i = 1, \ldots, n^c$. Thus, the new vector of observations $\mathbf{Y}^{(l,j)} = (Y_{i1}^{(l,j)}, \ldots, Y_{in^c}^{(l,j)}, Y_{n_i^c+1}, \ldots, Y_n)$ is a sample generated for the n^c censored cases and the observed values (uncensored cases), for $l = 1, \ldots, M$. Subsection 3.3 describes the details of the methods used to generate from the random variable $\mathbf{Y}^{(c)}$.

* **AE-step (Approximation Expectation)**
  Since we have the sequence $\mathbf{Y}^{(j,l)}$, at the j-th iteration, considering equations (25)-(26) and the results given in Basso et al. (2010), we replace the conditional expectations in (27) by the following stochastic approximations:

$$\mathcal{E}_{rsi}\left(\boldsymbol{\omega}^{(j)}\right) = \mathcal{E}_{rsi}(\boldsymbol{\omega}^{(j-1)}) + \gamma_j \left[\frac{1}{m}\sum_{l=1}^{m} \mathrm{E}[U_i T_i^r Y_i^{s(l,j)} | V_i, \rho_i, \boldsymbol{\omega}^{(j)}] - \mathcal{E}_{rsi}(\boldsymbol{\omega}^{(j-1)})\right],$$

  for $r, s = 0, 1, 2$.

An advantage of the SAEM algorithm is that even though it performs a Monte Carlo E-step, it requires a small and fixed Monte Carlo sample size, making it much faster than MCEM. Some authors claim that $m = 10$ is large enough, but to be more conservative, we chose $m = 20$.

- **CM-step:** Maximize $Q(\boldsymbol{\omega}|\boldsymbol{\omega}^{(j)})$ with respect to $\boldsymbol{\omega}$ obtaining $\boldsymbol{\omega}^{(j+1)}$, which leads to the following expressions:

$$\begin{aligned}
\boldsymbol{\beta}^{(j+1)} &= \left(\sum_{i=1}^{n} \mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i \mathbf{x}_i^\top)\right)^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i \mathcal{E}_{01i}(\boldsymbol{\omega}^{(j)}) - \Delta \mathbf{x}_i \mathcal{E}_{10i}(\boldsymbol{\omega}^{(j)})\right]; \\
\Delta^{(j+1)} &= \frac{\mathcal{E}_{11i}(\boldsymbol{\omega}^{(j)}) - \mathcal{E}_{10i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)})}{\mathcal{E}_{20i}(\boldsymbol{\omega}^{(j)})}; \\
\tau^{(j+1)} &= \frac{1}{n}\left(\sum_{i=1}^{n} \left[\mathcal{E}_{02i}(\boldsymbol{\omega}^{(j)}) - 2\mathcal{E}_{01i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)}) + \mathcal{E}_{00i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)})^2 \right.\right. ; \\
&\quad \left.\left. - 2\Delta^{(j+1)}\mathcal{E}_{11i}(\boldsymbol{\omega}^{(j)}) + 2\Delta^{(j+1)}\mathcal{E}_{10i}(\boldsymbol{\omega}^{(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)}) + (\Delta^{(j+1)})^2 \mathcal{E}_{20i}(\boldsymbol{\omega}^{(j)})\right]\right).
\end{aligned}$$

- **CML-step:** We estimates $\boldsymbol{\nu}$ by maximizing the actual marginal log-likelihood function, obtaining

$$\boldsymbol{\nu}^{(j+1)} = \text{argmax}_{\boldsymbol{\nu}} \left\{ \sum_{i=1}^{n} \log\left[F_{SMSN}\left(v_i|\boldsymbol{\theta}; H\right)\right]^{\rho_i} + \sum_{i=1}^{n} \log\left[f_{SMSN}(v_i|\boldsymbol{\theta}; H)\right]^{1-\rho_i} \right\}.$$

Note that $\sigma^{2(j+1)}$ and $\lambda^{(j+1)}$ can be recovered using (22). The more efficient CML-step can be easily accomplished by using, for instance, the *optim* routine in the *R* software (R Development Core Team, 2015).

Thus, considering $\boldsymbol{\theta}^{(j+1)} = \left(\boldsymbol{\beta}^{(j+1)\top}, \sigma^{2(j+1)}, \lambda^{(j+1)}, \nu^{(j+1)}\right)^{\top}$, this process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_{obs})$, like

$$||\ell(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}_{obs}) - \ell(\boldsymbol{\theta}^{(j)}|\mathbf{y}_{obs})|| \quad \text{or} \qquad ||\ell(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}_{obs})/\ell(\boldsymbol{\theta}^{(j)}|\mathbf{y}_{obs}) - 1||,$$

is small enough. We have adopted this strategy to update the estimate of $\boldsymbol{\nu}$, by direct maximization of the marginal log-likelihood, circumventing the computation of $\mathrm{E}_{\boldsymbol{\theta}^{(j)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\text{obs}_i}]$.

In order to make our proposed algorithm more informative for the reader, in Figure 1 we present a flow diagram, which reports all the steps needed to implement the SAEM algorithm.

3.3. Computational aspects

The convergence of the SAEM algorithm is ensured by a careful choice of the simulation data method. Thus, in this subsection, we describe two simulation methods to generate random samples from the random variable $Y \sim \text{TSMSN}(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$. We concentrate on the truncated skew normal (TSN), truncated skew-t (TST), truncated skew slash (TSSL) and truncated skew contaminated normal (TSCN) distributions.

We use the sampling/importance resampling method (Method 1), proposed by Rubin (1987) and Rubin et al. (1988), to generate samples from the TSN and TST models. For the TSS and TSCN models we use the stochastic representation of a SMSN random variable, given in Lemma 1 (Method 2). In the following, we present a brief description of those two methods:

- **Method 1**

  The sampling/importance resampling (SIR) method is useful to generate an approximate independent and identically distributed (*i.i.d.*) sample of size m, from the target density $f(y)$ where $y \in \mathcal{S}_Y \subseteq \mathbb{R}$. Thus, let $g(y)$ a proposal density with the same support $\mathcal{S}_Y$. The method consists of two steps:

  - **Step 1.**(*Sampling*) Generate a random sample $Y_1,\ Y_2, \cdots, Y_J$ from $g(y)$ and construct weights

$$W(Y_j) = \frac{f(Y_j)}{g(Y_j)}, \quad j = 1, \ldots, J$$

    and probabilities

$$\pi_j = \frac{W(Y_j)}{\sum_{j=1}^{J} W(Y_j)}, \quad j = 1, \ldots, J.$$

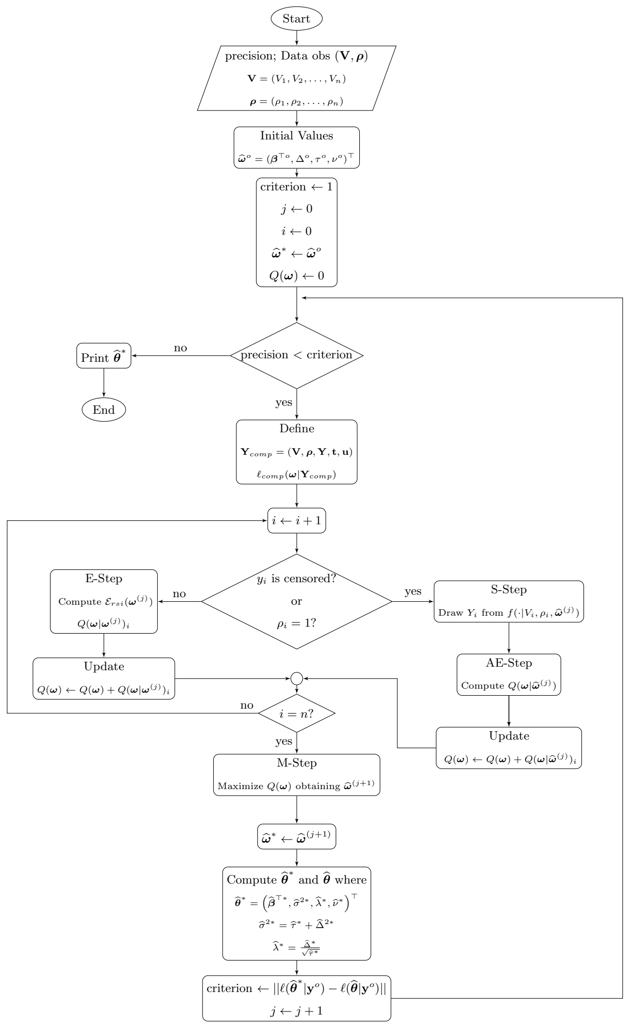

Figure 1: Flow diagram of the SAEM algorithm.

– **Step 2.**(*Importance resampling*) Draw m values $(m << J)$ $Y_1^*, ..., Y_m^*$ from the J values $Y_1, Y_2, \dots, Y_J$ with respective probabilities $\pi_1,\ \pi_2, \dots \pi_J$. In practice, Rubin (1987) suggested $J/m = 20$.

For the TSN model, the target density $f(\cdot)$ is a truncated skew normal distribution TSN $\left(\mu, \sigma^2, \lambda; \lfloor a, b \rfloor\right)$ and as proposal density $g(\cdot)$, we utilize truncated normal distribution, TN $\left(\mu, \sigma^2; \lfloor a, b \rfloor\right)$. For TST model, the target density $f(\cdot)$ is a truncated skew-t distribution TST $\left(\mu, \sigma^2, \lambda, \nu; \lfloor a, b \rfloor\right)$ and as the proposal density $g(\cdot)$, we utilize the truncated t distribution, Tt $\left(\mu, \sigma^2, \nu; \lfloor a, b \rfloor\right)$.

- **Method 2**

  In this case, we need to generate samples from the random variable $Y \sim$ TSMSN$(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$.

Then, since U is a positive random variable, we have that

$$a \ < \ Y \ < \ b,$$

which implies

$$(a-\mu)U^{1/2} \ < \ (Y-\mu)U^{1/2} \ < \ (b-\mu)U^{1/2}.$$

Considering the stochastic representation given in (2), we have that $Z = (Y-\mu)U^{1/2}$, where $Z \sim \mathrm{SN}(0, \sigma^2, \lambda)$. Thus,

$$(a-\mu)U^{1/2} \ < \ Z \ < \ (b-\mu)U^{1/2}.$$

Therefore, the algorithm to generate random samples of TSSL and TSCN models is as follows:

- **Step 1.** Generate a random sample $U_1, U_2, \ldots, U_m$ from $H(\,\cdot\,|\boldsymbol{\nu})$.
- **Step 2.** Generate a random sample $Z_1, Z_2, \ldots, Z_m$ from $\mathrm{TSN}(0, \sigma^2, \lambda; [\gamma_1, \gamma_2])$, where $\gamma_1 = (a-\mu)U^{1/2}$ and $\gamma_2 = (b-\mu)U^{1/2}$, using Method 1.
- **Step 3.** Using the stochastic representation given in (2), set $Y = \mu + U^{1/2}Z$.

Consequently, we draw $y_i^{(j)}$ from $f(y_i|\boldsymbol{\omega}^{(j)}, V_i, \rho_i)$ in the S-step.

3.4. Model selection

Because there is no universal criterion for mixture model selection, we chose three criteria to compare the models considered in this work. These are the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the efficient determination criterion (EDC) (Bai et al., 1989). Like the more popular AIC and BIC criteria, EDC has the form

$$-2\ell(\widehat{\boldsymbol{\theta}}) + \rho c_n,$$

where $\ell(\boldsymbol{\theta})$ is the actual log-likelihood, ρ is the number of free parameters that have to be estimated in the model and the penalty term c_n is a convenient sequence of positive numbers. Here, we use $c_n = 0.2\sqrt{n}$, a proposal that was considered in Basso et al. (2010) and Cabral et al. (2012). We have $c_n = 2$ for AIC, $c_n = \log n$ for BIC, where n is the sample size.

4. Approximated standard errors

Standard errors of the ML estimates can be approximated by the inverse of the observed information matrix, but there is generally no closed form. Thus, we consider the same strategy used by Meilijson (1989), Lin (2010) and Garay et al. (2015b) to get approximate standard errors of the parameter estimates based on the empirical information matrix.
Let $(\mathbf{V}, \boldsymbol{\rho})$ be the vector of observed data. So, considering $\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma^2, \lambda, \boldsymbol{\nu}\right)$, $\mathbf{Y}_{comp} = (\mathbf{V}^\top, \boldsymbol{\rho}^\top, \mathbf{Y}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$ and relations described in the Equation (22), the empirical information matrix is defined as

$$\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{V}, \boldsymbol{\rho}\right) = \sum_{i=1}^{n} \mathbf{s}\left(V_i, \rho_i|\boldsymbol{\theta}\right)\mathbf{s}^\top\left(V_i, \rho_i|\boldsymbol{\theta}\right) - \frac{1}{n}\mathbf{S}\left(\mathbf{V}, \boldsymbol{\rho}|\boldsymbol{\theta}\right)\mathbf{S}^\top\left(\mathbf{V}, \boldsymbol{\rho}|\boldsymbol{\theta}\right),$$

where $\mathbf{S}^\top(\mathbf{V},\boldsymbol{\rho}|\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{s}(V_i,\rho_i|\boldsymbol{\theta})$. It is noted from the result of Louis (1982) that, the individual score can be determined as

$$\mathbf{s}(V_i,\rho_i|\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta}|V_i,\rho_i)}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\theta}}|V_i,\rho_i,\boldsymbol{\theta}\right]. \tag{29}$$

Thus, substituting the ML estimates of $\boldsymbol{\theta}$ in (29), the empirical information matrix $\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{V},\boldsymbol{\rho})$ is reduced to

$$\mathbf{I}_e\left(\widehat{\boldsymbol{\theta}}|\mathbf{V},\boldsymbol{\rho}\right) = \sum_{i=1}^n \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top, \tag{30}$$

where $\widehat{\mathbf{s}}_i = \left(\widehat{\mathbf{s}}_{\boldsymbol{\beta}_i}, \widehat{\mathbf{s}}_{\sigma_i^2}, \widehat{\mathbf{s}}_{\boldsymbol{\lambda}_i}, \widehat{\mathbf{s}}_{\boldsymbol{\nu}_i}\right)$ is an individual score vector and

$$\begin{aligned}
\widehat{\mathbf{s}}_{\boldsymbol{\beta}_i} &= \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\beta}}|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right] = \frac{1+\widehat{\lambda}^2}{\widehat{\sigma}^2}\left(\mathbf{x}_i\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} - \widehat{\sigma}\frac{\widehat{\lambda}}{\sqrt{1+\widehat{\lambda}^2}}\mathbf{x}_i\mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\right),\\
\widehat{\mathbf{s}}_{\sigma_i^2} &= \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \sigma^2}|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right] = -\frac{1}{2\widehat{\sigma}^2} + \frac{1+\widehat{\lambda}^2}{2\widehat{\sigma}^4}\left(\mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)\\
&\quad -\frac{\widehat{\lambda}\sqrt{(1+\widehat{\lambda}^2)}}{2\widehat{\sigma}^3}\left(\mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right),\\
\widehat{\mathbf{s}}_{\lambda_i} &= \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \lambda}|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right] = \frac{\widehat{\lambda}}{1+\widehat{\lambda}^2} + \frac{\widehat{\lambda}}{\widehat{\sigma}^2}\left(\mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)\\
&\quad +\frac{1+2\widehat{\lambda}^2}{\widehat{\sigma}\sqrt{1+\widehat{\lambda}^2}}\left(\mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right) - \widehat{\lambda}\mathcal{E}_{20i}(\widehat{\boldsymbol{\theta}}),\\
\widehat{\mathbf{s}}_{\boldsymbol{\nu}_i} &= \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\nu}}|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[\frac{\partial \log\left(f(U_i|\boldsymbol{\nu})\right)}{\partial \boldsymbol{\nu}}|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right],
\end{aligned}$$

where $\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})$ is the log-likelihood formed from the single complete observation i and $\mathcal{E}_{rsi}(\boldsymbol{\omega}^{(k)}) = \mathrm{E}[U_iT_i^rY_i^s|V_i,\rho_i,\boldsymbol{\omega}^{(k)}]$. It is important to stress that the standard error of $\boldsymbol{\nu}$ depends heavily on the calculation of $\mathrm{E}\left[\log(U_i)|\mathbf{y}_{obs_i},\widehat{\boldsymbol{\theta}}\right]$, which relies on computationally intensive Monte Carlo integrations. In our analysis, we focus solely on comparing the standard errors of $\boldsymbol{\beta}^\top$, σ^2 and λ.

5. Simulation studies

In order to examine the performance of our proposed models and algorithm, we present three simulation studies. The first compare the performance of the estimates for SMSN-CR models in the presence of outliers on the response variable. The second study shows that our proposed SAEM algorithm estimates do provide good asymptotic properties. In the the third study we show the consistency of the approximate standard errors for the ML estimates of parameters. All computational procedures were implemented using the `R` software (R Development Core Team, 2015). We performed all Monte Carlo simulation studies considering the model SMSN-CR, defined by combining (18)–(20) where $\boldsymbol{\beta}^\top = (\beta_1,\beta_2) = (1,4)$, $\sigma^2 = 2$, $\lambda = 4$ and $\mathbf{x}_i^\top = (1,x_i)$. The values x_i, $i = 1,\ldots,n$, were generated independently from a uniform distribution on the interval (2,20) and those values were kept constant throughout the experiment. For all simulation studies, were considered a random sample with censoring levels $p = 0\%$, 8%, 20% and 35% (i.e., 0%, 8%, 20% and 35% of the observations in each dataset were censored respectively). In addition, we also choose the parameters $m = 20$, $c = 0.3$ and $S = 400$ for the SAEM implementation.

5.1. Robustness of the SAEM estimates (Simulation study 1)

The purpose of this simulation study is to compare the performance of the estimates for some censored regression models in the presence of outliers on the response variable. We consider the different cases of the SMSN-CR models with fixed $\boldsymbol{\nu}$, i.e., SN-CR, ST-CR ($\nu = 3$), SSL-CR ($\nu = 3$) and SCN-CR ($(\nu, \gamma) = (0.1, 0.1)$).

For this case, we generated 200 samples of size $n = 300$ under the SN-CR model with $\varepsilon_i \sim \text{SN}(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda)$ and four percent of left censored values for the response variable, in each sample. To assess how much the SAEM estimates are influenced by the presence of outliers, we replaced the observation y_{150} by $y_{150}(\vartheta) = y_{150} + \vartheta$, with $\vartheta = 1, 2, \ldots, 10$. For each replication, we obtained the parameter estimates with and without outliers, under the four SMSN-CR models. We are interested in evaluating the relative change in the estimates as a ϑ function. Given $\boldsymbol{\theta} = (\beta_1, \beta_2, \sigma^2, \lambda)$, the relative change is defined by

$$RC\left(\widehat{\theta}_i(\vartheta)\right) = \left|\frac{\left(\widehat{\theta}_i(\vartheta) - \widehat{\theta}_i\right)}{\widehat{\theta}_i}\right|,$$

where $\widehat{\theta}_i(\vartheta)$ and $\widehat{\theta}_i$ denote the SAEM estimates of θ_i with and without perturbation, respectively.

Figure 2 show the average values of the relative changes undergone by all the parameters, for the censoring level of 8%. We note that for all parameters, the average relative changes suddenly increase under SN-CR model as the ϑ value grows. In contrast, for the SMSN-CR models with heavy tails, namely the ST-CR, SSL-CR and SCN-CR, the measures vary little, indicating they are more robust than the SN-CR model in the ability to accommodate discrepant observations. We also conducted simulations with three censoring rates ($p = 0\%$, 20% and 35%), obtaining similar results, as shown in Figures 6, 7 and 8 in Appendix A.

5.2. Asymptotic properties (Simulation study 2)

In this simulation study, the main focus is to evaluate the finite-sample performance of the parameter estimates. To do so we generated left-censored samples from the SMSN-CR model with the different censoring levels $p = 8\%$, 20% and 35% and sample sizes fixed at $n = 50, 150, 300, 450, 600$ and 750. For each combination of sample size and censoring level, we generated 500 samples from the SMSN-CR models, under four different situations: SN-CR, ST-CR ($\nu = 3$), SSL-CR ($\nu = 4$) and SCN-CR ($\boldsymbol{\nu}^\top = (0.1, 01)$).

As in Garay et al. (2015b), to evaluate the estimates obtained by the proposed SAEM algorithm, we compared the bias (Bias) and the mean square error (MSE) for each parameter over the 500 replicates. They are defined as

$$\text{Bias}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}(\widehat{\theta}_i^{(j)} - \theta_i) \quad \text{and} \quad \text{MSE}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}(\widehat{\theta}_i^{(j)} - \theta_i)^2,$$

where $\widehat{\theta}_i^{(j)}$ is the estimate of θ_i from the j-th sample for $j = 1, \ldots, 500$.

Analyzing figures 3 and 4, for the censoring level $p = 8\%$, it can be seen that the Bias and MSE tend to zero in all SMSN-CR models when n increases. Thus, as a general rule the results indicate that

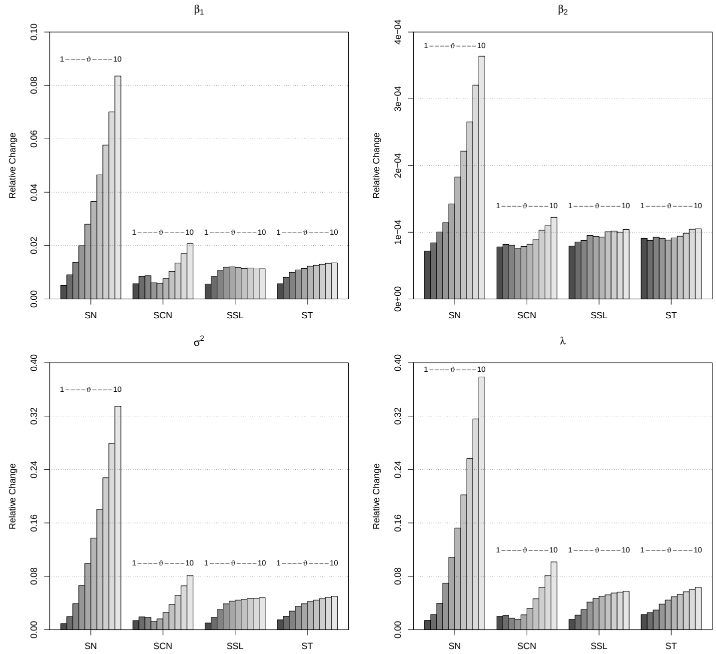

Figure 2: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 8\%$.

the ML estimates of the SMSN-CR models do provide good asymptotic properties. We also performed simulations with two higher censoring rates ($p = 20\%$ and 35%) and the patterns of convergence still behaved well (See Figures 9 - 12 given in Appendix B for more details).

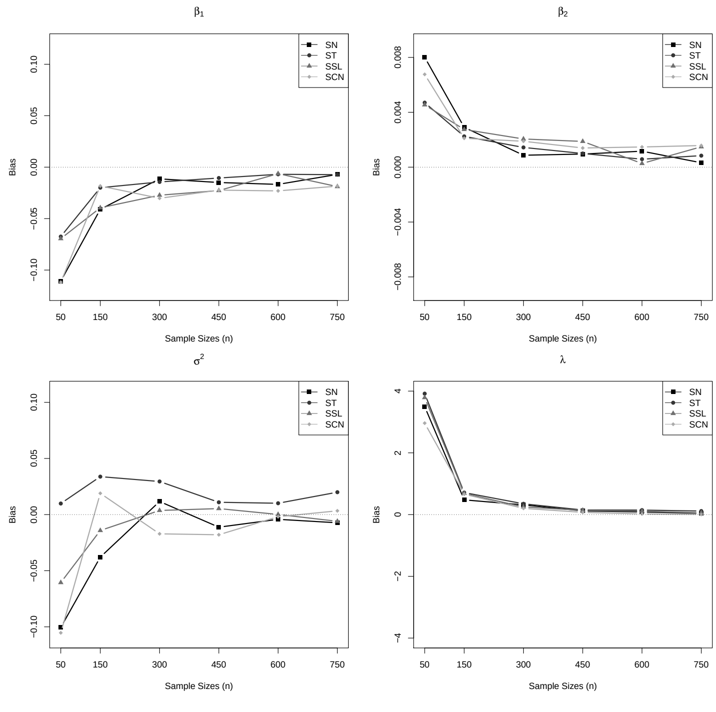

Figure 3: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN models with level of censoring $p = 8\%$.

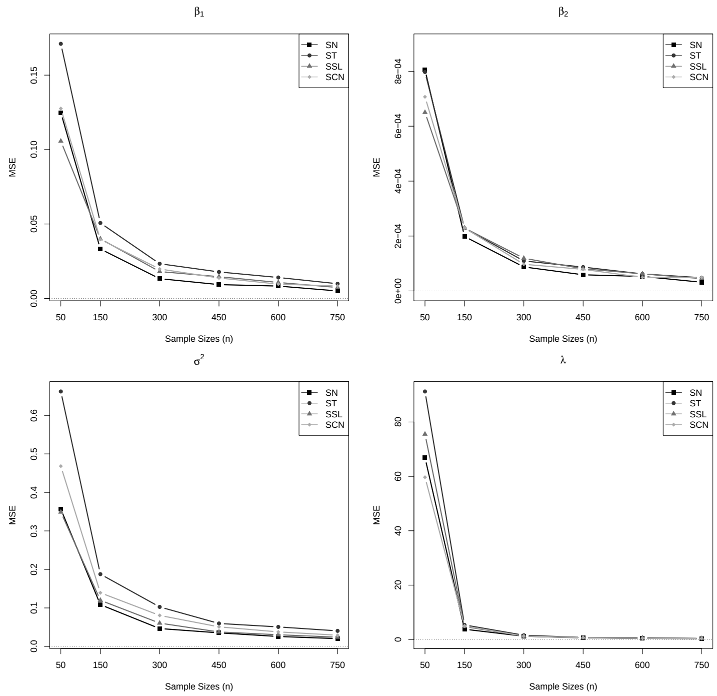

Figure 4: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN models with level of censoring $p = 8\%$.

5.3. Consistency of the estimates of the standard errors (Simulation study 3)

The design considered in this simulation study is the same as used in Subsection 5.1. Here, we examine the consistency of the approximation method, suggested in Section 4, to get the standard errors (SE) of ML estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma^2}, \widehat{\lambda})$ for the SMSN-CR models, considering four censoring levels $p = 0\%,\ 8\%,\ 20\%$ and 35%.

We generated 500 random samples of size $n = 450$ for the different SMSN-CR models: SN-CR, ST-CR ($\nu = 3$) , SSL-CR ($\nu = 4$) and SCN-CR ($\boldsymbol{\nu}^\top = (0.1, 01)$). For each sample, we obtained the ML estimates of $\boldsymbol{\theta} = (\beta_1, \beta_2, \sigma^2, \lambda)$, their SE using the technique proposed in Section 4 and the 95% normal approximation confidence intervals for each parameter, i.e., $\widehat{\theta} \pm 1.96\text{SE}$.

Considering all the ML estimates obtained (across 500 samples), we computed:

- the Monte Carlo standard deviation of $\widehat{\theta}_i$, defined by

$$\text{MC-Sd} = \sqrt{\frac{1}{499}\left[\sum_{j=1}^{500}\left(\widehat{\theta}_i^{(j)}\right)^2 - 500\left(\overline{\widehat{\theta}_i}\right)^2\right]} \quad \text{where} \quad \overline{\widehat{\theta}_i} = \frac{1}{500}\sum_{j=1}^{500}\widehat{\theta}_i^{(j)};$$

- the average values of the approximate standard errors of the SAEM estimates obtained through the method described in Subsection 3.2 using the empirical information matrix, denoted by AV-SE, and

- the percentage of times that the confidence intervals cover the true value of the parameter (COV MC).

Table 2: Simulation study 3. Results based on 500 simulated samples. MC Sd, AVE SE and COV MC are the respective average of the standard deviations, the average of the approximate standard error obtained through the information-based method and the coverage probability from fitting SMSN-CR models under various levels of censoring proportion.

Cens. Level	$\widehat{\theta}_i$	SN-CR			ST-CR		
		MC-Sd	AV-SE	COV MC	MC-Sd	AV-SE	COV MC
0%	$\widehat{\beta}_1$	0.0897	0.0847	92.60%	0.1099	0.1126	95.40%
	$\widehat{\beta}_2$	0.0070	0.0066	92.80%	0.0072	0.0076	97.00%
	$\widehat{\sigma}^2$	0.1803	0.1828	95.40%	0.2422	0.2411	95.40%
	$\widehat{\lambda}$	0.7739	0.7590	95.60%	0.7978	0.7884	95.00%
8%	$\widehat{\beta}_1$	0.0952	0.0981	96.00%	0.1331	0.1290	94.60%
	$\widehat{\beta}_2$	0.0076	0.0077	94.40%	0.0093	0.0089	93.80%
	$\widehat{\sigma}^2$	0.1878	0.1878	92.80%	0.2443	0.2484	95.60%
	$\widehat{\lambda}$	0.8127	0.7856	95.80%	0.8468	0.8460	95.00%
20%	$\widehat{\beta}_1$	0.1319	0.1314	94.20%	0.1612	0.1562	94.00%
	$\widehat{\beta}_2$	0.0095	0.0096	95.00%	0.0106	0.0105	94.20%
	$\widehat{\sigma}^2$	0.1827	0.2035	97.60%	0.2902	0.2653	92.60%
	$\widehat{\lambda}$	0.7963	0.8822	97.60%	0.9836	0.9104	92.80%
35%	$\widehat{\beta}_1$	0.1922	0.1881	94.60%	0.2452	0.2344	90.80%
	$\widehat{\beta}_2$	0.0134	0.0130	94.40%	0.0158	0.0153	93.40%
	$\widehat{\sigma}^2$	0.2185	0.2262	95.20%	0.2866	0.2835	93.40%
	$\widehat{\lambda}$	0.8434	0.9606	94.60%	0.7929	0.9635	92.20%
Cens. Level	$\widehat{\theta}_i$	SCN-CR			SSL-CR		
		MC-Sd	AV-SE	COV MC	MC-Sd	AV-SE	COV MC
0%	$\widehat{\beta}_1$	0.0982	0.0995	95.40%	0.1001	0.0953	92.80%
	$\widehat{\beta}_2$	0.0072	0.0074	96.40%	0.0076	0.0075	94.40%
	$\widehat{\sigma}^2$	0.2260	0.2184	94.40%	0.2030	0.1921	92.80%
	$\widehat{\lambda}$	0.8461	0.8028	95.00%	0.8115	0.7716	95.00%
8%	$\widehat{\beta}_1$	0.1152	0.1151	93.80%	0.1188	0.1176	94.40%
	$\widehat{\beta}_2$	0.0088	0.0086	95.20%	0.0088	0.0088	93.20%
	$\widehat{\sigma}^2$	0.2251	0.2244	93.80%	0.1940	0.1988	95.60%
	$\widehat{\lambda}$	0.8343	0.8303	94.20%	0.7410	0.8023	96.60%
20%	$\widehat{\beta}_1$	0.1372	0.1405	92.60%	0.1484	0.1523	94.80%
	$\widehat{\beta}_2$	0.0103	0.0102	93.20%	0.0105	0.0109	94.60%
	$\widehat{\sigma}^2$	0.2498	0.2431	95.80%	0.2080	0.2129	94.60%
	$\widehat{\lambda}$	0.8607	0.8978	95.20%	0.8307	0.8785	97.00%
35%	$\widehat{\beta}_1$	0.2032	0.2122	88.20%	0.2037	0.2125	93.20%
	$\widehat{\beta}_2$	0.0139	0.0142	90.60%	0.0139	0.0145	93.40%
	$\widehat{\sigma}^2$	0.2657	0.2653	93.80%	0.2304	0.2354	93.20%
	$\widehat{\lambda}$	0.8156	0.9845	93.80%	0.8894	0.9914	96.40%

Table 2 shows that in general, the COV MC for the parameters is quite stable for the censoring levels $p = 0\%$, 8%, and 20%, but it tends to be lower than the nominal level (95%) when considering a high level of censoring, say $p = 35\%$. This table also provides the average values of the approximate standard errors of the EM estimates obtained through the information-based method described in Subsection 3.2 (AV SE) and the Monte Carlo standard deviation (MC Sd) for the parameters. Table 2 also reveals that the estimation method of the standard errors provides relatively close results for the SMSN models, indicating that the proposed empirical information matrix(Equation 30) is reliable.

6. Application

In this section we provide an application of the results derived in the previous sections using the data described by Mroz (1987). The dataset consists of 753 married white women with ages between 30 and 60 years old in 1975, with 428 women who worked at some point during that year. The response variable is the wage rate, which represents a measure of the wage of the housewife known 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. The variables involved in the study were:

- y_i: defined as the average hourly earnings (wage rates);
- x_{i1}: wife's age;
- x_{i2}: wife's years of schooling;
- x_{i3}: the number of children younger than six years old in the household;
- x_{i4}: the number of children between the ages of six and nineteen.

These data were analyzed by Arellano-Valle et al. (2012) using the Student-t censored regression model; by Garay et al. (2015b) considering SMN-CR models and, more recently by Massuia et al. (2015) to evaluated the performance of the SMSN-CR models from a Bayesian perspective. Here, we revisit this dataset in order to evaluate the performance of the proposed SAEM algorithm to obtain ML estimates.

Table 3 contains the ML estimates for the parameters of the four models, i.e., SN-CR, ST-CR, SSL-CR and SCN-CR models, together with their corresponding standard errors calculated via the empirical information matrix. For the ST and SSL models, the estimated value of ν is small, indicating the in adequacy of the skew-normal (and normal) assumption for the wage rates dataset. Moreover, the results obtained under SN-CR and ST-CR models are consistent with those presented in Massuia et al. (2015). The SCN-CR and SSL-CR models presented estimates for λ closed to zero, indicating coherence with the results presented in Garay et al. (2015b). Table 4 compares the fit of the four SMSN models using the model selection criteria discussed in Subsection 3.4. Note that the SMSN distributions with heavy tails have better fit than the SN model. Particularly, the ST distribution fits the data better than the other three distributions. The comparison process is conducted now considering the symmetric SMN distributions (vide Garay et al. (2015b)) and we observe that according model selection criteria the ST-CR model still presents a better overall fit than the other four models (see Table 5 given in Appendix D).

In order to study departures from the error assumption as well as presence of outliers, we analyzed the transformation of the martingale type residual (MT), denoted by r_{MT_i}, proposed by Barros et al. (2010) for censored models. These residuals are defined by

$$r_{MT_i} = \text{sign}(r_{M_i})\sqrt{-2[r_{M_i} + \rho_i \log(\rho_i - r_{M_i})]}, \quad i = 1, \ldots, n.$$

Table 3: Wage rate data. Parameter estimates of the SMSN-CR models and SE for Wage rate data.

	SN-CR		**ST-CR**		**SCN-CR**		**SSL-CR**	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-1.3355	1.7627	-4.1685	1.4392	-1.3291	1.5026	-1.3489	1.4221
β_2	-0.1185	0.0272	-0.0722	0.0223	-0.1061	0.0229	-0.1053	0.0226
β_3	0.6917	0.0809	0.6541	0.0576	0.6490	0.0611	0.6434	0.0620
β_4	-3.2502	0.4345	-2.5956	0.3291	-3.0685	0.3642	-3.0480	0.3628
β_5	-0.2602	0.1433	-0.2676	0.1136	-0.3016	0.1199	-0.2901	0.1176
σ^2	32.8512	2.0202	19.4969	3.1730	11.8519	3.7339	6.7930	1.1830
λ	1.5454	0.4412	-1.6976	0.2942	0.1273	1.4542	-0.2144	0.3698
ν	-	-	2.5000	-	0.0537	-	1.45	-
γ	-	-	-	-	0.0645	-	-	-

Table 4: Wage rate data. Model selection criteria (values in bold correspond to the best model).

Criteria	SN-CR	ST-CR	SCN-CR	SSL-CR
log-likelihood	-1470.617	**-1410.583**	-1430.992	-1435.426
AIC	2955.234	**2837.166**	2879.984	2884.852
BIC	2987.602	**2874.159**	2921.601	2917.220
EDC	2979.651	**2865.071**	2911.378	2909.269

where $r_{M_i} = \rho_i + \log(S(y_i;\widehat{\boldsymbol{\theta}}))$ is the martingale residual, with $\rho_i = 0, 1$ indicating whether the observation is censored or not, respectively. $S(y_i, \widehat{\boldsymbol{\theta}})$ is the SAEM estimate of the survival function of y – see more details in Ortega et al. (2003) and Garay et al. (2015b). The normal probability plot of the MT residuals with generated envelopes is presented in Figure 5. From this figure, we note that the SMSN-CR models with heavy tails present better fit than the SN-CR model.

7. Conclusions

We have proposed a linear regression models with censored responses based on scale mixtures of skew-normal distributions, denoted by SMSN-CR, as a replacement for the conventional choice of normal (or symmetric) distribution for censored linear models. Our results generalize the works of Barros et al. (2010), Arellano-Valle et al. (2012), Massuia et al. (2014) and Garay et al. (2015b) from a frequentist point of view. In the context of SMSN-CR models, a Bayesian analysis was developed recently by Massuia et al. (2015). However, to the best of our knowledge, there are no previous studies of a likelihood based perspective related to this topic.

In order to explore the performance of our proposed models and SAEM algorithm, we developed three simulation studies. The study compared the performance of the estimates for SMSN-CR models in the presence of outliers on the response variable. The second study showed that our proposed SAEM algorithm estimates do provide good asymptotic properties. The third study showed the consistency of the approximate of standard errors for the ML estimates of parameters. We also applied our method to the wage rate dataset of Mroz (1987), in order to illustrate how the procedure developed can be used to

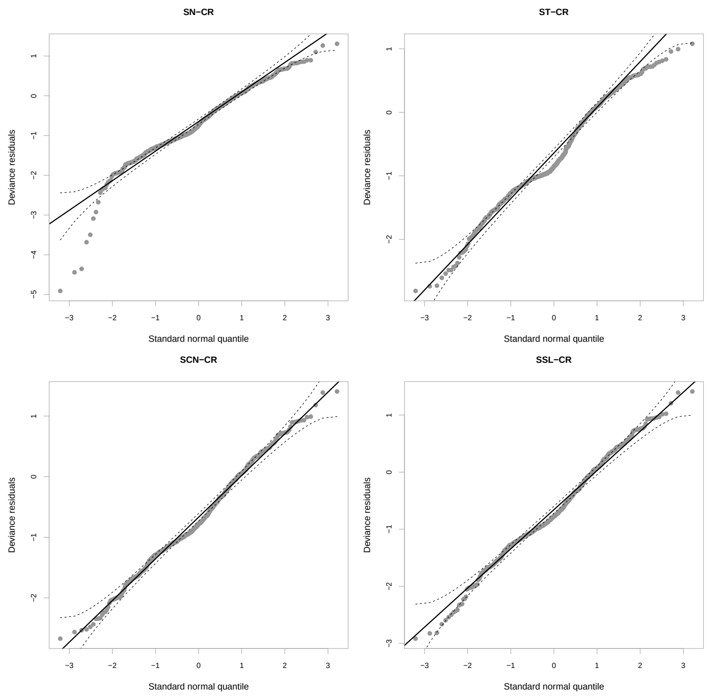

Figure 5: Wage rate data. Envelopes of the MT residuals for the SMSN-CR models

evaluate model assumptions and obtain robust parameter estimates. As expected, our proposed SMSN-CR with heavy tails models, as ST-CR, SSL-CR and SCN-CR models, present better results than the SN-CR model. It is interesting to note that the ST-CR model still presents a better overall fit than the symmetrical SMN-CR models.

Due to recent advances in computational technology, it is worth carrying out some extensions of the current work, for example, diagnostics analysis in the SMSN-CR models or censored nonlinear regression models (SMSN-NLCR) as in Garay et al. (2011). Another interesting topic for further research would be to generalize our results considering irregularly observed longitudinal data using the SMSN multivariate class of distributions, as in Garay et al. (2015a).

Acknowledgments

The research of Víctor H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2014/02938-9 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). The research of Aldo M. Garay was supported by Grant 2013/21468-0 from FAPESP-Brazil and Thalita B. Mattos was supported by CAPES.

Appendix A: Complementary results of the simulation study 1

In this appendix, we present the results of the simulation study 1 for different levels of censoring: $p = 0\%$, 20% and 35%.

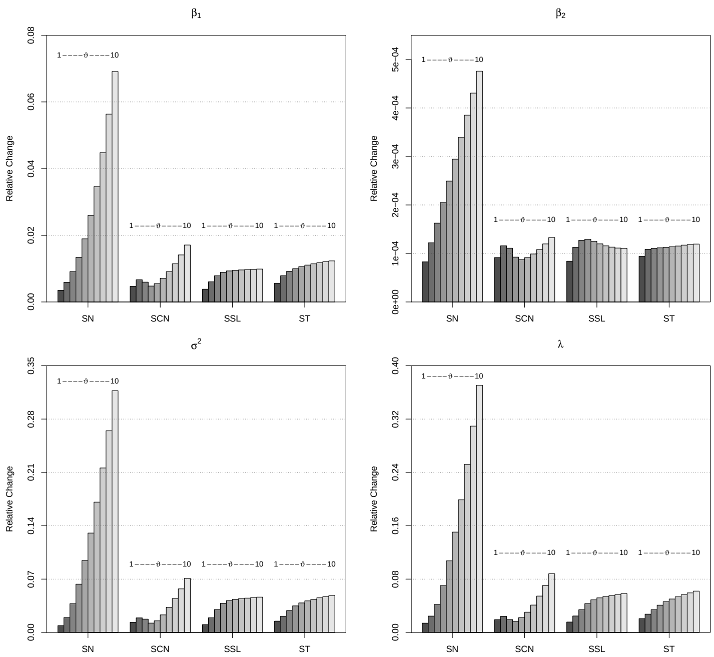

Figure 6: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 0\%$.

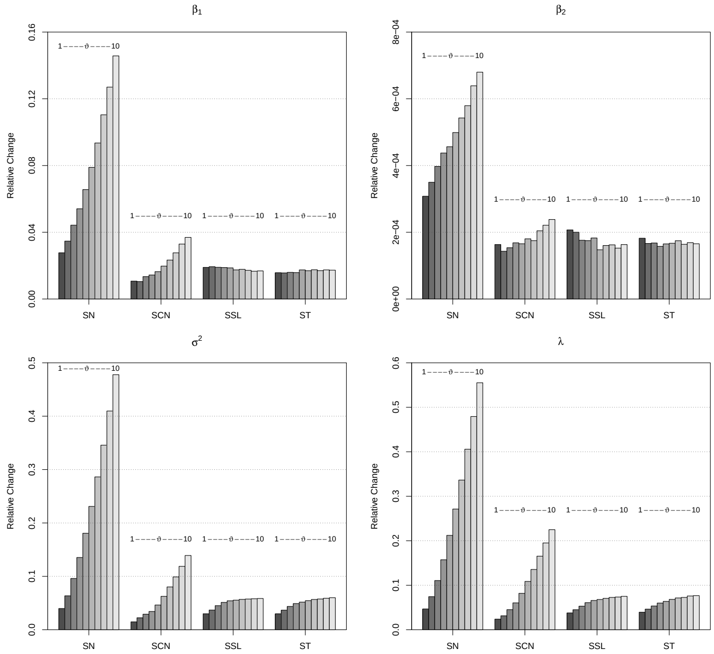

Figure 7: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 20\%$.

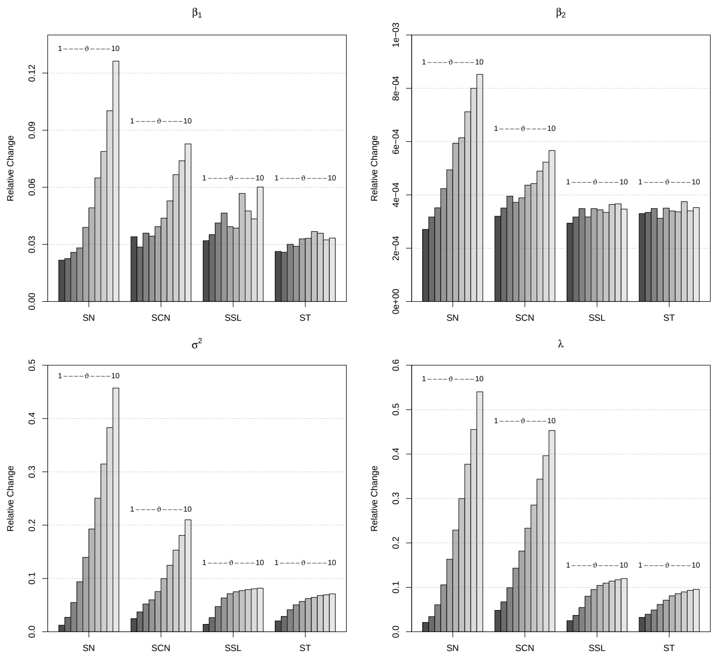

Figure 8: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 35\%$.

Appendix B: Complementary results of the simulation study 2

Here we show the Bias and MSE of parameters $\boldsymbol{\theta}$, for the levels of censoring $p = 20\%$ and 35%, respectively.

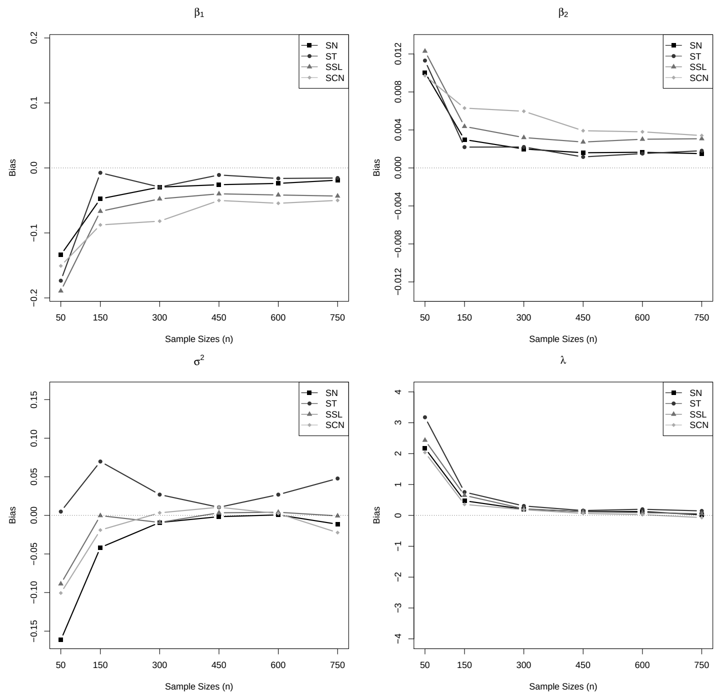

Figure 9: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 20\%$.

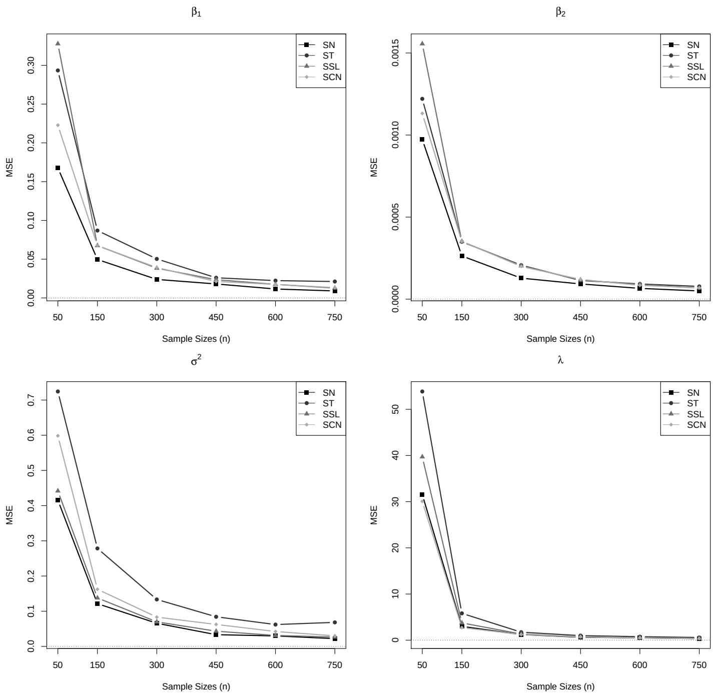

Figure 10: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 20\%$.

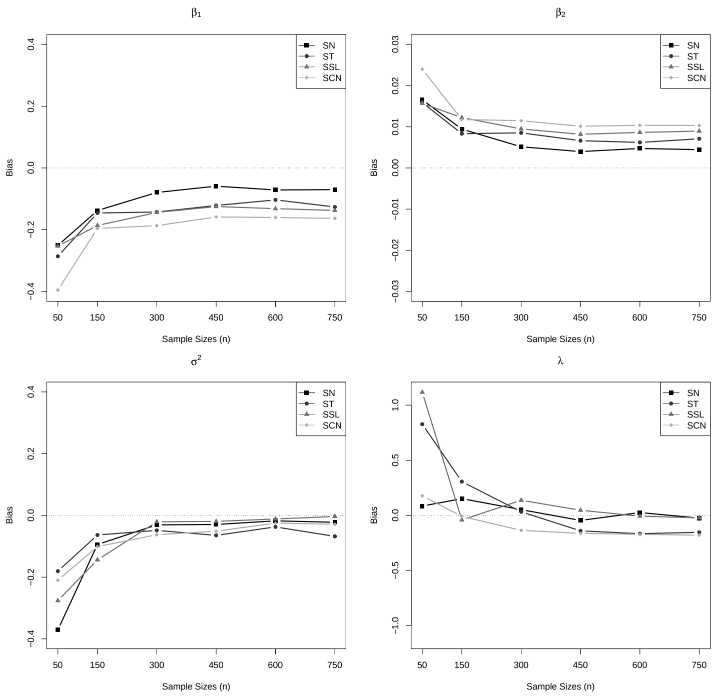

Figure 11: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 35\%$.

Appendix C: Complementary results of the application

In this appendix, we describe the summary of convergence for the parameters, $\boldsymbol{\beta}$, σ^2, λ, $\boldsymbol{\nu}$, for the SMSN-CR models. The vertical dashed line delimits the beginning of the almost sure convergence, as

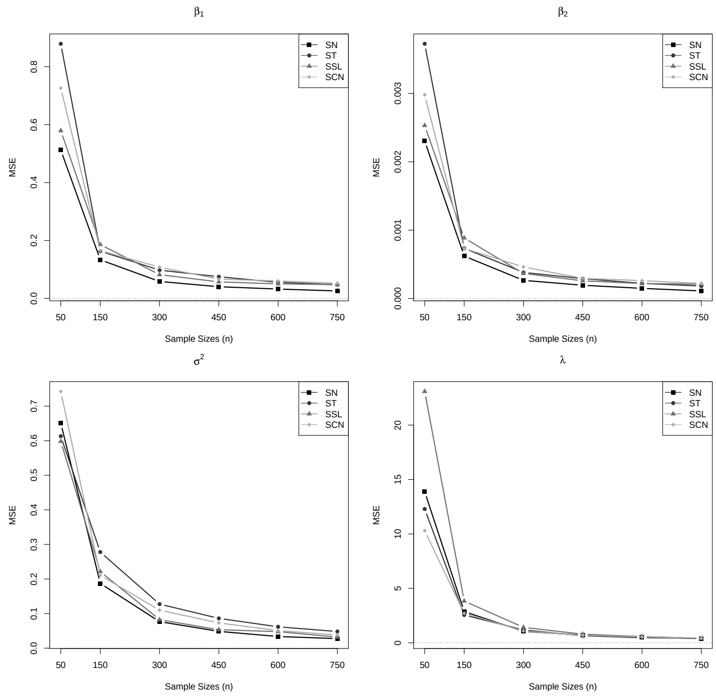

Figure 12: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 35\%$.

defined by the cutoff point, c.

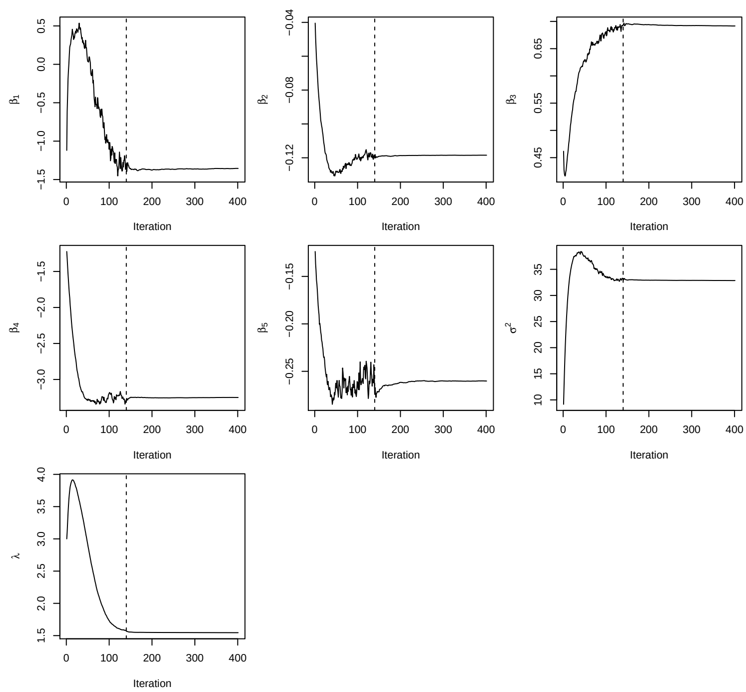

Figure 13: Wage rate data. Graphical summary of convergence for the parameters from SN-CR model, $m = 20$, $c = 0.35$ and $S = 400$.

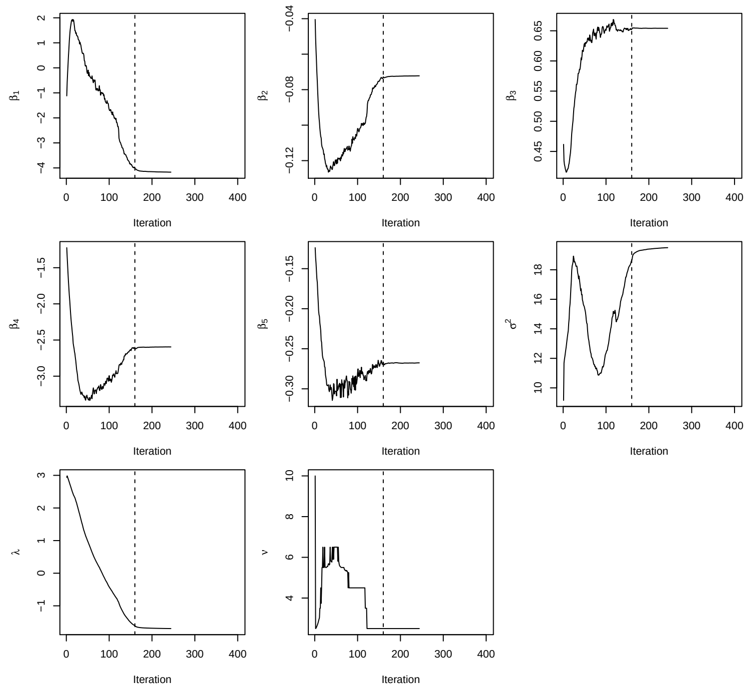

Figure 14: Wage rate data. Graphical summary of convergence for the parameters from ST-CR model, $m = 20$, $c = 0.40$ and $S = 400$.

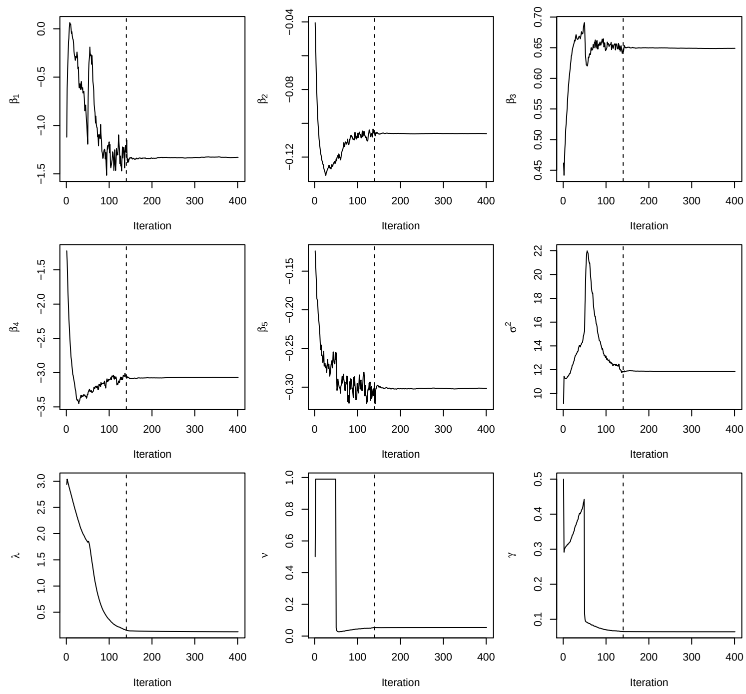

Figure 15: Wage rate data. Graphical summary of convergence for the parameters from SCN-CR model, $m = 20$, $c = 0.35$ and $S = 400$.

Appendix D: *Wage rate* data under SMN-CR models

In this appendix, we present the comparison between the SMN-CR models, considering the *wage rate* dataset.

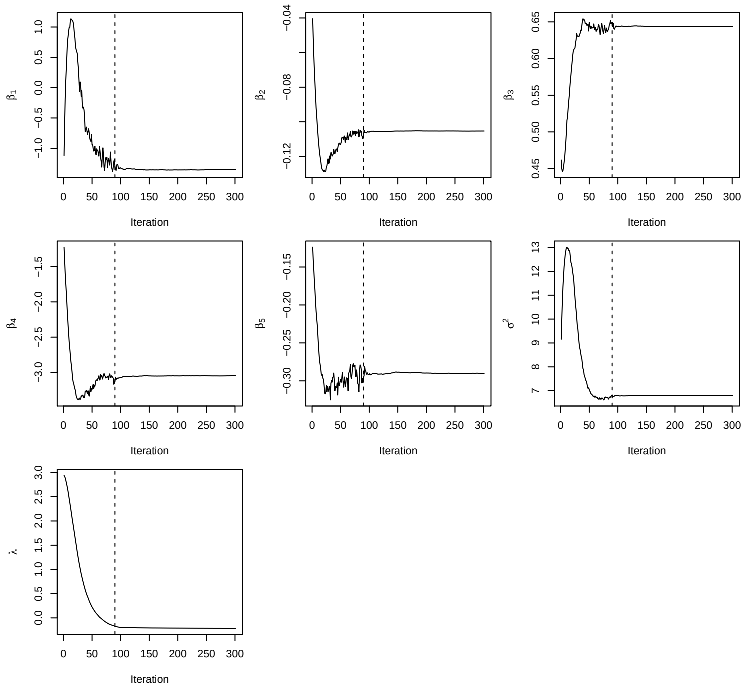

Figure 16: Wage rate data. Graphical summary of convergence for the parameters from SSL-CR model, $m = 20$, $c = 0.30$ and $S = 300$.

Table 5: Wage rate data. Values of some model selection criteria for SMN-CR models

Criteria	N-CR	T-CR	CN-CR	SL-CR
log-likelihood	-1481.6550	-1440.1450	-1432.0850	-1436.2860
AIC	2975.3110	2894.2910	2880.1710	2886.5730
BIC	3003.0550	2926.6590	2917.1630	2918.9410
EDC	2996.2400	2918.7080	2908.0760	2910.9900

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Cont. 19, 716–723.

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.

Bai, Z. D., Krishnaiah, P. R., Zhao, L. C., 1989. On rates of convergence of efficient detection criteria in signal processing with white noise. IEEE Trans. Info. Theory 35, 380–388.

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.

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.

Couvreur, C., 1996. The EM algorithm: a guided tour. In: Proceedings of the 2d IIEE European Workshop on Computationaly Intensive Methos in Control and Signal Processing.

Delyon, B., Lavielle, M., Moulines, E., 1999. Convergence of a stochastic approximation version of the EM algorithm. Annals of Statistics 27 (1), 94–128.

Fernández, C., Steel, M. J. F., 1999. Multivariate Student-t regression models: Pitfalls and inference. Biometrika 86, 153–167.

Galarza, C. E., Bandyopadhyay, D., Lachos, V. H., 2015. Quantile regression for linear mixed models: A stochastic approximation em approach. Tech. Rep. 2, Universidade Estadual de Campinas.

Garay, A. M., Castro, L. M., Leskow, J., Lachos, V. H., 2015a. Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. Statistical Methods in Medical Research, DOI: 10.1177/0962280214551191.

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.

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

Jank, W., 2006. Implementing and diagnosing the stochastic approximation em algorithm. Journal of Computational and Graphical Statistics 15, 803–829.

Kuhn, E., Lavielle, M., 2004. Coupling a stochastic approximation version of EM with an MCMC procedure. ESAIM: Probability and Statistics 8, 115–131.

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., 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.

Lin, T. I., 2010. Robust mixture modeling using multivariate skew-t distributions. Statistics and Computing 20 (3), 343–356.

Louis, T. A., 1982. Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society, Series B 44, 226–233.

Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H., 2014. Influence diagnostics for Student-t censored linear regression models. Statistics 49, 1074–1094.

Massuia, M. B., Garay, A. M., Lachos, V. H., Cabral, C. R. B., 2015. Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. Tech. Rep. 3, Universidade Estadual de Campinas.

Meilijson, I., 1989. A fast improvement to the em algorithm to its own terms. J. R. Stat. Soc. Ser. B 51, 127–138.

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.

Ortega, E. M. M., Bolfarine, H., Paula, G. A., 2003. Influence diagnostics in generalized log-gamma regression models. Computational Statistics & Data Analysis 42, 165–186.

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., 1987. The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. Journal of the American Statistical Association, 543–546.

Rubin, D. B., et al., 1988. Using the sir algorithm to simulate posterior distributions. Bayesian statistics 3 (1), 395–402.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society 26, 24–36.

Vaida, F., 2005. Parameter convergence for EM and MM algorithms. Statistica Sinica 15 (3), 831–840.

Wei, G. C., Tanner, M. A., 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association 85 (411), 699–704.

Wu, C. J., 1983. On the convergence properties of the EM algorithm. The Annals of Statistics 11 (1), 95–103.