Linear Censored Regression Models with Scale Mixtures of Normal Distributions

Aldo M. Garay[a] **and Victor H. Lachos**[b1]
and Heleno Bolfarine[a] **and Celso R. B. Cabral**[c]
[a] Department of Statistics, São Paulo University, Brazil
[b] Department of Statistics, Campinas State University, Brazil
[c] Department of Statistics, Federal University of Amazonas, Brazil

Abstract

In the framework of censored regression models the random errors are routinely assumed to have a normal distribution, mainly for mathematical convenience. However, this method has been criticized in the literature because of its sensitivity to deviations from the normality assumption. In practice, data such as income or viral load in AIDS studies, often violate this assumption because of heavy tails. Here, we first establish a new link between the censored regression model and a recently studied class of symmetric distributions, which extend the normal one by the inclusion of kurtosis, called scale mixtures of normal (SMN) distributions. The Student-t, Pearson type VII, slash, contaminated normal, among others distributions, are contained in this class. Choosing a member of this class can be a good alternative to model this kind of data, because they have been shown its flexibility in several applications. In this work, we develop an analytically simple and efficient EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters, with standard errors as a by-product. The algorithm has closed-form expressions at the E-step, that rely on formulas for the mean and variance of certain truncated SMN distributions. The proposed algorithm is implemented in the R package `SMNCensReg`. Applications with simulated and a real data are reported, illustrating the usefulness of the new methodology.

Keywords: Censored regression model; EM-type algorithms; Scale mixtures of normal distributions; Outliers.

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

1 Introduction

Regression models with censored dependent variable (hereafter CR models) are applied in many fields, like econometric analysis, clinical essays, medical surveys, engineering studies, among others. For example, in econometrics, the study of the labor force participation of married women is usually conducted under the ordinary Tobit model (Greene, 2012). In this case, the observed response is the wage rate, which is typically considered as censored below zero, i.e., for working women, positive values for the wage rates are registered, whereas for the non-working women the observed wage rates are zero (see Mroz, 1987). In AIDS research, the viral load measures may be subjected to some upper and lower detection limits, below or above which they are not quantifiable. As a result, the viral load responses are either left or right censored depending on the diagnostic assays used (see Wu, 2010).

In general, for mathematical tractability reasons, it is assumed that the random errors have a normal distribution (Wei and Tanner, 1990). However, it is well-known that several phenomena are not always in agreement with this assumption, yielding data with a distribution with heavier tails. The problem of longer-than-normal tails (or outliers) can be circumvented by data transformations (namely, Box–Cox, etc.), which can render approximate normality with reasonable empirical results. However, some possible drawbacks of these methods are: (i) transformations provide reduced information on the underlying data generation scheme; (ii) component wise transformations may not guarantee joint normality; (iii) parameters may lose interpretability on a transformed scale and (iv) transformations may not be universal and usually vary with the data set. Hence, from a practical perspective, there is a necessity to seek an appropriate theoretical model that avoids data transformations, yet presenting a robustified "Gaussian" framework.

To deal with the problem of atypical observations in regression models with complete responses, proposals have been made in the literature to replace normality with more flexible classes of distributions. For instance, Lange et al. (1989) discussed the use of the Student-t distribution in multivariate regression models. In this case, the degrees of freedom parameter is the natural choice to control kurtosis. Ibacache-Pulgar and Paula (2011), proposed some local influence measures in Student-t partially linear models. Villegas et al. (2012) proposed the generalized symmetric linear models, in which a link function is defined to establish a relationship between the mean values of symmetric distributions and linear predictors.

Recently, Arellano-Valle et al. (2012) advocated the use of the Student-t distribution in the context of censored regression models. Massuia et al. (2012) developed diagnostic measures for this model, including the implementation of an interesting (and simple) EM (expectation-maximization) algorithm for ML (maximum likelihood) estimation. They demonstrated its robustness aspects against outliers through extensive simulations.

Although there are some proposals that overcome the problem of atypical observations in CR models, there are no studies taking into account, at the same time, censored responses and observational errors modeled by a distribution in the scale mixture of normal class, which is, maybe, the most important family of symmetric distributions. SMN distributions are extensions of the normal one, incorporating kurtosis. The Student-t (T), Pearson type VII (PVII), slash (SL), power exponential (PE), contaminated normal (CN) and, obviously, the normal (N) distributions are included in this class. Comprehensive surveys are available in Fang and Zhang (1990), Arellano-Valle (1994) and Meza et al. (2012), among others. In this paper, we propose a CR model where the observational errors have a SMN distribution (hereafter we will call it the SMN-CR model). A fully likelihood-based approach is carried out, including the implementation of an exact EM-type algorithm for ML estimation. As in Massuia et al. (2012), we show that the E-step reduces to computing the first two moments of certain truncated SMN distributions. The general formulas for these moments were derived in closed form by Genç (2012). The likelihood function and the asymptotic standard errors are easily computed as a by-product of the E-step and are used for monitoring convergence and for model selection using the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). The theoretical justification of the proposal rests on the facts that the SMN class stochastically attributes varying weights to each subject, i.e., lower weight for outliers and thus controls the influence of atypical observations on the overall inference. Moreover, every member of the SMN class tends to the normal case, for example, as the Student-t degrees of freedom tends to the infinity, it approaches normality.

The rest of the paper is organized as follows. Section 2 briefly outlines some preliminary properties of the SMN and truncated SMN distributions. The SMN-CR model is presented in Section 3, including the implementation of the ECME algorithm (Liu and Rubin, 1994) for ML estimation, which is a simple extension/modification of the EM algorithm. In Section 4, we derive approximate standard errors for the regression parameters of the SMN-CR model. Section 5, presents some simulation studies to compare the performance of our methods with other normality-based methods. In Section 6, advantages of the proposed methodology is illustrated through the analysis of a real data set on housewives wages, previously analyzed under normal errors. Section 7 concludes with a short discussion on the issues raised by our study and some possible directions for future research.

2 Preliminaries

Throughout 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. In general, 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.

We start by defining the SMN distributions, through their hierarchical formulation, and then we introduce some further properties.

Definition 1. *We say that a random variable X has a SMN distribution with location parameter μ and scale parameter $\sigma^2 > 0$ if it has the following stochastic representation:*

$$X = \mu + U^{-\frac{1}{2}} Z, \quad Z \perp U \tag{1}$$

where $Z \sim \mathrm{N}(0, \sigma^2)$, U is a positive random variable with cdf $H(\cdot|\boldsymbol{\nu})$ and $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of U.

We use the notation $X \sim \mathrm{SMN}(\mu, \sigma^2, \boldsymbol{\nu})$. When $\mu = 0$ and $\sigma^2 = 1$ we have the so-called standard SMN distribution. Note from (1) that $X|U = u \sim N(\mu, u^{-1}\sigma^2)$. Thus, integrating out U from the joint density of X and U will lead to the following marginal density of X:

$$f_{SMN}\left(x|\mu, \sigma^2, \boldsymbol{\nu}\right) = (2\pi\sigma^2)^{-\frac{1}{2}} \int_0^\infty u^{\frac{1}{2}} \exp\left\{-(u/2\sigma^2)(x-\mu)^2\right\} dH\left(u|\boldsymbol{\nu}\right), \tag{2}$$

where $H(\cdot|\boldsymbol{\nu})$ is the cdf of U, which determines the form of the SMN distribution. U is called the scale factor and $H(\cdot|\boldsymbol{\nu})$ is called the mixture distribution.

It is important to notice that exists a relation between SMN distributions and elliptical distributions. We say that the random variable X has a univariate elliptical distribution with location parameter μ and scale parameter σ^2, when its density is given by

$$f(x) = \sigma^{-1} g\left(z\right), \tag{3}$$

where $z = (x-\mu)^2/\sigma^2$ and $g : \mathbb{R} \to [0,\infty)$ satisfies $\int_0^\infty z^{-\frac{1}{2}} g(z) dz < \infty$. It easy to see that (2) has the form (3). The relation between SMN and elliptical distributions will be used in Section 4, to obtain standard errors for the regression parameters.

Definition 2. *Let $X \sim SMN(\mu, \sigma^2, \boldsymbol{\nu})$ and $a < b$ such that $P(a < X < b) > 0$. A random variable Y has a truncated SMN distribution in the interval (a, b) if it has the same distribution as $X|X \in (a, b)$. In this case we write $Y \sim TSMN_{(a,b)}(\mu, \sigma^2, \boldsymbol{\nu})$.*

As an obvious consequence of Definition 2, we can obtain the density of $Y \sim \mathrm{TSMN}_{(a,b)}(\mu, \sigma^2, \boldsymbol{\nu})$, given by

$$f_{\,\mathrm{TSMN}}(y|\mu, \sigma^2, \boldsymbol{\nu}; (a,b)) = f_{SMN}(y|\mu, \sigma^2, \boldsymbol{\nu}) \left[\mathcal{F}_{SMN}\left(\frac{b-\mu}{\sigma}\right) - \mathcal{F}_{SMN}\left(\frac{a-\mu}{\sigma}\right)\right]^{-1}, \tag{4}$$

with $a < y < b$ and $f_{\,\mathrm{TSMN}}(y|\mu, \sigma^2, \boldsymbol{\nu}; (a,b)) = 0$ otherwise, where $\mathcal{F}_{SMN}(\cdot)$ denotes the cdf of the standard SMN distribution.

Now we establish the following proposition, which is crucial to the development of our proposed theory. It is a natural extension of Theorem 1 (and Corollary 1) of Genç (2012). In what follows $\mathrm{E}[\cdot]$ denotes expectation, $\mathrm{E}_X[\cdot]$ denotes expectation relative to the distribution of X and, for the sake of notation simplicity, we denote all pdf's by $f(\cdot)$. Thus, for example, $f(u)$ denotes the pdf of U, $f(u, x)$ denotes the joint pdf of U and X and $f(u|X \in \mathcal{A})$ denotes the pdf of U given the event $(X \in \mathcal{A})$.

Proposition 1. *Let $X \sim \mathrm{SMN}(0, 1, \boldsymbol{\nu})$ with scale factor U and mixture distribution $H(\cdot|\boldsymbol{\nu})$. Then, for $a < b$,*

$$\begin{aligned}
\mathrm{E}\left[U^r|X \in (a,b)\right] &= \tau(a,b)\left[\mathrm{E}_\Phi\left(r,b\right) - \mathrm{E}_\Phi\left(r,a\right)\right];\\
\mathrm{E}\left[U^r X|X \in (a,b)\right] &= \tau(a,b) \times \left[\mathrm{E}_\phi\left(r-\frac{1}{2},a\right) - \mathrm{E}_\phi\left(r-\frac{1}{2},b\right)\right];\\
\mathrm{E}\left[U^r X^2|X \in (a,b)\right] &= \tau(a,b)\left[\mathrm{E}_\Phi\left(r-1,b\right) - \mathrm{E}_\Phi\left(r-1,a\right)\right.\\
&\quad \left. + a\mathrm{E}_\phi\left(r-\frac{1}{2},a\right) - b\mathrm{E}_\phi\left(r-\frac{1}{2},b\right)\right],
\end{aligned}$$

where

$$\tau(a,b) = \left(\mathcal{F}_{SMN}(b) - \mathcal{F}_{SMN}(a)\right)^{-1}; \tag{5}$$

$$\mathrm{E}_{\phi}(r,h) = \mathrm{E}\left[U^r \phi\left(h\, U^{\frac{1}{2}}\right)\right] = \int_0^{\infty} u^r \phi\left(h\, u^{\frac{1}{2}}\right) dH\left(u|\boldsymbol{\nu}\right); \tag{6}$$

$$\mathrm{E}_{\Phi}(r,h) = \mathrm{E}\left[U^r \Phi\left(h\, U^{\frac{1}{2}}\right)\right] = \int_0^{\infty} u^r \Phi\left(h\, u^{\frac{1}{2}}\right) dH\left(u|\boldsymbol{\nu}\right). \tag{7}$$

Proof. Let $\mathcal{A} = (a,b)$. From Definitions 1 and 2, we have that $X|U = u \sim \mathrm{N}(0,u^{-1})$, $X|X \in \mathcal{A} \sim \mathrm{TSMN}_{\mathcal{A}}(0,1,\boldsymbol{\nu})$ and $X|U = u, X \in \mathcal{A} \sim \mathrm{TN}_{\mathcal{A}}(0,u^{-1})$, that is, a truncated normal distribution in $\mathcal{A}$, being 0 and u^{-1} the mean and variance, respectively, before truncation. Then,

$$\begin{aligned}
\mathrm{E}\left[U^r X^s | X \in \mathcal{A}\right] &= \mathrm{E}_U\left[U^r\ \mathrm{E}_X\left[X^s|U, X \in \mathcal{A}\right]|X \in \mathcal{A}\right] \\
&= \int_0^{\infty} U^r\ \mathrm{E}_X\left[X^s|U, X \in \mathcal{A}\right] f(u|X \in \mathcal{A})du. \qquad (8)
\end{aligned}$$

The pdf in the integral sign takes the following form:

$$\begin{aligned}
f(u|X \in \mathcal{A}) &= \int f(u,x|X \in \mathcal{A})dx = \int f(u|X = x, X \in \mathcal{A}) f(x|X \in \mathcal{A})dx \\
&= \tau(a,b) \int f(u|X = x, X \in \mathcal{A}) f(x) \mathbb{I}_{\mathcal{A}}(x)dx \qquad (9) \\
&= \tau(a,b) \int f\left(u,x\right) \mathbb{I}_{\mathcal{A}}(x)dx \qquad (10) \\
&= \tau(a,b) \int_{\mathcal{A}} f\left(u\right) \phi\left(x|0,u^{-1}\right) dx = \tau(a,b) f\left(u\right) \int_{\mathcal{A}^*} \phi\left(z\right) dz \\
&= \tau(a,b) f\left(u\right) \left[\Phi\left(bu^{\frac{1}{2}}\right) - \Phi\left(au^{\frac{1}{2}}\right)\right],
\end{aligned}$$

where $\mathcal{A}^* = (au^{\frac{1}{2}}, bu^{\frac{1}{2}})$. Equation (9) is obtained using the pdf's expression of $X|X \in \mathcal{A}$. Equation (10) is consequence of the fact that, if $x \in \mathcal{A}$, then $(X \in \mathcal{A}, X = x) = (X = x)$, implying that $f(u,x) = f(u|X = x)f(x) = f(u|X = x, X \in \mathcal{A})f(x)$. If $x \notin \mathcal{A}$ then $\mathbb{I}_{\mathcal{A}}(x) = 0$ and the integrands in (9) and (10) are equal to zero. By (8) and Lemma 1 given in Appendix A, it follows that

- for $s = 0$,

$$\mathrm{E}\left[U^r | X \in \mathcal{A}\right] = \int_0^{\infty} u^r f(u|X \in \mathcal{A})du = \tau(a,b)\mathrm{E}_U\left\{U^r\left[\Phi\left(bU^{\frac{1}{2}}\right) - \Phi\left(aU^{\frac{1}{2}}\right)\right]\right\};$$

- for $s = 1$,

$$\begin{aligned}
\mathrm{E}\left[U^r X | X \in \mathcal{A}\right] &= \int_0^{\infty} \frac{u^r}{u^{\frac{1}{2}}} \frac{\phi\left(au^{\frac{1}{2}}\right) - \phi\left(bu^{\frac{1}{2}}\right)}{\Phi\left(bu^{\frac{1}{2}}\right) - \Phi\left(au^{\frac{1}{2}}\right)} f(u|X \in \mathcal{A})(u)du \\
&= \tau(a,b)\mathrm{E}_U\left\{U^{r-\frac{1}{2}}\left[\phi\left(aU^{\frac{1}{2}}\right) - \phi\left(bU^{\frac{1}{2}}\right)\right]\right\}.
\end{aligned}$$

- for $s = 2$,

$$
\begin{aligned}
\mathrm{E}\left[U^r X^2 | X \in \mathcal{A}\right] &= \int_0^\infty \left[u^{r-1} + \frac{a u^{r-\frac{1}{2}} \phi\left(a u^{\frac{1}{2}}\right) - b u^{r-\frac{1}{2}} \phi\left(b u^{\frac{1}{2}}\right)}{\Phi\left(b u^{\frac{1}{2}}\right) - \Phi\left(a u^{\frac{1}{2}}\right)} \right] f(u|X \in \mathcal{A}) du \\
&= \tau(a,b) \mathrm{E}_U \left\{ U^{r-1} \left[\Phi\left(b U^{\frac{1}{2}}\right) - \Phi\left(a U^{\frac{1}{2}}\right) \right] \right. \\
&\quad \left. + U^{r-\frac{1}{2}} \left[a\phi\left(a U^{\frac{1}{2}}\right) - b\phi\left(b U^{\frac{1}{2}}\right) \right] \right\}.
\end{aligned}
$$

□

When the distribution of U is available, this proposition gives closed form expressions for the expected values $\mathrm{E}\left[U^r X^s | X \in (a,b)\right]$, where $s = 0, 1, 2$ and $r \geq 1$.

Now we compute the quantities $E_\phi(r,h)$ and $E_\Phi(r,h)$ for some elements of the SMN family. They are useful for implementing the ECME algorithm. For the sake of completeness, a detailed proof of these results is sketched in Appendix B.

- *Pearson type VII distribution:* in this case we consider $U \sim Gamma(\nu/2, \delta/2)$, with $\nu > 0$ and $\delta > 0$, where $Gamma(a,b)$ denotes the Gamma distribution with mean a/b. The density of the random variable X, defined in (1), takes the form

$$
f_{PVII}(x|\nu,\delta) = \frac{1}{B(\nu/2, 1/2)\sqrt{\delta}} \left(1 + \frac{x^2}{\delta}\right)^{-\frac{\nu+1}{2}},
$$

where $\delta > 0$ and $\nu > 0$ are shape parameters and $B(a,b)$ represents the beta function. We use the notation $X \sim PVII(0,1;\nu,\delta)$. In this case, we have that

$$
\begin{aligned}
\mathrm{E}_\Phi(r,h) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{-r} F_{PVII}(h|\nu+2r, \delta); \\
\mathrm{E}_\phi(r,h) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}} \left(\frac{\delta}{2}\right)^{\frac{\nu}{2}} \left(\frac{h^2+\delta}{2}\right)^{-\frac{(\nu+2r)}{2}},
\end{aligned}
$$

where $\Gamma(a)$ is the gamma function and $F_{PVII}(\cdot)$ is the cdf of the Pearson type VII distribution. When $\delta = \nu$ we have the Student-t distribution with ν degrees of freedom. Also, we have the Cauchy distribution when $\delta = \nu = 1$.

- *Slash distribution:* here the distribution of the scale factor U is $Beta(\nu, 1)$, with $\nu > 0$. The density of the random variable X, defined in (1), is given by

$$
f_{sl}(x|\nu) = \nu \int_0^1 u^{\nu-1} \phi(x u^{\frac{1}{2}}) du.
$$

We use the notation $X \sim SL(0,1;\nu)$. In this case, we have that

$$\mathrm{E}_{\Phi}\left(r,h\right) = \left(\frac{\nu}{\nu+r}\right) F_{SL}(h|\nu+r);$$
$$\mathrm{E}_{\phi}\left(r,h\right) = \frac{\nu}{\sqrt{2\pi}}\left(\frac{h^2}{2}\right)^{-(\nu+r)}\Gamma\left(\nu+r,\frac{h^2}{2}\right),$$

where $\Gamma\left(a,b\right)=\int_0^b e^{-t}t^{a-1}dt$ is the incomplete gamma function, see Lemma 6 in Genç (2012), and $F_{SL}(\cdot)$ is the cdf of the slash distribution.

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

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

where $\mathbb{I}_B(\cdot)$ is the indicator function of the set B. It follows immediately that the density of the random variable X, defined in (1), is given by

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

So, we have that

$$\mathrm{E}_{\Phi}\left(r,h\right) = \gamma^r F_{CN}(h|\xi,\gamma)+\left(1-\gamma^r\right)\left(1-\xi\right)\Phi\left(h\right);$$
$$\mathrm{E}_{\phi}\left(r,h\right) = \xi\gamma^r\phi\left(h\sqrt{\gamma}\right)+\left(1-\xi\right)\phi\left(h\right),$$

where $F_{CN}(\cdot)$ is the cdf of the contaminated normal distribution.

As a direct consequence of Proposition 1, in Appendix A we present some important corollaries, which are useful for implementing the ECME algorithm.

3 The SMN censored linear regression model

3.1 The model

Consider first a linear regression model where the responses are observed with errors which are independent and identically distributed according to some SMN distribution. To be more precise, let us write

$$Y_i=\mathbf{x}_i^{\top}\boldsymbol{\beta}+\varepsilon_i,\ \ \varepsilon_i \overset{\text{iid.}}{\sim} \mathrm{SMN}(0,\sigma^2,\boldsymbol{\nu}),\ \ i=1,\ldots,n, \tag{11}$$

where the $Y_i's$ are responses, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression parameters and $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ is a vector such that x_{ij} is the value of the j-th explanatory variable for the subject i. By Definition 1, we have that $Y_i \sim \mathrm{SMN}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$, for $i = 1, \ldots, n$. We call it the SMN regression (SMN-R) model.

We are interested in the case where left-censored observations can occur. That is, the observations are of the form

$$Y_{\mathrm{obs}_i} = \begin{cases} \kappa_i & \text{if} \quad Y_i \leq \kappa_i; \\ Y_i & \text{if} \quad Y_i > \kappa_i, \end{cases} \tag{12}$$

$i = 1, \ldots, n$, for some threshold point κ_i. This is called the SMN-CR model. For convenience, we have chosen to work with the left censored case, but the results are easily extensible to other censoring types. If we make $\kappa_i = 0$ and assume that $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, which corresponds to $U_i = 1$ in Definition 1, $i = 1, \ldots, n$, we obtain the Tobit censored response model studied by Barros et al. (2010). In addition, if $U_i \sim \mathit{Gamma}(\nu/2, \nu/2)$ we obtain the Student-t censored regression model developed by Arellano-Valle et al. (2012) and Massuia et al. (2012).

Let $\boldsymbol{\theta} = (\beta^\top, \sigma^2, \boldsymbol{\nu})^\top$ be the vector with all parameters in the SMN-CR model. Supposing that there are (possibly) m censored values of the characteristic of interest, we can partition the observed sample $\mathbf{y}_{\mathrm{obs}}$ in two subsamples of m censored and $n - m$ uncensored values, such that $\mathbf{y}_{\mathrm{obs}} = \{\kappa_1, \ldots, \kappa_m, y_{m+1}, \ldots, y_n\}$. Then, the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}}) = \sum_{i=1}^{m} \log\left[\mathcal{F}_{SMN}\left(\frac{\kappa_i - \mathbf{x}_i^\top\boldsymbol{\beta}}{\sigma}\right)\right] + \sum_{i=m+1}^{n} \log\left[f_{SMN}(y_i|\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})\right]. \tag{13}$$

To estimate the parameters of the SMN-CR model, an alternative is to maximize this log-likelihood function directly, a procedure that can be quite cumbersome. Alternatively, the standard algorithm in this case is the so-called EM algorithm of Dempster et al. (1977) or some extension like the ECM (Meng and Rubin, 1993) or the ECME algorithms (Liu and Rubin, 1994). Our choice is to use the ECME algorithm, a classical, reliable, widespread tool to obtain maximum likelihood estimates.

3.2 Parameter estimation via an EM-type algorithm for the SMN-CR Model

In this section we develop an EM-type algorithm for maximum likelihood estimation of the parameters in the SMN-CR model. In order to do this, we need a representation of the model in terms of missing data. First, note that using Definition 1, we have the following hierarchical representation:

$$Y_i|U_i = u_i \sim \mathrm{N}\left(\mathbf{x}_i^\top \boldsymbol{\beta}, u_i^{-1}\sigma^2\right); \quad U_i \sim H(\cdot|\boldsymbol{\nu}). \tag{14}$$

If the observation i is censored, we can consider y_i as a realization of the latent unobservable variable $Y_i \sim \mathrm{SMN}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$, $i = 1, \ldots, m$. The key to the development of our EM-type algorithm is to consider the complete-data $\mathbf{Z} = \{\mathbf{y}_{\mathrm{obs}}, y_1, \ldots, y_m, u_1, \ldots, u_n\}$, that is, we treat the problem as if the missing data $\mathbf{y}_L = \{y_1, \ldots, y_m\}$ and $\mathbf{u} = \{u_1, \ldots, u_n\}$ were in fact observed. Then, using representation (14), we obtain the complete-data log-likelihood, given by

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{Z}) = &-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \frac{1}{2}\sum_{i=1}^{n}\log u_i \\ &- \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \sum_{i=1}^{n}\log h(u_i|\boldsymbol{\nu}),\end{aligned} \tag{15}$$

where $h(\cdot|\boldsymbol{\nu})$ is the density of the random variable U.

In what follows the superscript (k) indicates the estimate of the related parameter at stage k of the algorithm. In the E-step of the algorithm, we must obtain the so-called Q-function,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[\ell_c\left(\boldsymbol{\theta}|\mathbf{Z}\right)|\mathbf{y}_{\mathrm{obs}}\right],$$

where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}$ means that the expectation is being effected using $\boldsymbol{\theta}^{(k)}$ for $\boldsymbol{\theta}$. Observe that the expression of the Q-function is completely determined by the knowledge of the following expectations

$$\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s | y_{\mathrm{obs}_i}], \quad s = 0, 1, 2,$$

as well as

$$\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i | y_{\mathrm{obs}_i}] \quad \text{and} \quad \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\mathrm{obs}_i}].$$

Thus, dropping unimportant constants, the Q-function can be written in a synthetic form as

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = &-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top \boldsymbol{\beta} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top \boldsymbol{\beta})^2\right] \\ &+ \frac{1}{2}\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i|y_{\mathrm{obs}_i}] + \sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\mathrm{obs}_i}].\end{aligned} \tag{16}$$

At each step, the conditional expectations $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)})$ can be easily derived from the results given in Proposition 1. Thus, *for an uncensored observation* i, we have that $Y_{\text{obs}_i} = Y_i \sim \text{SMN}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$ and, therefore,

$$\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)}) = y_i^s \text{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i], \tag{17}$$

where $\text{E}_{\boldsymbol{\theta}^{(k)}[U_i|y_i]}$ can be obtained using results in Osorio et al. (2007). Thus, for example,

- If $Y_i \sim PVII(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \nu, \delta)$, we have

$$\text{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i] = \frac{(\nu+1)}{\delta + d^2(\boldsymbol{\theta}^{(k)}, y_i)};$$

- If $Y_i \sim SL(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \nu)$, we have

$$\text{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i] = \frac{\Gamma\left(\nu + 1.5, d^2(\boldsymbol{\theta}^{(k)}, y_i)/2\right)}{\Gamma\left(\nu + 0.5, d^2(\boldsymbol{\theta}^{(k)}, y_i)/2\right)};$$

- If $Y_i \sim CN(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \nu, \gamma)$, we have

$$\text{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i] = \frac{1 - \nu + \nu\gamma^{1.5} e^{0.5(1-\gamma)d^2(\boldsymbol{\theta}^{(k)}, y_i)}}{1 - \nu + \nu\gamma^{0.5} e^{0.5(1-\gamma)d^2(\boldsymbol{\theta}^{(k)}, y_i)}},$$

where $d(\boldsymbol{\theta}^{(k)}, y_i) = \left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}^{(k)}\right)/\sigma^{(k)}$.

For a censored observation i, we have $Y_i \le \kappa_i$, so that

$$\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)}) = \text{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s | Y_i \le \kappa_i], \tag{18}$$

which can be obtained for the different distributions using the results given in Proposition 1, along with the results given in equations (6) and (7) with $r = 1$.

When the M-step turns out to be analytically intractable, it can be replaced with a sequence of condicional maximization (CM) steps. The resulting procedure in known as *ECM algorithm* (Meng and Rubin, 1993). The *ECME algorithm* (Liu and Rubin, 1994), a faster extension of EM and ECM algorithm, is obtained by maximizing the constrained Q-function with some CM-steps that maximize the corresponding constrained actual marginal likelihood function, called *CML-steps*. Therefore, our EM-type algorithm (ECME) for the SMN-CR models can be summarized in the following way (see Appendix C for details):

E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, for $i = 1, \ldots, n$;

- If observation i is uncensored then, for $s = 0, 1, 2$, compute $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)})$ given in (17);

- If observation i is censored then, for $s = 0, 1, 2$, compute $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)})$ in (18).

CM-step: Update $\boldsymbol{\theta}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions,

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{n} \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)}); \tag{19}$$

$$\sigma^{2(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)})^2\right]. \tag{20}$$

CML-step: Update $\nu^{(k)}$ by maximizing the actual marginal log-likelihood function, obtaining

$$\begin{aligned}\boldsymbol{\nu}^{(k+1)} = \text{argmax}_{\nu}\Bigg\{&\sum_{i=1}^{m}\log\left[\mathcal{F}_{SMN}\left(\frac{\kappa_i - \mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)}}{\sigma^{(k+1)}}\right)\right] \\ &+ \sum_{i=m+1}^{n}\log\left[f_{SMN}(y_i|\mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \boldsymbol{\nu})\right]\Bigg\}.\end{aligned} \tag{21}$$

This process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_{obs})$, like

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

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 $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i|y_{\text{obs}_i}]$ and $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\text{obs}_i}]$.

4 Approximated standard errors for the fixed effects

Hereafter we denote the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})^\top$. In order to assess its variability, its variance-covariance matrix, estimated at convergence, is adjusted for the censored information using Louis' formula (Louis, 1982), see also Vaida et al. (2007, Sec. 2), Vaida and Liu (2009, Sec.2) and Matos et al. (2013, Sec.3). The log-likelihood formed from the single complete observation $\mathbf{Z}_i = (y_{\text{obs}_i}, y_i, u_i)^\top$ will be denoted by $\ell_c(\boldsymbol{\theta}|\mathbf{Z}_i)$. In what follows, $\text{Var}_{\boldsymbol{\theta}}$ means that the variance-covariance matrix is being effected using $\boldsymbol{\theta}$ as the real value of the parameter. The estimate of $\mathrm{Var}_{\boldsymbol{\theta}}[\widehat{\boldsymbol{\beta}}]$ is given by the inverse of the matrix

$$-\sum_{i=1}^{n}\left\{\mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2\ell(\boldsymbol{\theta}|y_{obs_i})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]-\mathrm{Var}_{\boldsymbol{\theta}}\left[\frac{\partial\ell_c(\boldsymbol{\theta}|\mathbf{Z}_i)}{\partial\boldsymbol{\beta}}|y_{obs_i}\right]\right\} \tag{22}$$

evaluated at $\widehat{\boldsymbol{\theta}}$. As the SMN distributions are also elliptical distributions, with

$$g(z)=\int_0^\infty \sqrt{u}\,\exp\{-(u/2)z\}dH(u),$$

see Section 2, it is possible to prove that the information matrix under the SMN-R model (11) (that is, the uncensored model) is block diagonal, with the block corresponding to $\boldsymbol{\beta}$ given by

$$-\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2\ell(\boldsymbol{\theta}|y_{obs_i})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]=\frac{4}{\sigma^2}\sum_{i=1}^{n}\left(\mathbf{x}_i\mathbf{x}_i^\top\right)d_{gi},$$

where $d_{gi}=\mathrm{E}[Z_i^2W_g^2(Z_i^2)]$, $Z_i\sim\mathrm{SMN}(0,1,\boldsymbol{\nu})$ and $W_g(z)=g'(z)/g(z)$, where $g'(z)$ is the first derivative of $g(\cdot)$ at z. For some distributions in the SMN family there are closed-form expressions for $q\,(z)=-2W_g\,(z)$:

- Normal distribution, $q\,(z)=1$;
- Student-t distribution, $q\,(z)=(\nu+1)\,/\,(\nu+z)$;
- Slash distribution, $q\,(z)=\frac{\Gamma(\nu+1.5,z/2)}{\Gamma(\nu+0.5,z/2)}$;
- Contaminated normal distribution,

$$q\,(z)=\frac{1-\nu+\nu\gamma^{\frac{3}{2}}\exp\{(1/2)\,(1-\gamma)\,z\}}{1-\nu+\nu\gamma^{\frac{1}{2}}\exp\{(1/2)\,(1-\gamma)\,z\}}.$$

For more details, see Osorio et al. (2007, Sec 2). It is possible to show that, in the case of the Student-t distribution, we have $d_{gi}=(\nu+1)\,/4\,(\nu+3)$ (Lange et al., 1989). For the slash and the contaminated normal (and other elliptical distributions), the computation of d_{gi} involves complex integrals, which can be solved using Monte Carlo techniques.

The information adjusted for censoring is given by

$$
\begin{aligned}
\sum_{i=1}^{m} \mathrm{Var}_{\boldsymbol{\theta}}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Z}_i)}{\partial \boldsymbol{\beta}}|Y_i \le \kappa_i\right] &= \sum_{i=1}^{m} \mathrm{Var}_{\boldsymbol{\theta}}\left[\frac{1}{\sigma^2}\mathbf{x}_i(U_iY_i - U_i\mathbf{x}_i^\top\boldsymbol{\beta})|Y_i \le \kappa_i\right] \\
&= \frac{1}{\sigma^4}\sum_{i=1}^{m}\left(\mathbf{x}_i\mathbf{x}_i^\top\right)\left[\mathrm{Var}_{\boldsymbol{\theta}}\left\{U_iY_i - U_i\mathbf{x}_i^\top\boldsymbol{\beta}|Y_i \le \kappa_i\right\}\right] \\
&= \frac{1}{\sigma^4}\sum_{i=1}^{m}\left(\mathbf{x}_i\mathbf{x}_i^\top\right)\left\{\mathrm{Var}_{\boldsymbol{\theta}}\left[U_iY_i^\top|Y_i \le \kappa_i\right]\right. \\
&+ \mathrm{Var}_{\boldsymbol{\beta}}\left[U_i\mathbf{x}_i^\top\boldsymbol{\beta}|Y_i \le \kappa_i\right] - 2\mathrm{Cov}_{\boldsymbol{\theta}}\left[U_iY_i, U_iY_i\mathbf{x}_i^\top\boldsymbol{\beta}|Y_i \le \kappa_i\right]\left.\right\},
\end{aligned}
$$

where Cov denotes covariance. These expressions can be obtained, for different SMN distributions, using Corollaries 2 and 3 in Appendix A.

5 Simulation studies

5.1 Robustness of the EM estimates (Simulation Study 1)

The goal of this section 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 cases normal, Student-t, contaminated normal and slash, and denote them by N-CR, T-CR, CN-CR and SL-CR, respectively. The computational procedures were implemented using the `R` software (R Core Team, 2013).

We performed a simulation study based on the N-CR model. Specifically, we considered model (11) with $\mathbf{X}_i^\top = (1, x_i)$ and $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, $i = 1, \ldots, n$. We generated 1000 artificial samples of size $n = 100$, considering $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (1, 4), \sigma^2 = 2$ and fixing the censoring level at $p = 8\%$ (that is, 8% of the observations in each data set were censored). The values x_i, $i = 1, \ldots, n$, were generated independently from a uniform distribution on the interval $(2, 20)$. These values were fixed throughout the simulations.

To assess how much the EM estimates are influenced by the presence of outliers, we replaced the observation y_{50} by $y_{50}(\vartheta) = y_{50} - \vartheta$, with $\vartheta = 1, 2, \ldots, 10$. Let $\widehat{\beta}_i(\vartheta)$ be the EM estimate of β_i after this contamination, $i = 1, 2$. We are particularly interested in the relative changes

$$RC(\beta_i) = |(\widehat{\beta}_i(\vartheta) - \widehat{\beta}_i)/\widehat{\beta}_i|.$$

We define the relative changes for σ^2 analogously.

Table 1: Simulation study 1. Average relative changes on estimates for different contaminations ϑ.

	SMN-CR models	ϑ									
		1	2	3	4	5	6	7	8	9	10
	Normal	0.0689	0.1335	0.1937	0.2495	0.3011	0.3485	0.3919	0.4314	0.4672	0.4994
	T($\nu = 3$)	0.0271	0.0460	0.0540	0.0543	0.0517	0.0488	0.0465	0.0447	0.0434	0.0425
	T($\nu = 6$)	0.0605	0.1039	0.1270	0.1341	0.1321	0.1262	0.1194	0.1129	0.1074	0.1029
	T($\nu = 8$)	0.0260	0.0482	0.0632	0.0709	0.0732	0.0725	0.0704	0.0678	0.0651	0.0625
$RC(\beta_1)$	T($\nu = 10$)	0.0367	0.0684	0.0913	0.1048	0.1109	0.1120	0.1103	0.1073	0.1037	0.1000
	T($\nu = 12$)	0.0078	0.0160	0.0237	0.0299	0.0345	0.0377	0.0398	0.0411	0.0419	0.0423
	slash($\nu = 3$)	0.0066	0.0140	0.0208	0.0255	0.0277	0.0278	0.0269	0.0257	0.0246	0.0237
	slash($\nu = 4$)	0.0273	0.0504	0.0674	0.0769	0.0786	0.0743	0.0669	0.0592	0.0529	0.0482
	CN($\nu = (0.3, 0.3)$)	0.0574	0.0822	0.0931	0.1132	0.1382	0.1671	0.1994	0.2337	0.2691	0.3054
	Normal	0.0008	0.0017	0.0025	0.0033	0.0040	0.0047	0.0054	0.0060	0.0067	0.0072
	T($\nu = 3$)	0.0005	0.0009	0.0011	0.0011	0.0010	0.0010	0.0009	0.0009	0.0009	0.0009
	T($\nu = 6$)	0.0003	0.0006	0.0008	0.0009	0.0009	0.0008	0.0008	0.0008	0.0007	0.0007
	T($\nu = 8$)	0.0001	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
$RC(\beta_2)$	T($\nu = 10$)	0.0002	0.0005	0.0007	0.0008	0.0009	0.0009	0.0009	0.0009	0.0008	0.0008
	T($\nu = 12$)	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
	slash($\nu = 3$)	0.0001	0.0002	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
	slash($\nu = 4$)	0.0005	0.0009	0.0013	0.0015	0.0016	0.0015	0.0014	0.0012	0.0011	0.0010
	CN($\nu = (0.3, 0.3)$)	0.0001	0.0003	0.0004	0.0004	0.0005	0.0006	0.0007	0.0008	0.0010	0.0011
	Normal	0.0129	0.0300	0.0550	0.0904	0.1373	0.1961	0.2668	0.3488	0.4419	0.5461
	T($\nu = 3$)	0.0178	0.0310	0.0397	0.0455	0.0498	0.0531	0.0554	0.0570	0.0582	0.0590
	T($\nu = 6$)	0.0160	0.0317	0.0457	0.0574	0.0670	0.0747	0.0807	0.0852	0.0885	0.0909
	T($\nu = 8$)	0.0155	0.0312	0.0463	0.0607	0.0735	0.0844	0.0932	0.1001	0.1054	0.1095
$RC(\sigma^2)$	T($\nu = 10$)	0.0152	0.0309	0.0475	0.0644	0.0801	0.0938	0.1053	0.1146	0.1219	0.1278
	T($\nu = 12$)	0.0149	0.0318	0.0497	0.0674	0.0846	0.1005	0.1143	0.1260	0.1356	0.1433
	slash($\nu = 3$)	0.0142	0.0306	0.0487	0.0648	0.0759	0.0815	0.0834	0.0839	0.0840	0.0839
	slash($\nu = 4$)	0.0142	0.0309	0.0502	0.0698	0.0861	0.0974	0.1034	0.1057	0.1064	0.1064
	CN($\nu = (0.3, 0.3)$)	0.0159	0.0309	0.0453	0.0625	0.0870	0.1206	0.1631	0.2132	0.2696	0.3318

For each replication we obtained the parameter estimates with and without outliers, under the following models: N-CR, T-CR with $\nu \in \{3, 6, 8, 10, 12\}$, SL-CR with $\nu \in \{3, 4\}$ and CN-CR with $\boldsymbol{\nu}^\top = (\xi, \gamma) = (0.3, 0.3)$. Table 1 and Figure 1 depict the average values of the relative changes across all samples. In the N-CR case, we observe that influence increases dramatically when ϑ increases. However, for the SMN-CR models with heavy tails, as the T-CR and the SL-CR with different values of ν, these measures vary little, which indicates that they are more robust than the N-CR model in the presence of discrepant observations. For the CN-CR model we can observe that, specially for the parameter σ^2, the relative change increases as ϑ increases.

5.2 Asymptotic properties (Simulation Study 2)

We also conducted a simulation study to evaluate the finite-sample performance of the parameter estimates. We generated artificial samples from the SMN-CR model (11), with $\mathbf{X}_i^\top = (1, x_i)$, $i = 1, \dots, n$.

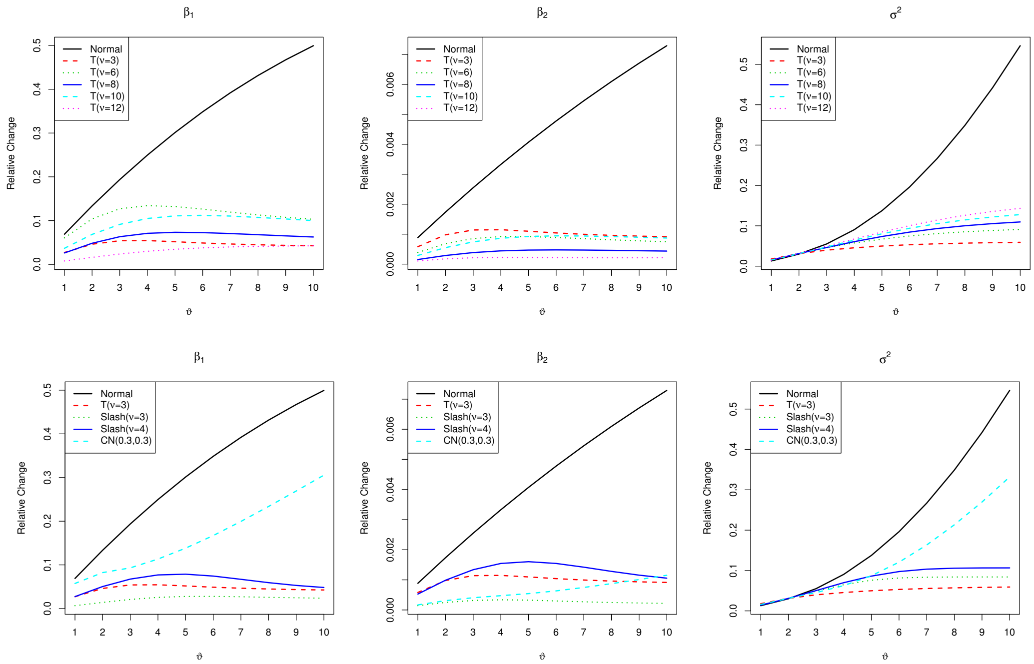

Figure 1: Simulation study 1. Average relative changes on estimates for different contaminations ϑ.

We considered the censoring levels $p = 10\%$, 25% and 45%. The sample sizes were fixed at $n = 50, 100, 150, 200, 300, 400, 500, 700$ and 800. The true values of the regression parameters were taken as $\beta_1 = 1.5,\ \beta_2 = 4$ and $\sigma^2 = 0.5$. As considered in Labra et al. (2012), the variable x_i ranges from 0.1 to 20 and these values were maintained throughout the experiment. For each combination of parameters, sample sizes and censoring levels, were generated 1000 samples from the SMN-CR model, under four different situations: N-CR, T-CR ($\nu = 3$), SL-CR ($\nu = 4$) and CN-CR $\big(\boldsymbol{\nu}^\top = (0.5, 0.5)\big)$.

In order to analyze the performance of the estimates obtained using our proposed EM-type algorithm, we computed, for each combination of sample size, level of censoring and parameter value, the bias and the mean squared error (MSE). For β_i, they are given, respectively, by

$$\text{Bias}\,(\beta_i) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\beta}_i^{(j)} - \beta_i\right);$$

$$\text{MSE}\,(\beta_i) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\beta}_i^{(j)} - \beta_i\right)^2,$$

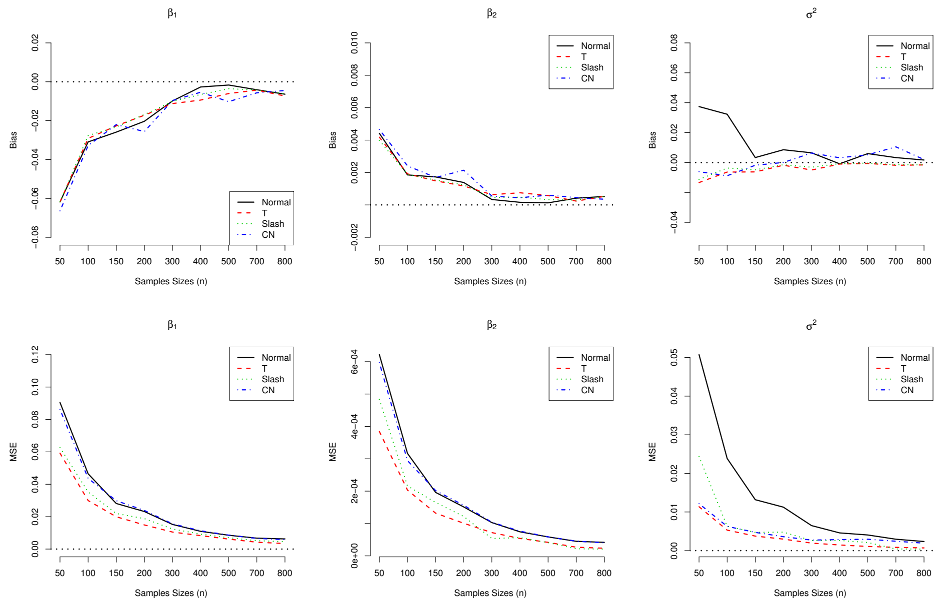

Figure 2: Simulation study 2. Average bias (first column) and MSE (second column) of parameter estimates in the SMN-CR models for $p = 10\%$.

where $\widehat{\beta}_i^{(j)}$ is the estimate of β_i for the j-th sample. We define bias and MSE for σ^2 in the same manner. The result considering $p = 10\%$ is shown in Figure 2. We can see a pattern of convergence to zero of the bias and MSE when n increases. As a general rule, we can say that bias and MSE tend to approach to zero when the sample size increases indicating that the estimates based on the proposed EM-type algorithm do provide good asymptotic properties. This same pattern of convergence to zero is repeated considering different levels of censoring p (see Appendix D for details).

5.3 Consistency of the estimates of the standard errors for the fixed effects (Simulation Study 3)

Now we show, via simulation study, that the method suggested in Section 4 to approximate the standard errors (SE) of the MLE of the regression parameters has good asymptotic properties. We fixed a SMN-CR model (N-CR, T-CR or SL-CR), a censoring level (5%, 10%, 20% or 50%). For each one of these twelve combinations of model and censoring level, we generated 1000 samples of size $n = 100$ with $\beta_1 = 2$, $\beta_2 = 1$ and $\sigma^2 = 1$. For each sample, we obtained the MLE's of β_1 and β_2, the

Table 2: Simulation study 3. MC SE of $\widehat{\beta}_i$, average values (across 1000 samples) of the SE computed using the information method (MC IM SE) and the percentual coverage of the resulting 95% confidence intervals (COV MC).

Censoring level	Measure	N-CR $\widehat{\beta}_1$	N-CR $\widehat{\beta}_2$	T-CR $\widehat{\beta}_1$	T-CR $\widehat{\beta}_2$	SL-CR $\widehat{\beta}_1$	SL-CR $\widehat{\beta}_2$
5%	MC SE	0.3445	0.1758	0.4044	0.2050	0.4041	0.2063
	MC IM SE	0.3650	0.1892	0.4354	0.2253	0.4279	0.2246
	COV MC	96.1%	96.6%	96.6%	97.6%	95.4%	95.6 %
10%	MC SE	0.3711	0.1884	0.4258	0.2116	0.4047	0.2034
	MC IM SE	0.3693	0.1899	0.4553	0.2358	0.4441	0.2325
	COV MC	93.0%	93.7%	96.2%	96.7%	96.2 %	97.1%
20%	MC SE	0.3505	0.1804	0.4056	0.2053	0.4107	0.2099
	MC IM SE	0.3707	0.1848	0.4577	0.2290	0.4497	0.2252
	COV MC	95.1%	94.9%	97.0%	96.9%	96.7 %	96.7%
50%	MC SE	0.3856	0.2114	0.4637	0.2589	0.4382	0.2361
	MC IM SE	0.3629	0.1748	0.4415	0.2089	0.4440	0.2121
	COV MC	93.3%	89.8%	93.2%	86.9%	95.3%	91.2%

estimates of their standard errors using the technique proposed in Section 4 and an approximate 95% confidence interval assuming asymptotic normality. Table 2 presents the sample standard errors of $\widehat{\beta}_i$, that is, the value

$$\text{MC SE} = \frac{1}{999}\left[\sum_{i=1}^{1000}\left(\widehat{\beta}_i\right)^2 - \frac{1}{1000}\left(\sum_{i=1}^{1000}\widehat{\beta}_i\right)^2\right],$$

average values (across 1000 samples) of the standard errors computed using the information method (IM MC SE) and the percentual coverage of the resulting 95% confidence intervals (COV MC). The results suggest that the approximation produced by the information method is satisfactory, and it does not depend on the censoring level or the chosen model.

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 data set consists of 753 married white women with ages between 30 and 60 years old in 1975, with 428 women that 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. It is important to stress that 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. Four predictor variables were considered: the wife's age, years of schooling, the number of children younger than six years old in the household and the number of children between six and nineteen years old. These data were analyzed by Arellano-Valle et al. (2012) using the T-CR model. We analyzed it with the aim of providing additional inferences by using the SMN distributions. We fitted a regression model with an intercept parameter β_1 and applied the EM-type algorithm for censored data explained in Section 3.2, considering again the N-CR, T-CR, SL-CR and CN-CR models for comparative purposes.

Table 3: Real data. EM estimates for fitting various SMN-CR models. SE are the estimated asymptotic standard errors.

	N-CR		T-CR		SL-CR		CN-CR	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-2.75101	1.732603	-1.04708	1.583697	-1.43588	1.498260	-1.29006	1.468302
β_2	-0.10455	0.027440	-0.11075	0.025402	-0.10717	0.023824	-0.10643	0.023358
β_3	0.72807	0.082547	0.64750	0.077664	0.65449	0.072178	0.64676	0.070814
β_4	-0.21426	0.152665	-0.29638	0.141302	-0.28434	0.132650	-0.29971	0.129894
β_5	-3.02635	0.434533	-3.16370	0.480036	-3.05183	0.399731	-3.06493	0.391158
σ^2	20.94012	-	10.63792	-	8.65565	-	11.169	-
ν	-	-	4.2000	-	2.1000	-	-	-
γ	-	-	-	-	-	-	0.1000	-
ξ	-	-	-	-	-	-	0.1000	-

Table 3 shows the parameter estimates, together with their corresponding standard errors (SE). Table 4 presents some model selection criteria, together with the values of the log-likelihood. The AIC (Akaike, 1974), BIC (Schwarz, 1978) and EDC (Bai et al., 1989) criteria indicate that the three models with longer tails than the N-CR model seem to produce more accurate estimates. The standard errors of the T-CR, SL-CR and CN-CR models are smaller than that of the N-CR model.

In order to identify atypical observations and/or model misspecification, we analyzed the transformation of the martingale residual, r_{MT_i}, proposed by Barros et al. (2010). These residuals are defined by

$$r_{MT_i} = \text{sign}(r_{M_i})\sqrt{-2\left[r_{M_i} + \delta_i \log\left(\delta_i - r_{M_i}\right)\right]},$$

$i = 1, \ldots, n$, where $r_{M_i} = \delta_i + \log S(y_i, \widehat{\boldsymbol{\theta}})$ is the martingale residual proposed by Ortega et al. (2003) – see more details in Therneau et al. (1990), with $\delta_i = 0, 1$ indicating whether the i-th observation is censored or not, respectively, $\text{sign}(r_{M_i})$ denoting the sign of r_{M_i} and $S(y_i, \widehat{\boldsymbol{\theta}}) = P_{\widehat{\boldsymbol{\theta}}}(Y_i > y_i)$ representing the survival function evaluated at y_i, supposing that it is being effected using the EM estimate $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$.

Table 4: Real data. Values of some model selection criteria.

Criterion	N-CR	T- CR	SL-CR	CN-CR
log-likelihood	-1481.6550	-1440.1450	-1439.5370	-1432.0850
AIC	2975.3110	2894.2910	2893.0750	2880.1710
BIC	3003.0550	2926.6590	2925.4430	2917.1630
EDC	2996.2400	2918.7080	2925.4430	2908.0760

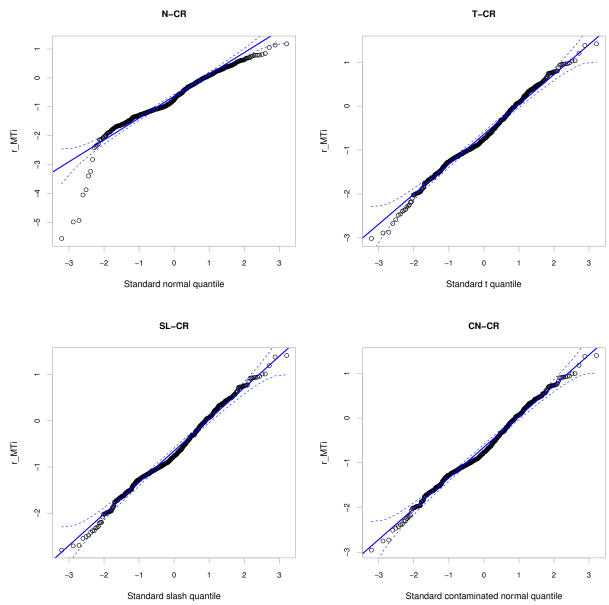

Figure 3: Real data. Envelopes of the martingale-type residuals, r_{MT_i}, for the SMN-CR models.

The plots of r_{MT_i} with generated confidence envelopes are presented in Figure 3. From this figure, we can see clearly that the SMN-CR models with heavy tails fit better the data than the N-CR model, since, in that cases, there are fewer observations which lie outside the envelopes.

The robustness of the three models with longer tails than the N-CR model can be assessed by considering the influence of a single outlying observation on the EM estimate of $\boldsymbol{\theta}$. In particular, we can assess how much the EM estimate of $\boldsymbol{\theta}$ is influenced by a change of ∇ units in a single observation y_i. Replacing y_i by $y_i(\nabla) = y_i + \nabla$, we define $\widehat{\beta}_i(\nabla)$ as the EM estimate of β_i after contamination, $i = 1, \ldots, 5$, and analyze the behavior of the relative changes, as we did in Section 5.1. In this study we contaminated the observations y_{750} (censored) and y_7 (uncensored), considering $\nabla \in \{0, 1, \ldots, 10\}$.

Figure 4 displays the results of the relatives changes of the estimates for different

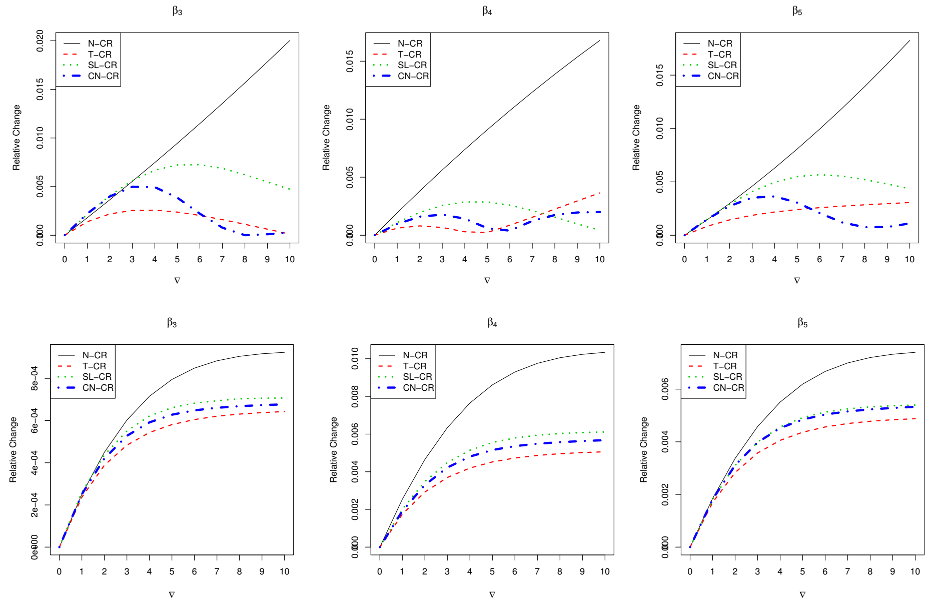

Figure 4: Real data. Relative changes on EM estimates from the SMN-CR models for different contaminations ∇ of the uncensored observation y_7 (first line) and the censored observation y_{750} (second line).

values of ∇. We omitted the plot concerning β_2 because the relative changes patterns are not so distinguishable in this case. As expected, the estimates from the models with longer tails than the N-CR model are less affected by variations on ∇, no matter if the observation is censored or not. Thus, it is clear that the SMN-CR models with heavy tails are more robust, providing more accurate estimates when the data have departures from normality.

7 Conclusions

We have proposed a robust approach to linear regression models with censored observations based on scale mixtures of normal distributions, called SMN-CR models. This offers a high degree of flexibility, allowing us to deal properly with censored data in the presence of outliers. A novel ECME algorithm to obtain approximated maximum likelihood estimates is developed using formulas for the moments of the truncated SMN distribution, leading to closed-form expressions for the E-step. We applied our methodology to real data set (freely downloadable from `R`) as well as to simulated data, in order to illustrate how the procedures can be used to evaluate model assumptions, identify outliers, and obtain robust parameter estimates. From these results, it is encouraging that the use of SMN-CR models with heavy tails offer a better fitting, a better protection against outliers and more precise inferences than the N-CR model.

Although the SMN-CR models considered here have shown great flexibility to model symmetric data, its robustness against outliers can be seriously affected by the presence of skewness. Recently, Lachos et al. (2010) proposed a remedy to accommodate skewness and heavy-tailedness simultaneously, using scale mixtures of skew-normal (SMSN) distributions. We conjecture that our methodology can be used under CR models, and should yield satisfactory results at the expense of additional complexity in implementation. An in-depth investigation of such extensions is beyond the scope of the present paper, but it is an interesting topic for further research. Finally, the proposed EM-type algorithm has been coded and implemented in the `R` package `SMNCensReg` (Garay et al., 2013), which is available for download at CRAN repository. A great advantage of this package is that all the censoring possibilities are taken into account: left, right and interval.

Acknowledgement

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 2011/17400-6 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). Celso R. B. Cabral was supported by CNPq (via the Universal and CT-Amazônia projects) and CAPES (via Project PROCAD 2007). The research of Aldo M. Garay is supported by Grant 161119/2012-3 from CNPq and Heleno Bolfarine was supported by CNPq.

Appendix A. Lemmas and corollaries

The following Lemmas, provided by Kim (2008) and Genç (2012), are useful for evaluating some integrals used in this paper as well as for the implementation of the proposed EM-type algorithm.

Lemma 1. *If* $Z \sim TN_{(a,b)}(0,1)$*, then*

$$(k+1)\,E\left[Z^k\right] - E\left[Z^{k+2}\right] = \frac{(b)^{k+1}\,\phi(b) - (a)^{k+1}\,\phi(a)}{\Phi(b) - \Phi(a)},$$

for $k = -1, 0, 1, 2, \ldots$

Proof. See Lemma 2.3 in Kim (2008). □

Lemma 2. *Let U be a positive random variable. Then $\mathcal{F}_{SMN}(a) = E_U\left[\Phi\left(aU^{\frac{1}{2}}\right)\right]$, where $\mathcal{F}_{SMN}(\cdot)$ denotes the cdf of a standard SMN random variable, that is, when $\mu = 0$ and $\sigma^2 = 1$.*

Proof. See Lemma 3 in Genç (2012). □

Lemma 3. *For $\nu > 0$,*

$$\int_0^u x^{v-1} e^{-\mu x} dx = \mu^{-v} \gamma^* (v, \mu u),$$

where $\gamma^*(a, x) = \int_0^x e^{-t} t^{a-1} dt$ is the incomplete gamma function.

Proof. See Lemma 6 in Genç (2012). □

The following Corollaries are a direct consequence of Proposition 1 given in Section 2. The proofs of Corollaries 2 and 3 follow direct from the definition of variance and the stochastic representation of a SMN random variable given in (1).

Corollary 1. *Let $Y \sim \mathrm{SMN}(\mu, \sigma^2, \boldsymbol{\nu})$ with scale factor U and $\mathcal{A} = (a, b)$. Then, for $r \geq 1$,*

$$\begin{aligned}
\mathrm{E}\left[U^r | Y \in \mathcal{A}\right] &= \mathrm{E}\left[U^r | X \in \mathcal{A}^*\right]; \\
\mathrm{E}\left[U^r Y | Y \in \mathcal{A}\right] &= \mu \mathrm{E}\left[U^r | X \in \mathcal{A}^*\right] + \sigma \mathrm{E}\left[U^r X | X \in \mathcal{A}^*\right]; \\
\mathrm{E}\left[U^r Y^2 | Y \in \mathcal{A}\right] &= \mu^2 \mathrm{E}\left[U^r | X \in \mathcal{A}^*\right] + 2\mu\sigma \mathrm{E}\left[U^r X | X \in \mathcal{A}^*\right] + \sigma^2 \mathrm{E}\left[U^r X^2 | X \in \mathcal{A}^*\right],
\end{aligned}$$

where $X \sim \mathrm{SMN}(0, 1, \boldsymbol{\nu})$ and $\mathcal{A}^* = (a^*, b^*)$, with $a^* = (a - \mu)/\sigma$ and $b^* = (b - \mu)/\sigma$.

Corollary 2. *Let $X \sim \mathrm{SMN}(0, 1, \boldsymbol{\nu})$ with scale factor U and $\mathcal{A} = (a, b)$. Then, for $r \geq 1$,*

$$\begin{aligned}
\mathrm{Var}\left[U^r | X \in \mathcal{A}\right] &= \tau(a, b)\left[\mathrm{E}_\Phi(2r, b) - \mathrm{E}_\Phi(2r, a) - \tau(a, b)\left\{\mathrm{E}_\Phi(r, b) - \mathrm{E}_\Phi(r, a)\right\}^2\right]; \\
\mathrm{Var}\left[U^r X | X \in \mathcal{A}\right] &= \tau(a, b)\left[\mathrm{E}_\Phi(2r - 1, b) - \mathrm{E}_\Phi(2r - 1, a)\right. \\
&\quad \left. + ab\mathrm{E}_\phi\left(2r - \frac{1}{2}, a\right)\mathrm{E}_\phi\left(2r - \frac{1}{2}, b\right) - \tau(a, b)\left\{\mathrm{E}_\Phi(r, b) - \mathrm{E}_\Phi(r, a)\right\}^2\right],
\end{aligned}$$

where $\tau(a, b)$ is given in (5).

Corollary 3. *Let $X \sim \text{SMN}(0, 1, \boldsymbol{\nu})$ with scale factor U, $\mathcal{A} = (a, b)$ and $Y = \mu + \sigma X$. Then, for $r \geq 1$,*

$$\begin{aligned}
\text{Var}\left[U^r | Y \in \mathcal{A}\right] &= \text{Var}\left[U^r | X \in \mathcal{A}^*\right]; \\
\text{Var}\left[U^r Y | Y \in \mathcal{A}\right] &= \mu^2 \text{Var}\left[U^r | X \in \mathcal{A}^*\right] + \sigma^2 \text{Var}\left[U^r X | X \in \mathcal{A}^*\right] \\
&\quad + 2\mu\sigma \text{Cov}\left[U^r, U^r X | X \in \mathcal{A}^*\right],
\end{aligned}$$

where $\mathcal{A}^* = (a^*, b^*)$, with $a^* = (a - \mu)/\sigma$ and $b^* = (b - \mu)/\sigma$.

Appendix B. Derivations of quantities $\boldsymbol{E}_\phi(r, h)$ and $\boldsymbol{E}_\Phi(r, h)$ for SMN distributions

In this Appendix, we calculate the expressions for the expected values $\text{E}_\phi(r, h)$ and $\text{E}_\Phi(r, h)$ given in Proposition 1.

Pearson type VII distribution (and the Student-t distribution)

In this case $U \sim Gamma(\nu/2, \delta/2)$, with $\nu > 0$ and $\delta > 0$. To facilitate the notation, let us make $\alpha_1 = (\nu + 2r)/2$ and $\alpha_2 = (h^2 + \delta)/2$. Then,

$$\begin{aligned}
\text{E}_\phi(r, h) &= \int_0^\infty \frac{\delta^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1} u^r}{\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \exp\left\{-\frac{u(h^2+\delta)}{2}\right\} du \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right) \delta^{\frac{\nu}{2}} \left(\frac{h^2+\delta}{2}\right)^{-\frac{\nu+2r}{2}}}{\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \int_0^\infty \frac{\alpha_2^{\alpha_1} u'^{\{\alpha_1 - 1\}}}{\Gamma(\alpha_1)} \exp\{-\alpha_2 u'\} du' \qquad (23) \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{\nu/2} \left(\frac{h^2+\delta}{2}\right)^{-\frac{\nu+2r}{2}},
\end{aligned}$$

where the integrand in (23) is the pdf of a random variable U' with distribution $Gamma\,(\alpha_1, \alpha_2)$.

$$
\begin{aligned}
\mathrm{E}_{\Phi}\,(r,h) &= \int_0^{\infty} \frac{u^{\frac{2r+\nu}{2}-1}\Phi\left(hu^{\frac{1}{2}}\right)\delta^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \exp\left\{-\frac{u\delta}{2}\right\} du \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{-r} \int_0^{\infty} \left(\frac{\delta}{2}\right)^{\alpha_1} \frac{\Phi\left(hu'^{\{\frac{1}{2}\}}\right) u'^{\{\alpha_1-1\}}}{\Gamma\left(\alpha_1\right)} \exp\left\{-\frac{u'\delta}{2}\right\} du' \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{-r} E_{U'}\left[\Phi\left(hU'^{\{\frac{1}{2}\}}\right)\right] \qquad (24) \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{-r} F_{PVII}(h|\nu+2r,\delta),
\end{aligned}
$$

where in (24) the expectation is computed with respect to $U' \sim Gamma\,(\alpha_1, \delta/2)$ and $F_{PVII}(\cdot)$ represents the cdf of the Pearson type VII distribution. Then, the result follows from Lemma 2. When $\delta = \nu$, i.e., the Student-t distribution, we have that $\mathrm{E}_{\phi}\,(r,h)$ and $\mathrm{E}_{\Phi}\,(r,h)$ are given by

$$
\begin{aligned}
\mathrm{E}_{\phi}\,(r,h) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \left(\frac{h^2+\nu}{2}\right)^{-\frac{(\nu+2r)}{2}}; \\
\mathrm{E}_{\Phi}\,(r,h) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{-r} F_{PVII}(h|\nu+2r,\nu).
\end{aligned}
$$

Slash distribution

In this case $U \sim Beta(\nu, 1)$, with positive shape parameter ν, and

$$
\begin{aligned}
E_{\phi}\,(r,h) &= \int_0^1 u^r \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{h^2}{2}u\right\} \nu u^{\nu-1} du \\
&= \frac{\nu}{\sqrt{2\pi}} \int_0^1 u^{\nu+r-1} \exp\left\{-\frac{h^2}{2}u\right\} du, \\
&= \frac{\nu}{\sqrt{2\pi}} \left(\frac{h^2}{2}\right)^{-(\nu+r)} \gamma^*\left(\nu+r, \frac{h^2}{2}\right), \qquad (25)
\end{aligned}
$$

where we used Lemma 3 to obtain equation (25).

$$
\begin{aligned}
E_{\Phi}\,(r,h) &= \int_0^1 u^r \Phi\left(hu^{\frac{1}{2}}\right) \nu u^{\nu-1} du \\
&= \frac{\nu}{\nu+r} \int_0^1 \Phi\left(hu'^{\{\frac{1}{2}\}}\right) u'^{\{\nu+r-1\}} (\nu+r)\, du' \qquad (26) \\
&= \frac{\nu}{\nu+r} F_{SL}(h|\nu+r), \qquad (27)
\end{aligned}
$$

where the integrand in (26) is the expectation of the random variable $\Phi(hU'^{\{\frac{1}{2}\}})$, with $U' \sim$ Beta$(\nu + r, 1)$. Using Lemma 2, we obtain equation (27), where $F_{SL}(\cdot)$ is the cdf of the slash distribution.

Contaminated normal distribution

$$
\begin{aligned}
\mathrm{E}_{\phi}\left(r, h\right) &= u^r \phi\left(hu^{\frac{1}{2}}\right)\left[\nu \mathbb{I}_{\{\gamma\}}(u) + (1-\nu)\mathbb{I}_{\{1\}}(u)\right] \\
&= \nu\gamma^r \phi\left(h\gamma^{\frac{1}{2}}\right) + (1-\nu)\phi\left(h\gamma^{\frac{1}{2}}\right); \\
\mathrm{E}_{\Phi}\left(r, h\right) &= u^r \Phi\left(hu^{\frac{1}{2}}\right)\left[\nu \mathbb{I}_{\{\gamma\}}(u) + (1-\nu)\mathbb{I}_{\{1\}}(u)\right] \\
&= \nu\gamma^r \Phi\left(h\gamma^{\frac{1}{2}}\right) + (1-\nu)\Phi\left(h\right) \\
&= \gamma^r\left[\nu\Phi\left(hu^{\frac{1}{2}}\right) + (1-\nu)\,\Phi\left(h\right)\right] + (1-\nu)\left(1-\gamma^r\right)\Phi\left(h\right) \\
&= \gamma^r F_{CN}(h|\nu,\gamma) + (1-\nu)\left(1-\gamma^r\right)\Phi\left(h\right),
\end{aligned}
$$

where $F_{CN}(\cdot)$ is the cdf of the contaminated normal distribution.

Appendix C. Details of the EM-type algorithm

In this Appendix, we derive the EM algorithm equations (19)–(21). Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\nu})$ be the vector with all parameters in the SMN-CR model and consider the notation given in Section 3.2. Denoting the complete-data likelihood by $L(\cdot|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u})$ and pdf's in general by $f(\cdot)$, we have that

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = f(\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) &= f(\mathbf{y}_{\text{obs}}, \mathbf{y}_L|\mathbf{u})f(\mathbf{u}) \\
&= f(\mathbf{y}|\mathbf{u})f(\mathbf{u}) = \prod_{i=1}^{n} f(y_i|u_i)h(u_i|\boldsymbol{\nu}).
\end{aligned}
$$

Dropping unimportant constants, the complete-data log-likelihood function is given by

$$
\begin{aligned}
&\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = \log(L(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u})) \\
&\qquad = \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\log u_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2 + \sum_{i=1}^{n}\log h(u_i|\boldsymbol{\nu}).
\end{aligned}
$$

The Q-function at the E-step of the algorithm is given by

$$
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[\ell_c\left(\boldsymbol{\theta}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_L, \mathbf{U}\right)|\mathbf{y}_{\text{obs}}\right],
$$

so we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \quad &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_iY_i^2|y_{\mathrm{obs}_i}]\right.\\
&\left.-2\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_iY_i|y_{\mathrm{obs}_i}]\mathbf{x}_i^\top\boldsymbol{\beta} + \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_{\mathrm{obs}_i}](\mathbf{x}_i^\top\boldsymbol{\beta})^2\right\}\\
&+\frac{1}{2}\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i|y_{\mathrm{obs}_i}] + \sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\mathrm{obs}_i}].
\end{aligned}
$$

The expectations $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_iY_i^s|y_{\mathrm{obs}_i}]$, $s = 0, 1, 2$, used in the E-step of the algorithm, are computed considering the two possible cases: (i) when the observation i is uncensored and (ii) otherwise. In the former case we solve the problem using results obtained by Osorio et al. (2007). In the later case we use Proposition 1. Then, we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \quad &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top\boldsymbol{\beta})^2\right]\\
&+\frac{1}{2}\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i|y_{\mathrm{obs}_i}] + \sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\mathrm{obs}_i}].
\end{aligned}
$$

In the CM-step, we take the derivatives of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ with respect to $\boldsymbol{\beta}$ and σ^2, i.e.,

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\mathbf{x}_i\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)}) - \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}\right];\\
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top\boldsymbol{\beta})^2\right].
\end{aligned}
$$

The solution of $\dfrac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\beta}} = 0$ is

$$
\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{n}\mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)}).
$$

The solution of $\dfrac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\sigma^2} = 0$ is

$$
\sigma^{2(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top\boldsymbol{\beta}^{(k+1)})^2\right].
$$

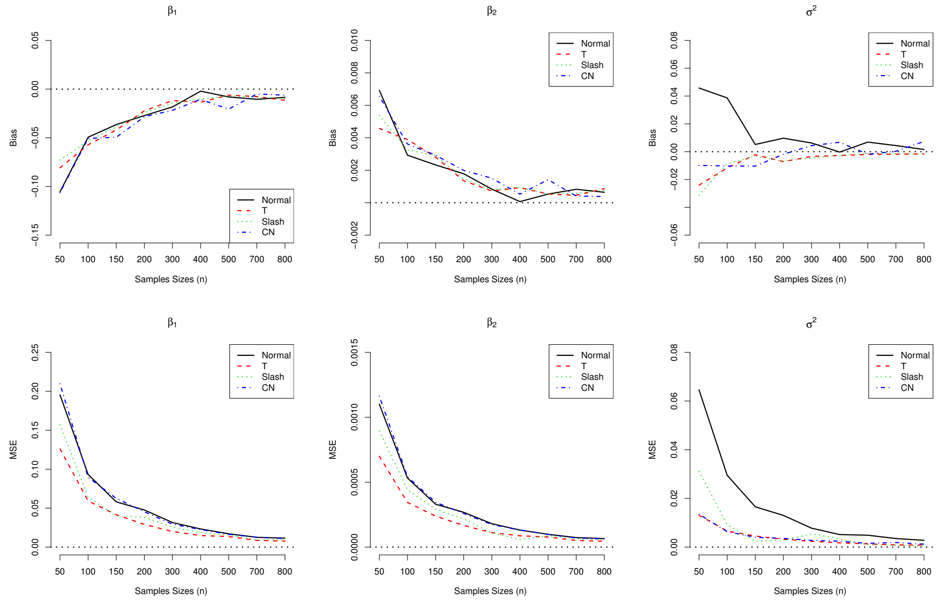

Figure 5: Simulation study 2. Average bias (first line) and average MSE (second line) of the estimates of β_1, β_2 and σ^2 from the SMN-CR models for level of censoring $p = 25\%$.

For the CML-step, we estimate $\boldsymbol{\nu}$ by maximizing the marginal log-likelihood, circumventing the (in general) complicated task of computing $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i | y_{\text{obs}_i}]$ and $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\{h(U_i|\boldsymbol{\nu})\}|y_{\text{obs}_i}]$, i.e.,

$$\begin{aligned}\boldsymbol{\nu}^{(k+1)} \quad &= \operatorname{argmax}_{\boldsymbol{\nu}} \left\{ \sum_{i=1}^{m} \log \left[\mathcal{F}_{SMN} \left(\frac{\kappa_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(k+1)}}{\sigma^{(k+1)}} \right) \right] \right. \\ &\left. + \sum_{i=m+1}^{n} \log \left[f_{SMN}(y_i | \mathbf{x}_i^\top \boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \boldsymbol{\nu}) \right] \right\}.\end{aligned}$$

Appendix D. Complementary results of the simulation studies: asymptotic properties

Figures 5 and 6 depict the average bias and the average MSE of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\sigma^2}$ for the levels of censoring $p = 25\%$ and $p = 45\%$, respectively.

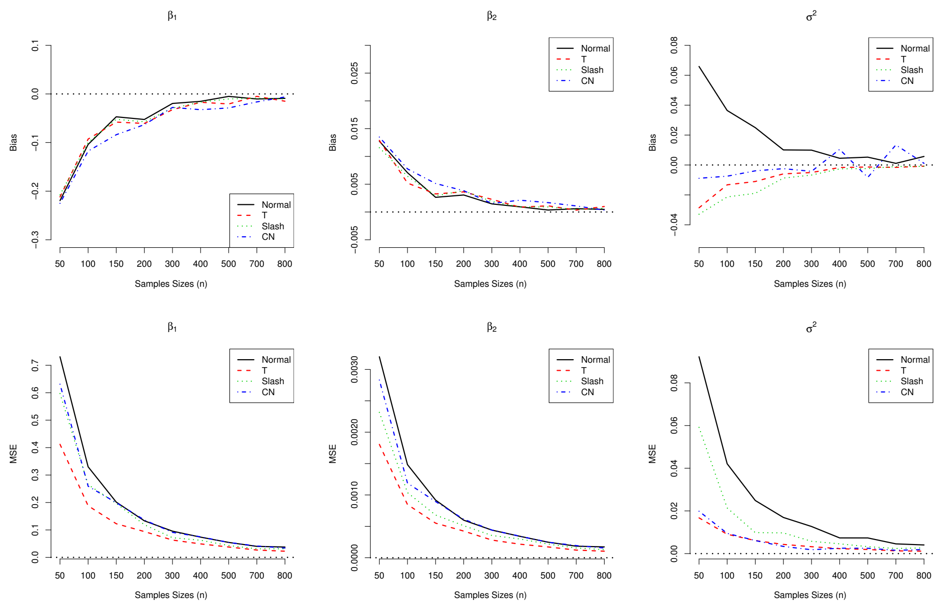

Figure 6: Simulation study 2. Average bias (first line) and MSE (second line) of estimates of β_1, β_2 and σ^2 from the SMN-CR models for different levels of censoring $p = 45\%$.

References

Akaike, H., 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19, 716–723.

Arellano-Valle, R., Castro, L., González-Farías, G., Muñoz-Gajardo, K., 2012. Student-t censored regression model: properties and inference. Statistical Methods & Applications 21, 453–473.

Arellano-Valle, R. B., 1994. Distribuições elípticas: Propriedades, inferência e aplicações a modelos de regressão. Ph.D. thesis, Instituto de Matemática e Estatística, Universidade de São Paulo, in portuguese.

Bai, Z. D., Krishnaiah, P. R., Zhao, L. C., 1989. On rates of convergence of efficient detection criteria in signal processing with white noise. Information Theory, IEEE Transactions on 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.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

Fang, K. T., Zhang, Y. T., 1990. Generalized Multivariate Analysis. Springer.

Garay, A. M., Lachos, V., Massuia, M. B., 2013. SMNCensReg: Fitting univariate censored regression model under the scale mixture of normal distributions. R package version 2.1.
URL http://CRAN.R-project.org/package=SMNCensReg

Genç, A. I., 2012. Moments of truncated normal/independent distributions. Statistical PapersDOI: 10.1007/s00362-012-0459-9.

Greene, W., 2012. Econometric Analysis. Prentice Hall.

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

Kim, H. J., 2008. Moments of truncated Student- distribution. Journal of the Korean Statistical Society 37, 81–87.

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., Little, R., Taylor, J., 1989. Robust statistical modeling using t distribution. Journal of the American Statistical Association 84, 881–896.

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

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

Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H., 2012. Influence diagnostics for Student-t censored linear regression models. Tech. Rep. 09-23, Instituto de Matemática e Estatística, Universidade de São Paulo.

Matos, L. A., Prates, M. O., Chen, M. H., Lachos, V. H., 2013. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Statistica Sinica 23, 1323–1342.

Meng, X. L., Rubin, B. D., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika 80, 267–278.

Meza, C., Osorio, F., la Cruz, R. D., 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.

Osorio, F., Paula, G. A., Galea, M., 2007. Assessment of local influence in elliptical linear models with longitudinal structure. Computational Statistics & Data Analysis 51, 4354–4368.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org`

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

Therneau, T. M., Grambsch, P. M., Fleming, R. T., 1990. Martingale-based residuals for survival models. Biometrika, 147–160.

Vaida, F., Fitzgerald, A. P., DeGruttola, V., 2007. Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. Computational Statistics & Data Analysis 51, 5718–5730.

Vaida, F., Liu, L., 2009. Fast Implementation for Normal Mixed Effects Models With Censored Response. Journal of Computational and Graphical Statistics 18, 797–817.

Villegas, C., Paula, G., Cysneiros, F., Galea, M., 2012. Influence diagnostics in generalized symmetric linear models. Computational Statistics & Data Analysis 59, 161–170.

Wei, C. G., Tanner, M. A., 1990. Posterior computations for censored regression data. Journal of the American Statistical Association 85, 829–839.

Wu, L., 2010. Mixed Effects Models for Complex Data. Chapman & Hall/CRC.