CrossMark

Linear censored regression models with scale mixtures of normal distributions

Aldo M. Garay[1] · **Victor H. Lachos**[1] ·
Heleno Bolfarine[2] · **Celso R. B. Cabral**[3]

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. 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. 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 set are reported, illustrating the usefulness of the new methodology.

✉ Aldo M. Garay
    medina_garay@yahoo.com

[1] Departamento de Estatística, Universidade Estadual de Campinas, Rua Sérgio Buarque de Holanda, 651 – Cidade Universitária Zeferino Vaz Campinas, São Paulo, SP, CEP 13083-859, Brazil

[2] Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil

[3] Departamento de Estatística, Universidade Federal do Amazonas, Manaus, AM, Brazil

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 truncated regression models. More recently, Massuia et al. (2014) developed diagnostic measures for censored regression models using the Student-t distribution, including the implementation of an interesting (and simple) expectation-maximization (EM) algorithm for maximum likelihood (ML) 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 kur-

tosis. 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. (2014), 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 (SE) 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 Sect. 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 Sect. 4, we derive approximate SE for the regression parameters of the SMN-CR model. Sect. 5, presents some simulation studies to compare the performance of our methods with other normality-based methods. In Sect. 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 $X \sim N(\mu, \sigma^2)$ 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 $Z \sim N(0, \sigma^2)$, 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:

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

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 there 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

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

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 Sect. 4, to obtain SE for the regression parameters.

Definition 2 Let $X \sim \mathrm{SMN}(\mu, \sigma^2, \boldsymbol{\nu})$ 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 $Y \sim \mathrm{TSMN}_{(a,b)}(\mu, \sigma^2, \boldsymbol{\nu})$.

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

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

and $f_{\mathrm{TSMN}}(y|\mu, \sigma^2, \boldsymbol{\nu}; (a, b)) = 0$ otherwise, where $F_{SMN}(\cdot)$ 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 E[·] 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, $f(u, x)$ 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* $a < b$, *the* $\mathrm{E}\left[U^r X^s | X \in (a, b)\right]$ *for* $r \geq 1$ *and* $s = 0, 1, 2$ *is given by:*

$$E\left[U^r|X \in (a, b)\right] = \tau(a, b)\left[E_\Phi (r, b) - E_\Phi (r, a)\right],$$

$$E\left[U^r X|X \in (a, b)\right] = \tau(a, b)\left[E_\phi\left(r - \frac{1}{2}, a\right) - E_\phi\left(r - \frac{1}{2}, b\right)\right],$$

$$E\left[U^r X^2|X \in (a, b)\right] = \tau(a, b)\left[E_\Phi (r - 1, b) - E_\Phi (r - 1, a) + aE_\phi\left(r - \frac{1}{2}, a\right)\right.$$
$$\left. - bE_\phi\left(r - \frac{1}{2}, b\right)\right],$$

where

$$\tau(a, b) = (F_{SMN} (b) - F_{SMN} (a))^{-1}; \tag{5}$$

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

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

Proof Let $\mathcal{A} = (a, b)$. From Definitions 1 and 2, we have that $X|U = u \sim \mathrm{N}(0, u^{-1})$, $X|X \in \mathcal{A} \sim \mathrm{TSMN}_\mathcal{A}(0, 1, \boldsymbol{v})$ 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,

$$E\left[U^r X^s|X \in \mathcal{A}\right] = E_U\left[U^r E_X\left[X^s|U = u, X \in \mathcal{A}\right]|X \in \mathcal{A}\right]$$
$$= \int_0^\infty u^r E_X\left[X^s|U = u, X \in \mathcal{A}\right] f(u|X \in \mathcal{A})du. \tag{8}$$

The pdf in the integral sign takes the following form:

$$f(u|X \in \mathcal{A}) = \int f(u, x|X \in \mathcal{A})dx \tag{9}$$

$$= \int f(u|X = x, X \in \mathcal{A}) f(x|X \in \mathcal{A})dx$$

$$= \tau(a, b) \int f(u|X = x, X \in \mathcal{A}) f(x)\mathbb{I}_\mathcal{A}(x)dx \tag{10}$$

$$= \tau(a, b) \int f(u, x) \mathbb{I}_\mathcal{A}(x)dx \tag{11}$$

$$= \tau(a, b) \int_\mathcal{A} f(u) \phi\left(x|0, u^{-1}\right) dx = \tau(a, b) f(u) \int_{\mathcal{A}^*} \phi(z) dz$$

$$= \tau(a, b) f(u)\left[\Phi\left(bu^{\frac{1}{2}}\right) - \Phi\left(au^{\frac{1}{2}}\right)\right],$$

where $\mathcal{A}^* = (au^{\frac{1}{2}}, bu^{\frac{1}{2}})$ and $\mathbb{I}_A(\cdot)$ is the indicator function of the set A. Equation (10) is obtained using the pdf's expression of $X|X \in \mathcal{A}$. Equation (11) 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 (10) and (11) are equal to zero. By (8) and Lemma 1 given in Appendix A, it follows that

– for $s = 0$,

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

– 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 $s = 2$,

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

$\square$

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

$$E_\Phi(r, h) = \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\delta}{2}\right)^{-r} F_{PVII}(h|\nu + 2r, \delta);$$

$$E_\phi(r, h) = \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}}\left(\frac{\delta}{2}\right)^{\frac{\nu}{2}}\left(\frac{h^2+\delta}{2}\right)^{-\frac{(\nu+2r)}{2}},$$

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

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

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

$$E_\Phi(r, h) = \left(\frac{\nu}{\nu+r}\right) F_{SL}(h|\nu + r);$$

$$E_\phi(r, h) = \frac{\nu}{\sqrt{2\pi}}\left(\frac{h^2}{2}\right)^{-(\nu+r)}\Gamma\left(\nu + r, \frac{h^2}{2}\right),$$

where $\Gamma(a, b) = \int_0^b e^{-t}t^{a-1}dt$ 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

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

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

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

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

As a direct consequence of Proposition 1, in Appendix A we present an important corollary, which is 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

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \text{SMN}(0, \sigma^2, \boldsymbol{\nu}), \quad i = 1, \ldots, n, \tag{12}$$

where the Y_i 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 $Y_i \sim \text{SMN}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$, 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

$$Y_{\text{obs}_i} = \begin{cases} \kappa_i & \text{if } Y_i \leq \kappa_i; \\ Y_i & \text{if } Y_i > \kappa_i, \end{cases} \tag{13}$$

$i = 1, \ldots, n$, 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 $\epsilon_i \sim \text{N}(0, \sigma^2)$, 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 $U_i \sim Gamma(\nu/2, \nu/2)$ we obtain the Student-t censored regression model developed by Massuia et al. (2014).

It is important to emphasize the difference between censored and truncated data. Citing Lee and Scott (2012), data are said to be censored when the exact values of measurements are not reported. For example, the needle of a scale that does not provide a reading over 200 kg will show 200 kg for all the objects that weigh more than the limit. Data are said to be truncated when the number of measurements outside a certain range is not reported.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\nu})^\top$ 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 $\mathbf{y}_{\text{obs}}$ in two subsamples of m censored and $n - m$ uncensored values, such that $\mathbf{y}_{\text{obs}} = \{\kappa_1, \ldots, \kappa_m, y_{m+1}, \ldots, y_n\}$. Then, the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) = \sum_{i=1}^m \log\left[F_{SMN}\left(\frac{\kappa_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right] + \sum_{i=m+1}^n \log\left[f_{SMN}(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}) \right]. \tag{14}$$

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. Alterna-

tively, 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:

$$
Y_i | U_i = u_i \sim \mathrm{N}\left(\mathbf{x}_i^\top \boldsymbol{\beta}, u_i^{-1} \sigma^2\right); \quad U_i \sim H(\cdot | \boldsymbol{v}). \tag{15}
$$

If the observation i is censored, we can consider y_i as a realization of the latent unobservable variable $Y_i \sim \mathrm{SMN}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \boldsymbol{v})$, $i = 1, \ldots, m$. 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 (15), we obtain the complete-data log-likelihood, given by

$$
\ell_c(\boldsymbol{\theta}|\mathbf{z}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\log(u_i)
$$
$$
- \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \sum_{i=1}^{n}\log\left(f(u_i|\boldsymbol{v})\right), \tag{16}
$$

where $f(\cdot|\boldsymbol{v})$ 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,

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

where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}$ means that the expectation is being affected 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

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

as well as

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

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

$$Q\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}\big) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\Big[\mathcal{E}_{2i}\big(\boldsymbol{\theta}^{(k)}\big) - 2\mathcal{E}_{1i}\big(\boldsymbol{\theta}^{(k)}\big)\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{0i}\big(\boldsymbol{\theta}^{(k)}\big)\big(\mathbf{x}_i^\top\boldsymbol{\beta}\big)^2\Big]$$

$$+ \frac{1}{2}\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\left(U_i\right)|y_{\mathrm{obs}_i}] + \sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log\left(f\left(U_i|\boldsymbol{\nu}\right)\right)|y_{\mathrm{obs}_i}]. \qquad (17)$$

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

$$\mathcal{E}_{si}\big(\boldsymbol{\theta}^{(k)}\big) = y_i^s\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i], \qquad (18)$$

where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i]$ 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

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

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

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

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

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

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

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

$$\mathcal{E}_{si}\big(\boldsymbol{\theta}^{(k)}\big) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_iY_i^s|Y_i \leq \kappa_i], \qquad (19)$$

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

When the M-step turns out to be analytically intractable, it can be replaced with a sequence of conditional 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 (18);
- If observation i is censored then, for $s = 0, 1, 2$, compute $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)})$ in (19).

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

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

$$
\sigma^{2(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \left[\mathcal{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathcal{E}_{1i}(\boldsymbol{\theta}^{(k)}) \mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)} + \mathcal{E}_{0i}(\boldsymbol{\theta}^{(k)}) (\mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)})^2 \right]. \tag{21}
$$

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

$$
\boldsymbol{\nu}^{(k+1)} = \text{argmax}_{\nu} \left\{ \sum_{i=1}^{m} \log \left[F_{SMN} \left(\frac{\kappa_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)}}{\sigma^{(k+1)}} \right) \right] \right.
$$
$$
\left. + \sum_{i=m+1}^{n} \log \left[f_{SMN}(y_i|\mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \boldsymbol{\nu}) \right] \right\}. \tag{22}
$$

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

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

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 $\mathbb{E}_{\boldsymbol{\theta}^{(k)}}[\log(U_i)|y_{obs_i}]$ and $\mathbb{E}_{\boldsymbol{\theta}^{(k)}}[\log(f(U_i|\boldsymbol{\nu}))|y_{obs_i}]$.

4 Approximated standard errors for the fixed effects

Standard errors of the ML estimates can be approximated by the inverse of the observed information matrix, but there is generally no closed form, see Meilijson (1989) and

Lin (2009). Considering $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{v})$, the empirical information matrix is defined as

$$\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{y}_{obs}\right) = \sum_{i=1}^n \mathbf{v}\left(\mathbf{y}_{obs_i}|\boldsymbol{\theta}\right) \mathbf{v}^\top\left(\mathbf{y}_{obs_i}|\boldsymbol{\theta}\right)$$
$$-\frac{1}{n}\mathbf{V}\left(\mathbf{y}_{obs}|\boldsymbol{\theta}\right) \mathbf{V}^\top\left(\mathbf{y}_{obs}|\boldsymbol{\theta}\right),$$

where $\mathbf{V}^\top\left(\mathbf{y}_{obs}|\boldsymbol{\theta}\right) = \sum_{i=1}^n \mathbf{v}\left(\mathbf{y}_{obs_i}|\boldsymbol{\theta}\right)$. It is noted from the result of Louis (1982) that, the individual score can be determined as

$$\mathbf{v}\left(\mathbf{y}_{obs_i}|\boldsymbol{\theta}\right) = \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{y}_{obs_i})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{z}_i)}{\partial \boldsymbol{\theta}}|\mathbf{y}_{obs_i}, \boldsymbol{\theta}\right]. \tag{23}$$

Thus, substituting the ML estimates of $\boldsymbol{\theta}$ in (23), the empirical information matrix $\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{y}_{obs}\right)$ is reduced to

$$\mathbf{I}_e\left(\widehat{\boldsymbol{\theta}}|\mathbf{y}_{obs}\right) = \sum_{i=1}^n \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^\top, \tag{24}$$

where $\widehat{\mathbf{v}}_i = \left(\widehat{\mathbf{v}}_{\boldsymbol{\beta}i}, \widehat{\mathbf{v}}_{\sigma^2 i}, \widehat{\mathbf{v}}_{vi}\right)$ is an individual score vector and

$$\widehat{\mathbf{v}}_{\boldsymbol{\beta}i} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{z}_i)}{\partial \boldsymbol{\beta}}|\mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right] = \frac{1}{\sigma^2}\left(\mathbf{x}_i \mathcal{E}_{1i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{0i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right),$$

$$\widehat{\mathbf{v}}_{\sigma^2 i} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{z}_i)}{\partial \sigma^2}|\mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right]$$
$$= -\frac{1}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\left(\mathcal{E}_{2i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{1i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} + \mathcal{E}_{0i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})^2\right), \text{ and}$$

$$\widehat{\mathbf{v}}_{vi} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{z}_i)}{\partial \boldsymbol{v}}|\mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[\frac{\partial \log\left(f(U_i|\boldsymbol{v})\right)}{\partial \boldsymbol{v}}|\mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right], \tag{25}$$

where $\ell_c(\boldsymbol{\theta}|\mathbf{z}_i)$ is the log-likelihood formed from the single complete observation $\mathbf{z}_i = (y_{obs_i}, y_i, u_i)^\top$ and $\mathcal{E}_{si}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s|y_{obs_i}]$. It is important notice that the values of Eq. (25) depend of the distribution of U. Thus for example:

- *For the Student-t distribution:*
  We consider $U \sim Gamma(v/2, \delta/2)$, with $v > 0$, then

$$\widehat{\mathbf{v}}_{vi} = -\psi\left(\frac{\widehat{v}}{2}\right) + \frac{1}{2}\left(\log\left(\frac{\widehat{v}}{2}\right) + 1\right)$$
$$+ \frac{1}{2}\left(\mathrm{E}\left[\log(U_i)|\mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right] - \mathcal{E}_{0i}(\widehat{\boldsymbol{\theta}})\right),$$

  where $\psi(x)$ represents the digamma function of x.

- *For the Slash distribution:*
  We consider $U \sim Beta(\nu, 1)$ with positive shape parameter ν, then

$$\widehat{\mathbf{v}}_{\nu i} = \frac{1}{\widehat{\nu}} + \mathrm{E}\left[\log\left(U_i\right) | \mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right].$$

It is important to stress that the SE of ν depends heavily on the calculation of $\mathrm{E}\left[\log\left(U_i\right) | \mathbf{y}_{obs_i}, \widehat{\boldsymbol{\theta}}\right]$, which relies on computationally intensive Monte Carlo integrations. In our analysis, we focus solely on comparing the SE of β and σ^2.

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 (2015).

We performed a simulation study based on the N-CR model. Specifically, we considered model (12) 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 $n = 300$, considering $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (1, 4)$, $\sigma^2 = 2$ and fixing the left censoring level at $p = 8$, 20 and 35 % (that is, 8, 20 and 35 % of the observations in each data set were left censored, respectively). We generated independently the values x_i, for $i = 1, \ldots, n$, 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_{150} by $y_{150}(\vartheta) = y_{150} - \vartheta$, with $\vartheta = 1, 2, \ldots, 10$. Let $\widehat{\beta}_i(\vartheta)$ and $\widehat{\beta}_i$ be the EM estimates of β_i with and without contamination, respectively, $i = 1, 2$. We are particularly interested in the relative changes

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

We define the relative changes for σ^2 analogously.
For each replication we obtained the parameter estimates with and without outliers, under the following models: N-CR, T-CR and SL-CR, both with $\nu = 3$, and CN-CR with $\boldsymbol{\nu}^\top = (\xi, \gamma) = (0.3, 0.3)$. Table 1 and Fig. 1 depict the average values of the relative changes across all samples and different censoring levels. 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, 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 with censoring level $p = 20$ % and 35 % we can observe that the relative change increases as ϑ increases, specially for the parameter β_1.

Table 1 Simulation study 1

Cens. Level	RC(·)	SMN-CR models	ϑ									
			1	2	3	4	5	6	7	8	9	10
8 %	$RC(\beta_1)$	Normal	0.01475	0.03003	0.04585	0.06219	0.07905	0.09641	0.11427	0.13262	0.15144	0.17074
		T	0.01896	0.03184	0.03701	0.03765	0.03646	0.03486	0.03344	0.03235	0.03153	0.03090
		SL	0.00962	0.01840	0.02493	0.02826	0.02862	0.02700	0.02463	0.02238	0.02059	0.01921
		CN	0.00096	0.00158	0.00190	0.00210	0.00239	0.00287	0.00350	0.00427	0.00526	0.00647
	$RC(\beta_2)$	Normal	2.16e-04	0.00044	0.00067	0.00091	0.00116	0.00142	0.00169	0.00196	0.00224	0.00253
		T	2.94e-04	0.00049	0.00058	0.00059	0.00057	0.00054	0.00052	0.00050	0.00049	0.00048
		SL	1.19e-04	0.00023	0.00031	0.00036	0.00036	0.00034	0.00031	0.00029	0.00026	0.00025
		CN	9.06e-05	0.00015	0.00017	0.00018	0.00019	0.00021	0.00023	0.00025	0.00027	0.00028
	$RC(\sigma^2)$	Normal	0.00431	0.01012	0.01862	0.03061	0.04632	0.06578	0.08912	0.11603	0.14668	0.18094
		T	0.00574	0.00997	0.01271	0.01469	0.01607	0.01705	0.01768	0.01813	0.01846	0.01868
		SL	0.00462	0.00978	0.01506	0.01987	0.02326	0.02510	0.02591	0.02607	0.02605	0.02601
		CN	0.00498	0.00965	0.01474	0.02085	0.02939	0.04102	0.05560	0.07267	0.09204	0.11355
	$RC(\beta_1)$	Normal	1.43e-02	2.94e-02	4.53e-02	0.06187	0.07915	0.09714	0.11584	0.13525	0.15537	0.17620
		T	1.25e-05	3.17e-05	6.65e-05	0.00013	0.00027	0.00059	0.00147	0.00416	0.01272	0.03326
		SL	1.02e-02	1.95e-02	2.63e-02	0.02982	0.03026	0.02867	0.02630	0.02404	0.02222	0.02086
		CN	1.75e-02	3.17e-02	3.95e-02	0.04307	0.04921	0.05854	0.06966	0.08169	0.09432	0.10741
20 %	$RC(\beta_2)$	Normal	1.65e-04	3.42e-04	5.29e-04	7.27e-04	9.35e-04	1.15e-03	0.00138	1.62e-03	0.00187	0.00212
		T	1.91e-07	4.88e-07	1.02e-06	2.05e-06	4.24e-06	9.36e-06	0.00002	6.56e-05	0.00020	0.00052
		SL	1.08e-04	2.08e-04	2.82e-04	3.20e-04	3.26e-04	3.10e-04	0.00028	2.61e-04	0.00024	0.00022
		CN	1.06e-04	1.88e-04	2.42e-04	2.83e-04	3.28e-04	3.90e-04	0.00046	5.50e-04	0.00063	0.00073
	$RC(\sigma^2)$	Normal	5.07e-03	1.25e-02	2.33e-02	3.78e-02	0.05675	0.07982	0.10719	0.13880	0.17457	0.21454
		T	9.55e-06	2.39e-05	4.63e-05	8.31e-05	0.00014	0.00026	0.00049	0.00094	0.00145	0.00139

Table 1 continued

Cens. Level	RC(·)	SMN-CR models	ϑ									
			1	2	3	4	5	6	7	8	9	10
		SL	5.64e-03	1.17e-02	1.79e-02	2.32e-02	0.02675	0.02865	0.02929	0.02949	0.02949	0.02945
		CN	5.98e-03	1.16e-02	1.72e-02	2.44e-02	0.03476	0.04842	0.06550	0.08547	0.10798	0.13290
	$RC(\beta_1)$	Normal	0.03972	0.05949	0.06658	0.07525	0.08552	0.09737	0.11079	0.12576	0.14229	0.16034
		T	0.01307	0.04910	0.05653	0.04350	0.04734	0.04113	0.04599	0.04194	0.03876	0.03620
		SL	0.01087	0.01947	0.02485	0.02644	0.02516	0.02278	0.02041	0.01848	0.01688	0.01571
		CN	0.00044	0.00174	0.00572	0.01951	0.03802	0.05857	0.06217	0.08615	0.09152	0.11027
35%	$RC(\beta_2)$	Normal	3.33e-06	1.28e-05	4.20e-05	0.00014	0.00041	0.00090	0.00146	0.00181	0.00185	0.00192
		T	2.76e-04	4.57e-04	5.25e-04	0.00053	0.00051	0.00048	0.00046	0.00044	0.00043	0.00041
		SL	2.26e-04	4.15e-04	5.43e-04	0.00059	0.00058	0.00054	0.00048	0.00043	0.00040	0.00037
		CN	4.48e-05	7.48e-05	9.16e-05	0.00010	0.00010	0.00011	0.00012	0.00015	0.00021	0.00028
	$RC(\sigma^2)$	Normal	6.08e-03	0.01452	0.02675	0.04407	0.06698	0.09539	0.12879	0.16729	0.21089	0.25958
		T	7.78e-03	0.01371	0.01814	0.02106	0.02283	0.02402	0.02491	0.02554	0.02600	0.02632
		SL	6.23e-03	0.01379	0.02170	0.02862	0.03315	0.03532	0.03581	0.03579	0.03592	0.03594
		CN	6.76e-05	0.00021	0.00050	0.00109	0.00141	0.00152	0.00765	0.01702	0.02509	0.03424

Average relative changes on estimates for different contaminations ϑ and censoring level p %. The number (or vector) in parentheses following the model's name is the value of ν
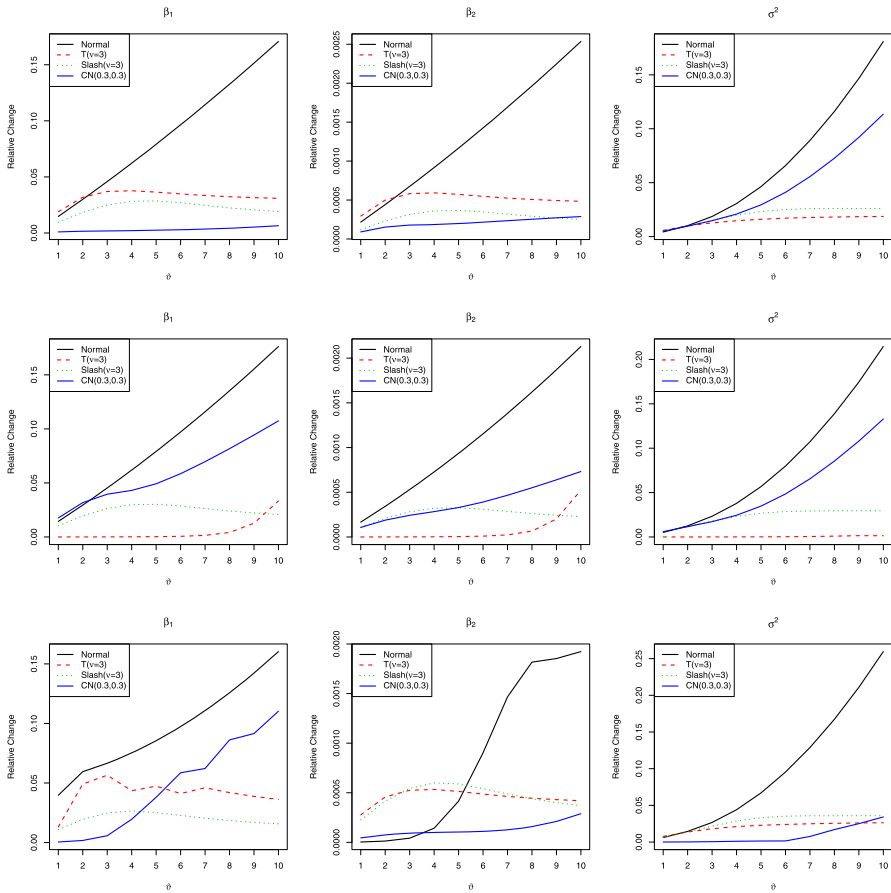
Fig. 1 Simulation study 1. Average relative changes on estimates for different contaminations ϑ and censoring level: $p = 8\%$ (*First line*), $p = 20\%$ (*Second line*) and $p = 35\%$ (*Third line*) respectively

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 (12), with $\mathbf{x}_i^\top = (1, x_i)$, $i = 1, \ldots, n$.

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, we 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 $\left(\boldsymbol{\nu}^\top = (0.5, 0.5)\right)$.

In order to analyze the performance of the estimates obtained using our proposed EM-type algorithm, we computed the bias and the mean squared error (MSE) for each

combination of sample size, level of censoring and parameter value. For β_i, they are given, respectively, by

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

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

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 Fig. 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 MLE's of the parameters (simulation study 3)

Now we show, via simulation study, that the method suggested in Sect. 4 to approximate the SE of the MLE of the regression parameters has good asymptotic properties. We fixed a SMN-CR model (N-CR and T-CR or SL-CR with $\nu = 4$ respectively) and a censoring level (5, 10, 20, 30 or 50\,%). For each one of these fifteen 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 $\boldsymbol{\theta} = (\beta_1, \beta_2, \sigma^2)$, the estimates of their SE using the technique proposed in Sect. 4 and an approximate 95\,% confidence interval assuming asymptotic normality. Table 2 presents the sample standard errors of $\widehat{\theta}$, i.e., the value

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

The results from this table show a reasonable MC coverage for both $\boldsymbol{\beta}$ and σ^2, although the values for σ^2 tend to be lower the nominal level (95\,%). Taking into account the moderate sample size (n=100), we consider these results quite satisfactory.

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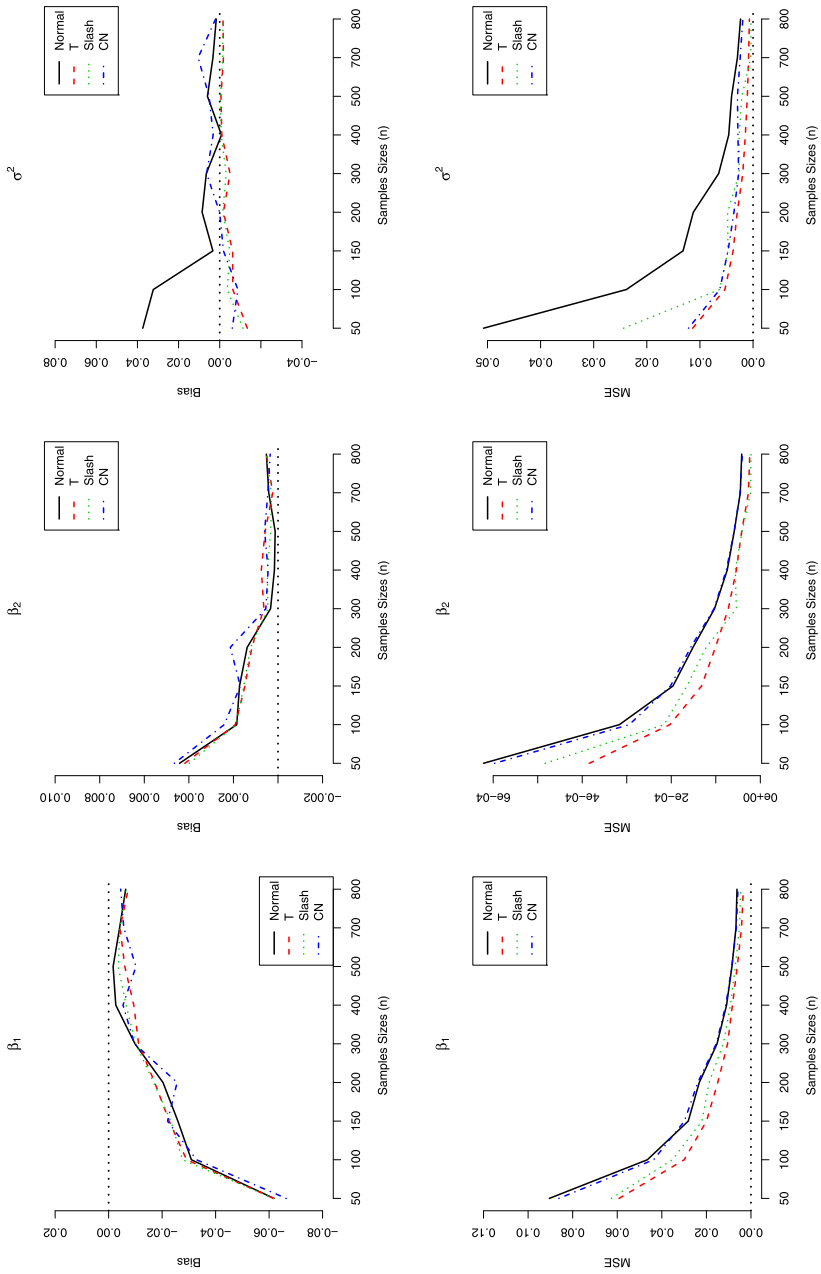
Fig. 2 Simulation study 2. Average bias (*first row*) and MSE (*second row*) of parameter estimates in the SMN-CR models for for $p = 10\%$

Table 2 Simulation study 3

Censoring level	Measure	N-CR			T-CR			SL-CR		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\sigma}^2$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\sigma}^2$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\sigma}^2$
5%	MC SE	0.2519	0.0204	0.1462	0.2678	0.0221	0.1961	0.2622	0.0209	0.1542
	MC IM SE	0.2613	0.0207	0.1530	0.2680	0.0220	0.1965	0.2594	0.0206	0.1593
	COV MC	95.6%	95.9%	93.8%	95.4%	94.4%	93.8%	94.1%	94.2%	94.7%
10%	MC SE	0.2860	0.0209	0.1491	0.3282	0.0248	0.1940	0.2828	0.0224	0.1620
	MC IM SE	0.2914	0.0213	0.1563	0.3336	0.0252	0.1987	0.2766	0.0221	0.1637
	COV MC	94.4%	94.5%	94.5%	95.2%	95.8%	94.7%	94.7%	94.3%	93.7%
20%	MC SE	0.3661	0.0278	0.1523	0.4055	0.0305	0.2005	0.3258	0.0277	0.1683
	MC IM SE	0.3748	0.0287	0.1658	0.4054	0.0306	0.2046	0.3441	0.0297	0.1721
	COV MC	95.5%	95.3%	94.1%	94.3%	93.2%	93.8%	95.1%	96.9%	94.0%
30%	MC SE	0.3756	0.0282	0.1688	0.3855	0.0300	0.2208	0.4171	0.0296	0.1752
	MC IM SE	0.3939	0.0298	0.1757	0.3837	0.0297	0.2229	0.4345	0.0308	0.1830
	COV MC	94.8%	95.2%	92.6%	93.5%	93.7%	93.2%	95.7%	95.1%	94.3%
50%	MC SE	0.7466	0.0469	0.1977	0.6934	0.0480	0.2567	0.8219	0.0533	0.2004
	MC IM SE	0.7746	0.0491	0.2135	0.7043	0.0483	0.2592	0.8336	0.0545	0.2197
	COV MC	94.6%	94.8%	93.3%	95.0%	94.3%	92.0%	95.2%	94.7%	93.3%

Sample standard errors of $\widehat{\theta}$ (MC SE), average values (across 1000 samples) of the standard errors computed using the information method (MC IM SE) and the percentual coverage of the resulting 95% confidence intervals (COV MC)

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 a truncated Student-t regression model. We analyzed it with the aim of providing additional inferences by using the SMN distributions in the context of censored models. We fitted a regression model with an intercept parameter β_1 and applied the EM-type algorithm for censored data explained in Sect. 3.2, considering again the N-CR, T-CR, SL-CR and CN-CR models for comparative purposes.

Table 3 shows the parameter estimates, together with their corresponding 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 SE 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 affected using the EM estimate $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$.

The plots of r_{MT_i} with generated confidence envelopes are presented in Fig. 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 Sect. 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 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

Table 3 Real data

Parameter	N-CR Estimate	SE	T-CR Estimate	SE	SL-CR Estimate	SE	CN-CR Estimate	SE
β_1	−2.7510	1.8899	−1.0470	1.4402	−1.1445	1.4377	−1.2901	1.4452
β_2	−0.1045	0.0282	−0.1107	0.0232	−0.1084	0.0232	−0.1064	0.0233
β_3	0.7281	0.0826	0.6475	0.0649	0.6434	0.0635	0.6468	0.0627
β_4	−0.2143	0.1493	−0.2964	0.1218	−0.2946	0.1210	−0.2997	0.1220
β_5	−3.0263	0.4195	−3.1637	0.3843	−3.0958	0.3785	−3.0649	0.3750
σ^2	20.9401	0.8106	10.6379	1.0117	6.7385	0.6355	11.1690	0.9941
ν	–	–	4.1993	–	1.4061	–	–	–
γ	–	–	–	–	–	–	0.1000	–
ξ	–	–	–	–	–	–	0.1000	–

EM estimates for fitting various SMN-CR models. SE are the estimated asymptotic standard errors

Table 4 Real data

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

Values of some model selection criteria
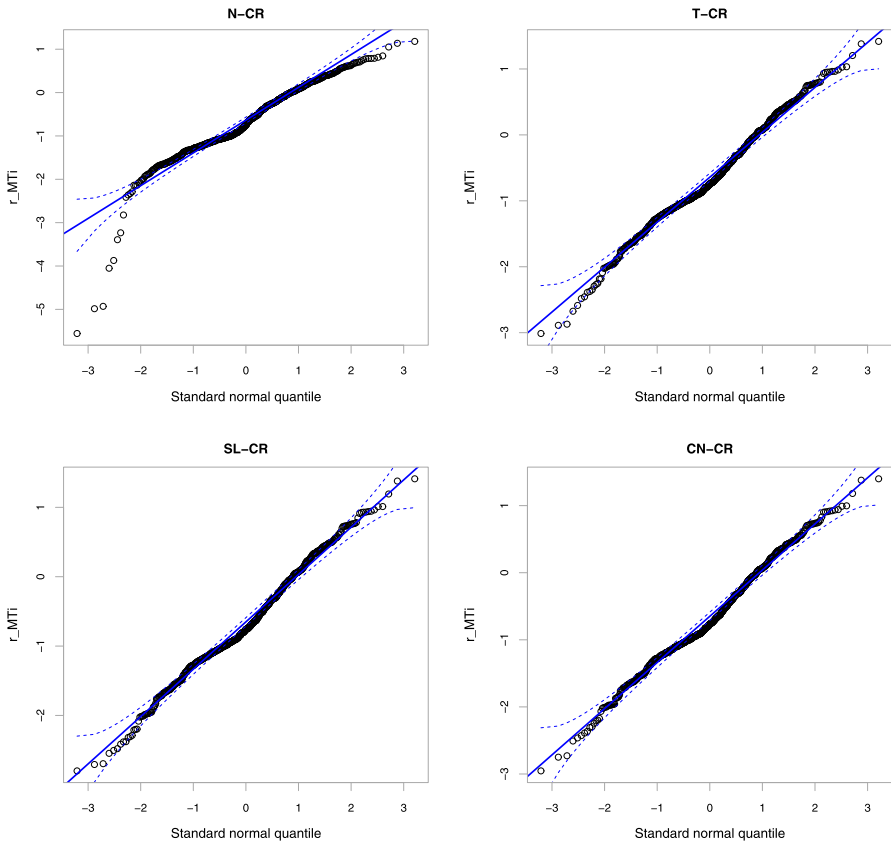
Fig. 3 Real data. Envelopes of the martingale-type residuals, r_{MT_i}, for the SMN-CR models

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 SMN distributions, called SMN-CR models. This offers a high degree
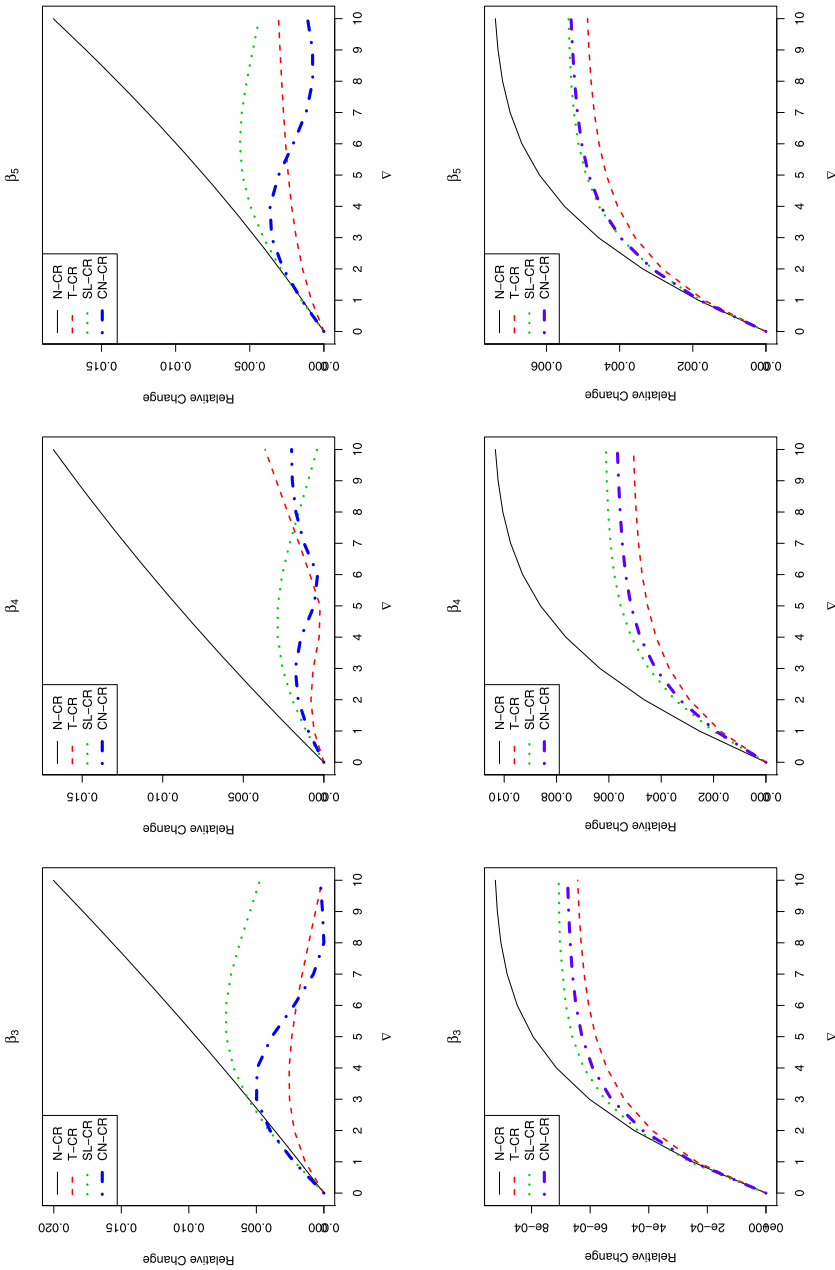
Fig. 4 Real data. Relative changes on EM estimates from the SMN-CR models for different contaminations ∇ of the uncensored observation y_7 (*first row*) and the censored observation y_{750} (*second row*)

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 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.

Appendix 1: Lemmas and corollary

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 $F_{SMN}(a) = E_U\left[\Phi\left(aU^{\frac{1}{2}}\right)\right]$, where $F_{SMN}(\cdot)$ 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).

The following Corollary is a direct consequence of Proposition 1 given in Sect. 2.

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

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

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

Appendix 2: Derivations of quantities $E_\phi(r, h)$ and $E_\Phi(r, h)$ for SMN distributions

In this Appendix, we calculate the expressions for the expected values $\mathrm{E}_\phi(r, h)$ and $\mathrm{E}_\Phi(r, h)$, for $r, h \geq 0$, 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,

$$
\begin{aligned}
\mathrm{E}_\phi(r, h) &= \int_0^\infty \frac{\delta^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1} u^r}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \exp\left\{-\frac{u(h^2 + \delta)}{2}\right\} du \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right) \delta^{\frac{\nu}{2}} \left(\frac{h^2+\delta}{2}\right)^{-\frac{\nu+2r}{2}}}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \int_0^\infty \frac{\alpha_2^{\alpha_1} u'^{\{\alpha_1-1\}}}{\Gamma(\alpha_1)} \exp\left\{-\alpha_2 u'\right\} du' \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{\nu/2} \left(\frac{h^2+\delta}{2}\right)^{-\frac{\nu+2r}{2}},
\end{aligned}
\tag{26}
$$

where the integrand in (26) is the pdf of a random variable U' with distribution *Gamma* (α_1, α_2).

$$
\begin{aligned}
\mathrm{E}_\Phi(r, h) &= \int_0^\infty \frac{u^{\frac{2r+\nu}{2}-1}\Phi\left(hu^{\frac{1}{2}}\right)\delta^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \exp\left\{-\frac{u\delta}{2}\right\} du \\
&= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\delta}{2}\right)^{-r} \int_0^\infty \left(\frac{\delta}{2}\right)^{\alpha_1} \frac{\Phi\left(hu'^{\{\frac{1}{2}\}}\right) u'^{\{\alpha_1-1\}}}{\Gamma(\alpha_1)} \exp\left\{-\frac{u'\delta}{2}\right\} du'
\end{aligned}
$$

$$= \frac{\Gamma \left(\frac{\nu+2r}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \left(\frac{\delta}{2} \right)^{-r} E_{U'} \left[\Phi \left(h U'^{\{\frac{1}{2}\}} \right) \right]$$

$$= \frac{\Gamma \left(\frac{\nu+2r}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \left(\frac{\delta}{2} \right)^{-r} F_{PVII}(h | \nu + 2r, \delta), \tag{27}$$

where in (27) 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 $E_\phi\,(r, h)$ and $E_\Phi\,(r, h)$ are given by

$$E_\phi\,(r, h) = \frac{\Gamma \left(\frac{\nu+2r}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right) \sqrt{2\pi}} \left(\frac{\nu}{2} \right)^{\frac{\nu}{2}} \left(\frac{h^2 + \nu}{2} \right)^{-\frac{(\nu+2r)}{2}};$$

$$E_\Phi\,(r, h) = \frac{\Gamma \left(\frac{\nu+2r}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \left(\frac{\nu}{2} \right)^{-r} F_{PVII}(h | \nu + 2r, \nu).$$

Slash distribution

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

$$E_\phi\,(r, h) = \int_0^1 u^r \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{h^2}{2} u \right\} \nu u^{\nu-1} du = \frac{\nu}{\sqrt{2\pi}} \int_0^1 u^{\nu+r-1} \exp \left\{ -\frac{h^2}{2} u \right\} du,$$

$$= \frac{\nu}{\sqrt{2\pi}} \left(\frac{h^2}{2} \right)^{-(\nu+r)} \gamma^* \left(\nu+r, \frac{h^2}{2} \right), \tag{28}$$

thus, considering $\gamma^*\,(a, x) = \int_0^x e^{-t} t^{a-1} dt$, we obtain Eq. (28).

$$E_\Phi\,(r, h) = \int_0^1 u^r \Phi \left(h u^{\frac{1}{2}} \right) \nu u^{\nu-1} du$$

$$= \frac{\nu}{\nu+r} \int_0^1 \Phi \left(h u'^{\{\frac{1}{2}\}} \right) u'^{\{\nu+r-1\}} (\nu+r)\,du' \tag{29}$$

$$= \frac{\nu}{\nu+r} F_{SL}(h | \nu + r), \tag{30}$$

where the integrand in (29) is the expectation of the random variable $\Phi(h U'^{\{\frac{1}{2}\}})$, with $U' \sim Beta(\nu+r, 1)$. Using Lemma 2, we obtain Eq. (30), where $F_{SL}(\cdot)$ is the cdf of the slash distribution.

Contaminated normal distribution

$$E_\phi\,(r, h) = u^r \phi \left(h u^{\frac{1}{2}} \right) \left[\nu \mathbb{I}_{\{\gamma\}}(u) + (1-\nu) \mathbb{I}_{\{1\}}(u) \right]$$

$$= \nu \gamma^r \phi \left(h \gamma^{\frac{1}{2}} \right) + (1-\nu) \phi \left(h \gamma^{\frac{1}{2}} \right);$$

$$E_\Phi(r, h) = u^r \Phi\left(hu^{\frac{1}{2}}\right)\left[v\mathbb{I}_{\{\gamma\}}(u) + (1-v)\mathbb{I}_{\{1\}}(u)\right]$$

$$= v\gamma^r \Phi\left(h\gamma^{\frac{1}{2}}\right) + (1-v)\Phi(h) = \gamma^r\left[v\Phi\left(hu^{\frac{1}{2}}\right) + (1-v)\Phi(h)\right]$$

$$+ (1-v)\left(1-\gamma^r\right)\Phi(h)$$

$$= \gamma^r F_{CN}(h|v, \gamma) + (1-v)\left(1-\gamma^r\right)\Phi(h),$$

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 Eqs. (20)–(22). Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{v})$ be the vector with all parameters in the SMN-CR model and consider the notation given in Sect. 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

$$L(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = f(\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = f(\mathbf{y}_{\text{obs}}, \mathbf{y}_L|\mathbf{u})f(\mathbf{u})$$

$$= f(\mathbf{y}|\mathbf{u})f(\mathbf{u}) = \prod_{i=1}^{n} f(y_i|u_i)f(u_i|\boldsymbol{v}).$$

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

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = \log(L(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}))$$

$$= \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\log(u_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}u_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2$$

$$+ \sum_{i=1}^{n}\log(f(u_i|\boldsymbol{v})).$$

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

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

so we have

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = -\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{E_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^2|y_{\text{obs}_i}]\right.$$

$$\left. - 2E_{\boldsymbol{\theta}^{(k)}}[U_i Y_i|y_{\text{obs}_i}]\mathbf{x}_i^\top\boldsymbol{\beta} + E_{\boldsymbol{\theta}^{(k)}}[U_i|y_{\text{obs}_i}](\mathbf{x}_i^\top\boldsymbol{\beta})^2\right\}$$

$$+ \frac{1}{2}\sum_{i=1}^{n}E_{\boldsymbol{\theta}^{(k)}}[\log(U_i)|y_{\text{obs}_i}] + \sum_{i=1}^{n}E_{\boldsymbol{\theta}^{(k)}}[\log(f(U_i|\boldsymbol{v}))|y_{\text{obs}_i}].$$

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

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

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.,

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

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

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

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

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

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

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.
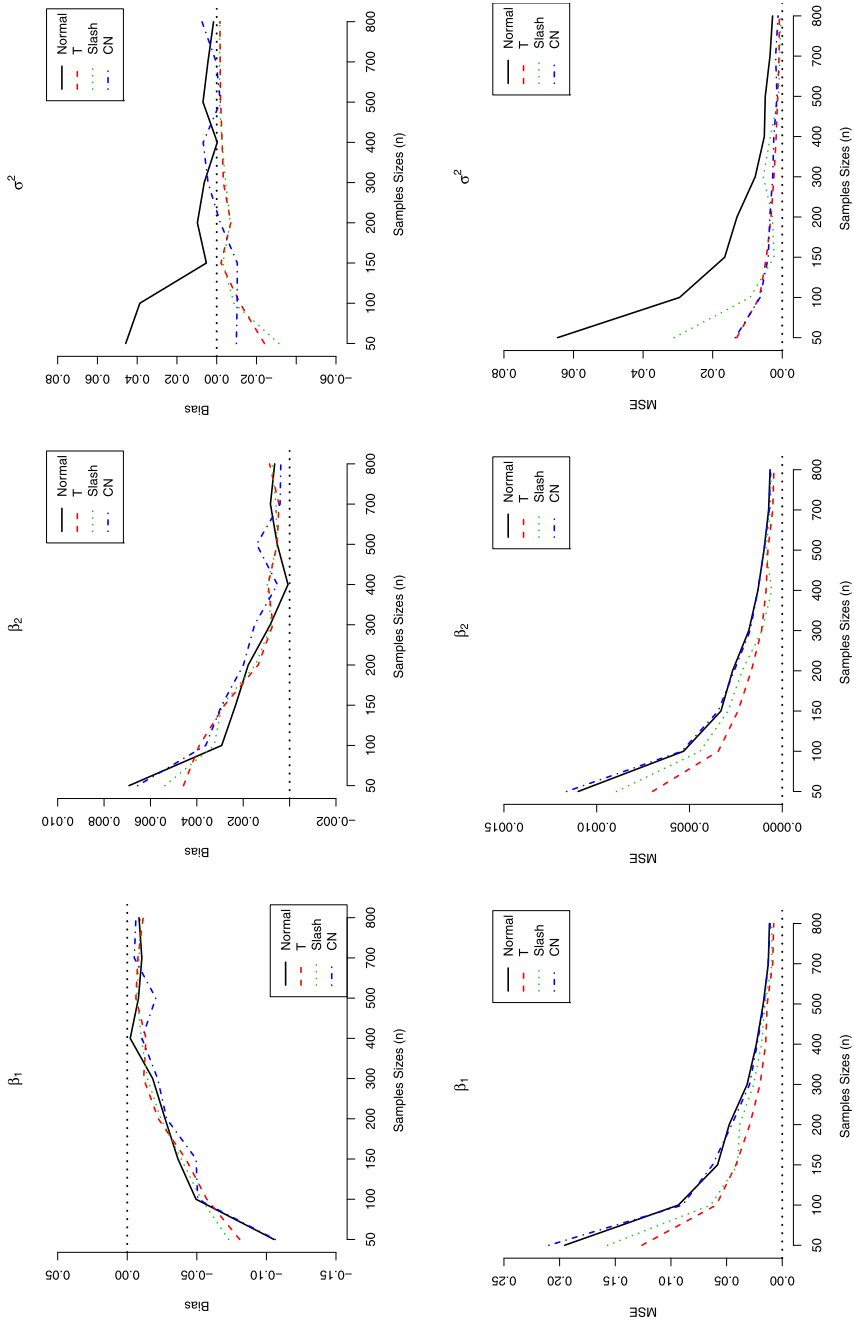
Fig. 5 Average bias (*first row*) and average MSE (*second row*) of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\sigma^2}$ from the SMN-CR models for level of censoring $p = 25\%$
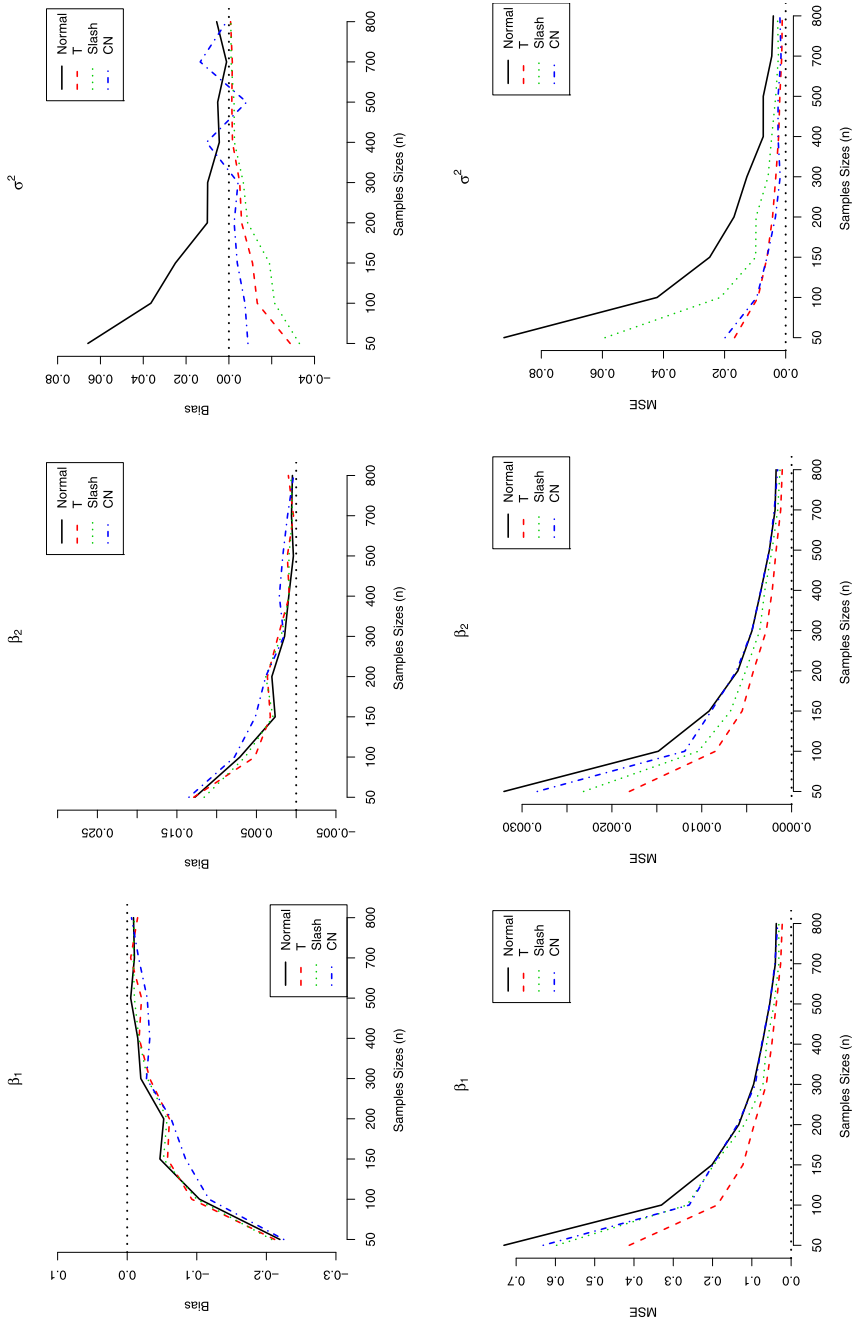
Fig. 6 Average bias (*first row*) and MSE (*second row*) of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\sigma^2}$ from the SMN-CR models for different levels of censoring $p = 45\,\%$

References

Akaike H (1974) A new look at the statistical model identification. Autom Control IEEE Trans 19:716–723

Arellano-Valle R, Castro L, González-Farías G, Muñoz-Gajardo K (2012) Student-t censored regression model: properties and inference. Stat Methods Appl 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 ZD, Krishnaiah PR, Zhao LC (1989) On rates of convergence of efficient detection criteria in signal processing with white noise. Inform Theory IEEE Trans 35:380–388

Barros M, Galea M, González M, Leiva V (2010) Influence diagnostics in the Tobit censored response model. Stat Methods Appl 19:716–723

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–38

Fang KT, Zhang YT (1990) Generalized multivariate analysis. Springer, Berlin

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 2.2. http://CRAN.R-project.org/package=SMNCensReg

Genç AI (2012) Moments of truncated normal/independent distributions. Stat Pap 54:741–764

Greene W (2012) Econometric analysis. Prentice Hall, New York

Ibacache-Pulgar G, Paula G (2011) Local influence for Student-t partially linear models. Comput Stat Data Anal 55:1462–1478

Kim HJ (2008) Moments of truncated Student- distribution. J Korean Stat Soc 37:81–87

Labra FV, Garay AM, Lachos VH, Ortega EMM (2012) Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. J Stat Plan Inference 142:2149–2165

Lachos VH, Ghosh P, Arellano-Valle RB (2010) Likelihood based inference for skew-normal independent linear mixed models. Stat Sin 20:303–322

Lange KL, Little R, Taylor J (1989) Robust statistical modeling using t distribution. J Am Stat Assoc 84:881–896

Lee G, Scott C (2012) EM algorithms for multivariate Gaussian mixture models with truncated and censored data. Comput Stat Data Anal 56:2816–2829

Liu C, Rubin DB (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika 80:267–278

Lin TI (2010) Robust mixture modeling using multivariate skew tădistributions. Stat Comput 20:343–356

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. J R Stat Soc 44:226–233

Massuia MB, Cabral CRB, Matos LA, Lachos VH (2014) Influence diagnostics for Student-t censored linear regression models. Statistics 85:1–21. doi:10.1080/02331888.2014.958489

Meilijson I (1989) A fast improvement to the EM algorithm to its own terms. J R Stat Soc Ser B 51:127–138

Meng XL, Rubin BD (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80:267–278

Meza C, Osorio F, la Cruz RD (2012) Estimation in nonlinear mixed-effects models using heavy-tailed distributions. Stat Comput 22:121–139

Mroz TA (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica 55:765–799

Ortega EMM, Bolfarine H, Paula GA (2003) Influence diagnostics in generalized log-gamma regression models. Comput Stat Data Anal 42:165–186

Osorio F, Paula GA, Galea M (2007) Assessment of local influence in elliptical linear models with longitudinal structure. Comput Stat Data Anal 51:4354–4368

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, http://www.R-project.org

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Therneau TM, Grambsch PM, Fleming RT (1990) Martingale-based residuals for survival models. Biometrika 77(1):147–160

Villegas C, Paula G, Cysneiros F, Galea M (2012) Influence diagnostics in generalized symmetric linear models. Comput Stat Data Anal 59:161–170

Wei CG, Tanner MA (1990) Posterior computations for censored regression data. J Am Stat Assoc 85:829–839

Wu L (2010) Mixed effects models for complex data. Chapman & Hall/CRC, Boca Raton