Communications for Statistical Applications and Methods
2022, Vol. 29, No. 3, 333–351

https://doi.org/10.29220/CSAM.2022.29.3.333
Print ISSN 2287-7843 / Online ISSN 2383-4757

The skew-*t* censored regression model: parameter estimation via an EM-type algorithm

Victor H. Lachos[1,a], Jorge L. Bazán[b], Luis M. Castro[cde], Jiwon Park[a]

[a]Department of Statistics, University of Connecticut, USA;
[b]Department of Applied Mathematics and Statistics, Universidade of São Paulo, Brazil;
[c]Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile;
[d]Millennium Nucleus Center for the Discovery of Structures in Complex Data, Santiago, Chile;
[e]Centro de Riesgos y Seguros UC, Pontificia Universidad Católica de Chile, Santiago, Chile

Abstract

The skew-*t* distribution is an attractive family of asymmetrical heavy-tailed densities that includes the normal, skew-normal and Student's-*t* distributions as special cases. In this work, we propose an EM-type algorithm for computing the maximum likelihood estimates for skew-t linear regression models with censored response. In contrast with previous proposals, this algorithm uses analytical expressions at the E-step, as opposed to Monte Carlo simulations. These expressions rely on formulas for the mean and variance of a truncated skew-*t* distribution, and can be computed using the `R` library `MomTrunc`. The standard errors, the prediction of unobserved values of the response and the log-likelihood function are obtained as a by-product. The proposed methodology is illustrated through the analyses of simulated and a real data application on Letter-Name Fluency test in Peruvian students.

Keywords: censored regression, EM-type algorithm, kurtosis, truncated moments, skewness, skew-*t* distribution

1. Introduction

Estimation of the censored regression (CR) model, or the Tobit model, has become quite common in the literature. However, as the Tobit model is based on normally distributed errors (N-CR), the maximum likelihood (ML) estimator is inconsistent if the underlying errors are not normally distributed. This inconsistency in the Tobit model led to the development of less sensitive estimators to the assumption of normality. Several authors have studied CR models involving response variables with heavier tails than the normal distribution in recent years. For instance, Arellano-Valle *et al.* (2012) and Massuia *et al.* (2015) have studied CR models based on the Student's-*t* distribution (T-CR). They demonstrated the robustness aspects of the T-CR model against outliers through extensive simulations by using the Expectation-Maximization (EM) algorithm, which is based on the first two moments of the truncated Student's-*t* distribution. However, the T-CR model is not appropriate when the data simultaneously present skewness and heavy tails.

Recently, Massuia *et al.* (2017) have established a new link between the CR model and asymmetrical heavy tails distributions by using the class of scale mixtures of skew-normal (SMSN) distributions

[1] Corresponding author: Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs CT-06269, USA. E-mail: hlachos@uconn.edu

Published 31 May 2022 / journal homepage: http://csam.or.kr

(SMSN-CR), which allows capturing, simultaneously, skewness and kurtosis and contains, as special cases, the normal (N), Student's-*t* (T), skew-*t* (ST) and skew-normal (SN) distributions. Under the Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) is introduced to carry out posterior inference. The method is implemented in the `R` package *BayesCR* (Garay *et al.*, 2017b). It is important to stress that the Massuia *et al.* (2017b)'s approach does not need the computation of the truncated moments of the SMSN family of distributions for implementing the estimation algorithm. The proposed MCMC procedure only needs to sample from a truncated normal distribution, conditional on the latent variables considered for each member of the SMSN class.

More recently and from the likelihood-based inference viewpoint, Mattos *et al.* (2018) proposed an efficient Monte Carlo EM (MCEM) algorithm to compute the ML estimates of the SMSN-CR model based on the stochastic approximation of the EM (SAEM) algorithm. However, by its nature, the SAEM algorithm is an expensive proposal, due to the combination of Monte Carlo simulations and other iterative procedures, which make this algorithm challenging to assess the convergence and time consuming. For this reason, our paper proposes an alternative analytically simple EM-type algorithm for computing ML estimates of the skew-*t* censored regression (ST-CR) model. We show that the E-step reduces to computing the first two moments of a certain truncated skew-*t* (TST) distribution. The general formulas for these moments were recently derived by Lachos *et al.* (2020), for which we will use the `MomTrunc` `R` package (Galarza *et al.*, 2020). The likelihood function is easily computed as a byproduct of the E-step and is used for monitoring convergence and for model selection. Furthermore, we consider a general information-based method for obtaining the asymptotic covariance matrix of the ML estimate. Our proposal has two advantages over the existing ones (*i.e.*, MCEM and SAEM algorithms). The first is that our EM-type algorithm is exact and does not require approximations at the E and M steps. The second one is that our approach is less time-consuming with the same precision (in terms of point estimation) related to its competitors.

The rest of the paper is organized as follows. In Section 2 we introduce some notation and outline the main results related to the SN, ST and TST distributions. In Section 3, the ST censored regression model (ST-CR) and related likelihood-based inference are presented, including the implementation of an EM-type algorithm called the Expectation/Conditional Maximization Either (ECME) algorithm (Liu and Rubin, 1994) for obtaining ML estimates of the parameters. Section 4 presents a simulation study to illustrate the performance of the proposed method. Section 5 discusses an application using a real data application on Letter-Name Fluency (LNF) test in Peruvian students, which is a standardized, individually administered test that provides a measure of Letter-Name Knowledge (LNK) and spelling abilities. Finally, Section 6 concludes with some discussion and possible directions for future research.

2. Notation and background

Throughout this paper, $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; and $\phi_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Phi}_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote its probability density function (pdf) and cumulative distribution function (cdf), respectively. When $p = 1$ we drop the index p. In this case, if $\mu = 0$ and $\sigma^2 = 1$, we write $\phi(\cdot)$ for the pdf and $\Phi(\cdot)$ for the cdf. In the same way, $\mathrm{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the p-variate Student's-t distribution with mean vector $\boldsymbol{\mu} \in \mathcal{R}^p$, scale matrix $\boldsymbol{\Sigma}$ (a positive definite matrix) and degrees of freedom $\nu \in (0, \infty)$; and $t_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $\boldsymbol{T}_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denote its pdf and cdf, respectively. Again, when $p = 1$ we drop the index p. In this case, if $\mu = 0$ and $\sigma^2 = 1$, we write $t(\cdot \mid \nu)$ for the pdf and $\boldsymbol{T}(\cdot \mid \nu)$ for the cdf. Finally, $\mathrm{Gamma}(\nu/2, \nu/2)$ denotes the Gamma distribution with scale and shape parameters equal to $\nu/2$.

We start defining the skew-normal (SN) distribution and then we introduce some useful properties.

As defined by Azzalini (1985), a random variable Z has a skew-normal distribution with location parameter $\mu \in \mathcal{R}$, scale parameter $\sigma^2 \in (0,\infty)$ and skewness parameter $\lambda \in \mathcal{R}$, denoted by $Z \sim \text{SN}(\mu,\sigma^2,\lambda)$, if its pdf is given by $\phi_{SN}(z \mid \mu,\sigma^2,\lambda) = 2\phi(z \mid \mu,\sigma^2)\Phi(\lambda(z-\mu)/\sigma)$. We denote the cdf of Z by $\Phi_{SN}(\cdot|\mu,\sigma^2,\lambda)$. If $\mu = 0$ and $\sigma^2 = 1$, we use $\phi_{SN}(\cdot \mid \lambda)$ and $\Phi_{SN}(\cdot \mid \lambda)$ for the pdf and cdf, respectively.

As proved by Azzalini and Dalla Valle (1996, Eqn. 2.11), the cdf of a skew-normal random variable is given by

$$\Phi_{SN}\left(z \mid \mu,\sigma^2,\lambda\right) = 2\boldsymbol{\Phi}_2\left(\frac{z-\mu}{\sigma}\mathbf{e}_1\middle|\mathbf{0},\boldsymbol{\Sigma}\right), \tag{2.1}$$

where $\mathbf{e}_1 = (1,0)^\top$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}, \quad \text{with } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}. \tag{2.2}$$

If $Z \sim \text{SN}(\mu,\sigma^2,\lambda)$, then a convenient stochastic representation is given by

$$Z = \mu + \Delta T + \Gamma^{\frac{1}{2}} T_1, \tag{2.3}$$

where $\Delta = \sigma\delta$, $\Gamma = (1-\delta^2)\sigma^2$, $T = |T_0|$, and T_0 and T_1 are independent standard normal random variables. Here, $|\cdot|$ denotes the absolute value. It is important to note that this stochastic representation is useful to generate random samples, obtaining moments as well as to derive other interesting properties.

The next definition introduces the stochastic representation of a skew-t random variable.

Definition 1. *Let $Z \sim SN(0,\sigma^2,\lambda)$ and $U \sim Gamma(\nu/2,\nu/2)$ assuming that Z and U are independent. We say that the distribution of $Y = \mu + U^{-1/2}Z$ is a skew-t distribution with location parameter $\mu \in \mathcal{R}$, scale parameter $\sigma^2 \in (0,\infty)$, shape parameter $\lambda \in \mathcal{R}$ and degrees of freedom $\nu \in (0,\infty)$. We use the notation $Y \sim ST(\mu,\sigma^2,\lambda,\nu)$.*

From Definition 1, the density of Y takes the following form (Azzalini and Capitanio, 2003)

$$\phi_{\text{ST}}\left(y \mid \mu,\sigma^2,\lambda,\nu\right) = 2t\left(y \mid \mu,\sigma^2,\nu\right)\boldsymbol{T}\left(\left(\frac{\nu+1}{d+\nu}\right)^{\frac{1}{2}} A\middle|\nu+1\right),$$

where $A = \lambda(y-\mu)/\sigma$ and $d = (y-\mu)^2/\sigma^2$. Some particular cases of the skew-t distribution are the skew-Cauchy distribution ($\nu = 1$) and the Student's-t distribution ($\lambda = 0$). Also, when $\nu \to \infty$, the skew-normal distribution arises as a limit case. Moreover, from Definition 1, the conditional distribution of Y given U is

$$Y \mid U = u \sim \text{SN}\left(\mu, u^{-1}\sigma^2, \lambda\right), \quad U \sim \text{Gamma}\left(\frac{\nu}{2},\frac{\nu}{2}\right). \tag{2.4}$$

Thus, from (2.1), we obtain the following expression for the cdf of a skew-t random variable:

$$\begin{aligned}\Phi_{\text{ST}}\left(y \mid \mu,\sigma^2,\lambda,\nu\right) &= 2\text{E}\left[\boldsymbol{\Phi}_2\left(U^{\frac{1}{2}}\frac{y-\mu}{\sigma}\mathbf{e}_1\right)\middle|\mathbf{0},\boldsymbol{\Sigma}\right] = 2\text{E}\left[P\left(\mathbf{X} \le U^{\frac{1}{2}}\frac{y-\mu}{\sigma}\mathbf{e}_1\middle|U\right)\right] \\ &= 2P\left(\frac{\mathbf{X}}{U^{\frac{1}{2}}} \le \frac{y-\mu}{\sigma}\mathbf{e}_1\right) = 2\,\boldsymbol{T}_2\left(\frac{y-\mu}{\sigma}\mathbf{e}_1\middle|\mathbf{0},\boldsymbol{\Sigma},\nu\right),\end{aligned}$$

where $\mathbf{e}_1 = (1,0)^\top$, $\mathbf{X} \sim \mathbf{N}_2(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{\Sigma}$ is given in (2.2).

In addition, from Basso *et al.* (2010), we have the following useful proposition related to a skew-*t* random variable.

Proposition 1. *Suppose $Y \sim ST(\mu, \sigma^2, \lambda, \nu)$. Then, for $r = 1, 2, \ldots$*

1. $Y \mid T = t,\ U = u \sim N\left(\mu + \Delta t, u^{-1}\Gamma\right),\ T \mid U = u \sim TN_{\lfloor 0,+\infty\rfloor}\left(0, u^{-1}\right),\ U \sim Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$,

2. $T \mid Y = y,\ U = u \sim TN_{\lfloor 0,+\infty\rfloor}\left(\mu_T, u^{-1}M_T^2\right)$,

3. $\gamma_r = E[U^r \mid Y = y] = \frac{2^r\Gamma\left(\frac{\nu+1+2r}{2}\right)(\nu+d)^{-r}}{\Gamma\left(\frac{\nu+1}{2}\right)} \frac{\boldsymbol{T}\left(\left(\frac{\nu+1+2r}{\nu+d}\right)^{\frac{1}{2}}A|\nu+1+2r\right)}{\boldsymbol{T}\left(\left(\frac{\nu+1}{d+\nu}\right)^{\frac{1}{2}}A|\nu+1\right)}$,

4. $\tau_r = E\left[U^{\frac{r}{2}}\frac{\phi\left(U^{\frac{1}{2}}A\right)}{\Phi\left(U^{\frac{1}{2}}A\right)}\middle| Y = y\right] = \frac{2^{\frac{r-1}{2}}\Gamma\left(\frac{\nu+1+r}{2}\right)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\left(\nu+d+A^2\right)^{\frac{\nu+1+r}{2}}} \frac{(\nu+d)^{\frac{\nu+1}{2}}}{\boldsymbol{T}\left(\left(\frac{\nu+1}{d+\nu}\right)^{\frac{1}{2}}A|\nu+1\right)}$,

5. $E[U\,T \mid Y = y] = \gamma_1\mu_T + M_T\tau_1$ *and* $E[U\,T^2 \mid Y = y] = \gamma_1\mu_T^2 + M_T^2 + \mu_T M_T\tau_1$,

where $M_T^2 = \Gamma/(\Gamma + \Delta^2)$, $\mu_T = \{\Delta/(\Gamma+\Delta^2)\}(y-\mu)$, $A = \lambda(y-\mu)/\sigma$, $d = (y-\mu)^2/\sigma^2$ and $TN_{\lfloor a,b\rfloor}(\mu, \sigma^2)$ denotes the truncated normal distribution in the interval $\lfloor a, b\rfloor$. Here $\lfloor a, b\rfloor$ means that each extreme of the interval can be either open or closed.

Now, we introduce a key concept in our development, namely the truncated skew-*t* (TST) distribution.

Definition 2. *Let $Y \sim ST(\mu, \sigma^2, \lambda, \nu)$, with $P(a < Y < b) > 0$ for some fixed $a < b$. A random variable X has a TST distribution in the interval $\lfloor a, b\rfloor$, denoted by $X \sim TST_{\lfloor a,b\rfloor}(\mu, \sigma^2, \lambda, \nu)$, if it has the same distribution as $Y|Y \in \lfloor a, b\rfloor$.*

As an obvious consequence of Definition 2, we have that the pdf of $X \sim \mathrm{TST}_{\lfloor a,b\rfloor}(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$ is given by:

$$\phi_{TST}\left(x \mid \mu, \sigma^2, \lambda, \nu; \lfloor a, b\rfloor\right) = \frac{\phi_{\mathrm{ST}}\left(x|\mu, \sigma^2, \lambda, \nu\right)}{\Phi_{\mathrm{ST}}\left(b|\mu, \sigma^2, \lambda, \nu\right) - \Phi_{\mathrm{ST}}\left(a|\mu, \sigma^2, \lambda, \nu\right)}\mathbb{I}_{\lfloor a,b\rfloor}(x),$$

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

An interesting property of the TST distribution is that it is a location-scale family. Indeed, let $X \sim \mathrm{TST}_{\lfloor \alpha,\beta\rfloor}(0, 1, \lambda, \nu)$, then $Y = \mu + \sigma X$ has a $\mathrm{TST}_{\lfloor a,b\rfloor}(\mu, \sigma^2, \lambda, \nu)$ distribution, where $a = \mu + \sigma\alpha$ and $b = \mu + \sigma\beta$. Consequently, for computing moments of Y, it is enough to compute the moments of X. Thus, the nth moment of Y is given by

$$\mathrm{E}\left[Y^n\right] = \sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!}\,\sigma^k\mu^{n-k}\mathrm{E}\left[X^k\right], \quad \text{for } n = 1, 2, 3\ldots.$$

Lachos *et al.* (2020) provide explicit expressions for the two first moments of the TST distribution. Let $X \sim TST_{[a,b]}(0, 1, \lambda, \nu)$, with $a < b$. Then,

$$\begin{aligned} E[X] &= \tau(a,b)\Big[L(1)\left\{\mathcal{E}_{\Phi}\left(-0.5, b_\lambda\right) - \mathcal{E}_{\Phi}\left(-0.5, a_\lambda\right)\right\} \\ &\quad - \left\{\mathcal{E}_{\phi_{SN}}\left(-0.5, b\right) - \mathcal{E}_{\phi_{SN}}\left(-0.5, a\right)\right\}\Big], \\ E[X^2] &= \tau(a,b)\Big[\left\{\mathcal{E}_{\Phi_{SN}}\left(-1, b\right) - \mathcal{E}_{\Phi_{SN}}\left(-1, a\right)\right\} - L(2)\left\{E_{\phi}\left(-1, b_\lambda\right) - \mathcal{E}_{\phi}\left(-1, a_\lambda\right)\right\} \\ &\quad - \left\{b\mathcal{E}_{\phi_{SN}}\left(-0.5, b\right) - a\mathcal{E}_{\phi_{SN}}\left(-0.5, a\right)\right\}\Big], \end{aligned}$$

where $\tau(a,b) = \{\Phi_{\mathrm{ST}}(b|\lambda,\nu) - \Phi_{\mathrm{ST}}(a|\lambda,\nu)\}^{-1}$, $a_\lambda = a(1+\lambda^2)^{1/2}$, $b_\lambda = b(1+\lambda^2)^{1/2}$, $L(s) = (2/\pi)^{1/2}\lambda/(1+\lambda^2)^{s/2}$, and

$$\begin{aligned} \mathcal{E}_{\phi_{SN}}(r,q) &= \frac{2^{r+1}\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+2r}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right)\left(q^2+\nu\right)^{\frac{\nu+2r}{2}}}\mathbf{T}\left(\left(\frac{2r+\nu}{q^2+\nu}\right)^{\frac{1}{2}}\lambda q\middle|2r+\nu\right), \\ \mathcal{E}_{\Phi_{SN}}(r,q) &= \frac{2^{r+1}\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^r}\mathbf{T}_2\left(\left(\frac{2r+\nu}{\nu}\right)^{\frac{1}{2}} q\mathbf{e}_1\middle|\mathbf{0}, \boldsymbol{\Sigma}, 2r+\nu\right), \\ \mathcal{E}_{\Phi}(r,q) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{2}{\nu}\right)^r\mathbf{T}\left(\left(\frac{2r+\nu}{\nu}\right)^{\frac{1}{2}} q\middle|2r+\nu\right), \\ \mathcal{E}_{\phi}(r,q) &= \frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\left(\frac{q^2+\nu}{2}\right)^{-\frac{\nu+2r}{2}}, \end{aligned}$$

with $\boldsymbol{\Sigma}$ given in (2.2). The first two moments of $Y = \mu + \sigma X$ are available through the `MomTrunc` R package (Galarza *et al.*, 2020). So far, this is the unique method to compute the moments of the TST, among others truncated skewed (multivariate) distributions.

3. The skew-*t* censored linear regression model

Let us consider a linear regression model where the responses are observed with errors which are independent and identically distributed (iid) according to some ST distribution, as follows:

$$Y_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \sigma\varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathrm{ST}(0, 1, \lambda, \nu),\ i = 1, \ldots, n, \tag{3.1}$$

where Y_i, $i = 1, \ldots, n$ are the 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 of covariates, such that x_{ij} is the value of the j^{th} explanatory variable for subject i. Under this setup, we have that $Y_i \overset{\text{ind}}{\sim} \mathrm{ST}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)$, $i = 1, \ldots, n$. To facilitate the mathematical derivations, we consider the case where left-censored observations can occur, that is, the observations are of the form:

$$Y_{\mathrm{obs}_i} = \begin{cases} \kappa_i, & \text{if } Y_i \le \kappa_i, \\ Y_i, & \text{if } Y_i > \kappa_i, \end{cases} \tag{3.2}$$

$i = 1, \ldots, n$, for some threshold point κ_i. The model defined in (3.1) and (3.2) is called the skew-*t* linear censored regression (ST-CR) model (Massuia *et al.*, 2017; Mattos *et al.*, 2018), for further details. Note that the right censored problem, as defined in the Introduction section, can be represented by a left censored problem by transforming the response Y_{obs_i} to $-Y_{\mathrm{obs}_i}$.

3.1. Parameter estimation via an EM-type algorithm

In what follows, we use the traditional convention denoting a random variable by an upper case letter and its realization by the corresponding lower case. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu)^\top$ be the vector with all parameters of the ST-CR model. Supposing that are 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 (log-like) function is given by

$$\begin{aligned}\ell(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}) &= \log\left(\prod_{i=1}^{n}\left[\Phi_{\text{ST}}\left(\kappa_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu\right)\right]^{\mathbb{I}_i}\left[\phi_{\text{ST}}\left(y_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu\right)\right]^{1-\mathbb{I}_i}\right)\\ &= \sum_{i=1}^{m}\log\left[\Phi_{\text{ST}}\left(\kappa_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu\right)\right] + \sum_{i=m+1}^{n}\log\left[\phi_{\text{ST}}\left(y_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu\right)\right],\end{aligned}$$

where $\mathbb{I}_i = 1$ if $y_i \leq \kappa_i$ and $\mathbb{I}_i = 0$ otherwise.

To estimate the parameters of the ST-CR model, an alternative is to maximize its log-likelihood function directly. However, this procedure can be quite cumbersome. Mattos *et al.* (2018) propose to compute the ML estimates by using SAEM algorithm. However, by its nature, MCEM is an expensive proposition, due to the combination of Monte Carlo simulation and other iterative procedures. Alternatively, in this work, we propose a simple EM-type algorithm (Dempster *et al.*, 1977) to obtain the ML estimates. To implement the EM algorithm, we need a representation of the model in terms of missing data. In the case of censoring, we can consider the unobserved y_i as a realization of the latent unobservable variable $Y_i \sim \text{ST}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)$, $i = 1, \ldots, m$. Here, the key is to consider the augmented data $\{\mathbf{y}_{\text{obs}}, y_1, \ldots, y_m, u_1, \ldots, u_n, t_1, \ldots, t_n\}$, that is, we treat the problem as if $\mathbf{y}_L = (y_1, \ldots, y_m)^\top$ were in fact observed. As a consequence, we can use the representation (2.4) to obtain the complete-data log-likelihood, given as

$$\ell_c\left(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}, \mathbf{t}\right) = C - \frac{n}{2}\log\Gamma + \sum_{i=1}^{n}\log u_i - \frac{1}{2\Gamma}\sum_{i=1}^{n}u_i\left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta} - \Delta t_i\right)^2 + \sum_{i=1}^{n}\log h(u_i|\nu),$$

where C is a constant that does not depend on the parameter of interest $\boldsymbol{\theta}$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$ and $h(\cdot|\nu)$ is the gamma density with scale and shape parameters equal to $\nu/2$.

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

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

where $\text{E}_{\boldsymbol{\theta}^{(k)}}$ means that the expectation is obtained using $\boldsymbol{\theta}^{(k)}$ instead of $\boldsymbol{\theta}$. Observe that the expression of the Q-function is completely determined by the knowledge of the expectations

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{(k)}\right) = \text{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^r Y_i^s \mid y_{\text{obs}_i}\right], \quad r,\ s = 0, 1, 2.$$

Thus, ignoring constants that do not depend on the parameter of interest, the Q-function can be written in a synthetic form as follows:

$$\begin{aligned}Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}\right) = -\frac{n}{2}\log\Gamma - \frac{1}{2\Gamma}\sum_{i=1}^{n}\Big[&\mathcal{E}_{02i}\left(\boldsymbol{\theta}^{(k)}\right) - 2\mathcal{E}_{01i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{00i}\left(\boldsymbol{\theta}^{(k)}\right)\left(\mathbf{x}_i^\top\boldsymbol{\beta}\right)^2\\ &+ \Delta^2\mathcal{E}_{20i}\left(\boldsymbol{\theta}^{(k)}\right) - 2\Delta\mathcal{E}_{11i}\left(\boldsymbol{\theta}^{(k)}\right) + 2\Delta\mathcal{E}_{10i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^\top\boldsymbol{\beta}\Big] + \sum_{i=1}^{n}\text{E}_{\boldsymbol{\theta}^{(k)}}\left[\log h(U_i|\nu) \mid y_{\text{obs}_i}\right].\end{aligned}$$

The M-step consists of maximization of $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ with respect to $\boldsymbol{\theta}$. To do so, we resort to a faster extension of the original EM called the ECME algorithm (Liu and Rubin, 1994), replacing the M step with a sequence of conditional maximization (CM) steps and maximizing the actual log-likelihood function with respect to ν. Due to the CM step below, it is not necessary to compute the expectations $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log h(U_i|\nu) \mid y_{\mathrm{obs}_i}]$. Thus, depending whether if the observation is censored or not and by using known properties of conditional expectation, the expectations involved in the Q-function will take specific, analytic and closed forms as follows:

1. *Uncensored observation case*. In this case, $Y_{\mathrm{obs}_i} = Y_i \sim \mathrm{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda, \nu)$ and from Proposition 1 (see also Basso *et al.*, 2010), we have that

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{(k)}\right) = y_i^s \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^r \mid y_i\right], \quad s,\ r = 0, 1, 2, \tag{3.3}$$

with

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i \mid y_i\right] &= \gamma_1\left(\boldsymbol{\theta}^{(k)}, y_i\right), \\
\mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i \mid y_i\right] &= \gamma_1\left(\boldsymbol{\theta}^{(k)}, y_i\right)\mu_{Ti}^{(k)} + M_T^{(k)}\tau_1\left(\boldsymbol{\theta}^{(k)}, y_i\right)
\end{aligned}$$

and

$$\mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i\, T_i^2 \mid y_i\right] = \gamma_1\left(\boldsymbol{\theta}^{(k)}, y_i\right)\mu_{Ti}^{2(k)} + M_T^{2(k)} + \mu_{Ti}^{(k)} M_T^{(k)}\tau_1\left(\boldsymbol{\theta}^{(k)}, y_i\right),$$

where $M_T^{2(k)}$, $\mu_{Ti}^{(k)}$, $\gamma_1(\boldsymbol{\theta}^{(k)}, y_i)$ and $\tau_1(\boldsymbol{\theta}^{(k)}, y_i)$ as defined in Proposition 1 with appropriate substitutions.

2. *Censored observation case*. Here, we have that $Y_{\mathrm{obs}_i} = \kappa_i$ if and only if $Y_i \le \kappa_i$. Then,

$$\begin{aligned}
\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{(k)}\right) &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^r Y_i^s \mid Y_i \le \kappa_i\right] \\
&= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[Y_i^s \mathrm{E}\left[U_i\left[T_i^r \mid U_i, Y_i\right] \mid Y_i\right] \mid Y_i \le \kappa_i\right], \quad r,\ s = 0, 1, 2,
\end{aligned} \tag{3.4}$$

where the conditional expectation in the second line of (3.4) is true because, if y_i were available, then it would be a realization of a $\mathrm{ST}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda, \nu)$ distribution and the expectation $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|Y_i, Y_i \le \kappa_i] = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|Y_i]$ in (3.3).

Finally, the expectation in (3.4) can be easily obtained by using the following Proposition:

Proposition 2. *For a censored observation* $i = 1, 2, \ldots, m$*, the conditional moments* $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{(k)})$, $r,\ s = 0, 1, 2$, *are given by*

$$\begin{aligned}
\mathcal{E}_{00i}\left(\boldsymbol{\theta}^{(k)}\right) &= E_{\boldsymbol{\theta}^{(k)}}[U_i \mid Y_i \le \kappa_i] = \frac{\Phi_{ST}\left(\kappa_i \mid \mathbf{x}_i^\top \boldsymbol{\beta}^{(k)}, \sigma^{2*(k)}, \lambda^{(k)}, \nu + 2\right)}{\Phi_{ST}\left(\kappa_i \mid \mathbf{x}_i^\top \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \lambda^{(k)}, \nu\right)}, \\
\mathcal{E}_{0ri}\left(\boldsymbol{\theta}^{(k)}\right) &= E_{\boldsymbol{\theta}^{(k)}}\left[U_i Y_i^r \mid Y_i \le \kappa_i\right] \\
&= \frac{\Phi_{ST}\left(\kappa_i \mid \mathbf{x}_i^\top \boldsymbol{\beta}^{(k)}, \sigma^{2*(k)}, \lambda^{(k)}, \nu + 2\right)}{\Phi_{ST}\left(\kappa_i \mid \mathbf{x}_i^\top \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \lambda^{(k)}, \nu\right)} E_{\boldsymbol{\theta}^{(k)}}\left[W_i^r\right],
\end{aligned}$$

$$
\begin{aligned}
\mathcal{E}_{10i}\left(\boldsymbol{\theta}^{(k)}\right) &= E_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i \mid Y_i \leq \kappa_i\right] \\
&= \frac{\Delta^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}\left(\mathcal{E}_{01i}\left(\boldsymbol{\theta}^{(k)}\right) - \mathcal{E}_{00i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}\right) + \sqrt{\frac{\Gamma^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}}\, c\left(\nu\right) W_{\Phi}(\boldsymbol{\theta}^{(k)}\right), \\
\mathcal{E}_{20i}(\boldsymbol{\theta}^{(k)}) &= E_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^2 \mid Y_i \leq \kappa_i\right] \\
&= \left(\frac{\Delta^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}\right)^2 \left(\mathcal{E}_{02i}\left(\boldsymbol{\theta}^{(k)}\right) - 2\mathcal{E}_{01i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)} + \left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 \mathcal{E}_{00i}\left(\boldsymbol{\theta}^{(k)}\right)\right) \\
&\quad + \sqrt{\frac{\Gamma^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}}\left(\frac{\Delta^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}\right) c(\nu) W_{\Phi}\left(\boldsymbol{\theta}^{(k)}\right)\left(E_{\boldsymbol{\theta}^{(k)}}(S_i) - \mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}\right) + \frac{\Gamma^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}, \\
\mathcal{E}_{11i}\left(\boldsymbol{\theta}^{(k)}\right) &= E_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i Y_i \mid Y_i \leq \kappa_i\right] \\
&= \left(\frac{\Delta^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}\right)\left(\mathcal{E}_{02i}\left(\boldsymbol{\theta}^{(k)}\right) - \mathcal{E}_{01i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}\right) + \sqrt{\frac{\Gamma^{(k)}}{\Delta^{2(k)}+\Gamma^{(k)}}} W_{\Phi}\left(\boldsymbol{\theta}^{(k)}\right) E_{\boldsymbol{\theta}^{(k)}}(S_i),
\end{aligned}
$$

where

$$
\begin{aligned}
W_{\Phi}\left(\boldsymbol{\theta}^{(k)}\right) &= \frac{\Phi_{ST}\left(\kappa_i \mid \mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}, \sigma_{**}^{2(k)}, 0, \nu+1\right)}{\Phi_{ST}\left(\kappa_i | \mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \lambda^{(k)}, \nu\right)}, \\
W_i &\sim TST_{\lfloor -\infty, \kappa_i \rfloor}\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}, \sigma^{2*(k)}, \lambda^{(k)}, \nu+2\right), \quad S_i \sim TST_{\lfloor -\infty, \kappa_i \rfloor}\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}^{(k)}, \sigma_{**}^{2(k)}, 0, \nu+1\right), \\
\sigma^{2*} &= \frac{\nu}{\nu+2}\sigma^{2(k)}, \quad \sigma_{**}^{2(k)} = \frac{\nu}{(\nu+1)(1+\lambda^{2(k)})}\sigma^{2(k)}, \quad \textit{and} \quad c(\nu) = \frac{2\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu(1+\lambda^2)\pi}}.
\end{aligned}
$$

Proof: The proof follows from the conditional expectation property given in (3.4) along with the conditional expectations given in Proposition 1. □

Thus, the EM algorithm for the ST-CR model, defined in (3.1) and (3.2), can be summarized in the following way:

1. E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$. For $i = 1, \ldots, n$.
   - If the observation i is uncensored then, for $r, s = 0, 1, 2$, compute $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{(k)})$ given in (3.3);
   - If the observation i is censored then, for $r, s = 0, 1, 2$, compute $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{(k)})$ in (3.4) by using Proposition 2.
2. CM-step: Update $\boldsymbol{\theta}^{(k)}$ by maximizing $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions

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

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

$$\nu^{(k+1)} = \arg\max_{\nu} \left\{ \sum_{i=1}^{n} \log \left[\Phi_{\mathrm{ST}} \left(\kappa_i \mid \mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \lambda^{(k+1)}, \nu \right) \right] + \sum_{i=m+1}^{n} \log \phi_{\mathrm{ST}} \left(y_i \mid \mathbf{x}_i^{\top} \boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \lambda^{(k+1)}, \nu \right) \right\}.$$

This process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}})$, like $|\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}_{\mathrm{obs}}) - \ell(\boldsymbol{\theta}^{(k)}|\mathbf{y}_{\mathrm{obs}})|$ or $|\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}_{\mathrm{obs}})/\ell(\boldsymbol{\theta}^{(k)}|\mathbf{y}_{\mathrm{obs}}) - 1|$, is small enough. The optimization step related to ν can be easily accomplished by using the `optim` routine in the `R` software after a double integration procedure. Note that $\sigma^{2(k)}$ and $\lambda^{(k)}$ can be obtained from $\Delta^{(k)}$ and $\Gamma^{(k)}$, by considering

$$\sigma^{2(k)} = \Gamma^{(k)} + \Delta^{2(k)} \quad \text{and} \quad \lambda^{(k)} = \frac{\Delta^{(k)}}{\sqrt{\Gamma^{(k)}}}. \tag{3.5}$$

Assuming complete data, *i.e.*, ignoring censoring, we used ordinary least squares (OLS) estimators as initial estimates of $\boldsymbol{\beta}^{(0)}$ and the moment estimator for $\sigma^{2(0)}$ by using the `R` function `lm()`. For $\lambda^{(0)}$, we considered twice the signal of the skewness coefficient of the residuals, and finally, $\nu^{(0)}$ was fixed at 3.

3.2. Model selection

To select an appropriate model from the candidates, we adopted the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978), the consistent AIC (CAIC) (Bozdogan, 1987) and the Hannan-Quinn information criterion (HQIC) (Burnham and Anderson, 2002).

Like the more popular AIC and BIC, which are the two most widely used model selection indices based on penalized likelihood and applicable for both nested and non-nested models, the model selection criteria considered in this work have the form $-2\ell(\hat{\boldsymbol{\theta}}) + \rho c_n$, where $\hat{\boldsymbol{\theta}}$ is the ML estimate obtained via the EM algorithm, $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{obs}})$ is the actual log-likelihood, ρ is the number of free parameters that have to be estimated in the model and the penalty term c_n is a convenient sequence of positive numbers that depends on the sample size n. Specifically we have AIC $= -2\ell(\hat{\boldsymbol{\theta}}) + 2\rho$, BIC $= -2\ell(\hat{\boldsymbol{\theta}}) + \rho \log(n)$, HQIC $= -2\ell(\hat{\boldsymbol{\theta}}) + 2\rho \log(\log(n))$ and CAIC $= -2\ell(\hat{\boldsymbol{\theta}}) + \rho(\log(n) + 1)$. A lower AIC, BIC, HQIC or CAIC value indicates that a closer fit of the model to the data.

3.3. Approximate standard errors

The standard errors of the ML estimates can be approximated by the inverse of the observed information matrix. Unfortunately, in our case, there is no a closed-form available for this matrix. Thus, we consider the same strategy used by Garay *et al.* (2017a) to get approximate standard errors of the parameter estimates based on the empirical information matrix by assuming ν known. Let $\mathbf{Y}_{\mathrm{obs}}$ be the vector of observed data. Then, considering $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \lambda)$, $\mathbf{Y}_{comp} = (\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_L, \mathbf{U}, \mathbf{T})^{\top}$ and relations described in the Equation (3.5), the empirical information matrix is defined as

$$\mathbf{I}_e\left(\boldsymbol{\theta} \mid \mathbf{y}_{\mathrm{obs}}\right) = \sum_{i=1}^{n} \mathbf{s}\left(y_{\mathrm{obs}_i} \mid \boldsymbol{\theta}\right) \mathbf{s}^{\top}\left(y_{\mathrm{obs}_i} \mid \boldsymbol{\theta}\right) - \frac{1}{n} \mathbf{S}\left(\mathbf{y}_{\mathrm{obs}} \mid \boldsymbol{\theta}\right) \mathbf{S}^{\top}\left(\mathbf{y}_{\mathrm{obs}} \mid \boldsymbol{\theta}\right),$$

where $\mathbf{S}^\top(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}(y_{\text{obs}_i} \mid \boldsymbol{\theta})$. It can be noted from the result of Louis (1982) that the individual score can be determined as

$$\mathbf{s}(y_{\text{obs}_i} \mid \boldsymbol{\theta}) = \frac{\partial Q_i\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}\right)}{\partial \boldsymbol{\theta}}, \quad i = 1, \ldots, n.$$

Thus, substituting the ML estimates of $\boldsymbol{\theta}$ in (3.6), the empirical information matrix $\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}})$ is reduced to

$$\mathbf{I}_e\left(\hat{\boldsymbol{\theta}} \mid \mathbf{Y}_{\text{obs}}\right) = \sum_{i=1}^{n} \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top,$$

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

$$\hat{\mathbf{s}}_{\boldsymbol{\beta}_i} = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\left(\mathbf{x}_i \mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} - \hat{\sigma}\frac{\hat{\lambda}}{\sqrt{1+\hat{\lambda}^2}}\mathbf{x}_i\mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\right),$$

$$\hat{\mathbf{s}}_{\sigma_i^2} = -\frac{1}{2\hat{\sigma}^2} + \frac{1+\hat{\lambda}^2}{2\hat{\sigma}^4}\left(\mathcal{E}_{02i}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2\right) - \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{2\hat{\sigma}^3}\left(\mathcal{E}_{11i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right),$$

$$\hat{\mathbf{s}}_{\lambda_i} = \frac{\hat{\lambda}}{1+\hat{\lambda}^2} - \frac{\hat{\lambda}}{\hat{\sigma}^2}\left(\mathcal{E}_{02i}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01i}(\hat{\boldsymbol{\theta}})\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2\right) + \frac{1+2\hat{\lambda}^2}{\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\left(\mathcal{E}_{11i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right) - \hat{\lambda}\mathcal{E}_{20i}\left(\hat{\boldsymbol{\theta}}\right),$$

for $i = 1, \ldots, n$, where the conditional expectations $\mathcal{E}_{rsi}(\omega^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i T_i^r Y_i^s \mid Y_i \leq \kappa_i]$, $r, s = 0, 1, 2$, can be easily obtained directly from the proposed EM algorithm.

4. Simulation studies

In this section, we present three simulations studies. In the first one, we study the performance of EM estimates under different censoring proportions. The second simulation study investigates whether the model comparison measures, AIC, BIC, CAIC, and HQIC determine the best-fitting model to the simulated data. The third study compares the performance of ML estimates obtained through EM and SAEM algorithms. For each scenario, 100 Monte Carlo samples were generated, and the data were artificially generated from the ST-CR model defined in (3.1) and (3.2), with $\boldsymbol{x}_i^T = (1, x_{i1}, x_{i2})$, such that $x_{i1} \sim U(1, 5)$ and $x_{i2} \sim U(-1, 1)$, for $i = 1, \ldots, n$.

4.1. Performance of EM estimates

This first simulation study is built to analyze the impact of the censoring level on the estimates of the ST-CR model. We consider three censoring levels, say, 5%, 10% and 20%. In a first part of this simulation study, each Monte Carlo sample is composed of a $n = 1000$ random draws from the ST-CR model, with the following parameters: $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (1, 2, 3)^T$, $\sigma^2 = 1$, $\lambda = -2$ and $\nu = 4$. From the generated data, in order to define the censored cases, we follow two steps: First, we sorted the observations in ascending order, and then the cases were defined as censored if the percentage of the observations below the fixed threshold was equal to the corresponding censoring level.

We computed the average values (MC Mean) and standard deviations (MC sd) across the estimates of the 100 Monte Carlo samples. Also, the average values of the approximate standard errors of the

Table 1: Results based on 100 simulated ST samples with $n = 1000$

	Cens		
	5%	10%	20%
β_0 (1)	1.03	1.03	1.03
IM SE	0.08	0.08	0.09
MC sd	0.09	0.09	0.09
β_1 (2)	2.00	2.00	2.00
IM SE	0.02	0.02	0.02
MC sd	0.02	0.02	0.03
β_2 (3)	3.00	3.00	2.99
IM SE	0.04	0.04	0.05
MC sd	0.17	0.17	0.15
σ^2(1)	1.09	1.09	1.09
IM SE	0.10	0.10	0.11
MC sd	0.04	0.04	0.04
λ (−2)	−2.21	−2.21	−2.18
IM SE	0.28	0.28	0.29
MC sd	0.37	0.36	0.31
ν (4)	4.27	4.33	4.37
IM SE	-		-
MC sd	0.76	0.77	0.79

The reported values are the MC means, and the MC sd are the standard deviations from fitting the ST-CR with different settings of censoring proportions. IM SE is the average value of the approximate standard error obtained through the information-based method. Cens indicates the censoring rate.

EM estimates were computed through the method described in Subsection 3.3. Table 1 shows the results for different censoring levels. We can see from this table that the EM-type algorithm recovers the original values of the parameters for all levels of censoring, closely. We also observe from this table that the estimates of the standard errors (IM SE) and MC standard deviations (MC sd) give relatively close results, showing that the proposed asymptotic approximation for the variance of the EM estimates is reliable.

The second part of this study is devoted to analyze the finite sample properties of the EM estimates under the same parameters setting. In this case, we fix the censoring level increasing the sample size, say, $n = 200$, 500, and 1000 for each Monte Carlo sample. Figure 1 shows boxplots of the EM estimates under the ST-CR model with censoring rate of 10%. We can see that the increased sample size corresponds to a decreasing bias and variability of the parameter estimates revealing that the ML estimates obtained via the proposed EM algorithm have consistent asymptotic properties. This tendency remains when we change the censoring rate.

4.2. Regression models comparison

In this simulation scheme, we compare the performance of some classic model comparison criteria to select the appropriate model among four different right-censored regression models, namely, the normal (N-CR), Student's-*t* (T-CR), skew-normal (SN-CR) and skew-*t* (ST-CR) models. In order to enhance the reliability of the model selection, we used well-known criteria, AIC, BIC, CAIC, and HQIC, explained in Subsection 3.2 to select the proper model. Table 2 shows the arithmetic average of these comparison measures. Note that all the measures favor the ST-CR model, indicating that the ST-CR model is generally more robust to deviations from the model assumptions and fits better than other candidates when neither is the true generating model. Figure 2 represents the AIC, BIC, CAIC, and HQIC values for each sample and model with a left-censoring level of 10%. From this figure, we can see that in 100% of the cases all the model selection criteria select the ST-CR model, as expected.

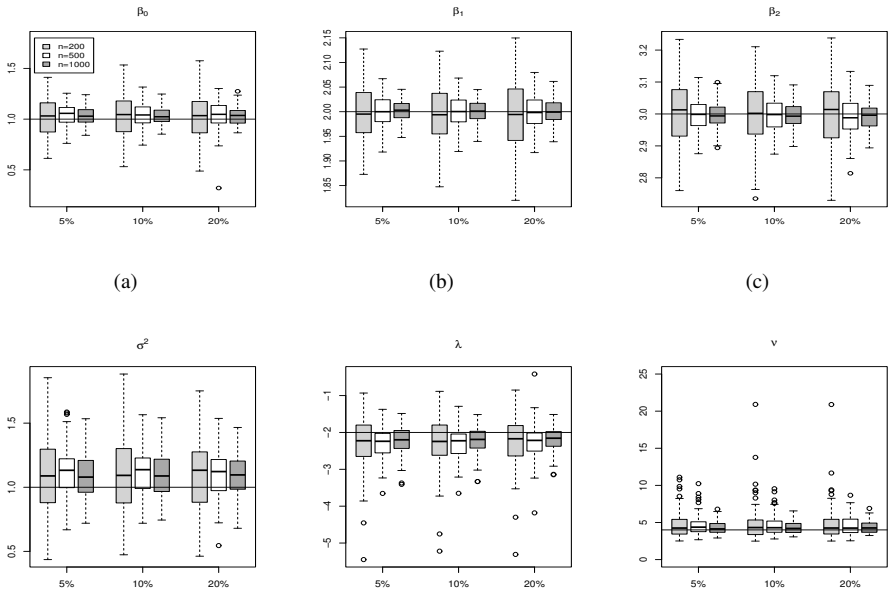

Figure 1: *Boxplot of the estimates of $\beta_0, \beta_1, \beta_2, \sigma_2, \lambda$ and ν (line indicates the true value of the parameter) for the ST-CR model. Legend in (a): the first three boxplots correspond to the left-censoring level of $c = 5\%$ with $n = 200$, 500, 1,000, respectively. The boxplots for $c = 10\%, 20\%$ are defined as similar way.*

Table 2: Model selection criteria for comparing N-CR, SN-CR, T-CR and ST-CR models under three different censoring proportions

Censoring	Criteria	N-CR	SN-CR	T-CR	ST-CR
5%	log-like	−1400.080	−1320.192	−1320.621	−1274.855
	AIC	2808.160	2652.385	2651.242	2561.711
	BIC	2827.791	2681.831	2675.781	2591.157
	CAIC	2831.791	2687.831	2680.781	2597.157
	HQIC	2815.621	2663.577	2660.568	2572.903
10%	log-like	−1327.306	−1261.528	−1257.535	−1217.715
	AIC	2662.612	2535.056	2525.070	2447.429
	BIC	2682.243	2564.502	2549.609	2476.876
	CAIC	2686.243	2570.502	2554.609	2482.876
	HQIC	2670.073	2546.247	2534.396	2458.621
20%	log-like	−1188.030	−1139.859	−1132.080	−1099.499
	AIC	2384.060	2291.719	2274.161	2210.997
	BIC	2403.691	2321.165	2298.699	2240.444
	CAIC	2407.691	2327.165	2303.699	2246.444
	HQIC	2391.521	2302.911	2283.487	2222.189

4.3. Comparison between SAEM and EM algorithms

In this simulation study, we compare the ML estimates obtained via the proposed EM-algorithm and SAEM (Mattos *et al.*, 2018). The parameter setting is the same as the previous simulation with censoring level 20% and the sample size for each Monte Carlo sample is $n = 500$. The initial values of

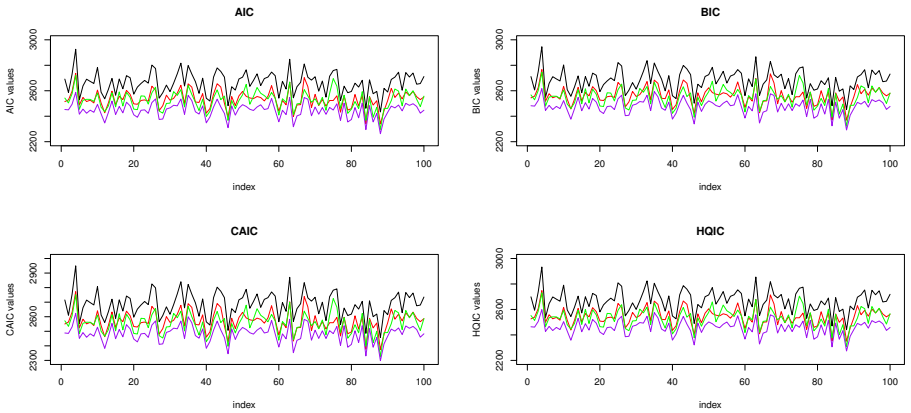

Figure 2: *AIC, BIC, CAIC and HQIC values for 100 samples and left-censoring level of* 10%. *Black curve: N-CR, red curve: T-CR, green curve: SN-CR and purple curve: ST-CR.*

Table 3: MC means and MC standard deviations (in parentheses) of the ML estimates obtained through the EM and the SAEM algorithms

Algorithm	β_1	β_2	σ^2	λ	ν	log-like	Time (min)
SAEM	1.9866	2.9997	1.1900	−2.2938	5.7270	−541.8204	1.4450
	(0.0364)	(0.0672)	(0.1727)	(0.4411)	(1.6433)	(18.4327)	(0.7873)
EM	1.9880	3.0022	1.2102	−2.3056	6.1596	−542.0523	0.5565
	(0.0367)	(0.0673)	(0.1859)	(0.4230)	(1.7823)	(18.6224)	(0.5450)

the EM algorithm were obtained as discussed in Subsection 3.1 with a maximum number of iterations $max.iter = 400$. In addition, we chose the parameters $m = 20$ (Monte Carlo sample size), $c = 0.3$ (cut-off point) and $S = 400$ (maximum number of iterations) for the SAEM implementation. The results are given in Table 3, where we can see that the ML estimates of the parameters obtained through the EM and the SAEM algorithms are close, as expected. However, the SAEM algorithm requires, in average, 3 times longer that the EM algorithm to reach the convergence. Consequently, we can conclude that our proposal provides the same accuracy as other competitors but with less computational time for obtaining ML estimates.

5. Application

LNK has been identified to be an important predictor of learning to read, spelling abilities, phonological awareness and intelligence (see, for example, Foulin (2005), Ritchey and Speece (2006) and references there in). However, not only LNK is considered as a predictor of spelling achievements, but also the speed of children in letter naming. This is another letter-name related skill closely associated to reading achievement (Cronin and Carver, 1998). A frequently used procedure considered by school teachers to measure LNK is based on LNF. In this case, teachers administer timed 1-minute fluency assessments to children, and then compare the results with established norms in order to determine how the students are performing in this task and if they are at risk for future academic problems. Observe that LNF is a continuous variable related to the average of letters read correctly in an interval of time and not a discrete variable.

In general, LNF can be considered a measure of risk (Marston and Magnusson, 1988) because it is highly predictive of later reading success. It is also included as an indicator for students with

Table 4: LNF data. Summary statistics of LNF response and proportions for explanatory variables by uncensored and censored groups

Variable	Statistics	Uncensored	Censored	Total
LNF	n	479	32	511
	mean	30.674	39.844	31.249
	median	28	31	29
	sd	16.228	20.718	16.669
	min	1	28	1
	max	98	99	99
	skewness	1.265	2.124	1.412
	kurtosis	2.567	3.075	3.131
Zone	rural	0.313	0.312	0.313
Grade	3rd	0.290	0.281	0.290
Gender	female	0.501	0.500	0.501

lower scores, who may require additional instructional support on their Early Literacy Skills (ELS). Additionally, LNF has been recognized as the best predictor of future reading and spelling abilities in children, and analyzing it, a benchmark can be obtained to determine the minimum level of satisfactory progress of spelling achievements by a student. However, LNF presents some challenges. Due to the time limit, some students may not complete the assessments while others will finish them before the set time. This situation generates that, if the teachers are interested in the average of the letters/words/sentences/paragraphs correctly read, the time limit restriction has to be taken into account. In other words, the students who finish the task in less than 1-minute could read more letters/words/sentences/paragraphs, and the reported mean corresponding to the correctly read objects could not be the real one. The situation described above corresponds to the typical case of right censored observations.

In this paper, we analyze LNF data set from the Early Grade Reading Assessment (EGRA), which is part of (RTI-FDA, 2008) instrument. Here, the response variable LNF is defined as the number of correctly letters read by the students in one minute and it presents right censored observations. Consequently, if we are interested in the mean of the LNF response for one group, this quantity could be underestimated due to the presence of censored observations. For that reason, a censored regression model able to take into account observation lying below or above a threshold could be more appropriated for estimating the true mean of the LNF response for different groups of interest.

Table 4 shows a summary of the response variable in the presence (Censored) and absence (Uncensored) of censoring. Note that 6.26% of the observations are censored, and the mean and the standard deviation of them are higher in comparison with uncensored observations. Therefore, the mean of 31.3 for the LNF response (letters correctly read per minute) showed in Table 4 seems to be underestimated. Further, we can observe some degree of right skewness and kurtosis on the response variable reveling a departure from the normal distribution and then other distribution must be considered for the data.

Additionally, three socio-demographic covariates for the students are considered in the LNF data, namely, the *Zone* where the respondent lives (0 = Urban, 1 = Rural), *Grade* (0 = 2nd grade, 1 = 3rd grade) and *Gender* (0 = Male, 1 = Female). Note that the proportions for rural zone, 3rd grade and female on the censored and uncensored groups are similar. Since the explanatory variables are all dummy variables, we expect to understand the effects of these covariates on the LNF response.

To investigate the behavior of the LNF in the Peruvian sample, and taking into account the censoring effect, we fit four different right-censored regression models, namely, the normal (N-CR), Student's-t (T-CR), skew-normal (SN-CR) and skew-t (ST-CR) models. Particularly, the proposed

Table 5: LNF data: Model selection criteria

Model	# parameters	log-like	AIC	BIC	CAIC	HQIC
N-CR	5	−2058.962	4127.923	4149.105	4154.105	4136.227
SN-CR	6	−2007.738	4027.475	4052.894	4058.894	4037.44
T-CR	6	−2023.104	4058.208	4083.626	4089.626	4068.173
ST-CR	7	−1995.336	4004.672	4034.327	4041.327	4016.298

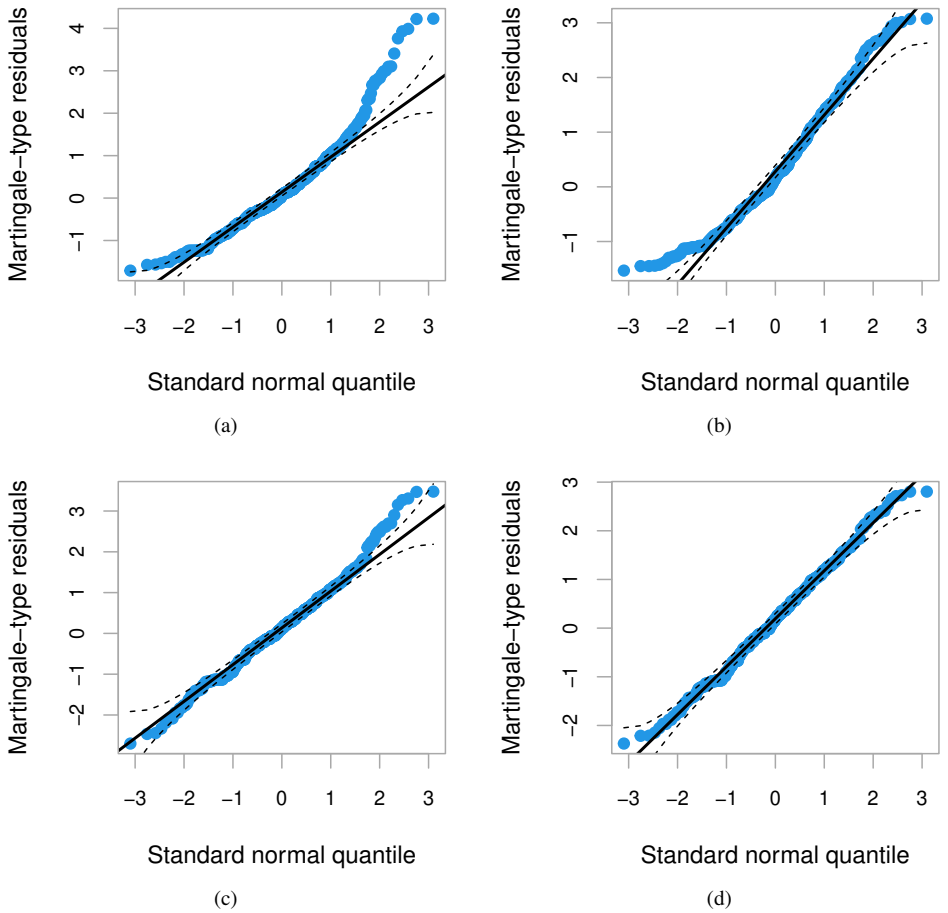

Figure 3: *LNF data: Martingale-type residuals under the (a) N-CR, (b) T-CR, (c) SN-CR and (d) ST-CR models. The envelope is obtained through simulation.*

censored model is given by:

$$Y_i = \beta_0 + \beta_1 \text{Zone}_i + \beta_2 \text{Grade}_i + \beta_3 \text{Gender}_i + \sigma\epsilon_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where the error term ϵ_i is independent for all $i = 1, \ldots, n$ and follows some of the distributions previously mentioned with location and scale parameters equal to 0 and 1, respectively.

Table 5 compares the fit of the four proposed models using the model selection criteria discussed in Subsection 3.2. Considering these results, we observe that the ST-CR model outperforms all its competitors (N-CR, SN-CR and T-CR). This conclusion is also corroborated by Figure 3, where we

Table 6: LNF data. ML estimates under SN-CR and ST-CR models. SE is the corresponding standard error

	SN-CR		ST-CR	
Parameter	Estimate	SE	Estimate	SE
Intercept β_0	12.706	1.323	14.813	1.349
Zone β_1	−4.675	1.510	−4.713	1.291
Grade β_2	5.972	1.323	6.192	1.253
Gender β_3	−0.159	1.206	0.032	1.137
σ^2	659.596	44.616	377.295	44.522
λ	3.800	0.599	2.710	0.509
ν	-	-	5.184	-

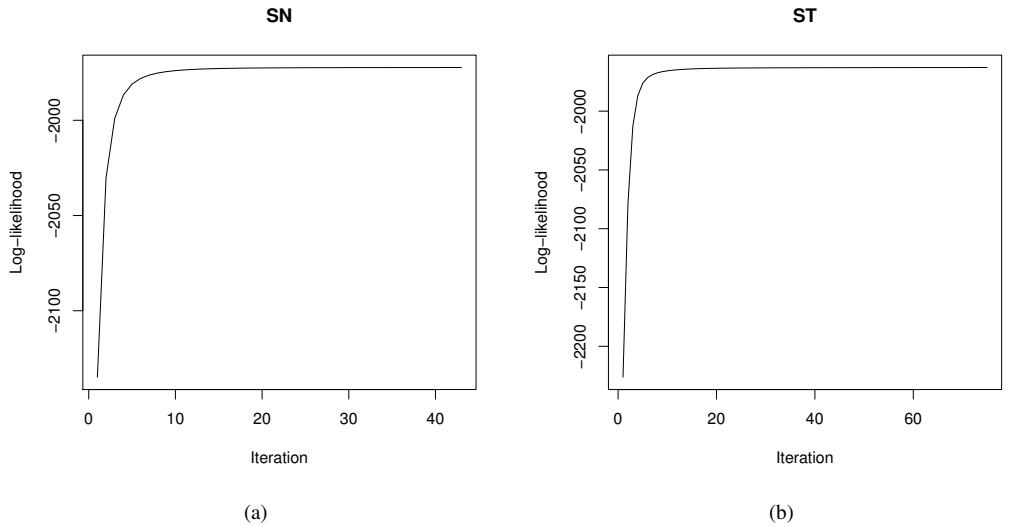

Figure 4: *LNF data: Plot of the log-likelihood vs. Iterations of the EM-type algorithm under the (a) SN-CR and (b) ST-CR models.*

present the plot of the Martingale-type residuals including a simulated envelope (Mattos *et al.*, 2018) for the T-CR, SN-CR and ST-CR models (See also Figure 1). It can be seen that the skew-*t* distribution accommodates the observations in a better way than its competitors.

Table 6 reports the ML estimates for the parameters of the best two models according to the previous analyses, *i.e.*, the SN-CR and ST-CR models, along with their corresponding standard errors calculated via the empirical information matrix (Subsection 3.3).

Using a tolerance of $\epsilon = 10^{-5}$ as stopping criterion, the algorithm attained convergence in 32 iterations and 2.58 seconds for the SN-CR model, and 27.61 seconds and 25 iterations for the ST-CR model, this on an Intel Core i7-6700, CPU @ 3.40GHz computer with 8 GB of RAM. The monotone convergence of the parameter estimates via the proposed EM algorithm is illustrated in Figure 4, where we can see that the log-likelihood increases in successive iterates of the EM algorithm.

Note from Table 6 that the covariate *Gender* can be considered non significant. The covariate *Zone* has a negative effect in favor of rural schools and covariate *Grade* has a positive effect in the students of the 3rd grade. As a conclusion, it can be stated that students from urban schools and the 3rd grade read correctly more letters than students from rural schools and the second grade.

By considering these results, we fitted an additional ST-CR model without the covariate *Gender*.

The resulting fitted model is the following:

$$\hat{Y}_i \sim \text{ST}\left(\hat{\mu} = 14.831 - 4.714\text{Zone}_i + 6.192\text{Grade}_i,\ \hat{\sigma}^2 = 377.164,\ \hat{\lambda} = 2.709,\ \hat{\nu} = 5.182\right),$$

$i = 1, \ldots, 511$, where the model comparison criteria are: Log-Likelihood = −1995.335, AIC = 4002.671, BIC = 4028.089, CAIC = 4034.089 and HQ = 4012.635.

It is worth mentioning that, under this model, the mean of the number of correct letters read by the Peruvian students in the sample, under the censored model, is 32.29, while when the censoring mechanism is omitted, the mean is 31.2 That is, if censored is not taken into consideration for modeling the LNF, the mean of the number of correct letters can be sub estimated. Since this kind of test is used frequently, we recommend to incorporate the censoring pattern in order to estimate the statistics of Fluency conveniently. Additionally, the summary of the estimated response variable is Min=1 and Max = 122.81 where quartile 1 is 21, median is 29 and quartile 3 is 42. Values which can be used to propose different criteria to classify the LNF.

6. Conclusions

In this paper, a novel exact EM-type algorithm for skew-*t* censored linear regression model is developed. In contrast with previous developments (MCEM and MCMC algorithms), the proposed EM-type algorithm uses analytical expressions at the E-step, that rely on formulas of the mean and variance of a truncated skew-*t* distribution. These formulas have been developed by Lachos *et al.* (2020) and are available in the `R` package `MomTrunc` (Galarza *et al.*, 2018). As an added benefit of the proposal, the EM likelihood sequence is monotonic and the difficulties in assessing convergence, which face Monte Carlo algorithms, are avoided. Furthermore, simulations studies and the analysis of the LNF dataset provides strong evidence about the implementation of the EM-type algorithm for fitting the ST-CR model. The method proposed in this paper is implemented in `R`, and the code is available for download from `GitHub` repository (`https://github.com/hlachos/skewt-censored`) .

Finally, some extensions of the current work includes the multivariate ST-CR model and finite mixture of censored skew-*t* models (Azzalini and Capitanio, 1999; Lachos *et al.*, 2017). An in-depth investigation of such extensions is beyond the scope of the present paper, but certainly an interesting topic for future research.

Acknowledgement

We thank the associate editor and two anonymous referees for their important comments and suggestions which lead to an improvement of this paper. Jorge L. Bazán acknowledges support from FAPESP-Brazil (Grant 2021/11720-0). L. M. Castro acknowledges support from Grant FONDECYT 1220799 from the Chilean government.

References

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

Arellano-Valle RB, Castro LM, González-Farías G and Muñoz-Gajardo KA (2012). Student-t censored regression model: properties and inference, *Statistical Methods & Applications*, **21**, 453–473.

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

Azzalini A and Capitanio A (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society: Series B*, **61**, 579–602.

Azzalini A and Capitanio A (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 367–389.

Azzalini A and Dalla Valle A (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.

Basso RM, Lachos VH, Cabral CR, and Ghosh P (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions, *Computational Statistics & Data Analysis*, **54**, 2926–2941.

Bozdogan H (1987). Model selection and Akaikeś Information Criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**, 345–370.

Burnham KP and Anderson DR (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlag.

Cronin V and Carver P (1998). Phonological sensitivity, rapid naming and beginning reading, *Applied Psycholinguistics*, **19**, 447–461.

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

Foulin JN (2005). Why is letter-name knowledge such a good predictor of learning to read?, *Reading and Writting*, **38**, 129–155.

Galarza CM, Kan R, and Lachos VH (2020). MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions, R package version 5.69, http://cran.r-project.org/package=MomTrunc

Garay AM, Lachos VH, Bolfarine H, and Cabral CRB (2017a). Linear censored regression models with scale mixtures of normal distributions, *Statistical Papers*, **58**, 247–278.

Garay AW, Massuia MB, and Lachos VH (2017b). *BayesCR: Bayesian Analysis of Censored Regression Models Under Scale Mixture of Skew Normal Distributions*. R package version 2.1, http://cran.r-project.org/package=BayesCR

Lachos VH, Garay A, and Cabral CR (2020). Moments of truncated skew-normal/independent distributions, *Brazilian Journal of Probability and Statistics*, **34**, 478–494.

Lachos VH, Moreno EJL, Chen K, and Cabral CRB (2017). Finite mixture modeling of censored data using the multivariate Student-t distribution, *Journal of Multivariate Analysis*, **159**, 151–167.

Liu C and Rubin DB (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, *Biometrika*, **81**, 633–648.

Louis TA (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**, 226–233.

Marston D and Magnusson D (1988). *Alternative Educational Delivery Systems: Enhancing Instructional Options for All Students*, (Ed. Graden J. and Zins, J. and Curtis, M.), Pages = 137–172, Publisher = National Association of School Psychology, Title = Curriculum-based measurement: District level implementation, Washington, DC.

Massuia MB, Cabral CRB, Matos LA and Lachos VH (2015). Influence diagnostics for Student-t censored linear regression models, *Statistics*, **49**, 1074–1094.

Massuia MB, Garay AM, Lachos VH and Cabral CRB (2017). Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions, *Statistics and its Interface*, **10**, 425-439,

Mattos TdB, Garay AM, and Lachos VH (2018). Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions, *Journal of Applied Statistics*,

45, 2039–2066.

Ritchey K and Speece D (2006). From letter names to word reading: The nascent role of sublexical fluency, *Contemporary Educational Psychology*, **31**, 301-327.

RTI-FDA (2008). Snapshot of School Management Effectiveness: Peru Pilot Study (Technical report), USAID.

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

Received October 19, 2021; Revised February 24, 2022; Accepted April 19, 2022