Received: Added at production | Revised: Added at production | Accepted: Added at production
DOI: xxx/xxxx

ARTICLE TYPE

A skew-t quantile regression for censored and missing data

Christian E. Galarza Morales[1] | Victor H. Lachos[2] | Marcelo Bourguignon[3]

[1]Departamento de Estadística, Escuela Superior Politecnica del Litoral, Guayaquil, Ecuador
[2]Department of Statistics, University of Connecticut, Storrs CT-06269, USA
[3]Departamento de Estatística, Universidade do Rio Grande do Norte, Brazil

Correspondence
Victor H. Lachos
Email: hlachos@uconn.edu

Quantile regression has emerged as an important analytical alternative to the classical mean regression model. However, the analysis could be complicated by the presence of censored measurements due to a detection limit of equipment in combination with unavoidable missing values arising when, for instance, a researcher is simply unable to collect an observation. Another complication arises when measures depart significantly from normality, for instance, in the presence of skew heavy-tailed observations. For such data structures, we propose a robust quantile regression for censored and/or missing responses based on the skew-t distribution. A computationally feasible EM-based procedure is developed to carry out maximum likelihood estimation within such a general framework. Moreover, the asymptotic standard errors of the model parameter are explicitly obtained via the information-based method. We illustrate our methodology by using simulated data and two real data sets.

KEYWORDS:
Censored regression models, EM algorithm, Quantile regression model, Student's-t distribution.

1 | INTRODUCTION

In many studies limited or censored data are collected. This occurs, in several practical situations, for reasons such as limitations of measuring equipment or from experimental design. Hence, the exact true value is recorded only if it falls within an interval range, so, the responses can be either left, right or interval censored. Moreover, missing values are often encountered due to various reasons, such as, withdrawal from a study, information is not reported or in experimental studies a researcher is simply unable to collect an observation. An additional complication arises in the presence of heavy-tails or atypical observations.

In general, censored regression (CR) models are based on the development of the so called Tobit model, which is constructed in terms of the normal assumption. In recent years, several authors have studied CR models for statistical modeling of censored dataset involving observed variables with heavier tails than the normal distribution. For instance, Massuia, Cabral, Matos, and Lachos (2015) proposed an extension of the CR model with normal errors (N-CR) to Student's-t (T-CR) errors. More recently, Massuia, Garay, Cabral, and Lachos (2017) and Mattos, Garay, and Lachos (2018) proposed a robust CR model where the observational errors follow a scale mixture of skew-normal (SMSN-CR) distribution proposed by Branco and Dey (2001). Note however that the majority of these methods focus on mean regression which is not a good measure of centrality when the conditional distribution of the response variable is skewed or multimodal, and therefore the mean regression estimator may be inadequate to make inferences about the shapes of these distributions. In contrast to the mean regression model, quantile regression (QR) can provide an overall assessment of the covariate effects at different quantiles of the outcome.

Inferences for QR models are based on the well-known asymmetric Laplace distribution (ALD), which posses a hierarchical representation that facilitates the implementation of efficient algorithms, such as, Gibbs sampling and the EM algorithm, as described in Lachos, Chen, Abanto-Valle, and Azevedo (2015) and Galarza, Castro, Louzada, and Lachos (2020), respectively. A drawback of the ALD distribution is that it is not differentiable at zero, which could lead to problems of numerical instability. Wichitaksorn, Choy, and Gerlach (2014) introduced a generalized class of skew densities

(SKD) for the analysis of QR that provides competing solutions to the ALD-based formulation. The robust SKD class of distributions includes as special cases the skew-normal (SKN), skew Student's-t (SKT), skew-Laplace, skew-slash and the skew-contaminated normal distribution. This class of distributions was also studied by Galarza, Lachos, Barbosa Cabral, and Castro Cepero (2017), who introduced an efficient EM-type algorithm for QR considering the error to follow a member of the SKD class, so that, the SKD-QR model is defined. Moreover, the proposed inferential methods were implemented in the R package `lqr` (Galarza, Benites, &, Lachos 2015). Even though some solutions have been proposed in the literature to deal with the problem of the misspecification of the error distribution in QR, so far, to the best of our knowledge, there is no attempt on extending the SKD-QR model to deal with censoring and/or missing information.

In this article, we propose the use of the SKT distribution in the context of QR for censored data so that the SKT-QRC model is defined and a fully likelihood-based approach is carried out, including the implementation of an exact EM-type algorithm for the ML estimation. Like Lachos, Moreno, Chen, and Cabral (2017), we show that the E-step reduces to computing the first two moments of truncated univariate Student's t distributions. The general formulas for these moments were derived efficiently by Galarza, Kan, and Lachos (2020), for which we use the `MomTrunc` package in `R`. The likelihood function is easily computed as a byproduct of the E-step and is used for monitoring convergence and for model selection such as the AIC (Akaike 1974) and BIC (Schwarz 1978). Furthermore, we consider a general information-based method for obtaining the asymptotic covariance matrix of the ML estimate. The method proposed in this paper is implemented in the `R` package `LQR`, which is available for download from CRAN.

The rest of the paper proceeds as follows. Section 2 presents the construction of the SKD family of distributions as a scale mixture of skew normal distribution and some important propositions and properties of this family. Section 3 introduces the QR model and the EM algorithm for ML estimation as well as evaluation of standard errors. Section 4 presents simulation studies of finite sample performance and robustness of the proposed method. Two real examples are discussed in Section 5. Finally, Section 6 closes the paper, sketching some future research directions.

2 | PRELIMINARIES

In this section, we present a skew Student's-t (SKT) distribution introduced in Wichitaksorn et al. (2014), which its skewness parameter characterize the pth quantile, being useful in the framework of QR. The SKT distribution is said to belong to a family of zero quantile skewed distributions, namely SKD as discussed by Galarza et al. (2017). Thus, in the following we present some definitions where we explain first the fundamental concept of the limiting case, said SKN distribution, and the construction of a SKT random variate as a scale normal mixture of a SKN distribution.

2.1 | The SKN distribution

As defined in Wichitaksorn et al. (2014), we say that a random variable X has a skew-normal (SKN) distribution with location parameter μ, scale parameter $\sigma > 0$ and skewness parameter $p \in (0,1)$, if its probability density function (*pdf*) is given by

$$\tilde{\phi}(x|\mu,\sigma,p) = 2\left[p\,\phi\left(x\middle|\mu,\frac{\sigma^2}{4(1-p)^2}\right)\mathbb{I}\{x\le\mu\} + (1-p)\,\phi\left(x\middle|\mu,\frac{\sigma^2}{4p^2}\right)\mathbb{I}\{x>\mu\}\right], \tag{1}$$

where $\phi(\cdot|\mu,\sigma^2)$ represents the *pdf* of the normal distribution with mean μ and variance σ^2 ($N(\mu,\sigma^2)$) and $\mathbb{I}\{\cdot\}$ denotes the indicator function. By convention, we shall write $X \sim \mathrm{SKN}(\mu,\sigma,p)$. Note that, $P(X\le\mu) = p$ and $P(x>\mu) = 1-p$, which allows a direct application to QR problems. When $p = 0.5$ we have the symmetric $N(\mu,\sigma^2)$ distribution. Also, the *pdf* in (1) is constructed as a mixture of two truncated normal distributions with weights p and $1-p$ respectively. Galarza et al. (2017) conveniently expressed the SKN's pdf as

$$\tilde{\phi}(x|\mu,\sigma,p) = \frac{4p(1-p)}{\sqrt{2\pi\sigma^2}}\exp\left\{-2\rho_p^2\left(\frac{x-\mu}{\sigma}\right)\right\},$$

where $\rho_p(\cdot)$ is the so called check (or loss) function in quantile regression framework, which is defined by $\rho_p(u) = u(p-\mathbb{I}\{u<0\})$. As a finite mixture of normal distributions, the SKN distribution inherits some convenient properties from the normal distribution. For instance, the SKN also belong to the location scale family. Furthermore, its moments can be easily computed as a mixture of two truncated normal moments. Other properties of the SKN variate such as stochastic representation and moments have been extensively studied in Galarza et al. (2017).

2.2 | The SKT distribution

Following the same idea as in (1), the Student's-t version can be written as a mixture of two symmetric Student-t densities. It is said that Y follows a SKT distribution, say $Y \sim \mathrm{SKT}(\mu,\sigma,p,\nu)$, if Y has a pdf given by (Galarza et al. 2017)

$$\tilde{t}(x|\mu,\sigma,p) = \frac{4p(1-p)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi\sigma^2}}\left\{1+\frac{4}{\nu}\rho_p^2\left(\frac{y-\mu}{\sigma}\right)\right\}^{-\frac{\nu+1}{2}},$$

with location, scale and skewness parameters μ, σ and p respectively, and degrees of freedom ν. The neat formula for the pdf of a SKT distribution was described in Galarza et al. (2017). Figure 1 shows the densities for the normal (a) and the Student-t (b) distribution with degrees of freedom $\nu = 4$ under different settings of skewness $\mathsf{p} = \{0.1, 0.5, 0.9\}$. As natural, when $\mathsf{p} = 0.5$, we recover the symmetric Student-t distribution. Besides, when $\nu \to \infty$, it follows that the SKT converges to a SKN distribution.

Furthermore, a SKT variate holds a stochastic representation (SR) as a scale mixture of a SKN distribution (Andrews & Mallows 1974), that is, Y can be represented stochastically as as

$$\mathsf{Y} = \mu + \sigma \mathsf{U}^{-1/2}\mathsf{Z}, \tag{2}$$

where $\mathsf{Z} \sim \mathrm{SKN}(0, 1, \mathsf{p})$ and $\mathsf{U} \sim \mathsf{G}(\nu/2, \nu/2)$, this last being a positive random variable known as mixture variable. A direct consequence of this definition is that, as the SKN distribution, $\mathsf{P}(\mathsf{Y} \leq \mu) = \mathsf{p}$ and $\mathsf{P}(\mathsf{Y} > \mu) = 1 - \mathsf{p}$. It is straightforward that the conditional distribution of Y given $\mathsf{U} = \mathsf{u}$ is $\mathrm{SKN}(\mu, \mathsf{u}^{-1/2}\sigma, \mathsf{p})$. Then, integrating out U, we have that the marginal *pdf* of Y is

$$\tilde{t}(y|\mu,\sigma,p,\boldsymbol{\nu}) = \int_0^\infty \frac{4p(1-p)}{\sqrt{2\pi\sigma^2}} u^{-1/2} \exp\left\{-2u\rho_p^2\left(\frac{y-\mu}{\sigma}\right)\right\} \mathrm{G}(\nu/2,\nu/2)\mathrm{d}u, \tag{3}$$

where $\mathrm{G}(\alpha, \lambda)$ represents the pdf of the Gamma distribution with mean α/λ. The SR given in 2 is highly convenient in order to propose a EM-based estimation model as will be described in the next section. As natural, several distributions can be obtained from different specifications of the mixture variable U. For further details, we refer to Galarza et al. (2017).

In this work, we focus on the SKT case since this heavy-tailed distribution showed to have a quite good performance in terms of parameter recovery and model fitting, being preferred using likelihood-based criteria. It is worth mentioning that we also do prefer the SKT distribution to the other heavy-tailed contenders, as the Laplace is not differentiable at μ, the slash distribution lacks of a closed form for its pdf and the contaminated normal has one more parameter to deal with.

2.3 | Truncated SKT distribution

Let $\mathsf{Y} \sim \mathsf{SKT}(\mu, \sigma, \mathsf{p}, \nu)$ as before. First, for two real numbers a, b such that $\mathsf{a} < \mathsf{b}$, we denote the the probability $\mathbb{P}(\mathsf{a} \leq \mathsf{Y} \leq \mathsf{b})$ as $\tilde{\mathsf{T}}(\mathsf{a}, \mathsf{b}|\mu, \sigma, \mathsf{p}, \nu)$, that is,

$$\tilde{T}(a,b|\mu,\sigma,p,\nu) = \int_a^b \tilde{t}(y|\mu,\sigma,p,\nu)\mathrm{d}y.$$

In the event that Y is restricted to take values in the interval (a, b), we may define its truncated version, say, $\mathsf{W} \stackrel{\mathsf{d}}{=} \mathsf{Y} \mid (\mathsf{a} \leq \mathsf{Y} \leq \mathsf{b})$, which is said to follow a truncated SKT (TSKT) distribution, denoted by $\mathsf{W} \sim \mathsf{TSKT}(\mu, \sigma, \mathsf{p}, \nu; (\mathsf{a}, \mathsf{b}))$, where its pdf is given by

$$f(w|\mu,\sigma,p,\nu;(a,b)) = \frac{\tilde{t}(w|\mu,\sigma,p,\nu)}{\tilde{T}(a,b|\mu,\sigma,p,\nu)}\mathbb{I}\{a \leq w \leq b\}. \tag{4}$$

Since the SKT pdf is also a mixture of two Student's-t components, the probability $\tilde{\mathsf{T}}(\mathsf{a}, \mathsf{b}|\mu, \sigma, \mathsf{p}, \nu)$ in (4), can be computed as a linear combination of two Student's-t probabilities, that is, $\mathbb{P}(\mathsf{a} \leq \mathsf{Y} \leq \mathsf{b}) = \mathsf{p}\mathcal{P}_1 + (1 - \mathsf{p})\mathcal{P}_2$, where $\mathcal{P}_1 = \mathbb{P}(\mathsf{a} \leq \mathsf{Y}_1 \leq \min(\mu, \mathsf{b}))$ and $\mathcal{P}_2 = \mathbb{P}(\max(\mathsf{a}, \mu) \leq \mathsf{Y}_2 \leq \mathsf{b})$, with Y_1 and Y_2 distributed as $\mathsf{Y}_1 \sim \mathsf{t}(\mu, \frac{\sigma^2}{4(1-\mathsf{p})^2}, \nu)$ and $\mathsf{Y}_2 \sim \mathsf{t}(\mu, \frac{\sigma^2}{4\mathsf{p}^2}, \nu)$, respectively. In general, the kth moment of W is given by

$$\mathbf{E}[\mathsf{W}^\mathsf{k}] = \frac{1}{\mathsf{p}\mathcal{P}_1 + (1-\mathsf{p})\mathcal{P}_2}\left\{\mathsf{p}\mathcal{P}_1\mathbf{E}[\mathsf{Y}_1^\mathsf{k}|\mathsf{a} \leq \mathsf{Y}_1 \leq \min(\mu,\mathsf{b})] + (1-\mathsf{p})\mathcal{P}_2\mathbf{E}[\mathsf{Y}_2^\mathsf{k}|\max(\mathsf{a},\mu) \leq \mathsf{Y}_2 \leq \mathsf{b}]\right\}, \tag{5}$$

which only depends on the kth moment of two doubly truncated Student-t distributions. For instance, the `R` package `MomTrunc` provides doubly truncated moments for several skew-elliptical multivariate distributions, including the univariate Student's-t as particular case. For a discussion of the various properties of the truncated multivariate Student's-t distribution, we refer to Ho, Lin, Chen and Wang (2012).

3 | SKT QUANTILE REGRESSION FOR CENSORED DATA

First, consider y_i, $\mathsf{i} = 1, \ldots, \mathsf{n}$, to be a fully observed response variable and $\mathbf{x}_\mathsf{i}$ a $\mathsf{k} \times 1$ vector of covariates for the ith observation, and let $\mathsf{Q}_{\mathsf{y}_\mathsf{i}}(\mathsf{p}|\mathbf{x}_\mathsf{i})$ be the pth $(0 < \mathsf{p} < 1)$ QR function of y_i given $\mathbf{x}_\mathsf{i}$. Suppose that the relationship between this quantile and $\mathbf{x}_\mathsf{i}$ can be modeled as $\mathsf{Q}_{\mathsf{y}_\mathsf{i}}(\mathsf{p}|\mathbf{x}_\mathsf{i}) = \mathbf{x}_\mathsf{i}^\top\boldsymbol{\beta}_\mathsf{p}$, where $\boldsymbol{\beta}_\mathsf{p}$ is a $(\mathsf{k} \times 1)$ vector of unknown parameters of interest. Then, we consider the quantile regression model given by

$$\mathsf{y}_\mathsf{i} = \mathbf{x}_\mathsf{i}^\top\boldsymbol{\beta}_\mathsf{p} + \epsilon_\mathsf{i}, \quad \mathsf{i} = 1, \ldots, \mathsf{n}, \tag{6}$$

where ϵ_i is the error term whose distribution (with density, say, $f_p(\cdot)$) is restricted to have the pth quantile equal to zero, that is, $\int_{-\infty}^{0} f_p(\epsilon_i)d\epsilon_i = p$, and consequently $P(y_i \le \mathbf{x}_i^\top \boldsymbol{\beta}_p) = p$. The density $f_p(\cdot)$ is often left unspecified in the classical literature. Thus, quantile regression estimation for $\boldsymbol{\beta}_p$ proceeds by minimizing

$$\widehat{\boldsymbol{\beta}}_p = \arg\min_{\boldsymbol{\beta}_p \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_p\big(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p\big), \tag{7}$$

where $\rho_p(\cdot)$ is the same check function in section before and $\widehat{\boldsymbol{\beta}}_p$ is the quantile regression estimate for $\boldsymbol{\beta}_p$ at the pth quantile. The case where $p = 0.5$ corresponds to the median regression.It is important to stress that there is a connection between the minimization of the sum in (7) and the maximum-likelihood theory, since to minimize (7) is equivalent to maximize the likelihood when data follows a distribution belonging to the family of zero conditional quantile SKD; for our particular case, a SKT distribution. It can be observed that the check function in (3) is inversely proportional to the *pdf* and therefore to the likelihood.

For example, suppose that $\epsilon_i \sim \mathrm{SKN}(0, \sigma, p)$, $i = 1, \ldots, n$ are independent random variables (and so $y_i \sim \mathrm{SKN}(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma, p)$). Then the likelihood function is given by

$$L(\boldsymbol{\beta}_p, \sigma|\mathbf{y}) = \frac{4^n p^n (1-p)^n}{(2\pi\sigma^2)^{n/2}} \exp\left\{-2\sum_{i=1}^{n} \rho_p^2\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p}{\sigma}\right)\right\},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top$.

For a fixed value of σ, the maximization of the resulting likelihood in the SKN family with respect to the parameter $\boldsymbol{\beta}_p$ is equivalent to the minimization of the objective function in (7). Therefore, the relationship between the check function and this family of distributions can be used to reformulate the QR method within the likelihood framework.

Besides, for $Y \sim \mathrm{SKT}(\mu, \sigma, p, \nu)$, it follows from Galarza et al. (2017) that the distance $D = \rho(\frac{Y-\mu}{\sigma})$ is a positive random variable following a half-Student-t distribution with $\sigma^2 = 1/4$ and same degrees of freedom ν, that is, $D \sim \mathrm{Ht}(1/4, \nu)$.

Note that the larger d_i, the more influential will be the observation i over the likelihood and therefore we might consider it as an influential value. Also, we use these distributions to create envelopes plots in order to graphically assess the fit of several distributions to the data as seen in the application section. In next subsection, we propose an estimation method for the censored QR model based on the EM algorithm for obtaining the ML estimates.

We are interested in the case where censored and/or missing observations can occur, that is, the response Y_i may not be fully observed due to censoring. So, we define (V_i, C_i) the observed data for the ith observation, with V_i being either an uncensored observation ($V_{ik} = V_{0i}$) or the interval censoring level ($V_{ik} \in [V_{1ik}, V_{2ik}]$), and C_i is a censoring indicator, satisfying

$$C_i = \begin{cases} 1 & \text{if } V_{1i} \le Y_i \le V_{2i}\,, \\ 0 & \text{if } Y_i = V_{0i}\,, \end{cases} \tag{8}$$

for all $i \in \{1, \ldots, n\}$, i.e., $C_i = 1$ if Y_i is located within a specific interval. In this case, the model in (6) considering $\epsilon_i \sim \mathrm{SKT}(\mu, \sigma, p, \nu)$ along with (8) defines the Student-t interval censored quantile regression model (hereafter, the SKT-QRC model). For instance, left censoring structure causes truncation from the lower limit of the support of the distribution, since we only know that the true observation Y_i is greater than or equal to the observed quantity V_{1i}. Moreover, missing observations can be handled easily by considering $V_{1i} = -\infty$ and $V_{2i} = +\infty$.

3.1 | Parameter estimation via the EM 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 $\boldsymbol{\theta} = (\boldsymbol{\beta}_p^\top, \sigma^2)^\top$ be the vector with all parameters of the SKT-QRC model, assuming that ν is known. Supposing that are m censored/missing values of the characteristic of interest, we can partition the observed sample, namely $\mathbf{y}_{\mathrm{obs}}$, in two subsamples of m censored and $n - m$ uncensored values. Then, the marginal log-likelihood function is given by

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}_{obs}) = \sum_{i=1}^{m} \log\left[\tilde{T}(v_{1i}, v_{2i} \mid \mathbf{x}_i^\top \boldsymbol{\beta}_p, \sigma, \nu)\right] + \sum_{i=m+1}^{n} \log\left[\tilde{t}(y_i \mid \mathbf{x}_i^\top \boldsymbol{\beta}_p, \sigma, \nu)\right]. \tag{9}$$

To estimate the parameters of the SKT-QRC model, an alternative is to maximize its log-likelihood function directly. However, this procedure can be quite cumbersome in terms of computational effort and we should carry mechanisms for dealing with the censored and missing data. Alternatively, our choice is to use the EM algorithm, a classical, reliable, widespread and general framework developed by Dempster, Laird, and Rubin (1977) to obtain ML estimates when the data has missing/censored observations and/or latent variables. The main features of EM algorithm is the ease of implementation and the stability of monotone convergence.

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 $Y_i \sim \mathrm{SKT}(\mathbf{x}_i^\top \boldsymbol{\beta}_p, \sigma, \nu)$, $i = 1, \ldots, m$. Here, the key is to consider the augmented data $\mathbf{Y}_{\mathrm{comp}} = (\mathbf{V}, \mathbf{C})^\top$ where $\mathbf{V} = (V_1, \ldots, V_n)^\top$ and $\mathbf{C} = (C_1, \ldots, C_n)^\top$, that is, we treat the problem as if the latent variables $\mathbf{Y}_{\mathrm{comp}}$ and

$\mathbf{U} = (\mathsf{U}_1, \ldots, \mathsf{U}_n)^\top$ were in fact observed. As a consequence, in light of the hierarchical representation given in (2), the QR model defined in (6) can be expressed as

$$\mathsf{Y}_i|\mathsf{U}_i = \mathsf{u}_i \sim \mathsf{SKN}(\mathbf{x}_i^\top \boldsymbol{\beta}_p, \mathsf{u}_i^{-1/2}\sigma, \mathsf{p}),$$
$$\mathsf{U}_i \sim \mathsf{G}(\nu/2, \nu/2).$$

Then, based on the SR above, the complete data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}_p, \sigma)$ given $(\mathbf{V}, \mathbf{C}, \mathbf{U})$, ignoring some additive constant terms, is given by $\ell_c(\boldsymbol{\theta}|\mathbf{V}, \mathbf{C}, \mathbf{U}) = \sum_{i=1}^n \ell_c(\boldsymbol{\theta}|\mathsf{v}_i, \mathsf{c}_i, \mathsf{u}_i)$, where

$$\ell_c(\boldsymbol{\theta}|\mathsf{v}_i, \mathsf{c}_i, \mathsf{u}_i) = \log\phi\left(\mathsf{y}_i \,\middle|\, \mathbf{x}_i^\top\boldsymbol{\beta}_p, \frac{\mathsf{u}_i^{-1}\sigma^2}{4(1-\mathsf{p})^2}\right)\mathbb{I}\{\mathsf{y}_i \le \mathbf{x}_i^\top\boldsymbol{\beta}_p\} + \log\phi\left(\mathsf{y}_i \,\middle|\, \mathbf{x}_i^\top\boldsymbol{\beta}_p, \frac{\mathsf{u}_i^{-1}\sigma^2}{4\mathsf{p}^2}\right)\mathbb{I}\{\mathsf{y}_i > \mathbf{x}_i^\top\boldsymbol{\beta}_p\} + \log\, \mathsf{h}(\mathsf{u}_i|\nu),$$

for $i = 1, \ldots, n$ and where $h(\cdot|\nu)$ represents the Gamma density with scale and shape parameters equal to $\nu/2$. Denoting by $\xi_i = (1 - \mathsf{p})\,\mathbb{I}\{\mathsf{y}_i \le \mathbf{x}_i^\top\boldsymbol{\beta}_p\} + \mathsf{p}\,\mathbb{I}\{\mathsf{y}_i > \mathbf{x}_i^\top\boldsymbol{\beta}_p\}$, the complete log-likelihood $\ell_c(\boldsymbol{\theta}|\mathbf{V}, \mathbf{C}, \mathbf{U})$ can be written as

$$\ell_c(\boldsymbol{\theta}|\mathbf{V}, \mathbf{C}, \mathbf{U}) = \sum_{i=1}^n \log\phi\left(\mathsf{y}_i \,\middle|\, \mathbf{x}_i^\top\boldsymbol{\beta}_p, \frac{\sigma^2}{4\xi_i^2\mathsf{u}_i}\right) + \sum_{i=1}^n \log\, \mathsf{h}(\mathsf{u}_i|\nu). \tag{10}$$

In what follows the superscript (k) will indicate the estimate of the related parameter at the stage k of the algorithm. The E step of the EM algorithm requires evaluation of the so-called Q-function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathbf{E}[\ell_c(\boldsymbol{\theta}|\mathbf{V}, \mathbf{C}, \mathbf{U})|\mathbf{y}, \boldsymbol{\theta}^{(k)}]$. Assuming for the moment that ν is known, and ignoring constants do not depend on the parameter of interest, the Q-function can be written as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \propto -n\log\sigma - \frac{2}{\sigma^2}\sum_{i=1}^n\left\{\xi_i^2\left(\widehat{\mathsf{uy}_i^2} - (\mathsf{x}_i^\top\boldsymbol{\beta}_p)(2\widehat{\mathsf{uy}_i} - \mathsf{x}_i^\top\boldsymbol{\beta}_p\widehat{\mathsf{u}_i})\right)\right\} + \sum_{i=1}^n \mathbf{E}\left[\log\, \mathsf{h}(\mathsf{u}_i|\nu)|\mathsf{y}_i, \boldsymbol{\theta}^{(k)}\right]. \tag{11}$$

Observe that the expression of the Q-function (11) is completely determined by the knowledge of the expectations of the form $\widehat{\mathsf{uy}_i^r} = \mathbf{E}[\mathsf{U}_i\mathsf{Y}_i^r|\mathsf{V}_i, \mathsf{C}_i, \boldsymbol{\theta}^{(k)}]$, for $r = \{0, 1, 2\}$. It is important to stress that, depending whether if the observation is censored or not, the expectations involved in the Q-function will take specific closed forms as follows:

1. Uncensored case. In this case, $\mathsf{Y}_{\mathrm{obs}_i} = \mathsf{Y}_i \mid (\mathsf{V}_i, \mathsf{C}_i) \sim \mathsf{SKT}(\mathbf{x}_i^\top\boldsymbol{\beta}_p, \sigma, \nu)$ and from STAT, it follows that $\mathsf{U}_i|\mathsf{Y}_i \sim \mathsf{G}\left((\nu+1)/2, (\nu + 4\xi_i^2 z_i^2)/2\right)$, where $z_i = (y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_p)/\sigma_i$. Then

$$\widehat{\mathsf{u}_i} = \mathbf{E}[\mathsf{U}_i|\mathsf{Y}_i, \boldsymbol{\theta}^{(k)}] = \frac{\nu+1}{\nu + 4\xi_i^2 z_i^2},$$
$$\widehat{\mathsf{uy}_i^r} = \mathbf{E}[\mathsf{U}_i\mathsf{Y}_i^r|\mathsf{Y}_i, \boldsymbol{\theta}^{(k)}] = \widehat{\mathsf{u}_i}\mathsf{y}_i^r.$$

2. Censored case. Here, we have that $\mathsf{Y}_{\mathrm{obs}_i} = \mathsf{v}_{0i} \in (\mathsf{v}_{1i}, \mathsf{v}_{2i})$ and then $\mathsf{Y}_i \mid (\mathsf{V}_i, \mathsf{C}_i) \sim \mathsf{TSKT}(\mathbf{x}_i^\top\boldsymbol{\beta}_p, \sigma, \nu; (\mathsf{v}_{1i}, \mathsf{v}_{2i}))$. Hence, it is easy to show that

$$\widehat{\mathsf{u}_i} = \mathbf{E}[\mathsf{U}_i|\mathsf{V}_i, \mathsf{C}_i, \boldsymbol{\theta}^{(k)}] = \frac{\tilde{\mathsf{T}}(\mathsf{v}_{1i}, \mathsf{v}_{2i} \mid \mathbf{x}_i^\top\boldsymbol{\beta}_p, \sqrt{\nu/(\nu+2)}\sigma, \nu+2)}{\tilde{\mathsf{T}}(\mathsf{v}_{1i}, \mathsf{v}_{2i} \mid \mathbf{x}_i^\top\boldsymbol{\beta}_p, \sigma, \nu)},$$
$$\widehat{\mathsf{uy}_i^r} = \mathbf{E}[\mathsf{U}_i\mathsf{Y}_i^r|\mathsf{V}_i, \mathsf{C}_i, \boldsymbol{\theta}^{(k)}] = \widehat{\mathsf{u}_i}\widehat{\mathsf{y}_i^r},$$

with

$$\widehat{y_i^r} = \mathbf{E}[Y_i^r|v_{1i} \le Y_i \le v_{2i}], \tag{12}$$

where $\mathsf{Y}_i \sim \mathsf{TSKT}(\mathbf{x}_i^\top\boldsymbol{\beta}_p, \sqrt{\nu/(\nu+2)}\sigma, \nu+2; (\mathsf{v}_{1i}, \mathsf{v}_{2i}))$, which can be computed using (5).

The proposed EM algorithm can be summarized in the following steps:

1. ***E-step***: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, compute the expectations $\widehat{\mathsf{uy}_i^r}$ for $r = 0, 1, 2$ based on $\widehat{\mathsf{y}_i}$, $\widehat{\mathsf{y}_i^2}$ and $\widehat{\mathsf{u}_i}$, this last using the corresponding expression for censored/missing and observed data.

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

$$\widehat{\boldsymbol{\beta}_p}^{(k+1)} = (\mathbf{X}^\top\widehat{\Omega}^{(k)}\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\Omega}^{(k)}\widehat{\mathbf{w}}^{(k)},$$
$$\widehat{\sigma^2}^{(k+1)} = \frac{4}{n}\mathrm{tr}\left\{\widehat{\Omega}^{(k)}\left(\widehat{\mathbf{w}\mathbf{w}^\top}^{(k)} - \left(2\widehat{\mathbf{w}}^{(k)} - \mathbf{X}\boldsymbol{\beta}_p^{(k+1)}\right)\left(\mathbf{X}\boldsymbol{\beta}_p^{(k+1)}\right)^\top\right)\right\},$$

where $\widehat{\mathbf{w}} = (\widehat{y_1}, \ldots, \widehat{y_m}, y_{m+1}, \ldots, y_n)^\top$, $\widehat{\mathbf{w}\mathbf{w}^\top} = \mathrm{diag}(\widehat{y_1^2}, \ldots, \widehat{y_m^2}, y_{m+1}^2, \ldots, y_n^2)$, $\mathbf{X}$ is the design matrix and Ω being a $n \times n$ diagonal matrix, with elements $\xi_i^2\,\widehat{\mathsf{u}_i}$, $i = 1, \ldots, n$. After the M-step, we will update the parameter $\boldsymbol{\nu}$ by maximizing the marginal log-likelihood function (9), obtaining

$$\widehat{\nu}^{(k+1)} = \arg\max{}_\nu \sum_{i=1}^n \ell(\boldsymbol{\beta}_p^{(k+1)}, \sigma^{2(k+1)}, \nu \mid \mathbf{y}_{\mathrm{obs}}).$$

In practice, the EM algorithm iterates until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, like $||\ell(\boldsymbol{\theta}^{(k+1)})-\ell(\boldsymbol{\theta}^{(k)})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)})/\ell(\boldsymbol{\theta}^{(k)})-1||$, is small enough. We have use the solution for (7) as an initial estimate of β_p, and ordinary least square for σ^2 reaching convergence in a fraction of a second.

3.2 | 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 Galarza et al. (2017) to get approximate standard errors of the parameter estimates based on the empirical information matrix. Let $\mathbf{Y}_{\rm obs}$ be the vector of observed data. Then, considering $\boldsymbol{\theta}=(\boldsymbol{\beta},\sigma^2)$, $\mathbf{Y}_{\mathsf{comp}}=(\mathbf{Y}_{\rm obs},\mathbf{V},\mathbf{C},\mathbf{U})^\top$, the empirical information matrix is defined in Meilijson (1989) as

$$\mathbf{I}_e\left(\boldsymbol{\theta}\mid\mathbf{y}_{\rm obs}\right)=\sum_{i=1}^{n}\mathbf{s}\left(y_{{\rm obs}_i}\mid\boldsymbol{\theta}\right)\mathbf{s}^\top\left(y_{{\rm obs}_i}\mid\boldsymbol{\theta}\right)-\frac{1}{n}\mathbf{S}\left(\mathbf{y}_{\rm obs}\mid\boldsymbol{\theta}\right)\mathbf{S}^\top\left(\mathbf{y}_{\rm obs}\mid\boldsymbol{\theta}\right), \tag{13}$$

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

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

Thus, substituting the ML estimates of $\boldsymbol{\theta}$ in (13), the empirical information matrix $\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{y}_{\rm obs}\right)$ is reduced to

$$\mathbf{I}_e\big(\widehat{\boldsymbol{\theta}}\mid\mathbf{Y}_{\rm obs}\big)=\sum_{i=1}^{n}\widehat{s}_i\widehat{s}_i^\top,$$

where $\widehat{s}_i=\big(\frac{\partial Q_i}{\partial\beta_p}^\top,\frac{\partial Q_i}{\partial\sigma}\big)^\top$ is an individual score vector with elements

$$\begin{aligned}\frac{\partial Q_i}{\partial\beta_p}&=\frac{4}{\sigma^2}\widehat{u}_i\xi_i^2\left(\widehat{y_i}\mathbf{x}_i-\mathbf{x}_i^\top\mathbf{x}_i\boldsymbol{\beta}_p\right),\\ \frac{\partial Q_i}{\partial\sigma}&=-\frac{1}{\sigma}+\frac{4}{\sigma^3}\widehat{u}_i\xi_i^2\left(\widehat{y_i^2}-\mathbf{x}_i\boldsymbol{\beta}_p(2\widehat{y}_i-\mathbf{x}_i\boldsymbol{\beta}_p),\right)\end{aligned}$$

for $i=1,\ldots,n$, where the conditional expectations are recycled from the proposed EM algorithm. For the estimation of standard errors based on the observed log-likelihood using the outer-score, Hessian matrix, and sandwich-type methods, we refer to Wang and Lin (2016).

4 | SIMULATION STUDIES

In this section, two simulation studies are conducted in order to evaluate the finite sample performance and the asymptotic properties of the ML estimates obtained using the proposed EM algorithm. The computational procedure is implemented using `R` software R Core Team (2019) by means of our proposed package `LQR`. For both studies, we consider the linear model in (6). Our interest is to estimate the fixed effects parameters $\boldsymbol{\beta}$ for a grid of quantiles $p=\{0.25,0.50,0.75\}$, and under different scenarios of censoring (0%, 5%,10 % and 15%), that is, we vary its percentage from no censoring at all, up to 15%.

The simulated dataset was generated as follows. We considered a $n\times 3$ design matrix $\mathbf{x}_i^\top$ for the fixed effects $\boldsymbol{\beta}$, where the first column corresponds to the intercept and the other two columns were generated randomly from a uniform distribution so that both take values in $(2,20)$. The fixed effect parameters were chosen as $\beta_0=-3$, $\beta_1=-2$, $\beta_2=1$. For each level of censoring, we adopt the following mechanism: half of the observations were interval censored as $y_i\in(y_i\pm{\rm sd}(Y))$ while the remainder were considered as missing data. It is worth mentioning that an interval censoring scheme was preferred in this study, since a one-sided censoring would impair the parameter recovery for extreme quantiles (for instance, a left-censoring scheme would impair only first quantile while the third will stand unaffected). Regarding the missingness mechanism, we focus on the missing completely at random (MCAR), that is, half of the proportion of missing data is randomly selected and then deleted. Finally, we consider five different distributions for the error. Further details for each study are given next.

4.1 | Simulation study 1: asymptotic properties

In this first study, we intend to show how the ML estimates behave asymptotically, that is, we study the behavior in terms of precision while increasing the sample size. We considered different sample sizes, say, $n=25,50,100,200$ and 400, as well as the three quartiles and censoring levels defined before. In particular for this first study, we consider the error term ϵ_i to follow a SKT distribution, what we will call the true model. That is, ϵ_i has been generated independently from a $\text{SKT}(0,1,\nu,p)$ distribution, where p stands for the quantile to be estimated.

For all scenarios, we compute the mean square error (MSE), the bias (Bias) and the Monte Carlo standard deviation (MC-SD) for each parameter over the $m = 500$ replicates. For a vector of parameter θ, these quantities are defined, respectively, by

$$\text{MC-SD}(\widehat{\theta}) = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}\left(\widehat{\theta}^{(j)} - \overline{\widehat{\theta}}\right)^2} \text{ and } \text{MSE}(\widehat{\theta}) = \text{MC-SD}^2(\widehat{\theta}) + \text{Bias}^2(\widehat{\theta}),$$

where $\text{Bias}(\widehat{\theta}) = \overline{\widehat{\theta}} - \theta$, $\overline{\widehat{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\widehat{\theta}^{(j)}$ is the Monte Carlo mean and $\theta^{(j)}$ is the estimate of θ from the j-th sample, with $j = 1, \ldots, m$.

Figure 2 shows the obtained results of the MSE for the regression parameter estimates for different quantiles and censoring levels under our SKT-QRC model. As seen, the MSE for all $\boldsymbol{\beta}$ parameters tends to approach zero with increasing sample size (n), revealing that the ML estimates obtained via the proposed model are conformable to the expected asymptotic properties. As natural, we see that the MSEs are larger for extreme quantiles, taking values symmetrically along quantiles (for instance, between 25 and 75, or 5 and 95). Besides, MSE increases along the percentage of censoring, however this difference becomes unnoticed for n large, exposing the good performance of our proposal to deal with censored/missing data in large samples.

Furthermore, Figure 3 shows the boxplots for the absolute error, namely $\text{Bias}(\hat{\theta}^{(j)}) = \hat{\theta}^{(j)} - \theta$, along the quartiles 1,2 and 3, for the highest level of censoring, 15%. It is easy to see that even for this high percentage of censoring (the worst case), parameter recovery is quite good for all quantiles and βs, becoming better as the sample size increases. Finally, note that for the true model, the ML estimates for the intercept β_0 for all quantiles are unbiased. This is due to the zero-quantile property of the SKT distribution.

4.2 | Simulation study 2: Robustness

In second place, we aim to show the robustness of the proposed SKT-QRC model in order to recover the ML estimates when the error distributions does not follow a SKT distribution. We used the same settings of quantiles and censoring levels as the study before, but now fixing the sample size to $n = 200$ and $m = 500$ MC replicates. Here, we considered four different distributions for the error term, chosen to have different characterizations of tail thickness, skewness and kurtosis. We considered a $N(0, 1)$, a $t_{(4)}$, a $\chi^2_{(4)}/2$, and a mixture $0.5N(-2, 0.36) + 0.5N(2, 0.36)$, namely, henceforth, Normal, Student-t, χ^2 and mixture. As the χ^2 is a positive random variable, it was shifted to have zero median.

For all scenarios, we also compute the bias and MC-SD and MSE. In addition, we also compute the average of the standard deviations (IM-SD) obtained via the observed information matrix derived in Subsection 3.2 and 95% coverage probability (MC-CP) defined as $\text{CP}(\hat{\theta}) = \frac{1}{m}\sum_{j=1}^{m} I(\theta \in [\hat{\theta}_{\text{LCL}}, \hat{\theta}_{\text{UCL}}])$, where I is the indicator function such that θ lies in the interval $[\hat{\theta}_{\text{LCL}}, \hat{\theta}_{\text{UCL}}]$, with $\hat{\theta}_{\text{LCL}}$ and $\hat{\theta}_{\text{UCL}}$ the estimated lower and upper bounds of the 95% CI, respectively.

Tables 1 and 2 summarize the results for these quantities for the regression parameters β_1, β_2. Given that the intercept β_0 estimation is biased in the QR framework ($p \neq 0.5$), we excluded this parameter since it does not make sense to study its MC-CP, for instance.

It can be observed that the BIAS under all scenarios is practically zero, evidencing that the estimators are unbiased. In addition, both tables present the IM-SD, MC-SD and MC-CP for β_2 and β_3 across various quantiles, censoring levels and error distributions. The estimates of MC-SD and IM-SD are very close (also for β_0 not reported here), hence we can infer that the asymptotic approximation of the parameter standard errors are reliable. Furthermore, as expected, we observe that the 95% MC-CP remains lower for extreme quantiles but highly accurate for the median. Finally, it is worth to note that for $p = 0.5$, our SKT-QRC reduces to a Student-t median regression model, since the SKT boils down to a symmetric Student's-t distribution.

5 | REAL-WORLD ANALYSIS

In this section, we conduct two applications of the proposed algorithms and method to real data for illustrative purposes. The proposed model is applied to the the wage rate of married women (economics data) and ambulatory expenditures (financial data). The required numerical evaluations for data analysis were implemented here by using the `R` software (R Core Team 2019).

5.1 | Wages data

In this section, we use a data set previously analyzed by Mroz (1987) consisting of observations, made in 1975, on 753 married women. Our goal is to model the wage rate of these women (wage rates data, hereafter) through the following four covariates: `age`, years of schooling (`educ`), number of hours worked/100 (`hours`) and the number of children younger than 6 years old (`kidslt6`). The model is:

$$\texttt{wages}_i = \beta_0 + \beta_1\texttt{age}_i + \beta_2\texttt{educ}_i + \beta_3\texttt{hours}_i + \beta_4\texttt{kidslt6}_i + \epsilon_i,$$

where ϵ_i follows a $\mathrm{SKT}(0, 1, \nu, \mathrm{p})$ distribution with p equal to the quantile of the response to be predicted, for $i = 1, \ldots, 753$. Among the 753 observations, the first 428 are for women with hours worked greater than zero, while the remaining are for those who did not work for pay. This application is based on left censoring, which is to stress that all husbands in the data set worked for pay in 1975, so there are no censored covariates. We applied the EM algorithm for censored data explained in Section 3, for the dense grid of quantiles $\mathrm{p} = \{0.05, 0.10, \ldots, 0.95\}$.

Besides, we fit the censored mean-based t regression (T-CR) model in Massuia et al. (2015) to compare our results for the particular case of median regression. Table 3 shows the estimates for both regression models. As expected, we find similar results for the mean and median censored regression in terms of estimation (EM) of the regression parameters β's, their standard errors (SE) and the estimation of the degrees of freedom ν. Furthermore, the estimated values of ν for all quantiles were small ($\nu \leq 4$), indicating the heavy-tailness of the wages rate data. For this dataset, the log-likelihood value and Akaike information criterion (AIC) were practically the same for mean and median regression, so these are not shown. Thus, we conclude that the SKT-QRC and the T-CR models are competitive in this case.

We should interpret how the parameters values changes along quantiles meaning that, for this, Figure 4 shows the point estimates and 95% confidence interval for model parameters under the SKT-QRC model for different values of the quantiles. From Figure 4, we can observe some interesting evidences which can not be detected by the T-CR model.

For instance, It can be seen that as p increases the coefficient of years of schooling and the coefficient of number of hours worked/100 increase as well. Also, note that the number of hours worked/100 is significant for all quantiles modeled, even though it takes small values for lower quantiles. Besides, the effect of `age` (measured through the parameter β_1) does not change drastically along quantiles, looking significant only for qunatiles around $\mathrm{p} \in (0.5, 0.8)$. We see that the number of children under 6 years old, makes no difference for women with lower wage rates but it becomes significantly important factor as women earn more. This can be seen because its 95% CI band departs from zero as quantiles increases. Furthermore, its negative sign indicates that higher the number of children, lower the wage rate. Finally, For each β in Figure 4, it is also shown a 95% confidence bar for the mean regression parameter, which look consistent with the corresponding median section of our 95% quantile band.

5.2 | Ambulatory expenditures data

The second application concerns a study of ambulatory expenditures aiming to adjust a response with missing values. The data are taken from Cameron and Trivedi (2009), which were re-analyzed by Marchenko and Genton (2012) in the framework of selection models. In our analysis, we choose the logarithm of ambulatory expenditures (ambexp) as the outcome variable as in Marchenko and Genton (2012). We consider a design matrix X = (1; age; female; educ; blhisp; totchr; ins), including the age, gender, education status, ethnicity, number of chronic diseases and insurance status, respectively. The dataset contains 3328 observations and there are 526 missing values of `ambexp`. More details about the data can be found in Chap. 16 of Cameron and Trivedi (2009). We consider that ϵ_i to follow a $\mathrm{SKT}(0, 1, \nu, \mathrm{p})$ distribution. Thus, the model is given by

$$\mathtt{ambexp}_i = \beta_0 + \beta_1 \mathtt{age}_i + \beta_2 \mathtt{female}_i + \beta_3 \mathtt{educ}_i + \beta_4 \mathtt{blhisp}_i + \beta_5 \mathtt{totchr}_i + \beta_6 \mathtt{ins}_i + \epsilon_i.$$

In addition, we also fit the T-CR mean regression model as defined in Massuia et al. (2015) for comparison.

Table 4 shows the estimates and corresponding standard errors, the maximized log-likelihood and the AIC. In this case, results for the mean and median censored regression differs a little in terms of point estimation. On the other hand, the standard errors are practicality the same for both models. Also note the big difference between the estimation of ν for both models. See that the error distribution for the median regression has lighter tails. We conjecture this is due to the robustness feature of quantile regression. Furthermore, the median regression model fits better the data in comparison with the mean regression model, based on the log-likelihood and the AIC.

Figure 5 displays the parameter estimates and its 95% confidence interval for the parameters assuming different values for the quantiles. Again, 95% confidence bars for the mean regression coefficient parameters are plotted. These last look shifted from our quantile band for all parameters, except β_6. This is due to the difference between the mean and median regression in terms of goodness-of-fit (e.g., see Table 4), for this data set. As natural, the intercept increases as p increases. Besides, it can be seen that the insurance status is not significant for central quantiles but its significance increases for the extremes. Furthermore, its effect becomes positive for lower quantiles and negative for higher quantiles, as expected. From Figure 5, we have that the effect of the gender and the educations status remain similar for all quantiles, both with a positive effect. Finally, we see that there is a negative relationship between the response variable and the ethnicity for all quantiles.

Figure 6 shows the boxplot for the estimated values of the degrees of freedom ν along quantiles. Note that the degrees of freedom ν were small for most of quantiles, evidencing the lack of the normal assumption. Extreme quantiles presented heavier tails. The greatest ν value was 15.3 for the median regression. This behavior is expected since it is usual to have less observations and outliers for extreme conditional quantiles.

6 | CONCLUSIONS

In this paper, a novel EM-type algorithm for the SKT-QRC model has been developed, which uses closed-form expressions at the E-step, that rely on formulas of the mean and variance of a truncated Student's-t distribution. The general formulas for these moments were derived efficiently by Galarza et al. (2020), for which we use the `MomTrunc` package in `R`. The analysis of two real data sets provide strong evidence about the usefulness and effectiveness of our proposal. Moreover, intensive simulation studies show the robustness aspects of the SKT-QRC model, as well as, the consistent properties of the EM estimates. The proposed EM algorithm was implemented as part of the R package `lqr` and is available for download at the CRAN repository. A promising avenue for future research is to consider a generalization for the multivariate Student-t QR model (Galarza et al. 2020), by including a different type of missing mechanisms as in Lin, Lachos, and Wang (2018) and (Lin & Wang 2020).

ACKNOWLEDGMENTS

We are grateful to two anonymous referees and the associate editor for very useful comments and suggestions, which greatly improved this paper. Dr. Galarza acknowledges support from FAPESP-Brazil (Grant 2015/17110-9 and Grant 2018/11580-1).

References

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

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 99–102.

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

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata* (Vol. 5). Stata press College Station, TX.

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

Galarza, C., Castro, L. M., Louzada, F., & Lachos, V. H. (2020). Quantile regression for nonlinear mixed effects models: a likelihood based perspective. *Statistical Papers*, *61*, 1281-1307.

Galarza, C., Lachos, V. H., Cabral, C. R. B., & Castro, C. L. (2017). Robust quantile regression using a generalized class of skewed distributions. *Stat*, *6*(1), 113–130.

Galarza, C. , Benites, L., & Lachos, V. H. (2015). lqr: Robust linear quantile regression [Computer software manual]. R package version 1.5. doi: 10.13140/RG.2.1.3162.9682

Galarza, C, Kan, R., & Lachos, V. H. (2020). MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions. *R Package Version 5.87 URL http://CRAN.R-project.org/package=MomTrunc*.

Ho, H. L., Lin, T.-I., Chen, H. Y., & Wang, W. L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, *142*, 25–40.

Lachos, V. H., Chen, M.-H., Abanto-Valle, C. A., & Azevedo, C. L. (2015). Quantile regression for censored mixed-effects models with applications to HIV studies. *Statistics and its Interface*, *8*(2), 203.

Lachos, V. H., Moreno, E. J. L., Chen, K., & Cabral, C. R. B. (2017). Finite mixture modeling of censored data using the multivariate student-t distribution. *Journal of Multivariate Analysis*, *159*. 151-167.

Lin, T.-I., Lachos, V. H., & Wang, W.-L. (2018). Multivariate longitudinal data analysis with censored and intermittent missing responses. *Statistics in Medicine*, *37*(19), 2822–2835.

Lin, T.-I., & Wang, W.-L. (2020). Multivariate-t linear mixed models with censored responses, intermittent missing values and heavy tails. *Statistical Methods in Medical Research*, *29*(5), 1288–1304.

Louis, T. (1982). Finding the observed information when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *44*, 226–232.

Marchenko, Y. V., & Genton, M. G. (2012). A Heckman selection-t model. *Journal of the American Statistical Association*, *107*(497), 304–317.

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

Massuia, M. B., Garay, A. M., Cabral, C. R. B.., & Lachos, V. (2017). Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. *Statistics and Its Interface*, *10*(3), 425–439.

Mattos, T. B., Garay, A. M., & Lachos, V. H. (2018). Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, *45*(11), 2039–2066.

Meilijson, I. (1989). A fast improvement to the EM algorithm to its own terms. *Journal of the Royal Statistical Society, Series B*, *51*, 127–138.

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, *55*, 765–799.

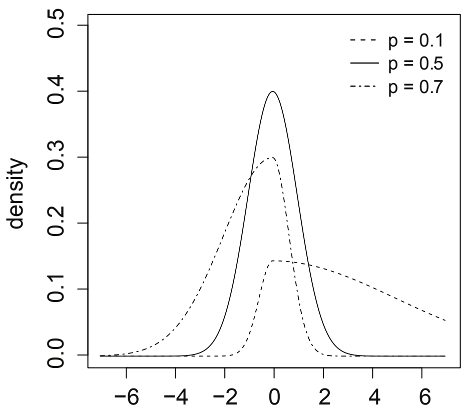
R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

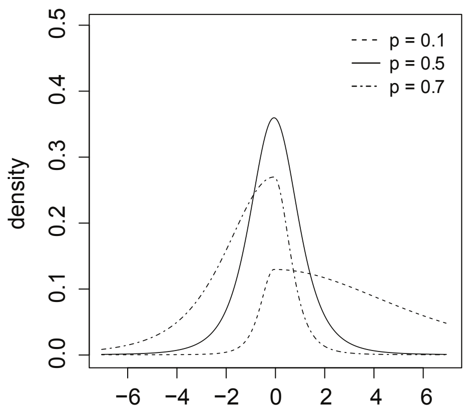
Wang, W. L., & Lin, T. -I. (2016). Maximum likelihood inference for the multivariate mixture model. *Journal of Multivariate Analysis*, *149*, 54–64.

Wichitaksorn, N., Choy, S., & Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics*, *42*(4), 579–596.

□



(a) Skew normal



(b) Skew Student-t

FIGURE 1 Density functions for the standard skewed normal and Student-t distribution with $\nu = 4$ under different values of the skewness parameter.

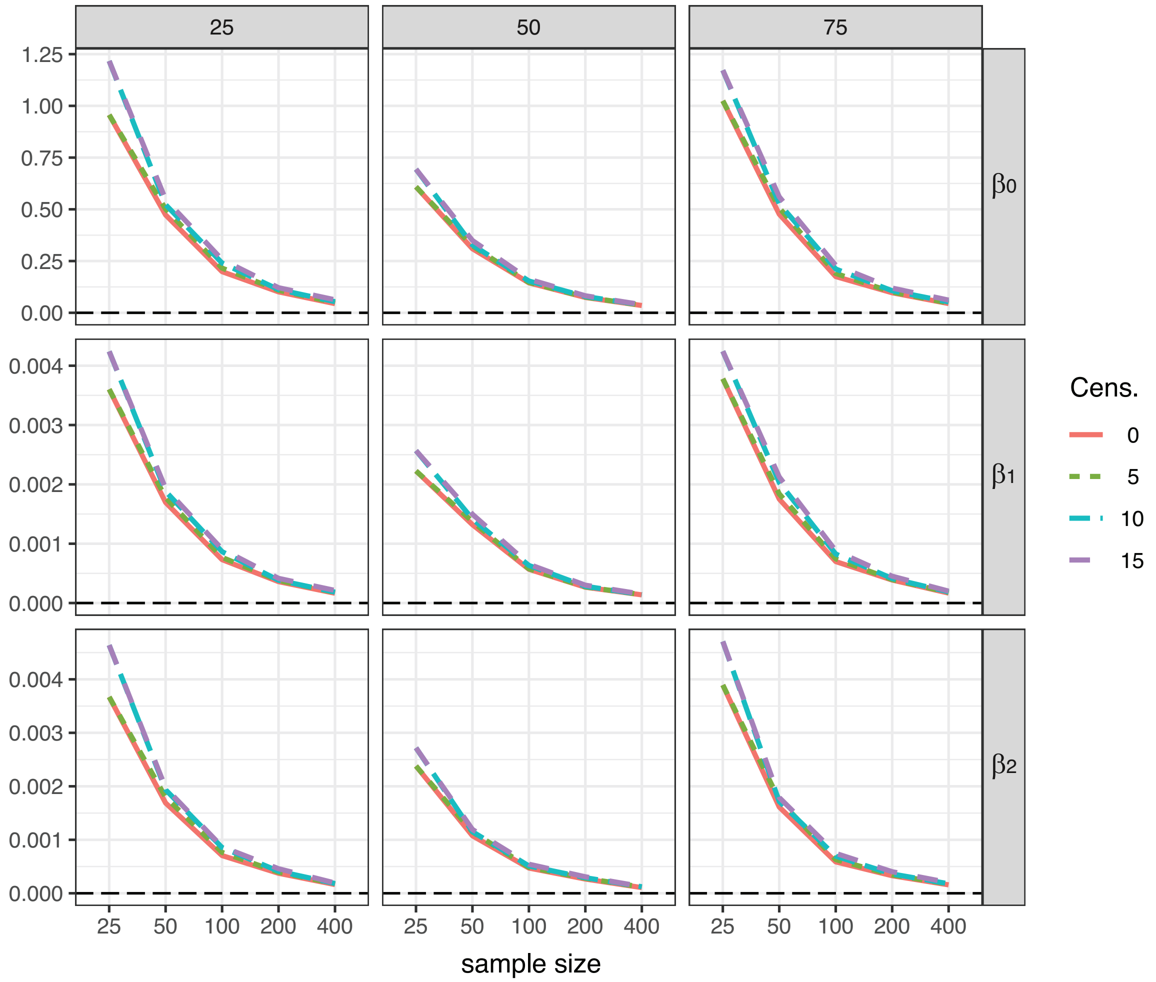


FIGURE 2 Simulation study 1: MSE for the fixed effects β_0, β_1 and β_2 for varying sample sizes over the quantiles $\mathrm{p} = \{0.25, 0.50, 0.75\}$.

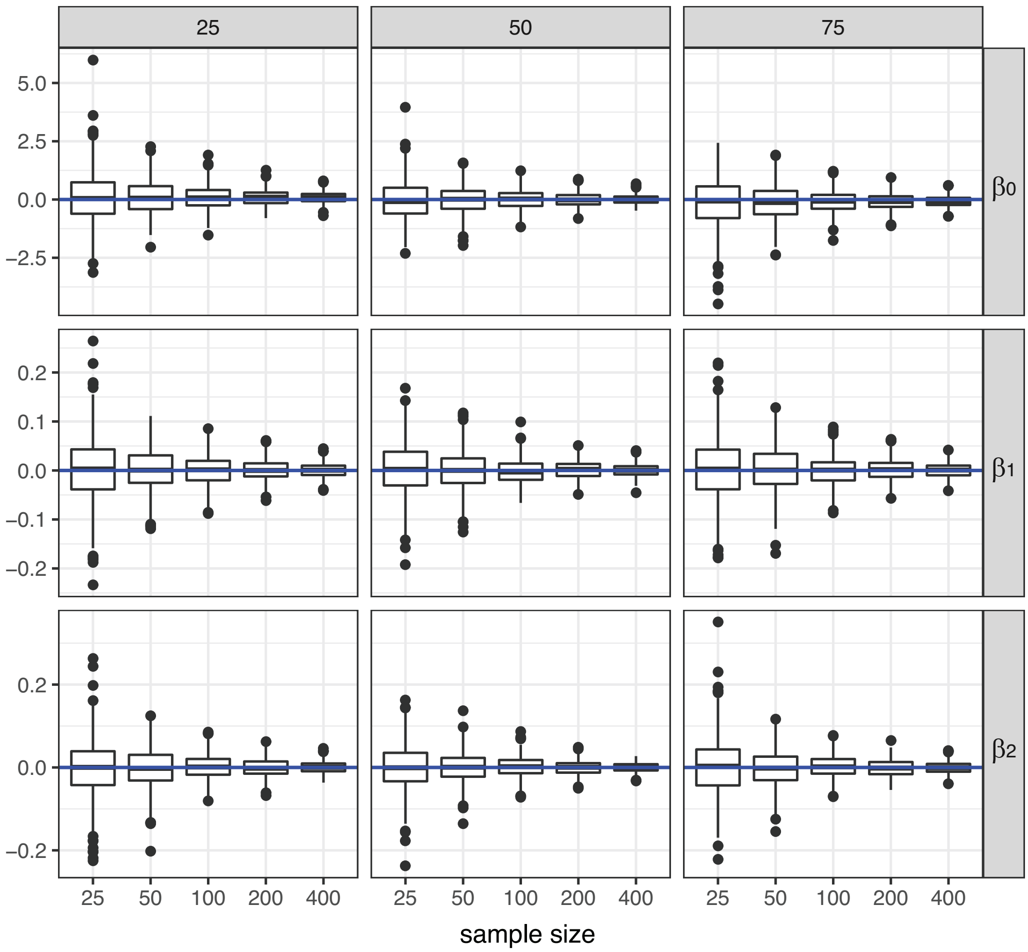


FIGURE 3 Simulation study 1: Box plots from 500 simulated estimates of the absolute error for the fixed effects β_0, β_1 and β_2 for varying sample sizes over the quantiles $\mathrm{p} = \{0.25, 0.50, 0.75\}$.

TABLE 1 Simulation study 2: Absolute bias (BIAS), Monte Carlo standard error (MC-SD), average standard deviations (IM-SD) and 95% coverage probability (MC-CP) for parameter estimates of β_1 and β_2 for the SKN and SKT error distributions.

			β_1				β_2			
Distr.	Cens.	Quantile	BIAS	MC-SD	IM-SD	MC-CP	BIAS	MC-SD	IM-SD	MC-CP
SKN	0	25	0.000	0.015	0.012	85%	0.000	0.016	0.011	84%
		50	0.001	0.013	0.013	94%	0.000	0.013	0.013	93%
		75	0.001	0.016	0.012	85%	-0.001	0.016	0.011	84%
	5	25	0.000	0.016	0.012	85%	0.000	0.016	0.011	82%
		50	0.001	0.014	0.014	94%	0.000	0.014	0.014	94%
		75	0.002	0.017	0.012	83%	-0.001	0.017	0.011	82%
	10	25	0.000	0.017	0.012	83%	0.000	0.017	0.011	80%
		50	0.001	0.014	0.014	93%	0.000	0.014	0.014	93%
		75	0.001	0.017	0.012	79%	0.000	0.017	0.011	80%
	15	25	0.000	0.016	0.012	84%	0.000	0.017	0.011	82%
		50	0.001	0.014	0.015	95%	0.000	0.015	0.014	95%
		75	0.001	0.018	0.012	80%	0.000	0.018	0.011	80%
SKT	0	25	−0.001	0.018	0.014	87%	0.000	0.019	0.014	83%
		50	−0.001	0.015	0.016	96%	0.000	0.015	0.016	95%
		75	−0.001	0.019	0.014	86%	0.000	0.019	0.014	86%
	5	25	−0.001	0.019	0.014	85%	0.001	0.019	0.014	84%
		50	−0.001	0.016	0.016	96%	0.000	0.016	0.016	95%
		75	−0.001	0.019	0.014	84%	0.000	0.019	0.014	84%
	10	25	−0.001	0.020	0.014	84%	0.001	0.019	0.013	81%
		50	−0.001	0.017	0.017	97%	0.001	0.016	0.016	95%
		75	−0.002	0.020	0.014	83%	0.000	0.019	0.014	85%
	15	25	−0.001	0.020	0.014	83%	0.001	0.020	0.013	80%
		50	−0.001	0.017	0.017	95%	0.000	0.017	0.017	94%
		75	−0.002	0.020	0.014	81%	−0.001	0.020	0.013	80%

TABLE 2 Simulation study 2: Absolute bias (BIAS), Monte Carlo standard error (MC-SD), average standard deviations (IM-SD) and 95% coverage probability (MC-CP) for parameter estimates of β_1 and β_2 for the SKT-QRC model for χ^2 and mixture error distributions.

			β_1				β_2			
Distr.	Cens.	Quantile	BIAS	MC-SD	IM-SD	MC-CP	BIAS	MC-SD	IM-SD	MC-CP
χ^2	0	25	0.000	0.013	0.014	97%	−0.001	0.012	0.014	98%
		50	−0.002	0.017	0.017	96%	−0.001	0.016	0.016	96%
		75	−0.003	0.026	0.016	78%	−0.001	0.025	0.016	79%
	5	25	0.000	0.013	0.014	97%	−0.001	0.012	0.014	98%
		50	−0.001	0.017	0.017	96%	−0.001	0.017	0.017	96%
		75	−0.003	0.027	0.016	77%	−0.002	0.025	0.016	78%
	10	25	0.000	0.013	0.014	96%	0.000	0.013	0.014	97%
		50	−0.002	0.018	0.018	96%	−0.001	0.018	0.017	94%
		75	−0.002	0.028	0.016	73%	−0.001	0.026	0.015	76%
	15	25	0.000	0.014	0.014	94%	0.000	0.014	0.014	95%
		50	−0.001	0.019	0.018	96%	0.000	0.018	0.018	96%
		75	−0.002	0.028	0.015	70%	−0.001	0.026	0.015	74%
Mixture	0	25	0.000	0.007	0.005	83%	0.000	0.007	0.005	84%
		50	0.000	0.006	0.006	96%	0.000	0.005	0.006	96%
		75	0.001	0.007	0.005	84%	0.000	0.006	0.005	86%
	5	25	0.000	0.007	0.005	84%	0.007	0.005	0.005	83%
		50	0.000	0.006	0.006	96%	0.006	0.006	0.006	96%
		75	0.001	0.007	0.005	84%	0.006	0.005	0.005	85%
	10	25	0.000	0.007	0.005	83%	0.000	0.007	0.005	81%
		50	0.000	0.006	0.006	96%	0.000	0.006	0.006	96%
		75	0.001	0.007	0.005	81%	0.000	0.007	0.005	83%
	15	25	0.000	0.007	0.005	81%	0.000	0.007	0.005	79%
		50	0.000	0.006	0.006	95%	0.000	0.006	0.006	95%
		75	0.001	0.007	0.005	81%	0.000	0.007	0.005	83%

TABLE 3 Parameter estimates and standard errors for SKT-QRC and T-CR (Massuia et al. 2017) models in wages data.

	Mean regression T-CR		Median regression SKT-QRC $(p = 0.5)$	
Parameter	EM	SD	EM	SD
β_0	−5.054	0.867	−4.812	0.820
β_1	−0.032	0.014	−0.031	0.013
β_2	0.452	0.044	0.431	0.042
β_3	0.275	0.013	0.272	0.012
β_4	−0.984	0.258	−0.941	0.249
ν	2.3		2.6	

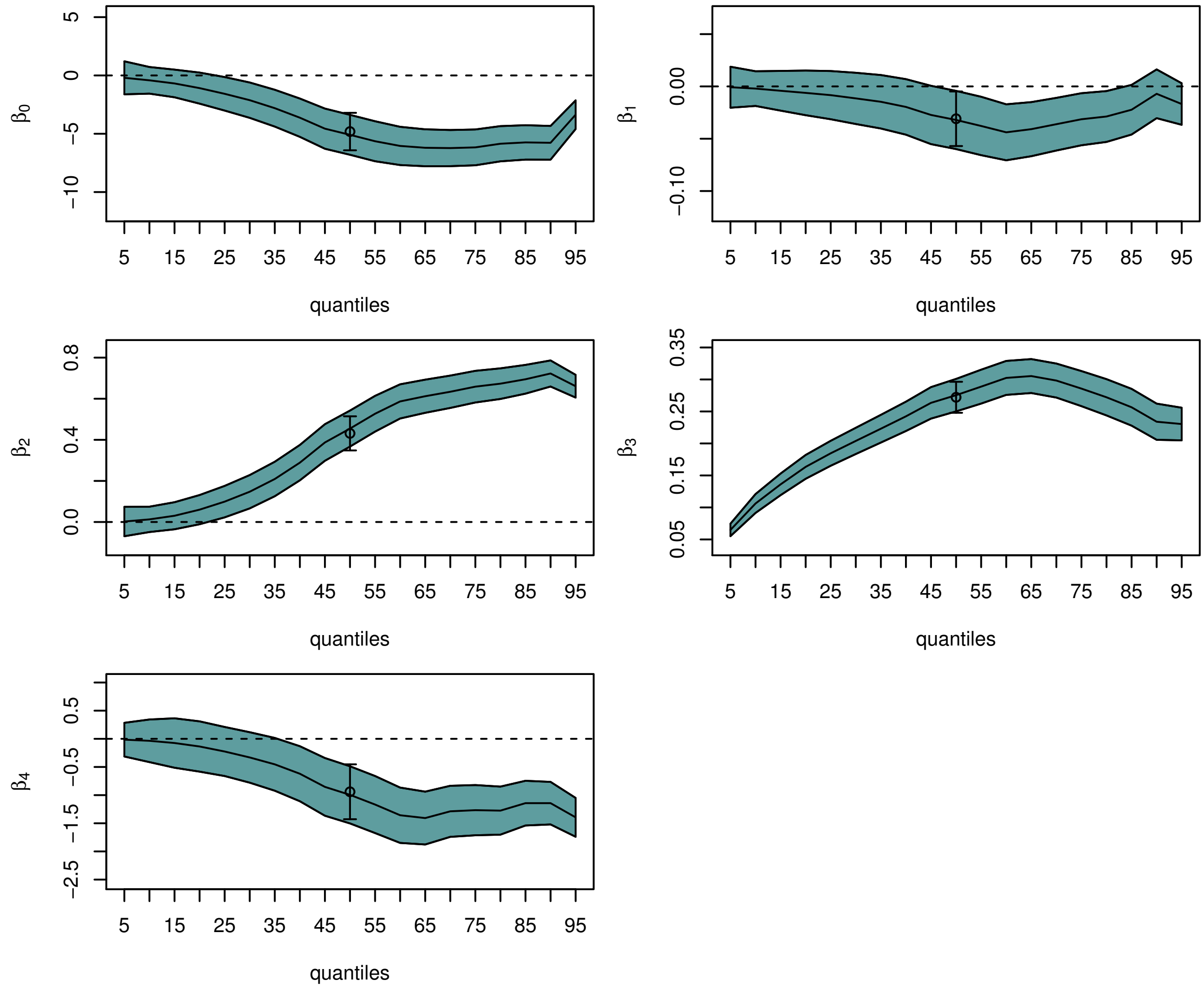


FIGURE 4 Wages data: 95% confidence bands for the quantile regression parameters along the grid of quantiles $\{0.05, \ldots, 0.95\}$. Interval bars represent the 95% CI for the mean of the regression parameters.

TABLE 4 Parameter estimates and standard errors for SKT-QRC and T-CR (Massuia et al. 2017) models in Ambulatory expenditures data.

	Mean regression		Median regression	
	T-CR		SKT-QRC ($p = 0.5$)	
Parameter	EM	SD	EM	SD
β_0	4.309	0.165	4.874	0.161
β_1	0.252	0.022	0.219	0.022
β_2	0.486	0.048	0.384	0.048
β_3	0.052	0.010	0.026	0.010
β_4	−0.289	0.055	−0.241	0.054
β_5	0.610	0.031	0.565	0.032
β_6	−0.029	0.050	−0.027	0.050
ν	1.01		15.3	
likelog	−6709.92		−4630.54	
AIC	13437.85		9279.09	

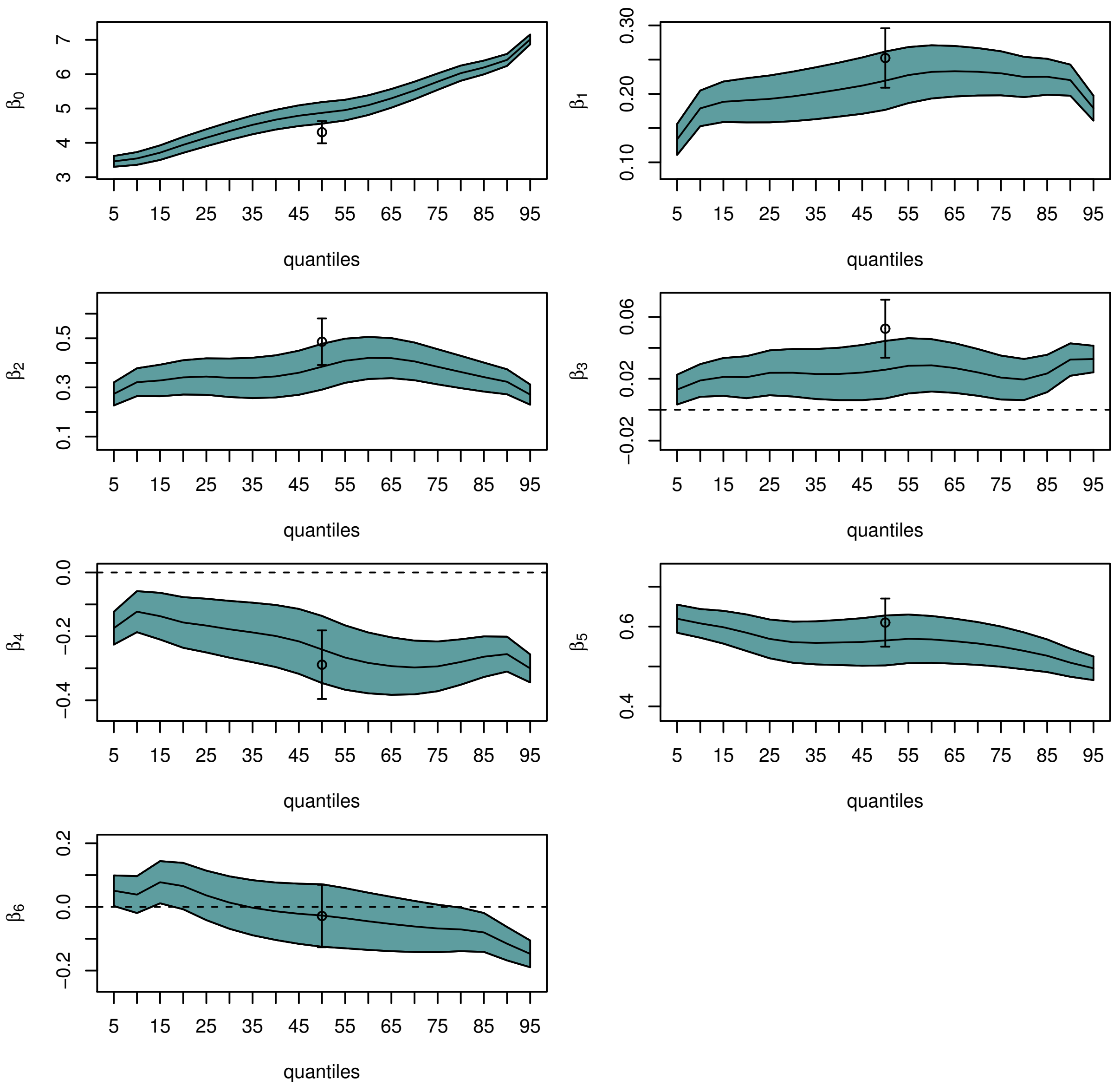


FIGURE 5 Ambulatory expenditures data: 95% confidence bands for the quantile regression parameters along the grid of quantiles $\{0.05, \ldots, 0.95\}$. Interval bars represent the 95% CI for the mean of the regression parameters.

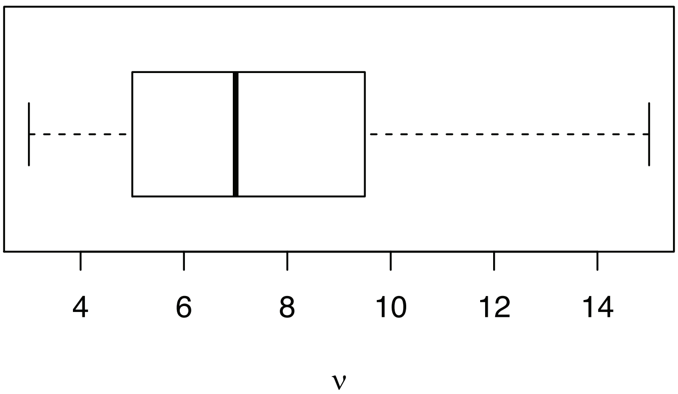


FIGURE 6 Boxplot for the estimated values of the degrees of freedom ν along quantiles.