Robust Quantile Regression using a Generalized Class of Skewed Distributions

Christian E. Galarza[a,b] Victor H. Lachos[b*] Celso R. B. Cabral[c]
Luis M. Castro[d]

[a]*Departamento de Matemáticas, Escuela Superior Politécnica del Litoral, ESPOL, Ecuador*
[b]*Departamento de Estatística, Universidade Estadual de Campinas, Brazil*
[c]*Departamento de Estatística, Universidade Federal do Amazonas, Brazil*
[d]*Departamento de Estadística and CI²MA, Universidad de Concepción, Chile*

Abstract

It is well known that the widely popular mean regression model could be inadequate if the probability distribution of the observed responses do not follow a symmetric distribution. To deal with this situation, the quantile regression turns to be a more robust alternative for accommodating outliers and the misspecification of the error distribution since it characterizes the entire conditional distribution of the outcome variable. This paper presents a likelihood-based approach for the estimation of the regression quantiles based on a new family of skewed distributions introduced by Wichitaksorn *et al.* (2014). This family includes the skewed version of Normal, Student-t, Laplace, contaminated Normal and slash distribution, all with the zero quantile property for the error term, and with a convenient and novel stochastic representation which facilitates the implementation of the EM algorithm for maximum-likelihood estimation of the pth quantile regression parameters. We evaluate the performance of the proposed EM algorithm and the asymptotic properties of the maximum-likelihood estimates through empirical experiments and application to a real life dataset. The algorithm is implemented in the R package `lqr()`, providing full estimation and inference for the parameters as well as simulation envelopes plots useful for assessing the goodness-of-fit.

Keywords Quantile regression model; EM algorithm; Scale mixtures of Normal distributions.

1 Introduction

Quantile regression (QR) models have become increasingly popular since the seminal work of Koenker & G Bassett (1978). In contrast to the mean regression model, QR belongs to a robust model family, which can give an overall assessment of the covariate effects at different quantiles of the outcome (Koenker, 2005). In particular, we can model the lower or higher quantiles of the outcome to provide a natural assessment of covariate effects specific for those regression quantiles. Unlike conventional models, which only address the conditional mean or the central effects of the covariates, QR models quantify the entire conditional distribution of the outcome variable. In addition, QR does not impose any distributional assumption on the error, except the requirement about the zero conditional quantile. The foundations of the methods for independent data are now consolidated, and some statistical methods for estimating and drawing inferences about conditional quantiles are provided by most of the available statistical programs (*e.g.*, `R`, SAS, Matlab and Stata). For instance, just to name a few of them, in the well-known `R` package `quantreg()` is

*Address for correspondence: Departamento de Estatística, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859. E-mail: `hlachos@ime.unicamp.br`

implemented as a variant of the Barrodale & Roberts (1977) simplex (BR) for linear programming problems described in Koenker & d'Orey (1987), where the standard errors are computed by the rank inversion method (Koenker, 2005). Another method implemented in this popular package is the Lasso Penalized Quantile Regression (LPQR), introduced by Tibshirani (1996), where a penalty parameter is specified to determine how much shrinkage occurs in the estimation process. As it can be seen, the QR model can be implemented in a wide range of different methodologies.

From a Bayesian point of view, Kottas & Gelfand (2001) considered the study of the median regression, which is a special case of QR. In this context, these authors discussed a non-parametric approach for the error distribution based on either Polya tree or Dirichlet process priors. Regarding general quantile regression, Yu & Moyeed (2001) proposed a Bayesian modeling approach by using the asymmetric Laplace distribution (ALD), Kottas & Krnjajić (2009) developed Bayesian semi-parametric models for quantile regression using Dirichlet process mixtures for the error distribution. Kozumi & Kobayashi (2011) developed a simple and efficient Gibbs sampling algorithm for fitting the quantile regression model based on a location-scale mixture representation of the ALD. From the classical viewpoint, Benites *et al.* (2013), Zhou *et al.* (2014) and Tian *et al.* (2014) adjusted a linear QR model based on EM algorithm for maximum likelihood (ML) assuming ALD errors. Particularly Benites *et al.* (2013) showed that their approach out-performed other common non-parametric estimators as those obtained via BR and LPQR algorithms. While ALD has the zero quantile property and a useful stochastic representation, it is not differentiable at zero, which could lead to problems of numerical instability. Thus the Laplace density is a pretty strong assumption in order to set a quantile regression model through the classical or Bayesian framework.

To overcome this deficiency, recently Wichitaksorn *et al.* (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 is constructed by mixing a skew-normal distribution (SKN) proposed by Fernández & Steel (1998) and the symmetric class of scale mixture of normal (SMN) distributions proposed by Andrews & Mallows (1974). The SKN distribution is obtained by partitioning two scaled mixture of normal (Gaussian) distributions, which have a skewness parameter defined in the interval $(0, 1)$, allowing direct application to parametric quantile regression. On the other hand, employing scale mixture of normals facilitates efficient estimation via Markov chain Monte Carlo (MCMC) methods and the EM algorithm. In fact, Wichitaksorn *et al.* (2014) adopt a MCMC approach as a natural solution to estimation and inference by using the marginal representation of the SKD class of distributions. In contrast to the marginal approach adopted in Wichitaksorn *et al.* (2014), in this paper a novel stochastic representation is proposed, which allows the study of many of its properties and also the implementation of an efficient (and easy) EM algorithm for ML estimation of the parameters at the pth level, with closed form expressions at the E- and M- steps. Therefore, the main contribution of this paper is to propose a robust method for drawing inferences about conditional quantiles in linear regression problems from a likelihood-based perspective. Moreover, the proposed EM-type algorithm has been coded and implemented in the R package `lqr()` (Galarza *et al.*, 2015), which is available for download at CRAN repository. A great advantage of this package is that it offers an automatic fit of all the SKD distributions taking into consideration.

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 the standard errors. Section 4 presents simulation studies of finite sample performance and robustness of our proposed method. An application of the EM algorithm to a dataset examining some characteristics of Australian athletes available from the Australian Institute of Sport (AIS) is presented in Section 5. Finally, Section 6 closes the paper, sketching some future research directions.

2 The SKD family of distributions

In order to define the SKD class of distributions, we first make some remarks related to the skew-normal (SKN) distributions as defined by Wichitaksorn *et al.* (2014). Thus, in the following we present some definitions where we explain first the fundamental concept of the SKN distribution and its relation with the SKD family of distributions.

2.1 Preliminaries

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

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

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 $X \sim \text{SKN}(\mu,\sigma,p)$. Note that, $P(X\leq\mu) = p$ and $P(x>\mu) = 1-p$, which allows a direct application to quantile regression 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. Therefore, it can be conveniently written as

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

where $\rho_p(\cdot)$ is the so called check (or loss) function defined by $\rho_p(u) = u(p - \mathbb{I}\{u<0\})$. It is important to remark that the SKN distribution is closed under location-scale transformations, *i.e.*, if $Z \sim \text{SKN}(0,1,p)$, then $X = \mu+\sigma Z \sim \text{SKN}(\mu,\sigma,p)$. Moreover, the SKN distribution has a useful stochastic representation, given in the next result. The proof can be found in the Appendix.

Lemma 1. *Let* $T_0 \sim \text{N}(0,1)$ *and* I *with probability function*

$$P\left(I = -\frac{1}{2(1-p)}\right) = p \quad \textit{and} \quad P\left(I = \frac{1}{2p}\right) = 1-p,$$

be independent. Then, the random variable with stochastic representation

$$X = \mu + \sigma I|T_0|,$$

follows a $\text{SKN}(\mu,\sigma,p)$ *distribution.*

2.2 Scale mixture of normal distributions

The family of scale mixture of normal (SMN) distributions (Andrews & Mallows, 1974; Lange & Sinsheimer, 1993) is a wide class of thick-tailed distributions including the normal one as a special case. This class of symmetric distributions also includes the Student-t (T), slash (S), contaminated normal (CN), among many others. The SMN class can be conveniently represented using the following stochastic representation,

$$W = \boldsymbol{\mu} + \sigma\kappa(U)^{1/2}T_0,$$

where μ is a location parameter, $\kappa(\cdot)$ is the weight function, U is positive random variable with *pdf* $h(u|\boldsymbol{\nu})$ and cumulative distribution function (*cdf*) $H(u|\boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of U and $T_0 \sim \text{N}(0,1)$, with U independent of T_0. Under this setup, the marginal *pdf* of Y is given by

$$f(w|\mu,\sigma,\boldsymbol{\nu}) = \int_0^\infty \phi\left(w|\mu,\kappa(u)\sigma^2\right)dH(u|\boldsymbol{\nu}).$$

We shall use $W \sim \text{SMN}(\mu,\sigma^2,\boldsymbol{\nu})$ to denote this class of distributions.

2.3 A family of zero-quantile skewed distributions

This new class of distributions, defined by Wichitaksorn *et al.* (2014), is constructed as a scale mixture of SKN distributions. We say that Y follows a skewed distribution (denoted by SKD) with location parameter μ, scale parameter σ and skewness parameter p, if Y can be represented stochastically as

$$Y = \mu + \sigma\kappa(U)^{1/2}Z, \tag{3}$$

where $Z \sim \text{SKN}(0, 1, p)$. A direct consequence of this definition is that, as the SKN distribution, $P(Y \leq \mu) = p$ and $P(Y > \mu) = 1 - p$.

From the stochastic representation (3), the conditional distribution of Y given $U = u$ is $\text{SKN}(\mu, \kappa(u)^{1/2}\sigma, p)$. Then, integrating out U, we have that the marginal *pdf* of Y is

$$f(y|\mu, \sigma, p, \boldsymbol{\nu}) = \int_0^\infty \frac{4p(1-p)}{\sqrt{2\pi\kappa(u)\sigma^2}} \exp\left\{-2\rho_p^2\left(\frac{y-\mu}{\kappa^{1/2}(u)\sigma}\right)\right\} dH(u|\boldsymbol{\nu}), \tag{4}$$

and consequently, several skewed and thick-tailed distributions can be obtained from different specifications of the weight function $\kappa(\cdot)$ and *pdf* $h(u|\boldsymbol{\nu})$. Alternatively, another stochastic representation for the SKD distribution can be provided. This representation is given in the following result, whose proof is given in Appendix.

Lemma 2. *Let* $T_0 \sim \text{N}(0, 1)$ *and* U *a positive random variable, with* U *independent of* T_0*. Let* I *be the discrete random variable defined in Lemma 1. Then,* $Y \sim \text{SKD}(\mu, \sigma, p, \boldsymbol{\nu})$ *can be stochastically represented as*

$$Y \stackrel{d}{=} \mu + \sigma\kappa(U)^{1/2}I|T_0|.$$

Table 1 presents some particular cases belonging to the SKD family, namely, the skewed Student-t ($\text{SKT}(\mu, \sigma, p, \nu)$); skewed Laplace ($\text{SKL}(\mu, \sigma, p)$); skewed slash ($\text{SKS}(\mu, \sigma, p, \nu)$) and skewed contaminated normal ($\text{SKCN}(\mu, \sigma, p, \nu, \gamma)$), respectively. Moreover, Figure 1 presents the *pdf* associated to these members considering $\mu = 0$ and $\sigma = 1$. In particular, this figure shows how the skewness changes for different values of parameter p. It is important to remark that, when $p = 0.5$ these distributions turn to be symmetrical.

Distribution	$\kappa(u)$	$h(u\|\boldsymbol{\nu})$	$f(y\|\mu, \sigma, \boldsymbol{\nu})$
skewed Student-t	u^{-1}	$\text{G}(\frac{\nu}{2}, \frac{\nu}{2})$	$\frac{4p(1-p)\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{2\pi\sigma^2}}\left\{\frac{4}{\nu}\rho_p^2\left(\frac{y-\mu}{\sigma}\right)+1\right\}^{-\frac{\nu+1}{2}}$
skewed Laplace	u	$\text{Exp}(2)$	$\frac{2p(1-p)}{\sigma}\exp\left\{-2\rho_p\left(\frac{y-\mu}{\sigma}\right)\right\}$
skewed slash	u^{-1}	$\text{Beta}(\nu, 1)$	$\nu\int_0^1 u^{\nu-1}\phi_{skd}(y\|\mu, u^{-1/2}\sigma, p)du$
skewed cont. normal	u^{-1}	$\nu\mathbb{I}\{u=\gamma\} + (1-\nu)\mathbb{I}\{u=1\}$ $0 \leq \nu, \gamma \leq 1,$	$\nu\phi_{skd}(y\|\mu, \gamma^{-1/2}\sigma, p) + (1-\nu)\phi_{skd}(y\|\mu, \sigma, p)$

Table 1: $\kappa(\cdot)$, $h(u|\boldsymbol{\nu})$ and *pdf* for some members of the SKD family. $\text{G}(\alpha, \beta)$ denotes the Gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$, $\text{Exp}(\beta)$ denotes the exponential distribution with mean β, $\text{Beta}(\alpha, \beta)$ denotes the Beta distribution and $\phi_{skd}(y|\mu, \sigma, p)$ denotes the *pdf* of the SKN distribution defined in (2).

An interesting result related to the moments of the SKD distributions is presented next.

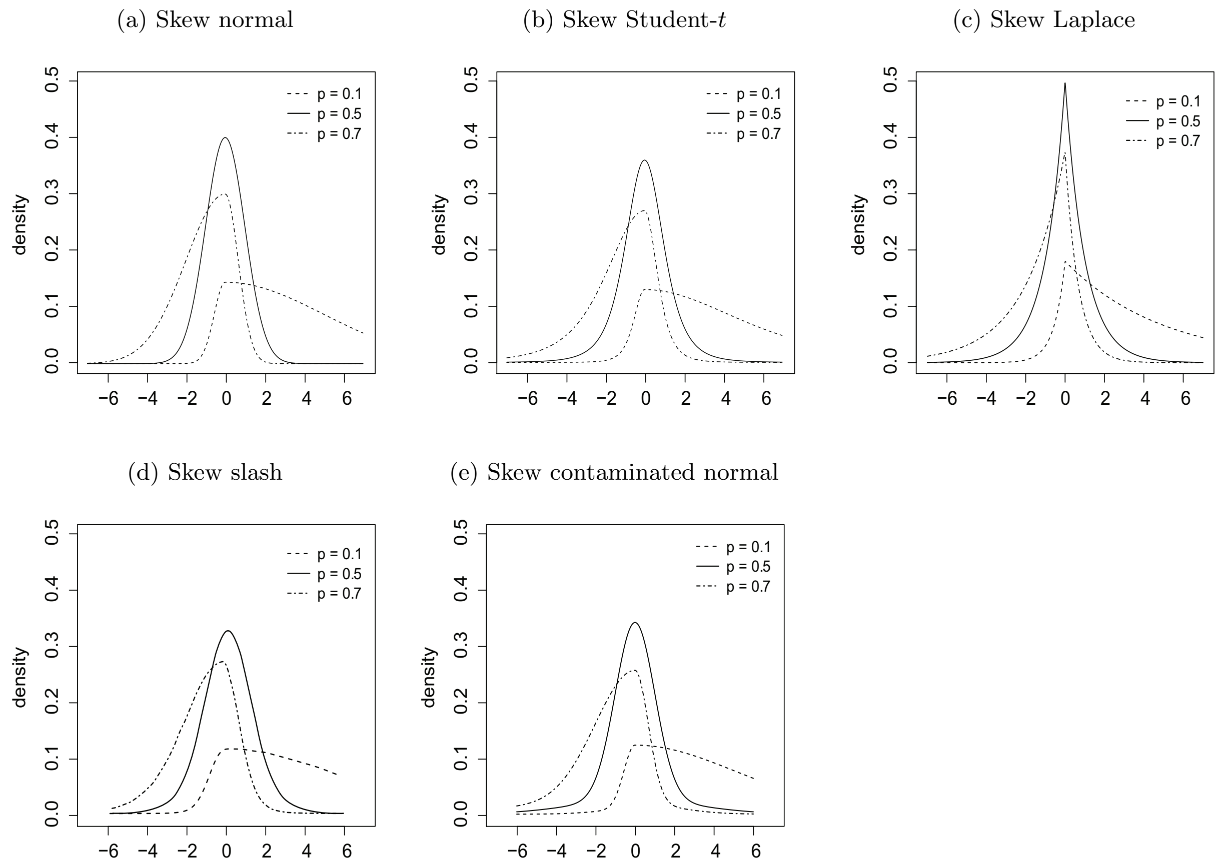

Figure 1: Density functions for the standard skewed normal, skewed Student-t, skewed Laplace, skewed slash and skewed contaminated normal distributions under different values of the skewness parameter. Parameters have been set as $\mu = 0$, $\sigma = 1$, $\gamma = 0.1$ and $\nu = (4, 2, 0.1)$ for the Student-t, slash and contaminated normal distribution, respectively.

Moment generating function for the SKD family

From the stochastic representation of Y given in (2), we have that

$$\mathrm{E}[(Y-\mu)^k] \quad = \sigma^k E[I^k]\, E[H^k]\, E[\kappa(U)^{k/2}],$$

where $H = |T_0|$. Moreover, from (3) we have

$$\mathrm{E}[I^k] \quad = \tfrac{(-1)^k p^{k+1} + (1-p)^{k+1}}{2^k p^k (1-p)^k}, \;\; k = 1, 2, \ldots.$$

In addition, since the weight function $\kappa(U)$ depends of the mixture distribution $h(u|\boldsymbol{\nu})$, $\mathrm{E}[\kappa(U)^{k/2}]$ is given iqual to 1 when $\kappa(U) = U^{-1}$ and $P(U = 1) = 1$; $(\nu/2)^{k/2}\Gamma((\nu - k)/2)/\Gamma(\nu/2)$ when $\kappa(U) = U^{-1}$ and $U \sim \mathrm{G}(\nu/2, \nu/2)$; $\sqrt{2^k}\,\Gamma((k+2)/2)$ when $\kappa(U) = U$ and $U \sim \mathrm{Exp}(2)$; $\nu/(\nu - k/2)$ when $\kappa(U) = U^{-1}$ and $U \sim \mathrm{Beta}(\nu, 1)$ and $\nu/\gamma^{k/2} + (1-\nu)$ when $\kappa(U) = U^{-1}$ and $h(u|\nu,\gamma) = \nu\mathbb{I}_{\{u=\gamma\}} + (1-\nu)\mathbb{I}_{\{u=1\}}$.

The moments $\mathrm{E}[H^k]$ are obtained using the moment generating function of a half normal distribution. This function is defined as $M_H(t) = 2\exp\{t^2/2\}[1 - \Phi(-t)]$. It is important to note that, after some algebra, these moments are

$$\mathrm{E}[H^k] = \begin{cases} (k-1)!!, & \text{for } k \text{ even;} \\ (k-1)!!\sqrt{2/\pi}, & \text{for } k \text{ odd,} \end{cases}$$

where $n!!$ denotes the double factorial function. Finally the kth centred moment of Y is given by

$$\mathrm{E}[(Y-\mu)^k] = \begin{cases} \sigma^k (k-1)!! \left[\frac{(-1)^k p^{k+1} + (1-p)^{k+1}}{2^k p^k (1-p)^k}\right] \mathrm{E}[\kappa(U)^{k/2}], & \text{for } k \text{ even;} \\ \sqrt{2/\pi}\sigma^k (k-1)!! \left[\frac{(-1)^k p^{k+1} + (1-p)^{k+1}}{2^k p^k (1-p)^k}\right] \mathrm{E}[\kappa(U)^{k/2}], & \text{for } k \text{ odd.} \end{cases}$$

3 Quantile regression using the SKD family

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

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_p + \epsilon_i, \;\; i = 1, \ldots, n, \tag{5}$$

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 $P(y_i \leq \mathbf{x}_i^\top \boldsymbol{\beta}_p) = p$. The density $f_p(\cdot)$ is often left unspecified in the classical literature. Thus, quantile regression estimation for $\boldsymbol{\beta}_p$ proceeds by minimizing

$$\widehat{\boldsymbol{\beta}}_p = \arg \min_{\boldsymbol{\beta}_p \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_p\big(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p\big), \tag{6}$$

where $\rho_p(\cdot)$ is known as the check function 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. Is important to stress that there is a connection between the minimization of the sum in (6) and the maximum-likelihood theory, since to minimize (6) is equivalent to maximize the likelihood when data follows a distribution belonging to the family of zero conditional quantile SKD introduced in Section 2.3. It can be observed that the check function in (4) is inversely proportional to the *pdf* and therefore to the likelihood. Particularly, in the case of the skewed Laplace distribution, the check function is linearly being not differentiable at zero. In this case, we cannot derive explicit solutions to the minimization problem and linear programming methods have to be applied to obtain quantile regression estimates for $\boldsymbol{\beta}_p$.

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 (6). Therefore, the relationship between the check function and this family of distributions can be used to reformulate the QR method within the likelihood framework. In order to do that, we propose the following useful result. Its proof is given in Appendix.

Lemma 3. *Let* $T_0 \sim \mathrm{N}(0,1)$ *and* $Y \sim \mathrm{SKD}(\mu, \sigma, p, \boldsymbol{\nu})$. *If* $D = \rho(\frac{Y-\mu}{\sigma})$, *then* D *can be represented stochastically as*

$$D \overset{d}{=} \tfrac{1}{2}\kappa(u)^{1/2}|T_0|. \tag{7}$$

Thus, from (7), we have that $D \overset{d}{=} |W|$ where $W = \frac{1}{2}\kappa(u)^{1/2} T_0$ belongs to the SMN class of distribution given in Section 2.2. Hence, D is the half-type version of a SMN random variable with location parameter $\mu = 0$, scale parameter $\sigma = 1/2$. Table 2 presents different probability distributions for D under specific members of the SKD family.

Distribution	Distribution of D	*pdf* of D
skewed normal	$\mathrm{HN}(\frac{1}{2}^2)$	$\frac{4}{\sqrt{2\pi}}\exp(-2d^2)$
skewed Student-t	$\mathrm{HT}(\frac{1}{2}^2,\nu)$	$\frac{4\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}}\left\{\frac{4}{\nu}d^2+1\right\}^{-\frac{\nu+1}{2}}$
skewed Laplace	$\mathrm{Exp}(\frac{1}{2})$	$2\,\exp(-2d)$
skewed slash	$\mathrm{HS}(\frac{1}{2}^2,\nu)$	$2\nu\int_0^1 u^{\nu-1}\phi(d\vert\mu,\frac{1}{4u})\,du$
skewed cont. normal	$\mathrm{HCN}(\frac{1}{2}^2,\nu,\gamma)$	$2\nu\phi(d\vert\mu,\frac{1}{4\gamma})+2(1-\nu)\phi(d\vert\mu,\frac{1}{4})$

Table 2: Probability distributions for the function D defined in Lemma 3. HT denotes the half-Student-t distribution, HS denotes the half-slash distribution and HCN denotes the half-contaminated normal distribution.

3.1 Parameter estimation via the EM algorithm

In this section, we propose an estimation method for the QR model based on the EM algorithm for obtaining the ML estimates.

The EM algorithm (Dempster *et al.*, 1977), is a powerful frequentist approach to estimate parameters via ML 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.

From the hierarchical representation given in (3), the QR model defined in (5) can be expressed as

$$\begin{aligned} Y_i|U_i=u_i &\sim \mathrm{SKN}(\mathbf{x}_i^\top\boldsymbol{\beta}_p,\sqrt{\kappa(u_i)}\sigma,p),\\ U_i &\sim h(u_i|\boldsymbol{\nu}),\end{aligned}$$

where $h(u|\boldsymbol{\nu})$ represents the mixture density. Let $\mathbf{y}=(y_1,\ldots,y_n)$ and $\mathbf{u}=(u_1,\ldots,u_n)$ be the observed and missing (latent) data, respectively. Then, the complete data log-likelihood function of $\boldsymbol{\theta}=(\boldsymbol{\beta}_p^\top,\sigma,\boldsymbol{\nu})$ given $(\mathbf{y},\mathbf{u})$, ignoring some additive constant terms, is given by $\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})=\sum_{i=1}^n\ell_c(\boldsymbol{\theta}|y_i,u_i)$, where

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|y_i,u_i) = & \sum_{\{y_i\le\mathbf{x}_i^\top\boldsymbol{\beta}_p\}}\log\phi\left(y_i\,\middle|\,\mathbf{x}_i^\top\boldsymbol{\beta}_p,\frac{\kappa(u_i)\sigma^2}{4(1-p)^2}\right)+\sum_{\{y_i>\mathbf{x}_i^\top\boldsymbol{\beta}_p\}}\log\phi\left(y_i\,\middle|\,\mathbf{x}_i^\top\boldsymbol{\beta}_p,\frac{\kappa(u_i)\sigma^2}{4p^2}\right)\\ &+\log h(u_i|\boldsymbol{\nu}),\end{aligned}$$

for $i=1,\ldots,n$. Denoting by $\xi_i=(1-p)\,\mathbb{I}\{y_i\le\mathbf{x}_i^\top\boldsymbol{\beta}_p\}+p\,\mathbb{I}\{y_i>\mathbf{x}_i^\top\boldsymbol{\beta}_p\}$, the expression $\ell_c(\boldsymbol{\theta}|y_i,u_i)$ can be rewritten as

$$\ell_c(\boldsymbol{\theta}|y_i,u_i)=\sum_{i=1}^n\log\phi\left(y_i\,\middle|\,\mathbf{x}_i^\top\boldsymbol{\beta}_p,\frac{\kappa(u_i)\sigma^2}{4\xi_i^2}\right)+\sum_{i=1}^n\log h(u_i|\boldsymbol{\nu}).$$

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 $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})=\mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})|\mathbf{y},\boldsymbol{\theta}^{(k)}]$. Thus, ignoring constants that does not the depend on the parameter of interest $\boldsymbol{\theta}$, the Q-function is given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \propto -n\log\sigma - 2\sum_{i=1}^{n}\left\{\widehat{\kappa^{-1}(u_i)}\xi_i^2 z_i^2\right\} + \sum_{i=1}^{n}\mathrm{E}\left[\log\, h(u_i|\boldsymbol{\nu})|y_i,\boldsymbol{\theta}^{(k)}\right], \tag{8}$$

with $z_i = (y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_p)/\sigma$. For evaluating (8), it is required to compute $\widehat{\kappa^{-1}(u_i)} = \mathrm{E}[\kappa^{-1}(U_i)|y_i,\boldsymbol{\theta}^{(k)}]$, that will depend of the weight function $\kappa(\cdot)$. The conditional distribution of the latent variable given the observed data $f(u_i|y_i,\boldsymbol{\theta}^{(k)})$ will depend on the functional form of $h(u_i|\boldsymbol{\nu})$. Table 3 shows the conditional *pdf* of U given Y for specific choices of $h(u_i|\boldsymbol{\nu})$.

Distribution	Distribution of U	Conditional distribution of $U\|Y$	$\widehat{\kappa^{-1}(u_i)}$
skewed Student-t	$\mathrm{G}(\frac{\nu}{2},\frac{\nu}{2})$	$\mathrm{G}\left(\frac{\nu+1}{2},\frac{\nu+4\xi_i^2z_i^2}{2}\right)$	$\frac{\nu+1}{\nu+4\xi_i^2z_i^2}$
skewed Laplace	Exp(2)	$\mathrm{GIG}\left(\frac{1}{2},2\xi_i^2z_i^2,\frac{1}{2}\right)$	$\frac{1}{2\xi_i\,\|z_i\|}$
skewed slash	$\mathrm{Beta}(\nu,1)$	$\mathrm{TG}\left(\nu+\frac{1}{2},2\xi_i^2z_i^2,1\right)$	$\left[\frac{\nu+\frac{1}{2}}{2\xi_i^2z_i^2}\right]\frac{\mathcal{F}(1\|\nu+\frac{3}{2},2\xi_i^2z_i^2)}{\mathcal{F}(1\|\nu+\frac{1}{2},2\xi_i^2z_i^2)}$
skewed cont. normal	$\nu\mathbb{I}\{u=\gamma\}+(1-\nu)\mathbb{I}\{u=1\}$ $0\le\nu,\gamma\le1$	$\frac{a\,\mathbb{I}\{u=\gamma\}+b\,\mathbb{I}\{u=1\}}{a+b}$	$\frac{a\gamma+b}{a+b}$

Table 3: Conditional distribution of U given Y for specific SKD distributions.

In Table 3, $\mathcal{F}(x|\alpha,1/\beta)$ represents the *cdf* of a Gamma $(\alpha,1/\beta)$ distribution. Moreover, expressions a and b are given by $a=\nu\phi\left(y_i|\mathbf{x}_i^\top\boldsymbol{\beta}_p,\frac{\gamma^{-1}\sigma^2}{4\xi_i^2}\right)$ and $b=(1-\nu)\phi\left(y_i|\mathbf{x}_i^\top\boldsymbol{\beta}_p,\frac{\sigma^2}{4\xi_i^2}\right)$. The notation $\mathrm{TG}(a,b,t)$ represents a random variable with $\mathrm{Gamma}(a,1/b)$ distribution truncated to the right at the value t. Finally, $GIG(\nu,a,b)$ denotes the Generalized Inverse Gaussian (GIG) distribution - see Barndorff-Nielsen & Shephard (2001) for more details. The pdf of the GIG distribution is given by

$$f(x|\nu,a,b) = \frac{(b/a)^\nu}{2K_\nu(ab)}x^{\nu-1}\exp\left\{-\frac{1}{2}\left(a^2/x+b^2x\right)\right\},\ x>0,\quad \nu\in\mathbb{R},\ a,b>0,$$

with $K_\nu(.)$ being the modified Bessel function of the third kind. The proposed EM algorithm can be summarized in the following steps:

1. ***E-step***: Given $\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}$, compute $\widehat{\kappa^{-1}(u_i)}$.
2. ***M-step***: Update $\boldsymbol{\theta}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_p^{(k+1)} &= (\mathbf{X}^\top\boldsymbol{\Omega}^{(k)}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{(k)}\mathbf{y},\\ \widehat{\sigma^2}^{(k+1)} &= \frac{4}{n}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_p^{(k+1)})^\top\boldsymbol{\Omega}^{(k)}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_p^{(k+1)}),\end{aligned}$$

where $\boldsymbol{\Omega}$ is a $n\times n$ diagonal matrix, with elements $\xi_i^2\,\widehat{\kappa^{-1}(u_i)}$, $i=1,\ldots,n$, $\mathbf{X}$ is the design matrix and $\mathbf{y}$ is the vector of observations. After the *M-step*, we will update the parameter $\boldsymbol{\nu}$ by maximizing the marginal log-likelihood function of $\mathbf{y}$, obtaining

$$\widehat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}}\sum_{i=1}^{n}\log f(y_i|\widehat{\boldsymbol{\beta}}_p^{(k+1)},\widehat{\sigma}^{(k+1)},\boldsymbol{\nu}).$$

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 ordinary least squares estimators (OLSE) as an initial estimate of $\boldsymbol{\beta}$, reaching convergence in a few seconds.

3.2 Standard error approximation

Louis' missing information principle (Louis, 1982) relates the score function of the incomplete data log-likelihood with the complete data log-likelihood through the conditional expectation $\boldsymbol{\nabla}_o(\boldsymbol{\theta}) = \mathrm{E}[\boldsymbol{\nabla}_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})|\mathbf{y}]$, where $\boldsymbol{\nabla}_o(\boldsymbol{\theta}) = \partial\ell_o(\boldsymbol{\theta}|\mathbf{y})/\partial\boldsymbol{\theta}$ and $\boldsymbol{\nabla}_c(\boldsymbol{\theta}) = \partial\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})/\partial\boldsymbol{\theta}$ are the score functions for the incomplete and complete data, respectively. As defined in Meilijson (1989), the empirical information matrix can be computed as

$$\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(y_i|\boldsymbol{\theta})\,\mathbf{s}^{\top}(y_i|\boldsymbol{\theta}) - \frac{1}{n}\,\mathbf{S}(\mathbf{y}|\boldsymbol{\theta})\,\mathbf{S}^{\top}(\mathbf{y}|\boldsymbol{\theta}), \tag{9}$$

where $\mathbf{S}(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n}\mathbf{s}(y_i|\boldsymbol{\theta})$ and $\mathbf{s}(y_i|\boldsymbol{\theta})$ is the empirical score function for the ith individual. Replacing $\boldsymbol{\theta}$ by its ML estimator $\widehat{\boldsymbol{\theta}}$ and considering $\boldsymbol{\nabla}_o(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, equation (9) takes the simple form

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(y_i|\widehat{\boldsymbol{\theta}})\,\mathbf{s}^{\top}(y_i|\widehat{\boldsymbol{\theta}}).$$

At the kth iteration, the empirical score function $\mathrm{s}(y_i|\boldsymbol{\theta})^{(k)}$ for the ith subject can be computed as

$$\mathbf{s}(y_i|\boldsymbol{\theta})^{(k)} \quad = \mathrm{E}\left[\mathbf{s}(y_i, u_i^{(k)}|\boldsymbol{\theta}^{(k)})|y_i\right],$$

where $u_i{}^{(k)}$, is the latent variable following the conditional distribution $f(u_i|y_i,\boldsymbol{\theta}^{(k-1)})$. Using Louis's method (Louis, 1982), the observed information matrix at iteration k, can be approximated as $\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\mathbf{y})^{(k)} = \sum_{i=1}^{n}\mathbf{s}(y_i|\boldsymbol{\theta})^{(k)}\mathbf{s}^{\top}(y_i|\boldsymbol{\theta})^{(k)}$. such that at convergence, $\mathbf{I}_e^{-1}(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = (\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y})\mid_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}})^{-1}$ is an estimate of the covariance matrix of the parameter estimates.

Thus, by taking partial derivatives of the complete log-likelihood function in (8) with respect to $\boldsymbol{\theta}$, we obtained the following elements of the score function:

$$\begin{aligned}\frac{\partial\ell_{c_i}}{\partial\boldsymbol{\beta}_p} &= \frac{4}{\sigma}\,[\boldsymbol{\Omega}_1\mathbf{X}]_i\\ \frac{\partial\ell_{c_i}}{\partial\sigma} &= \frac{4}{\sigma^3}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_p)^{\top}\boldsymbol{\Omega}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_p) - \frac{1}{\sigma}\end{aligned}$$

where $\boldsymbol{\Omega}_1$ is a $n\times n$ diagonal matrix with diagonal elements $\xi_i^2\, z_i\,\widehat{\kappa^{-1}(u_i)}$ and $[\,\cdot\,]_i$ denotes the ith element of the vector.

4 Simulation studies

In this section, a simulation study to evaluate the finite sample performance of ML estimates obtained using the proposed EM algorithm is performed. The computational procedure is implemented using `R` software R Core Team (2014) using the `lqr` package by Galarza *et al.* (2015). In particular, we consider the following linear model

$$y_i = \mathbf{x}_i^{\top}\boldsymbol{\beta} + \epsilon_i,\ \ i = 1,\ldots,n.$$

			Quantiles (%)					
			25		50		75	
Distribution	Parameter	n	BIAS	MC-SD	BIAS	MC-SD	BIAS	MC-SD
skewed normal	β_0	100	0.004	(0.060)	0.004	(0.053)	-0.013	(0.062)
		200	0.005	(0.044)	0.002	(0.037)	-0.003	(0.040)
		400	0.002	(0.029)	-0.001	(0.026)	-0.003	(0.029)
	β_1	100	0.000	(0.059)	-0.001	(0.049)	0.002	(0.058)
		200	-0.003	(0.040)	0.001	(0.034)	-0.002	(0.040)
		400	0.001	(0.028)	-0.000	(0.025)	0.000	(0.029)
	β_2	100	0.003	(0.060)	-0.002	(0.049)	0.001	(0.059)
		200	-0.000	(0.048)	0.000	(0.040)	-0.000	(0.046)
		400	-0.002	(0.030)	-0.000	(0.026)	-0.001	(0.028)
	σ	100	-0.009	(0.036)	-0.011	(0.036)	-0.007	(0.035)
		200	-0.006	(0.026)	-0.005	(0.025)	-0.004	(0.026)
		400	-0.001	(0.018)	-0.002	(0.017)	-0.001	(0.017)
skewed Student-t	β_0	100	0.012	(0.070)	-0.001	(0.060)	-0.011	(0.069)
		200	0.005	(0.051)	0.001	(0.042)	-0.004	(0.048)
		400	0.000	(0.033)	0.002	(0.029)	-0.003	(0.034)
	β_1	100	0.005	(0.071)	0.000	(0.060)	-0.002	(0.070)
		200	0.001	(0.055)	-0.001	(0.045)	-0.002	(0.050)
		400	0.002	(0.036)	0.001	(0.029)	0.000	(0.034)
	β_2	100	0.004	(0.080)	0.000	(0.067)	-0.003	(0.078)
		200	0.005	(0.057)	0.002	(0.048)	0.003	(0.053)
		400	0.001	(0.036)	-0.001	(0.030)	0.002	(0.038)
	σ	100	-0.008	(0.050)	-0.012	(0.046)	-0.010	(0.047)
		200	-0.003	(0.036)	-0.002	(0.032)	-0.002	(0.032)
		400	-0.002	(0.024)	-0.002	(0.023)	-0.002	(0.023)
skewed Laplace	β_0	100	0.009	(0.067)	-0.003	(0.052)	-0.006	(0.068)
		200	0.006	(0.042)	0.001	(0.039)	-0.002	(0.046)
		400	0.002	(0.031)	-0.001	(0.029)	-0.002	(0.032)
	β_1	100	-0.005	(0.066)	0.005	(0.055)	0.000	(0.068)
		200	0.001	(0.046)	-0.000	(0.039)	-0.002	(0.044)
		400	-0.001	(0.029)	-0.001	(0.027)	-0.000	(0.029)
	β_2	100	-0.002	(0.074)	0.002	(0.064)	-0.007	(0.078)
		200	-0.000	(0.045)	-0.000	(0.037)	0.000	(0.044)
		400	-0.001	(0.030)	-0.002	(0.027)	-0.001	(0.030)
	σ	100	-0.008	(0.052)	-0.007	(0.050)	-0.007	(0.051)
		200	-0.004	(0.035)	-0.005	(0.036)	-0.002	(0.035)
		400	-0.005	(0.026)	-0.001	(0.026)	-0.003	(0.025)

Table 4: **Simulation study**: Absolute bias (BIAS) and Monte Carlo standard error (MC-SD) for parameter estimates of the QR model for different samples size.

			$\widehat{\beta}_0$			$\widehat{\beta}_1$			$\widehat{\beta}_2$	
Distribution	Quantile (%)	MC-SD	IM-SD	MC-CP	MC-SD	IM-SD	MC-CP	MC-SD	IM-SD	MC-CP
skewed normal	25	0.027	0.026	0.93	0.028	0.027	0.93	0.027	0.028	0.95
	50	0.022	0.023	0.96	0.023	0.023	0.96	0.023	0.024	0.95
	75	0.027	0.026	0.94	0.028	0.027	0.93	0.026	0.028	0.95
skewed Student-t	25	0.036	0.034	0.94	0.035	0.033	0.95	0.037	0.038	0.95
	50	0.030	0.030	0.94	0.030	0.029	0.93	0.030	0.033	0.97
	75	0.034	0.034	0.93	0.033	0.034	0.96	0.038	0.038	0.96
skewed Laplace	25	0.031	0.029	0.93	0.031	0.030	0.94	0.031	0.028	0.93
	50	0.027	0.025	0.95	0.025	0.025	0.95	0.027	0.024	0.92
	75	0.031	0.029	0.94	0.033	0.030	0.93	0.031	0.028	0.94

Table 5: **Simulation study**: Monte Carlo standard deviation (MC-SD), mean standard deviation (IM-Sd) and Monte Carlo coverage probability (MC-CP) estimates of the fixed effects β_0, β_1 and β_2 for different quantiles ($n = 400$).

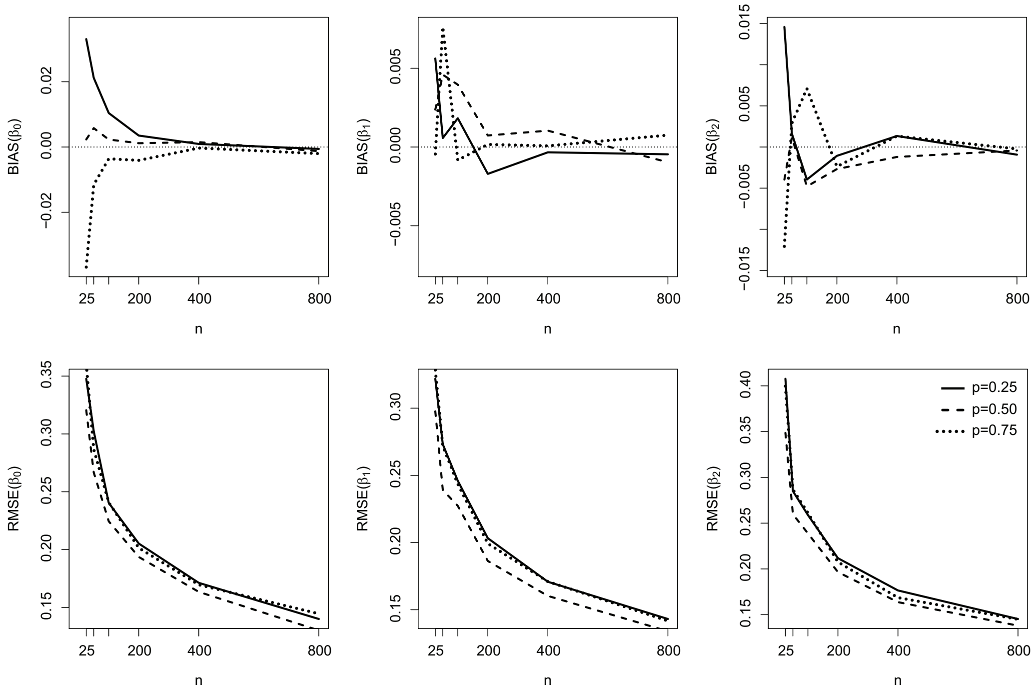

Figure 2: **Simulation study**: Bias (upper panel) and RMSE (lower panel) for the fixed effects β_0, β_1 and β_2 for varying sample sizes over the quantiles $p = \{0.25, 0.50, 0.75\}$.

Our interest is to estimate the fixed effects parameters $\boldsymbol{\beta}$ and nuisance parameter σ, for a grid of quantiles $p = \{0.25, 0.50, 0.75\}$. 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 from a bivariate normal $\mathrm{N}_2(\mathbf{0}, \mathbf{I}_2)$. The parameters were chosen as $\beta_0 = 2$, $\beta_1 = 3$, $\beta_2 = 5$, $\sigma = 0.50$ and $\nu = 4$ and the error term ϵ_i has been generated independently from an $\mathrm{SKD}(0, \sigma, \nu, p)$ distribution, where p stands for the quantile to be estimated. We considered different sample sizes, say, $n = 100$, 200 and 400. For each sample size, we generated $m = 500$ datasets. For the data simulation, we considered the skewed normal, skewed Student-t and skewed Laplace distributions.

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

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

where $\text{Bias}(\widehat{\theta}) = \overline{\widehat{\theta}} - \theta$, $\overline{\widehat{\theta}} = \dfrac{1}{m}\displaystyle\sum_{j=1}^{m}\widehat{\theta}^{(j)}$ is the Monte Carlo mean and $\theta^{(j)}$ is the estimate of θ from the j-th sample, with $j = 1 \ldots m$. 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 $\mathrm{CP}(\hat{\theta}) = \frac{1}{m}\sum_{j=1}^{m} I(\theta \in [\hat{\theta}_{LCL}, \hat{\theta}_{UCL}])$, where I is the indicator function such that θ lies in the interval $[\hat{\theta}_{LCL}, \hat{\theta}_{UCL}]$, with $\hat{\theta}_{LCL}$ and $\hat{\theta}_{UCL}$ the estimated lower and upper bounds of the 95% CI, respectively.

From Table 4 it can be observed that the Bias and MC-SD for the regression parameters β_0, β_1 and β_2 tends to approach zero when sample size is increased revealing that the ML estimates obtained have consistent asymptotic properties. Figure 2 (given in Appendix) shows the obtained results of BIAS and RMSE under the skew normal model. In addition, IM-SD, MC-SD and MC-CP for β_0, β_1 and β_2 are presented in Table 5 for different quantiles. Note that the values of MC-SD and IM-SD are very close indicating that the asymptotic approximation of the parameter standard errors are reliable.

5 Application

In this section we present an application based on a dataset from Cook & Weisberg (1994) on characteristics of Australian athletes available from the Australian Institute of Sport (AIS). We consider the variables body mass index (*bmi*), lean body mas (*lbm*) and gender (*sex*) associated with $n = 202$ Australian athletes. See Figure 3, where we also present the plot of the *bmi* versus the *lbm* and *sex*.

In order to illustrate the model proposed in Section 3, we consider the following quantile regression model

$$bmi_i = \beta_0 + \beta_1 lbm_i + \beta_2 sex_i + \epsilon_i,$$

where ϵ_i belongs to the SKD family for $i = 1, \ldots, 202$. Note that we have disconsidered the interaction (between *lbm* and *sex*) because it was found to be not significant in a preliminary analysis.

Using the `R` package `lqr()` (see Appendix), we fit five models as was described in Section 3, performing a median regression ($p = 0.5$). To compare them, we consider the Akaike (AIC; Akaike, 1974), Schwarz (BIC; Schwarz, 1978), Hannan-Quinn (HQ; Hannan & Quinn, 1979) information criteria and the value of the estimated log-likehood function. Table 6 presents the obtained results for each model comparison criterion. According to these measures, it can be concluded that the best model is the skewed slash been just a little better than skew Student-t model. Figure 4 shows the envelope plots for the residuals obtained after fitting the $p = 0.5$ quantile regression model under the skew Student-t and skew slash distribution respectively. In addition, it can be observed that the rest of the skewed and heavy-tailed models, say, the Student-t, Laplace and contaminated normal outperforms the skewed normal one. Consequently, these results provide evidence about

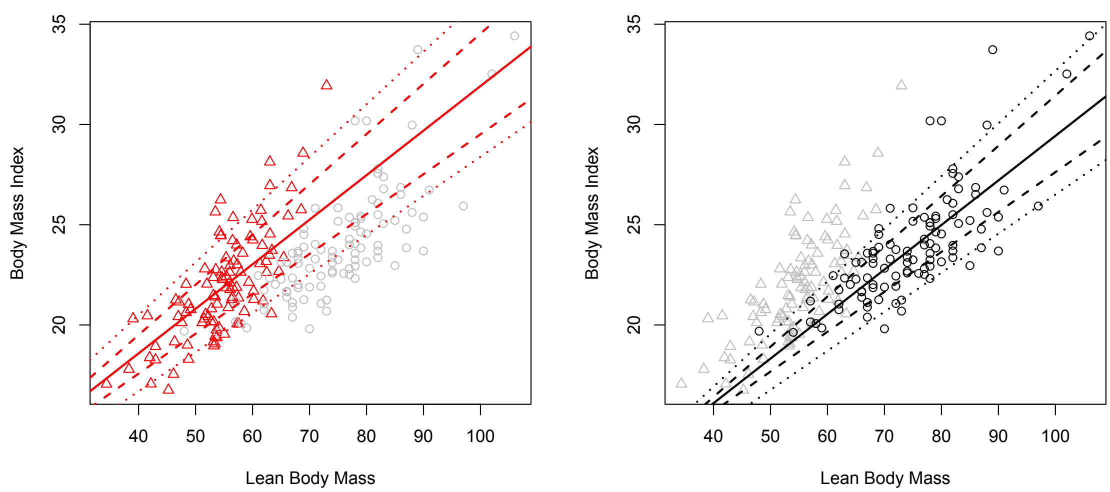

Figure 3: **Data analysis**: Fitted skewed slash QR overlayed with five different quantile regression lines over the grid $p = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ for the AIS data, by gender.

the presence of possible outliers and influence observations in the data that the skewed normal model cannot accommodate.

	skewed normal	skewed Student-t	skewed Laplace	skewed slash	skewed cont. normal
AIC	1095.56	810.99	821.78	**810.83**	814.11
BIC	1108.79	824.22	835.01	**824.06**	827.34
HQ	1100.91	816.34	827.14	**816.18**	819.46
log-likelihood	-543.78	-401.49	-406.89	**-401.41**	-403.05

Table 6: Model comparison criteria for the median ($p = 0.5$) linear regression model using the SKD family.

Under the skewed slash model, Table 7 presents the parameter estimation for the QR model. It is important to note that the significance of the fixed effects is verified for a significance level equal to $\alpha = 0.05$. Figure 3 shows the fitted quantile regression overlayed with five different quantile regression lines over the grid $p = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ for the skew slash model, by gender. For the Student-t model, estimates and standard errors model were very similar. Moreover, the point estimate of the (degrees of freedom) parameter ν is 7.98 indicating an important depart from normality and, therefore, a moderate tail behaviour.

Parameter	Estimate	Std. Error	z value	p-value
β_0	7.21	2.09	3.46	0.000
β_1	0.22	0.03	8.48	0.000
β_2	2.49	0.77	3.25	0.001
σ	1.31			
ν	2.07			

Table 7: Parameter estimation under skewed Slash QR model.

Finally, Figures 5 and 6 show the point estimates and 95% confidence interval for model parameters under the Student-t and slash QR model respectively for different values of the quantiles. Confidence intervals for both models looks really similar. Note that, in each case, the parameters estimates and also de confidence interval are far from zero confirming the conclusion obtained from the analysis of the p-value.

6 Conclusion

In this paper, we have proposed a likelihood-based approach for the estimation of the QR model based on a family of skewed distributions, namely, the SKD class of distributions as opposed to the use of the ALD distribution. By using the relationship between the QR check function and zero quantile property related to this family, we cast the QR problem into the usual likelihood framework. The newly stochastic representation of this class of distributions allows us to express the QR model based on a mixture of normal distributions, making easy the implementation of an EM algorithm for obtaining the ML estimates of the model parameters with closed form expressions at the E- and M- step. This EM algorithm was implemented as part of the `R` package `lqr()`. We hope that by making the code of our method available to the community, we will encourage to other researchers to use the EM algorithm and the SKD class of distributions in their studies of QR.

Finally, the proposed method can be extended to a more general framework, such as, censored regression models, measurement error models, nonlinear regression models, among many others, providing satisfactory results at the expense of additional complexity in implementation. An in-depth investigation of such extension is beyond the scope of the present paper, but certainly an interesting topic for future research.

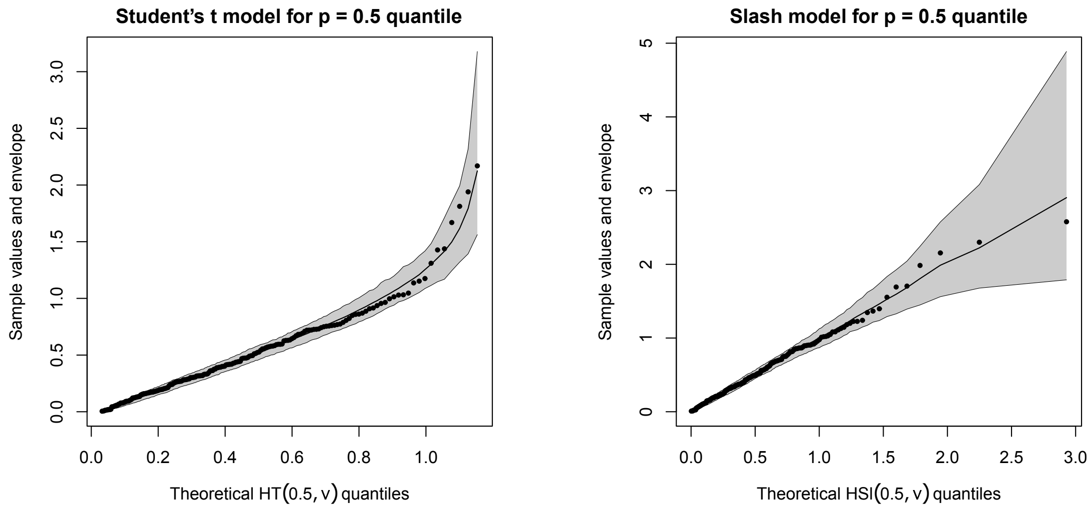

Figure 4: **Data analysis**: Envelopes for the residuals after fitting the $p = 0.5$ skewed Student-t and skewed slash QR model.

Acknowledgements

The research of V.H. Lachos was supported by Grant 306334/2015-1 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2014/02938-9 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). L.M. Castro acknowledges support from Grant FONDECYT 1130233 from the Chilean government and CONICYT-Chile through BASAL project CMM, Universidad de Chile. The research of C.E. Galarza was supported by Grant 2015/17110-9 from FAPESP-Brazil. C. R. Cabral was supported by CNPq-Brazil (Grants 308243/2012-9 and 447964/2014-3).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 379–397.

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B.*, **36**, 99–102.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, **63**, 167–241.

Barrodale, I. & Roberts, F. (1977). Algorithms for restricted least absolute value estimation. *Communications in Statistics-Simulation and Computation*, **6**, 353–363.

Benites, L., Lachos, V. H. & Vilca, F. (2013). Likelihood based inference for quantile regression using the asymmetric Laplace distribution. Technical Report 15, Universidade Estadual de Campinas.

Cook, R. D. & Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.

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.

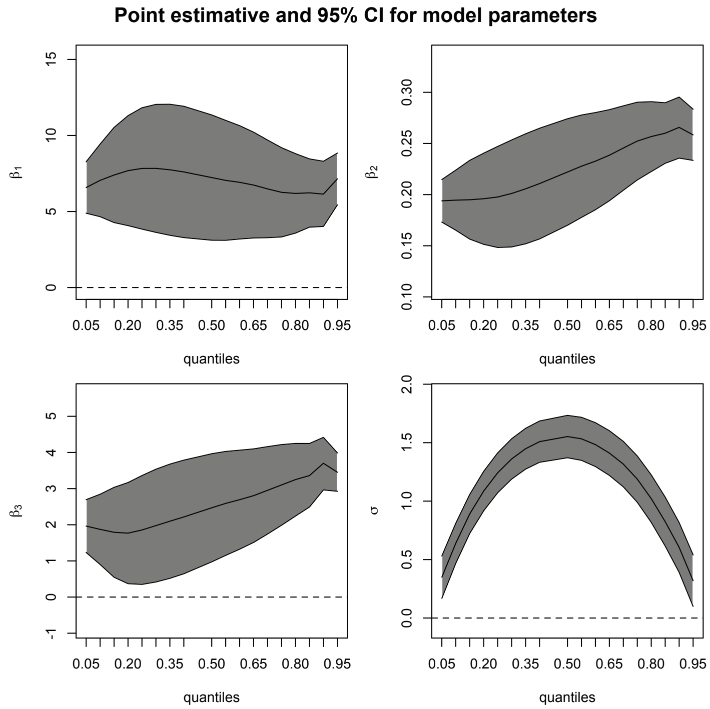

Figure 5: **Data analysis**: Point estimates and 95% confidence interval for model parameters under the skewed Student-t QR model.

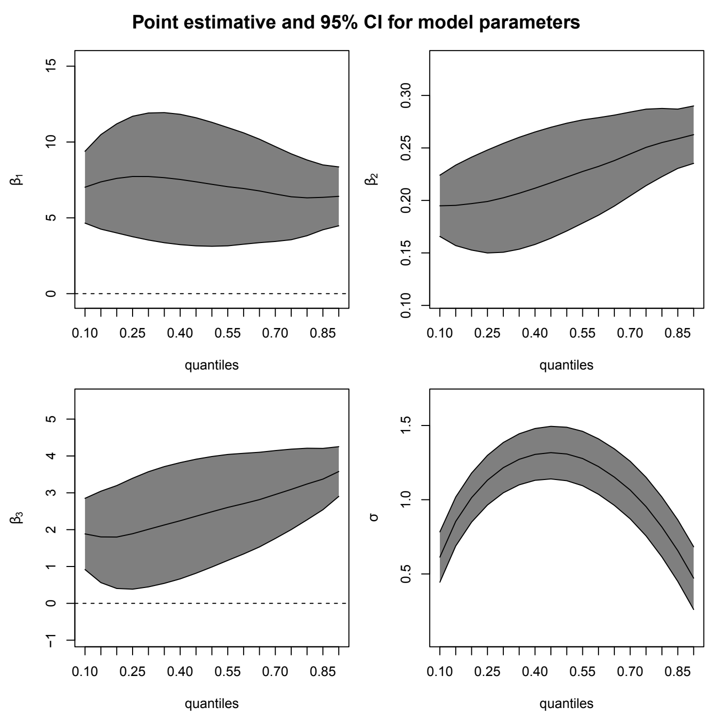

Figure 6: **Data analysis**: Point estimates and 95% confidence interval for model parameters under the skewed slash QR model.

Fernández, C. & Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**(441), 359–371.

Galarza, C. E., Benites, L. & Lachos, V. H. (2015). *lqr: Robust Linear Quantile Regression*. R package version 1.1.

Hannan, E. & Quinn, B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, **41**, 190–195.

Koenker, R. (2005). *Quantile Regression*, volume 38. Cambridge University Press.

Koenker, R. & G Bassett, J. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, **46**, 33–50.

Koenker, R. W. & d'Orey, V. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36**, 383–393.

Kottas, A. & Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468.

Kottas, A. & Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, **36**, 297–319.

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**, 1565–1578.

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

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

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

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

Tian, Y., Tian, M. & Zhu, Q. (2014). Linear Quantile Regression Based on EM Algorithm. *Communications in Statistics - Theory and Methods*, **43**(16), 3464–3484.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, pages 267–288.

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.

Yu, K. & Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–447.

Zhou, Y.-h., Ni, Z.-x. & Li, Y. (2014). Quantile Regression via the EM Algorithm. *Communications in Statistics - Simulation and Computation*, **43**(10), 2162–2172.

Appendix

Proof of Lemma 1

Let $H = |T_0|$, then we have that $X = \mu + \sigma I H$. In order to use the transformation method we define an auxiliar variable $K = I$, leading to $I = K$, $H = (X - \mu)/\sigma K$ and thus $|\mathbf{J}| = 1/\sigma K$, where $|\mathbf{J}|$ represents the determinant of the Jacobian of the transformation. The joint distribution for K and X can be computed as

$$f(k, x|\mu, \sigma) = \frac{P(I = k)}{\sigma|k|} f_H\left(\frac{x - \mu}{\sigma k}\right).$$

Using the fact that I and H are independent, then

$$f(k, x|\mu, \sigma) = \begin{cases} 2p(1-p)/\sigma \, f_H\left(-2(1-p)(x-\mu)/\sigma\right) & ; \quad x \leq 0,\ k = -1/2(1-p) \\ 2p(1-p)/\sigma \, f_H\left(2p(x-\mu)/\sigma\right) & ; \quad x > 0,\ k = 1/2p \end{cases}$$

where $P(I = k) = p\mathbb{I}\{k = -\frac{1}{2(1-p)}\} + (1-p)\mathbb{I}\{k = \frac{1}{2p}\}$ and H is a Half normal random variable with *pdf* given by

$$f_H(h) = \frac{\sqrt{2}}{\sqrt{\pi}} \exp\left\{-\frac{h^2}{2}\right\} \mathbb{I}\{h > 0\}.$$

Then, the *pdf* of X is obtained by marginalization, *i.e.*

$$\begin{aligned} f(x|\mu, \sigma) &= 2\left[\frac{\sqrt{2}p(1-p)}{\sqrt{\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma/2(1-p)}\right)^2\right\}\right] \mathbb{I}\{x \leq 0\} \\ &\quad +2\left[\frac{\sqrt{2}p(1-p)}{\sqrt{\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma/2p}\right)^2\right\}\right] \mathbb{I}\{x > 0\} \\ &= 2\left[p\,\phi\left(x \,\middle|\, \mu, \frac{\sigma^2}{4(1-p)^2}\right) \mathbb{I}\{x \leq 0\} + (1-p)\,\phi\left(y \,\middle|\, \mu, \frac{\sigma^2}{4p^2}\right) \mathbb{I}\{x > 0\}\right], \end{aligned}$$

concluding the proof.

Proof of Lemma 2

From Lemma 1, we have that $Y|U = u \sim \text{SKN}(\mu, \kappa(u)^{1/2}\sigma, p)$. Consequently, the marginal distribution of Y can be obtained from

$$f(y|\mu, \sigma, \boldsymbol{\nu}) = \int_0^\infty \frac{4p(1-p)}{\sqrt{2\pi\kappa(u)\sigma^2}} \exp\left\{-2\rho_p^2\left(\frac{y-\mu}{\kappa^{1/2}(u)\sigma}\right)\right\} dH(u|\boldsymbol{\nu})$$

corresponding to the *pdf* of the (4), concluding the proof.

Proof of Lemma 3

From the definition of check function $\rho_p(\cdot)$, the distance $D = \rho(\frac{Y-\mu}{\sigma})$ can be written as $D = I_2(\frac{Y-\mu}{\sigma})$ where I_2 is a discrete random variable such that $P(I_2 = p - 1) = p$ and $P(I_2 = p) = 1 - p$. Let I_2 be independent of Y. Note that $(Y - \mu)/\sigma$ represents a standardized SKD random variable. Therefore, from Lemma 2, it follows that $D \stackrel{d}{=} I_2\kappa(u)^{1/2}I|T_0|$. Finally, from the fact that $T = I_2 \times I$ is a degenerate random variable such that $P(T = 1/2) = 1$, we conclude the proof.

Sample output from R package lqr()

```
-------------------------------------------------------------
        Quantile Linear Regression using SKD family
-------------------------------------------------------------

Criterion = AIC
Best fit = Slash
Quantile = 0.5

--------------------------------
Model Likelihood-Based criterion
--------------------------------

          Normal Student-t   Laplace     Slash C. Normal
AIC    1095.5632  810.9939  821.7857  810.8339  814.1112
BIC    1108.7963  824.2269  835.0188  824.0670  827.3443
HQ     1100.9173  816.3480  827.1398  816.1880  819.4654
loglik -543.7816 -401.4969 -406.8929 -401.4169 -403.0556

---------
Estimates
---------

       Estimate Std. Error z value Pr(>|z|)
beta 1  7.21136    2.08532 3.45815  0.00054 ***
beta 2  0.22220    0.02619 8.48464  0.00000 ***
beta 3  2.48574    0.76500 3.24932  0.00116  **
---
Signif. codes:  0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

sigma = 1.30806
nu    = 2.0699
```