Computer Methods and Programs in Biomedicine 221 (2022) 106816

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data

Josmar Mazucheli[a], Bruna Alves[a], André F.B. Menezes[a], Víctor Leiva[b,*]

[a] Department of Statistics, Universidade Estadual de Maringá, Brazil
[b] School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile

ARTICLE INFO

Article history:
Received 19 October 2021
Revised 28 February 2022
Accepted 13 April 2022

Keywords:
GLM
Likelihood methods
Parametric regression
Quantile function
R software

ABSTRACT

Quantile regression allows us to estimate the relationship between covariates and any quantile of the response variable rather than the mean. Recently, several statistical distributions have been considered for quantile modeling. The objective of this study is to provide a new computational package, two biomedical applications, one of them with COVID-19 data, and an up-to-date overview of parametric quantile regression. A fully parametric quantile regression is formulated by first parameterizing the baseline distribution in terms of a quantile. Then, we introduce a regression-based functional form through a link function. The density, distribution, and quantile functions, as well as the main properties of each distribution, are presented. We consider 18 distributions related to normal and non-normal settings for quantile modeling of continuous responses on the unit interval, four distributions for continuous response, and one distribution for discrete response. We implement an R package that includes estimation and model checking, density, distribution, and quantile functions, as well as random number generators, for distributions using quantile regression in both location and shape parameters. In summary, a number of studies have recently appeared applying parametric quantile regression as an alternative to the distribution-free quantile regression proposed in the literature. We have reviewed a wide body of parametric quantile regression models, developed an R package which allows us, in a simple way, to fit a variety of distributions, and applied these models to two examples with biomedical real-world data from Brazil and COVID-19 data from US for illustrative purposes. Parametric and non-parametric quantile regressions are compared with these two data sets.

1. Introduction

The aim of traditional regression is to assess the effect of one or more explanatory variables (hereafter covariates) on the mean of the response variable (hereafter response) [109]. The idea of modeling the conditional mean using covariates is the core of the regression techniques. Under the assumption of normality and homoscedasticity of an error term, a traditional regression model is able to provide a parsimonious description of how the mean of the response depends on the values of the covariates [54]. In a parametric context, the use of traditional regression models is unfeasible when the underlying probability or statistical distribution (hereafter distribution) does not have a simple form for its mean, making it difficult to assess the effects of covariates on the mean response.

An attractive alternative beyond mean modeling is the quantile regression proposed in [58], where the mean is replaced by a defined set of quantiles that provide a better and complete view of the underlying relationships between the response and covariates. There is a lot of literature about quantile regression, which has been applied in areas as biology, economy, engineering, and medicine. A review of quantile regression has been provided in [46] for different types of data and application areas. Note that there are three approaches to quantile regression: (i) distribution-free [57], (ii) based on a pseudo-likelihood through the Laplace distribution [58]; and (iii) the parametric modeling using maximum likelihood (ML) methods. The approach mentioned in (i) above may be overly complicated or unnecessary for the relatively simple forms of quantile dependence that are often observed in real-world data. Besides the disadvantage of not exploring the parametric sta-

* Corresponding author.
E-mail address: victorleivasanchez@gmail.com (V. Leiva).

https://doi.org/10.1016/j.cmpb.2022.106816
0169-2607/

tistical modeling elements that can be considered, this approach can bring problems such as crossing quantile curves [36,98].

In the parametric approach indicated in (iii), the central idea is to insert a quantile parameter in a baseline distribution so obtaining a distribution parameterized in terms of a fixed quantile. By parameterizing a distribution in terms of its quantile function, one may interpret its location parameter as being a quantile of the distribution from which one formulates a regression for a fixed value of this. This strategy was adopted, among others, by [92,93] with the Kumaraswamy distribution [67] which claims that:

Employing the median-dispersion re-parameterizations of the Kumaraswamy distribution instead of the beta distribution in regression models may be preferable in at least three cases. First, the median of the dependent variable may be more interesting or relevant than its mean on theoretical grounds. Second, if the conditional distribution of the dependent variable is skewed, the median may be a more appropriate measure of central tendency than the mean. Third, by using the median as location parameter, Kumaraswamy regressions are likely to be much more robust to outliers than beta regressions.

It should be worth mentioning that, instead of “median-dispersion re-parameterizations”, we can consider “quantile re-parameterizations” for any distribution with a closed-form expression for the quantile function. In recent years, in addition to the Kumaraswamy distribution, several distributions have been used in their “quantile re-parameterization” forms. The veracity of this statement can be confirmed by many recently published works on this subject presented through the present paper.

Therefore, following the proposal considered in [58], the main objectives of our study are to provide a new computational package implemented in the R software, two biomedical applications, one of them with COVID-19 data, and an up-to-date review of the parametric quantile regression models obtained re-parameterizing a distribution in terms of a quantile. We expect this study to be a reference source, and to encourage the use of parametric quantile regression. Although we employ distributions with support on the unit interval, the quantile family of two-parameter distributions described in [118] and recently elaborated by [117] will not be considered in this survey. Also, the semiparametric quantile regression models using the quantile-based asymmetric densities family, introduced in [36], will also not be considered, but some comparison with semiparametric quantile structures are provided.

This paper is organized as follows. Section 2 identifies the distributions used in the analysis of Gaussian-related bounded responses on the unit interval, whereas Section 3 provides similarly the case of non-Gaussian-related bounded responses on this interval. In Section 4, we introduce distributions for continuous positive responses and the case of a discrete response. In Section 5, the regression formulation is presented as well as the ML estimation method while introducing an R package named `unitquatreg`, closing this section with two applications based on real-world biomedical data sets from Brazil and United States (US), including COVID-19 data, as illustration. Some concluding remarks are stated in Section 6. Parametric and non-parametric quantile regression models are compared with these two data sets and reported in an appendix.

2. Parametric quantile regressions for Gaussian-related bounded responses

In this section, in alphabetic order, we present the distributions utilized as baseline in quantile regression modeling of responses on the intervals (0,1) and [0,1] generated from Gaussian or normal distributions. The regression model is developed by first re-parameterizing, in terms of the 100τth quantile, with $0<\tau<1$, one parameter of a baseline distribution and then introducing a regression-based functional form through an appropriate link function. We investigate the relevant properties of the following distributions: exponentiated arcsech-normal hyperbolic [62], Johnson SB [18,120], unit-Birnbaum-Saunders [77,81,82], unit-half-normal [5,60], and Vasicek [75,80]. The main probabilistic features, such as probability density function (PDF), cumulative distribution function (CDF) and quantile function (QF) are introduced. Furthermore, for illustrative purposes and detecting its distributional shapes, we show the plots of the quantile re-parameterized PDFs.

2.1. The exponentiated arcsech-normal hyperbolic distribution

The exponentiated arcsech-normal hyperbolic model [62] is obtained from the transformation $F(y;\alpha,\theta)=G(y;\alpha,0)^{\theta}$, where G denotes here the arcsech-normal hyperbolic CDF. The corresponding PDF, CDF and QF of Y are written, respectively, as

$$f(y;\ \alpha,\ \theta) = \frac{2\ \theta}{\alpha\ y\ \sqrt{1\ -\ y^2}}\phi\left[\frac{1}{\alpha}\text{arcsech}(y)\right] \times\left\{2\ -\ 2\ \Phi\left[\frac{1}{\alpha}\text{arcsech}(y)\right]\right\}^{\theta\ -\ 1}, \tag{2.1}$$

$$F(y;\ \alpha,\ \theta) = \left\{2-2\Phi\left[\frac{1}{\alpha}\text{arcsech}(y)\right]\right\}^{\theta},$$
$$Q(\tau;\alpha,\ \theta) = \text{sech}\left[\alpha\Phi^{-1}\left(1-\frac{\tau^{1/\theta}}{2}\right)\right], \tag{2.2}$$

where $0<y<1$ and $\alpha,\ \theta>0$, with Φ^{-1} denoting the QF of the standard normal distribution corresponding to the inverse function of Φ, that is, the standard normal CDF obtained from the standard normal PDF, ϕ namely. In addition, $\text{arcsech}(y)=\log[1+(1-y^2)^{1/2})/y]$ and $\text{sech}(y)=2/[\exp(y)+\exp(-y)]$. When y tends to zero, since $\text{arcsech}(y)\sim\log(y)\to\infty$ and then the exponential term appears in $f(y;\alpha,\theta)$, we have $f(y;\ \alpha,\ \theta)\to 0$. When y tends to one, since $\text{arcsech}(1)=0$, we have $f(y;\ \alpha,\ \theta)\to\infty$. From the expression defined in (2.2), the parameter α can be re-parameterized as $\alpha=h^{-1}(\mu)=\text{arcsech}(\mu)/\Phi^{-1}[(2-\tau^{1/\theta})/2]$, such that μ is, for a fixed and known value τ, the $100\,\tau$th quantile of the distribution of Y. Fig. 1 shows some possible shapes of the re-parameterized exponentiated arcsech-normal hyperbolic PDF for selected values of μ, θ and τ. Possible shapes of this distribution are slanted as well as U and N shaped and increasing shapes. In [62], the exponentiated arcsech-normal hyperbolic model was applied to data that verified the relationship between reading accuracy with dyslexia and intelligence quotient. The model was compared with a quantile unit-Weibull regression, indicating that the exponentiated arcsech-normal hyperbolic regression has better modeling capabilities for this application.

2.2. The Johnson SB distribution

The Johnson SB model [18] is obtained from the transformation $Y=\{1+\exp[-(X-\alpha)/\theta]\}^{-1}$, where $X\sim\text{N}(0,1)$, which denotes a standard normal distributed random variable. The corresponding PDF, CDF and QF of Y are stated, respectively, as

$$f(y;\ \alpha,\ \theta) = \frac{\theta}{\sqrt{2\ \pi}}\frac{1}{y\ (1\ -\ y)} \times\exp\left\{-\frac{1}{2}\left[\alpha\ +\ \theta\ \log\left(\frac{y}{1\ -\ y}\right)\right]^2\right\}, \tag{2.3}$$

$$F(y;\alpha,\theta) = \Phi\left[\alpha+\theta\ \log\left(\tfrac{y}{1\ -\ y}\right)\right],$$
$$Q(\tau;\alpha,\theta) = \frac{\exp\left[\dfrac{\Phi^{-1}(\tau)\ -\ \alpha}{\theta}\right]}{1\ +\ \exp\left[\dfrac{\Phi^{-1}(\tau)\ -\ \alpha}{\theta}\right]}, \tag{2.4}$$

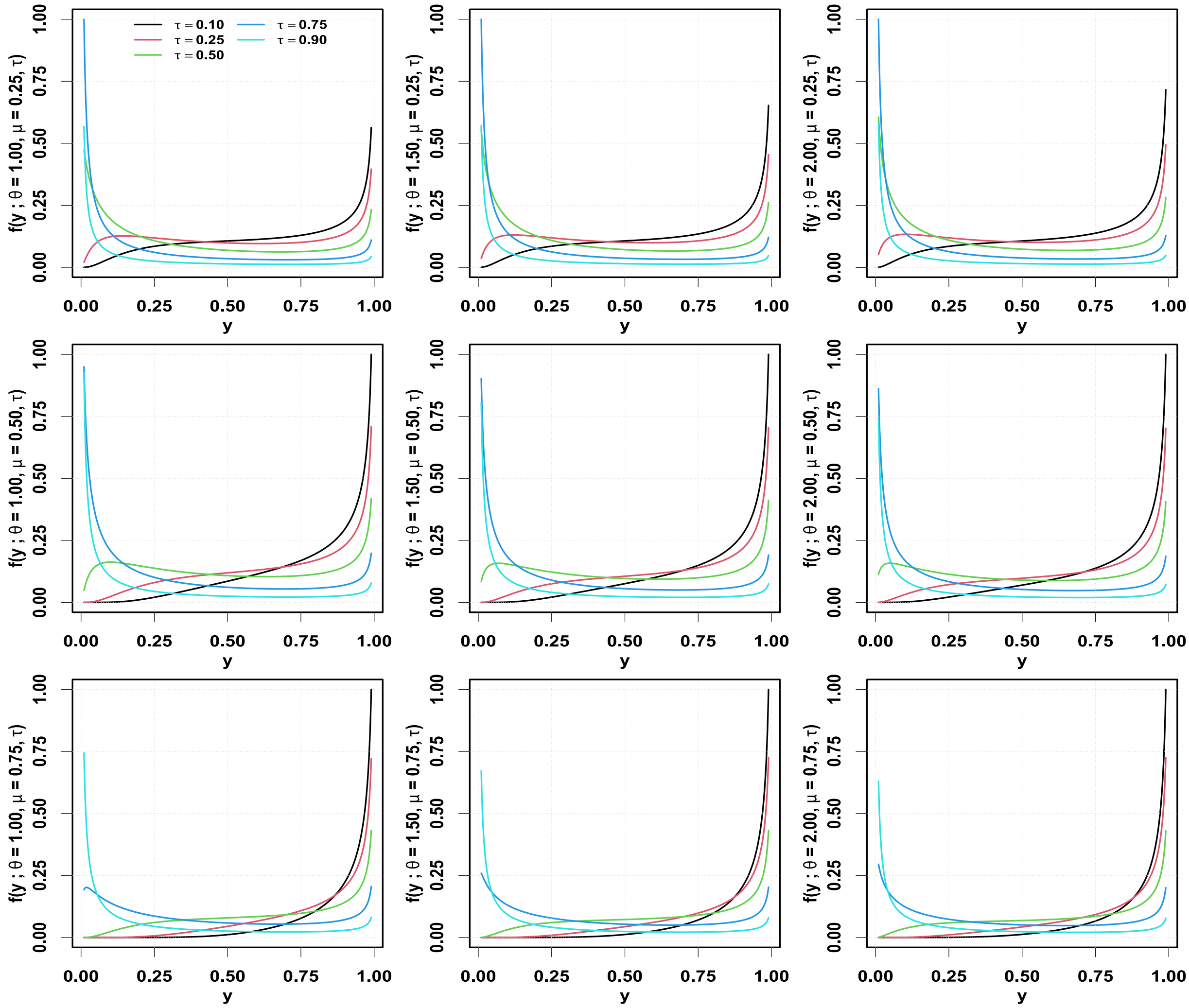

Fig. 1. Plots of the re-parameterized PDF stated from (2.1) for indicated values of μ, θ and τ.

where $0 < y < 1$, $\alpha \in \mathbb{R}$ and $\theta > 0$. For quantile regression, note that α presented in the expression defined in (2.4) must be reparametrized as $\alpha = h^{-1}(\mu) = \Phi^{-1}(\tau) - \theta \log[\mu/(1 - \mu)]$, where μ is, for a fixed and known value τ, the $100\,\tau$th quantile of the distribution of Y. Fig. 2 shows some possible shapes of the re-parameterized Johnson SB PDF for selected values of μ, θ and τ.

A quantile regression model considering the Johnson SB distribution has not been considered in the literature. However, tree quantile regression models have been proposed stating some transformations from a Johnson SB distributed random variable. The first model was presented with basis on the symmetric family of distributions [72,114]. This family is also called generalized Johnson SB distribution and is obtained by replacing $X \sim \mathrm{N}(0, 1)$ by $X \sim \mathrm{S}(0,\ 1;\ f_g)$ where $X \sim \mathrm{S}(0,\ 1;\ f_g)$ means that the random variable X follows the standardized form of the symmetric family of distributions for some PDF generating function f_g. The corresponding PDF, CDF and QF of $Y = \{1 + \exp[-(X - \alpha)/\theta]\}^{-1}$, where $X \sim \mathrm{N}(0, 1)$, are formulated, respectively, as

$$f(y;\ \alpha,\ \theta) = \frac{\theta\ f_g\left\{[\alpha + \theta\ t(y)]^2\right\}}{y\ (1 - y)}, \tag{2.5}$$

$$\begin{aligned} F(y;\ \alpha,\ \theta) &= \int_{-\infty}^{\alpha + \theta\ t(y)} f_g\left(u^2\right)\ \mathrm{d}u, \\ Q(\tau;\ \alpha,\ \theta) &= \left[1 + \exp\left(-\tfrac{(x_\tau - \alpha)}{\theta}\right)\right]^{-1}, \end{aligned} \tag{2.6}$$

where $0 < y < 1$, $\alpha > 0$, $\theta \in \mathbb{R}$, $t(y) = \log[y/(1-y)]$, and x_τ is the $100\,\tau$th quantile of $X \sim \mathrm{S}(0,\ 1;\ f_g)$.

The second model was introduced in [18] considering the power Johnson SB distribution, that is obtained from the Johnson SB distribution and the composition of a baseline standard power normal distribution [41] and the QF of the logistic distribution. The corresponding PDF, CDF and QF of Y are defined, respectively, as

$$f(y;\ \alpha,\ \theta,\ \delta) = \frac{\delta\ \theta\ \phi\left[\alpha + \theta\ \log\left(\frac{y}{1-y}\right)\right]\left\{\Phi\left[\alpha + \theta\ \log\left(\frac{y}{1-y}\right)\right]\right\}^{\delta - 1}}{y\ (1 - y)}, \tag{2.7}$$

$$\begin{aligned} F(y;\ \alpha,\ \theta,\ \delta) &= \left\{\Phi\left[\alpha + \theta\ \log\left(\frac{y}{1 - y}\right)\right]\right\}^{\delta}, \\ Q(\tau;\ \alpha,\ \theta,\ \delta) &= H\left(\frac{x_\tau(\delta) - \alpha}{\theta}\right), \end{aligned} \tag{2.8}$$

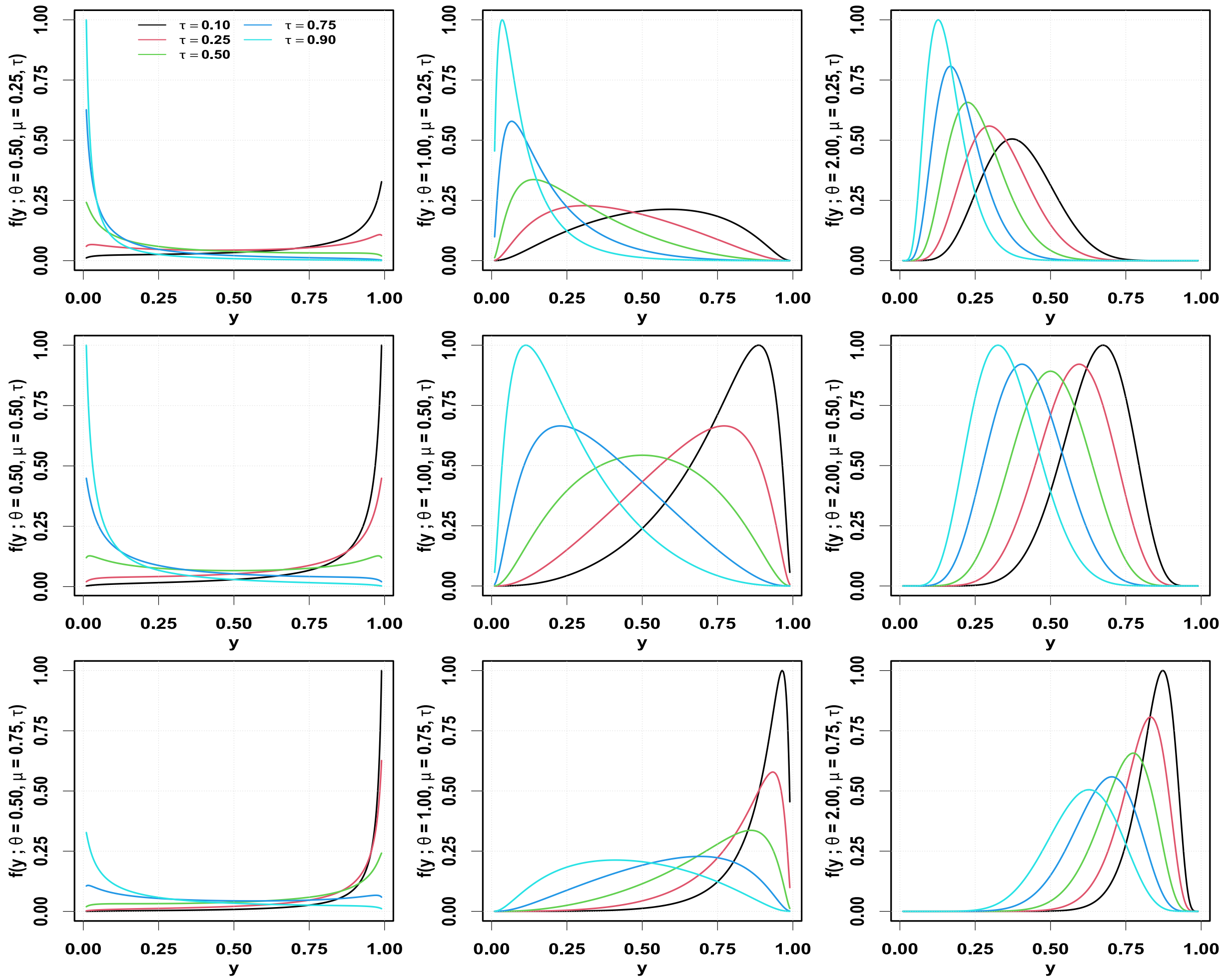

Fig. 2. Plots of the re-parameterized PDF stated from (2.3) for indicated values of μ, θ and τ.

where $0 < y < 1$, $\alpha > 0$, $\theta > 0$ and $\delta \in \mathbb{R}$, $H(z) = 1/[1 + \exp(-z)]$, with $x_\tau(\delta)$ being the 100 τth quantile of the power normal distribution. When $\delta = 1$, the power Johnson SB distribution reduces to the Johnson SB distribution. The lack of a simple formula for the mean of the power Johnson SB distribution inhibits the construction of a regression model, but its median has a simple form. From the expression defined in (2.8), the parameter α can be expressed as $\alpha = h^{-1}(\mu) = x_\tau(\delta) - \theta\, H^{-1}(\mu)$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y.

The third model considers the Johnson-t distribution. This distribution is obtained by replacing $X \sim \mathrm{N}(0, 1)$ by $X \sim t(\nu)$ when generating the Johnson SB distribution, where $X \sim t(\nu)$ means that the random variable X follows the Student-t distribution with $\nu > 0$ degrees of freedom. Then, by the transformation

$$Y = \left\{1 + \exp\left[-\left(\frac{X - \alpha}{\theta}\right)\right]\right\}^{-1},$$

we obtain that Y follows a Johnson-t distribution. The corresponding PDF, CDF and QF of Y are established, respectively, as

$$f(y;\ \alpha,\ \theta) = \frac{\theta\ \nu^{\frac{\nu}{2}} B\left(\frac{1}{2}, \frac{\nu}{2}\right)^{-1}}{y\ (1 - y)} \left\{\nu + [\alpha + \theta\ l(y)]^2\right\}^{\frac{-\nu-1}{2}}, \tag{2.9}$$

$$F(y;\ \alpha,\ \theta) = \frac{1}{2}\left\{1 + \operatorname{sign}[\alpha + \theta\ l(y)]\left[1 - I_{m(\alpha+\theta h(y))}\left(\frac{\nu}{2}, \frac{1}{2}\right)\right]\right\},$$

$$Q(\tau;\ \alpha,\ \theta) = \left[1 + \exp\left(-\frac{(Q_x(\tau) - \alpha)}{\theta}\right)\right]^{-1}, \tag{2.10}$$

where $0 < y < 1$, ν represents the degrees of freedom, $\alpha \in \mathbb{R}$ is the location parameter, and $\theta > 0$ is the dispersion parameter, with $Q_x(\tau)$ being the 100 τth quantile of the Student-t distribution. Also, note that $l(y) = \log[y/(1 - y)]$, $m(z) = z/(\nu + z^2)$, and $I_y(a, b) = B(y; a, b)/B(a, b)$ is the regularized incomplete beta function, with $B(a, b)$ and $B(y; a, b)$ being the incomplete and complete beta functions, respectively. From the expression defined in (2.10), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = -\log[(1 - \mu)/\mu]\ \theta - Q_x(\tau)$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Motivated by the presence of zeros or ones, a new class of zero-or-one inflated distributions was introduced using a mixture of two models: a generalized Johnson SB distribution and a degenerate distribution at a known value c, where $c = 0$ or $c = 1$, depending on the case. Thus, the PDF takes the form stated as $k(y;\ \nu,\ \mu,\ \theta\ ,\ \tau) = \nu$, if $y = c$ or $k(y;\ \nu,\ \mu,\ \theta,\ \tau) = (1 - \nu)\ f(y;\ \mu,\ \theta,\ \tau)$, if $y \in (0, 1)$, where f is the re-parameterized generalized Johnson SB PDF.

2.3. The unit-Birnbaum-Saunders distribution

The Birnbaum-Saunders distribution [77,76,81] is often considered as a life distribution due to its origins in fatigue of materials. Hence, this distribution assumes a prominent role in the areas of reliability and survival analysis, being a good alternative to traditional distributions. In addition, the Birnbaum-Saunders distribution has been considered as a model for tumor growth [68] among

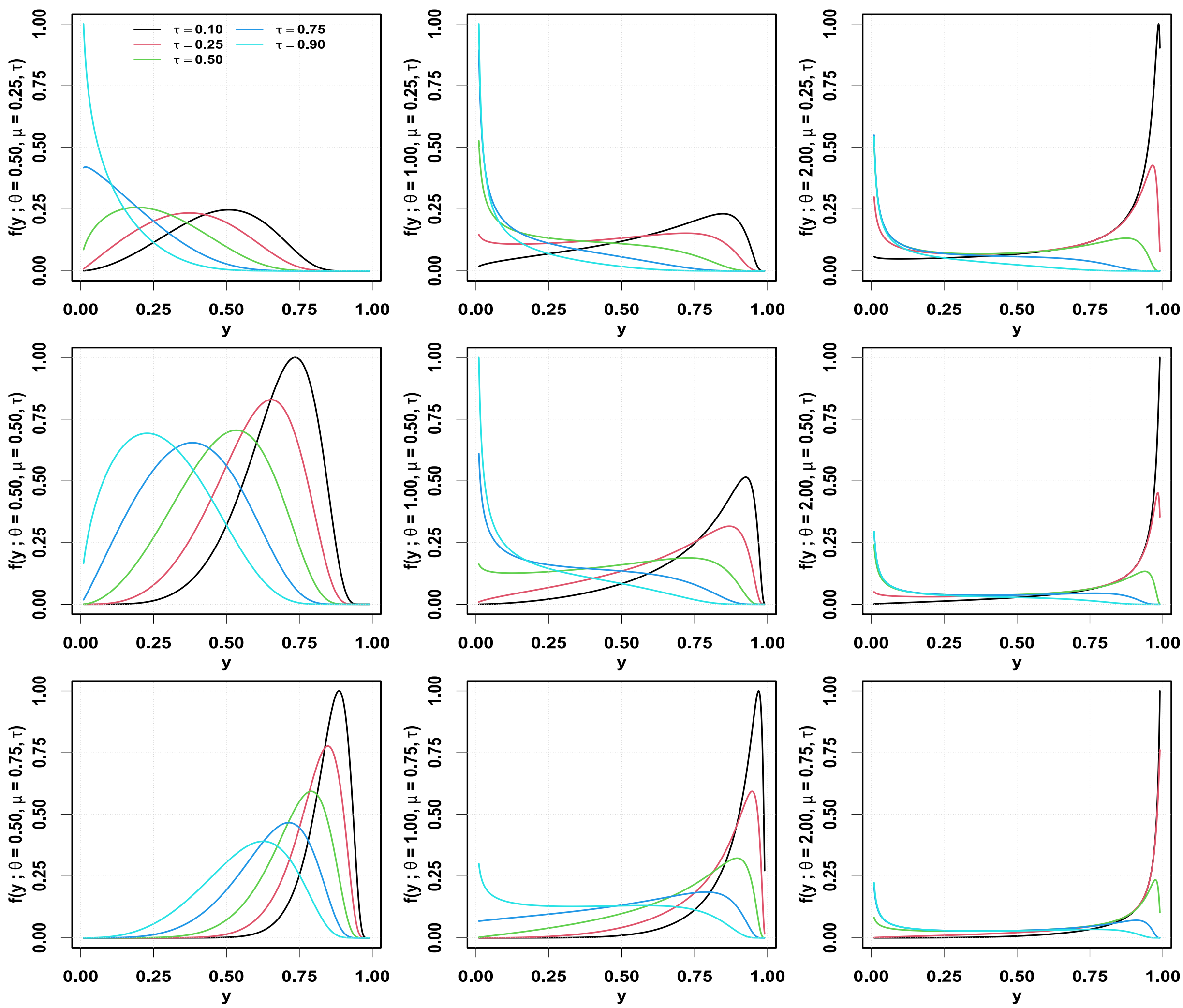

Fig. 3. Plots of the re-parameterized PDF stated from (2.11) for indicated values of μ, θ and τ.

other biomedical applications. The unit-Birnbaum-Saunders model [82] is obtained from the transformation $Y = \exp(-X)$, where $X \sim \mathrm{BS}(\alpha, \theta)$, which denotes a Birnbaum-Saunders distributed random variable [69]. The PDF, CDF and QF of X are written, respectively, as

$$f(y;\ \alpha,\ \theta) = \frac{1}{2y\alpha\theta\sqrt{2\pi}}\left\{\left[-\frac{\alpha}{\log(y)}\right]^{1/2} + \left(-\frac{\alpha}{\log(y)}\right)^{3/2}\right\} \exp\left\{\frac{1}{2\theta^2}\left[2 + \frac{\log(y)}{\alpha} + \frac{\alpha}{\log(y)}\right]\right\}, \tag{2.11}$$

$$F(y;\alpha,\theta) = 1 - \Phi\left(\frac{1}{\theta}\left\{\left[-\frac{\log(y)}{\alpha}\right]^{\frac{1}{2}} - \left[-\frac{\alpha}{\log(y)}\right]^{\frac{1}{2}}\right\}\right),$$

$$Q(\tau;\alpha,\theta) = \exp\left\{-\frac{2\alpha}{2 + [\theta\Phi^{-1}(1-\tau)]^2 - \theta\Phi^{-1}(1-\tau)\sqrt{4 + [\theta\Phi^{-1}(1-\tau)]^2}}\right\}, \tag{2.12}$$

where $0 < y < 1$, $\theta > 0$ and $\alpha > 0$.

Note that $\delta = \exp(-\alpha)$ is the median of the distribution of Y, since $F(\delta;\alpha,\theta) = 0.5$ and the rth moment, for $r \in \{1, 2, \dots\}$, of Y is given by

$$\mathrm{E}(Y^r) = \frac{\left(1 + 2\ r\ \alpha\ \theta^2 + \sqrt{1 + 2\ r\ \alpha\ \theta^2}\right)}{2\left(1 + 2\ r\ \alpha\ \theta^2\right)} \exp\left(\frac{1 - \sqrt{1 + 2\ r\ \alpha\ \theta^2}}{\theta^2}\right).$$

Observe that α nor θ have a direct interpretation in terms of the observed data. For example, α is no longer the median as in the distribution of X. However, from the expression defined in (2.12), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = \log(\mu)\ l(\theta,\ \tau)$, where

$$l(\theta,\ \tau) = -\frac{1}{2}\left\{2 + \left[\theta\ \Phi^{-1}(1-\tau)\right]^2 - \theta\ \Phi^{-1}(1-\tau)\sqrt{4 + \theta\ \Phi^{-1}(1-\tau)]^2}\right\},$$

such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y.

Plots of the PDF for the re-parameterized unit-Birnbaum-Saunders distribution using several values μ, θ and τ are given in Fig. 3. Note that increasing values of θ also increases the negative asymmetry. As μ increases, the variance decreases and the curves tend to become unimodal for all values of τ. In addition, observe that, when varying τ, the PDF takes different shapes.

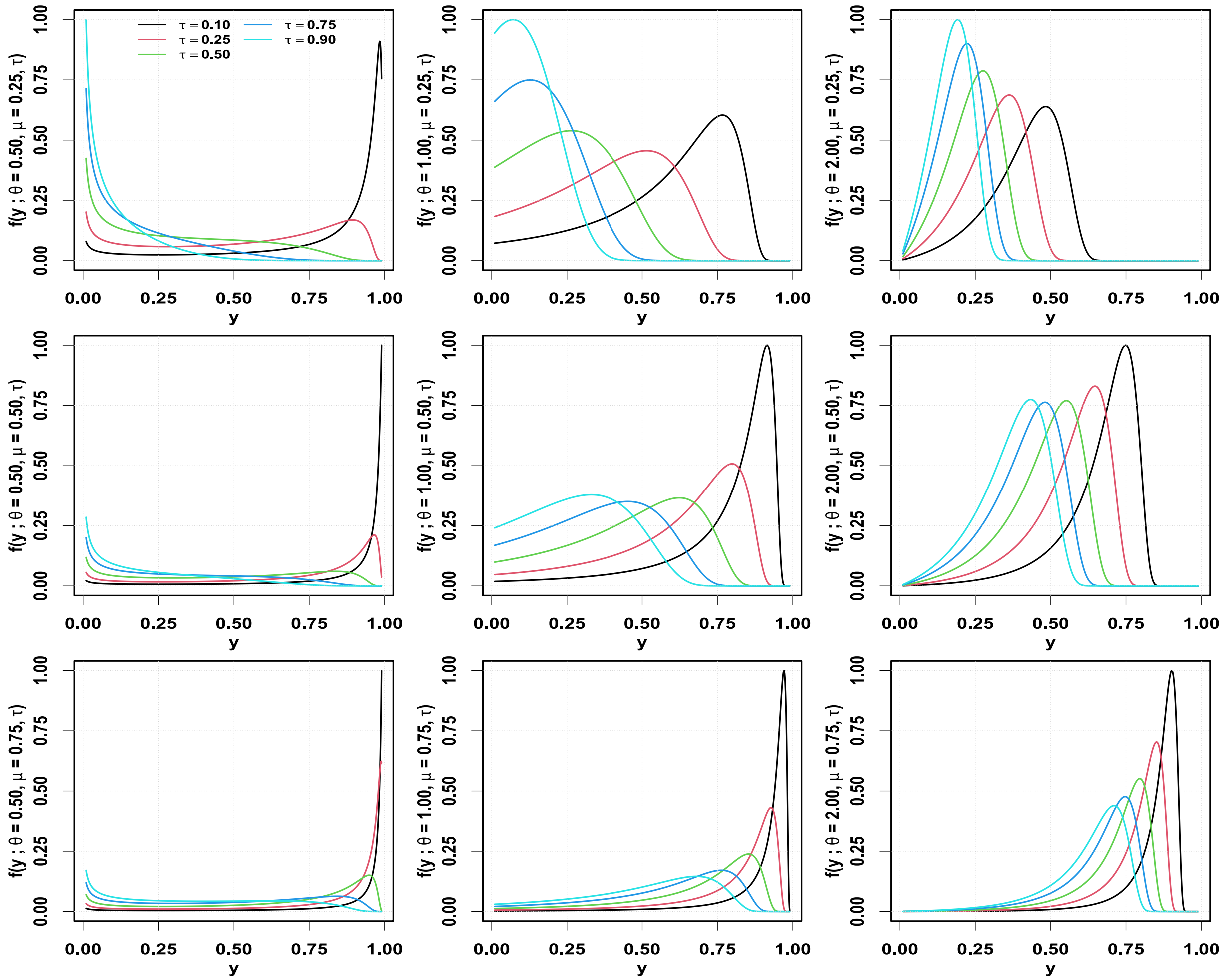

Fig. 4. Plots of the re-parameterized PDF stated from (2.15) for indicated values of μ, θ and τ.

Two examples with real data were provided in [77] related to political sciences and sports medicine. For comparison purposes, in addition to the unit-Birnbaum-Saunders quantile regression model, the Kumaraswamy, L-logistic, log-extended exponential-geometric, unit-Burr-XII, unit-Chen, unit-half-normal, and unit-Weibull quantile regression models were also considered. Both of these examples reported a performance superior to all of the competing models, giving evidence that the unit-Birnbaum-Saunders distribution is an excellent alternative for quantile modeling and for dealing with bounded data into the unit interval. For parameter estimation, model selection and diagnostics based on the unit-Birnbaum-Saunders distribution, the codes are available at https://github.com/AndrMenezes/unitBSQuantReg and by the `unitBSQuantReg` package of R [81].

2.4. The unit-half-normal distribution

The unit-half-normal model [5] is obtained from the transformation $Y = X/(1+X)$, where $X \sim \mathrm{HN}(\alpha)$, which denotes a half-normal distributed random variable [26]. The corresponding PDF, CDF and QF of Y are formulated, respectively, as

$$f(y;\alpha) = \frac{2}{\alpha(1-y)^2}\phi\left(\frac{y}{\alpha(1-y)}\right), \tag{2.13}$$

$$F(y;\alpha) = 2\Phi\left(\frac{y}{\alpha(1-y)}\right) - 1,$$

$$Q(\tau;\alpha) = \frac{\alpha\,\Phi^{-1}\left(\frac{\tau+1}{2}\right)}{1+\alpha\,\Phi^{-1}\left(\frac{\tau+1}{2}\right)}, \tag{2.14}$$

where $0 < y < 1$ and $\alpha > 0$. The rth moment of Y is given by

$$\mathrm{E}(Y^r) = \alpha^r \mathrm{E}\left(\frac{X}{1+\alpha X}\right), \quad r \in \{1, 2, \ldots\}.$$

From the expression defined in (2.14), the parameter α can be re-parameterized as

$$\alpha = h^{-1}(\mu) = \frac{\mu}{(1-\mu)\ \Phi^{-1}([\tau+1]/2)},$$

such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. In [5], the unit-half-normal distribution outperformed the fit obtained by the unit-logistic, unit-Lindley [76], Kumaraswamy and beta distributions considering image data.

An extension of the unit-half-normal distribution may be obtained taking the generalized half-normal distribution as baseline [21]. Considering the transformation $Y = X/(1+X)$, where now $X \sim \mathrm{GHN}(\alpha, \theta)$ denotes a generalized half-normal distributed random variable with CDF given by $F_X(x;\alpha,\theta) = 2\Phi[-(x/\alpha)^{\theta}] - 1$, the PDF, CDF and QF of Y are presented, respectively, as

$$f(y;\alpha,\theta) = \sqrt{\frac{2}{\pi}}\,\frac{\theta}{y(1-y)}\left[\frac{y}{\alpha(1-y)}\right]^{\theta}\exp\left(-\frac{1}{2}\left[\frac{y}{\alpha(1-y)}\right]^{2\theta}\right), \tag{2.15}$$

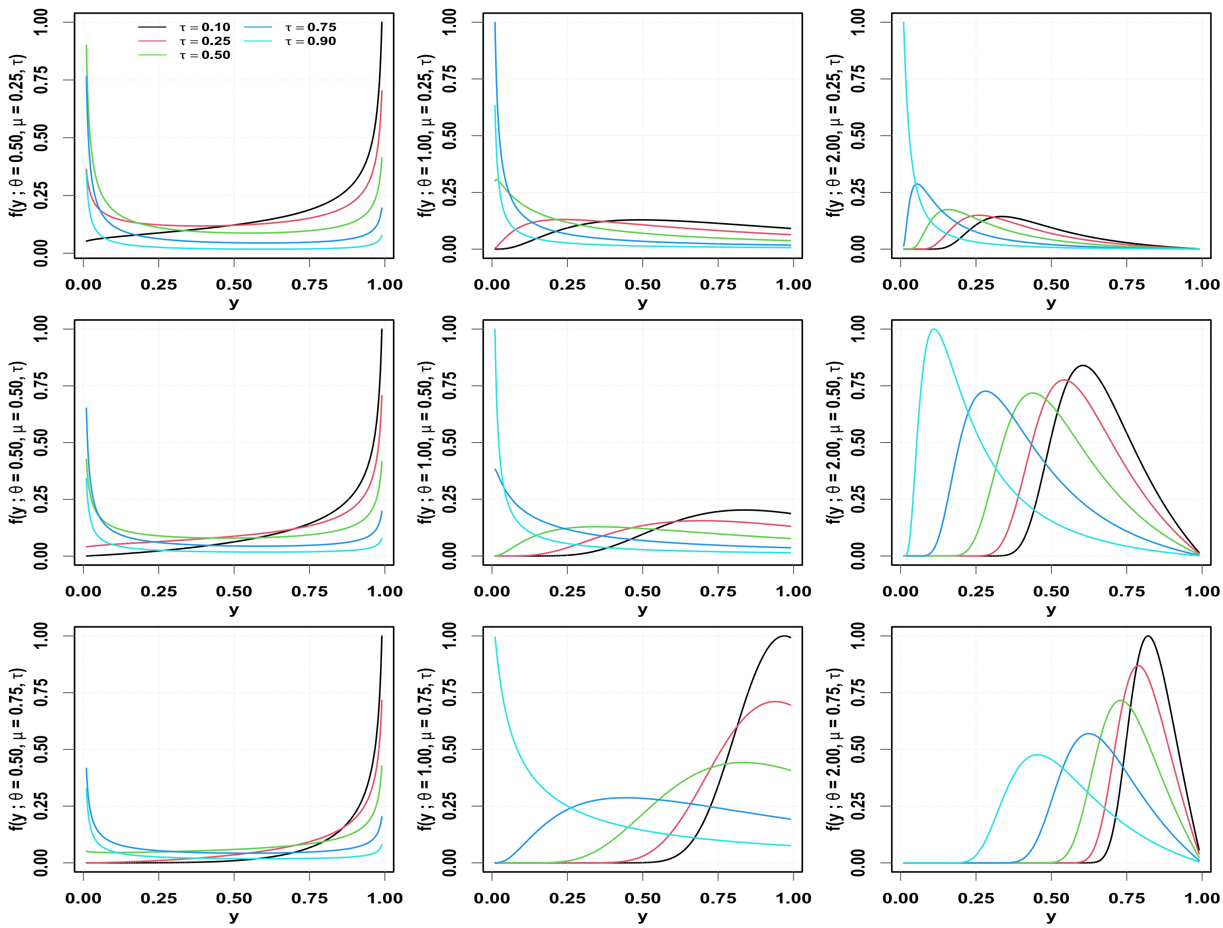

Fig. 5. Plots of the re-parameterized PDF stated from (2.17) for indicated values of μ, θ and τ.

$$F(y;\alpha,\theta) = 2\Phi\left[\left(\frac{y}{\alpha(1-y)}\right)^{\theta}\right] - 1,$$

$$Q(\tau;\alpha,\theta) = \frac{\alpha\left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right]^{\frac{1}{\theta}}}{1+\alpha\left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right]^{\frac{1}{\theta}}}, \quad 0 < y < 1. \tag{2.16}$$

To evaluate the effect of covariates on the quantile of the distribution, the parameter α can be expressed as

$$\alpha = h^{-1}(\mu) = \frac{\mu}{(1-\mu)[\Phi^{-1}([\tau+1]/2)]^{\frac{1}{\theta}}},$$

such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. For $\theta = 1$, we have the unit-half-normal distribution considered in [5]. In addition, from the transformation $Y = \exp(-X)$, where $X \sim \mathrm{GHN}(\alpha, \theta)$, we have another generalized unit-half-normal distribution with PDF, CDF and QF written, respectively, as

$$f(y;\alpha,\theta) = \sqrt{\frac{2}{\pi}}\frac{\theta}{y[-\log(y)]}\left[-\frac{\log(y)}{\alpha}\right]^{\theta}\exp\left(-\frac{1}{2}\left[-\frac{\log(y)}{\alpha}\right]^{2\theta}\right), \tag{2.17}$$

$$F(y;\alpha,\theta) = 2\Phi\left\{-\left[-\frac{\log(y)}{\alpha}\right]^{\theta}\right\},$$

$$Q(\tau;\alpha,\theta) = \exp\left\{-\alpha\left[-\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^{\frac{1}{\theta}}\right\}, \tag{2.18}$$

where $0 < y < 1$. From the expression defined in (2.18), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = -\log(\mu)[\Phi^{-1}(\tau/2)]^{-1/\theta}$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y.

Plots of the PDFs stated in (2.15) and (2.17) for the re-parameterized generalized unit-half-normal distribution, with several values of μ, θ and τ, are given in Figs. 5 and 4, respectively. In the second column of Fig. 5, we see the behavior of the unit-half-normal distribution, whose PDF shapes are unimodal and asymmetrical (skewed to the left and to the right). These shapes make the unit-half-normal distribution flexible to model proportion data in many applied sciences. Note that these two extensions of the unit-half-normal have not been considered in the literature.

2.5. The Vasicek distribution

The Vasicek distribution was proposed in [123] and used, for the first time, in [80] to model the mean and quantiles conditional on covariates. A random variable Y with bounded support (0,1) is Vasicek distributed if its PDF, CDF, and QF are written, respectively, as

$$f(y;\alpha,\theta) = \sqrt{\frac{1-\theta}{\theta}}\exp\left\{\frac{1}{2}\left[\Phi^{-1}(y)^2 - \left(\frac{\Phi^{-1}(y)\sqrt{1-\theta}-\Phi^{-1}(\alpha)}{\sqrt{\theta}}\right)^2\right]\right\}, \tag{2.19}$$

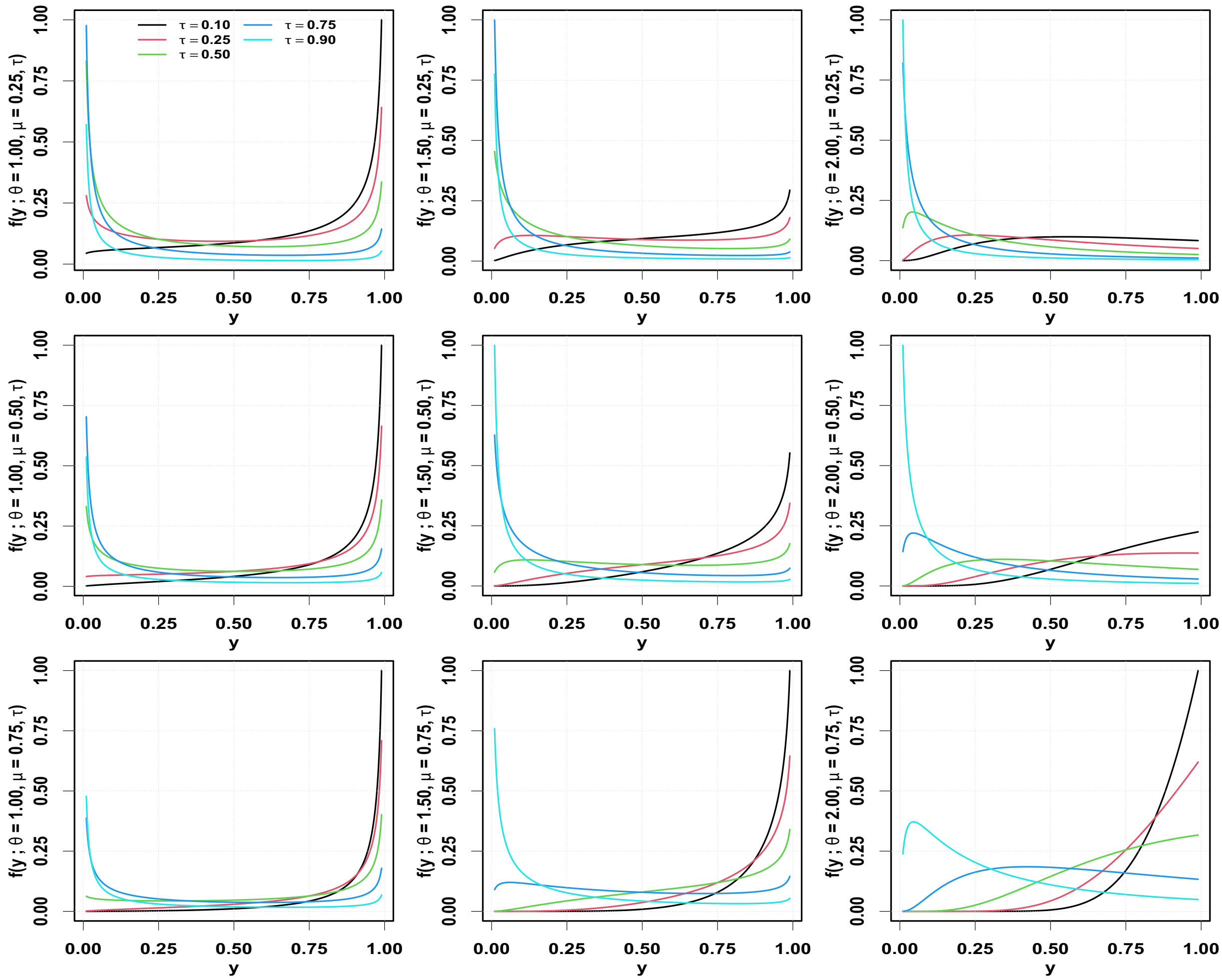

Fig. 6. Plots of the re-parameterized PDF stated from (2.19) for indicated values of μ, θ and τ.

$$F(y;\alpha,\theta) = \Phi\left(\frac{\Phi^{-1}(y)\sqrt{1-\theta} - \Phi^{-1}(\alpha)}{\sqrt{\theta}}\right),$$

$$Q(\tau;\alpha,\theta) = \Phi\left(\frac{\Phi^{-1}(\alpha) + \Phi^{-1}(\tau)\sqrt{\theta}}{\sqrt{1-\theta}}\right), \tag{2.20}$$

where $y > 0$ and $\alpha,\ \theta < 1$. Note that θ is a shape parameter and the mean and variance of Y are, respectively, given by

$$\mathrm{E}(Y) = \alpha, \quad \mathrm{Var}(Y) = \Phi_2\left(\Phi^{-1}(\alpha), \Phi^{-1}(\alpha), \theta\right) - \alpha^2,$$

where

$$\Phi_2(a,b,c) = \frac{1}{2\pi\sqrt{1-c^2}} \int_{-\infty}^{a}\int_{-\infty}^{b} \exp\left(-\frac{x^2 - 2cxy + y^2}{2(1-c^2)}\right) \mathrm{d}y\,\mathrm{d}x.$$

As stated in [123], the PDF defined in (2.19) is unimodal with mode at $\Phi[\Phi^{-1}(\alpha)(1-\theta)^{1/2}/(1-2\theta)]$, when $\theta < 0.5$; monotone when $\theta = 0.5$; and U-shaped when $\theta > 0.5$. Note that we can easily assess the effect of covariates on the mean of the distribution of Y through some appropriate link function for α. Moreover, from the expression defined in (2.20), we may re-parameterize α in terms of the $100\,\tau$th quantile, $\tau \in (0,1)$ namely, using the expression $\alpha = h^{-1}(\mu) = \Phi[\Phi^{-1}(\mu)(1-\theta)^{1/2} - \Phi^{-1}(\tau)\theta^{1/2}]$, where μ is, for a fixed value of τ, the 100τth quantile of the distribution of Y Fig. 6 shows some possible shapes of the re-parameterized PDF stated in (2.19) for selected values of μ, θ and τ.

To the best of our knowledge, the Vasicek distribution was used for the first time in [80] to estimate quantiles and means conditional on covariates. In that work, the authors presented applications to medical and political data. In the first application, the Vasicek quantile regression model outperformed the models based on the Johnson SB, Kumaraswamy, unit-logistic, and unit-Weibull distributions. In the second one, the Vasicek mean regression outperformed the fits obtained by beta [32] and simplex [119] regressions. Parameter estimation, model selection and diagnostics are available on the `vasicekreg` R package [75]. Notice that the literature on the Vasicek distribution is rather scarce and it is typically used to model economic data.

3. Parametric quantile regressions for non-Gaussian-related bounded responses

In this section, by using a similar presentation structure to Section 2, we introduce parametric quantile regression models for non-Gaussian-related bounded responses based on the arcsecant hyperbolic Weibull [64], Kumaraswamy [92], Lambert-uniform [49], L-logistic [99], log-extended exponential-geometric [52], transmuted unit-Rayleigh [63], unit-Bur-XII [61], unit-Chen [59], unit-Gompertz [79], unit-Gumbel [78], and unit-Weibull [85] distributions.

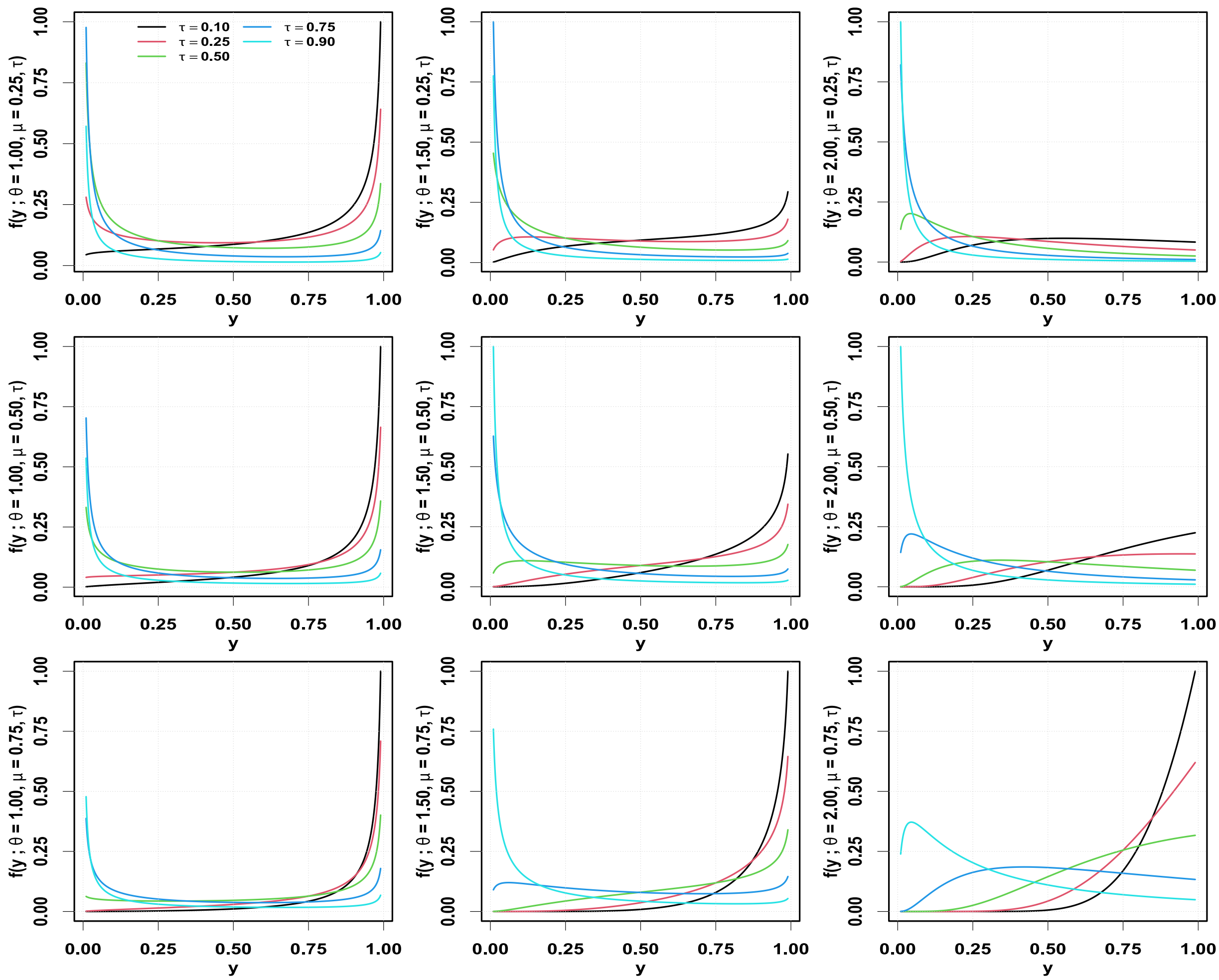

Fig. 7. Plots of the re-parameterized PDF stated from (3.21) for indicated values of μ, θ and τ.

3.1. The arcsecant hyperbolic Weibull distribution

The arcsecant hyperbolic Weibull model [64] is obtained from $Y = \text{sech}(X)$, where $X \sim \text{Weibull}(\alpha, \theta)$, which denotes a Weibull distributed random variable with CDF given by $F_X(x; \alpha, \theta) = \exp(-\alpha x^{\theta})$ [126]. The corresponding PDF, CDF and QF of Y are written, respectively, as

$$f(y; \alpha, \theta) = \frac{\alpha\theta}{y\sqrt{1-y^2}} \text{arcsech}(y)^{\theta-1} \exp\left[-\alpha\, \text{arcsech}(y)^{\theta}\right], \tag{3.21}$$

$$F(y; \alpha, \theta) = \exp\left[-\alpha\, \text{arcsech}(y)^{\theta}\right],$$
$$Q(\tau; \alpha, \theta) = \text{sech}\left(\left[-\alpha^{-1} \log(\tau)\right]^{\frac{1}{\theta}}\right), \tag{3.22}$$

where $0 < y < 1$ and $\text{arcsech}(y) = \log[(1 + (1 - y^2)^{1/2})/y]$. Note that $\alpha > 0$ is the rate parameter, while $\theta > 0$ is the shape parameter and does not has a direct interpretation in terms of the observed data.

For quantile regression, the parameter α presented in (3.22) can be re-parameterized as $\alpha = h^{-1}(\mu) = -\log(\tau)/\text{arcsech}(\mu)^{\theta}$, such that μ is, for a fixed and known value τ, the $100\,\tau$th quantile of the distribution of Y. Plots of the re-parameterized arcsecant hyperbolic Weibull distribution PDF for several values μ, θ and τ are given in Fig. 7.

3.2. The Kumaraswamy distribution

The Kumaraswamy model with support in the interval (a, b) was proposed in [67]. The particular case for the interval (0,1) can be obtained from the transformation $Y = \exp(-X)$, where $X \sim \text{EE}(\alpha, \theta)$, which denotes an exponentiated-exponential distributed random variable [42] with CDF given by $F_X(x; \alpha, \theta) = [1 - \exp(-\theta x)]^{\alpha}$. Considering the interval (0,1), the corresponding PDF, CDF and QF of Y are expressed, respectively, as

$$f(y; \alpha, \theta) = \alpha\theta y^{\theta-1}(1 - y^{\theta})^{\alpha-1}, \tag{3.23}$$

$$F(y; \alpha, \theta) = 1 - (1 - y^{\theta})^{\alpha},$$
$$Q(\tau; \alpha, \theta) = \left[1 - (1 - \tau)^{\frac{1}{\alpha}}\right]^{\frac{1}{\theta}}, \tag{3.24}$$

where $0 < y < 1$ and $\alpha, \theta > 0$ are shape parameters.

The mean and variance of Y are, respectively, given by

$$\text{E}(Y) = \alpha B\left(1 + \frac{1}{\theta}, \alpha\right),$$

$$\text{Var}(Y) = \alpha B\left(1 + \frac{2}{\theta}, \alpha\right) - \left[-\alpha B\left(1 + \frac{1}{\theta}, \alpha\right)\right]^2,$$

where, as mentioned, $B(a, b)$ is the beta function.

The available formula for $\text{E}(Y)$ makes a mean-based re-parameterization unfeasible and then α nor θ have a direct interpretation in terms of the observed data. For example, θ is no

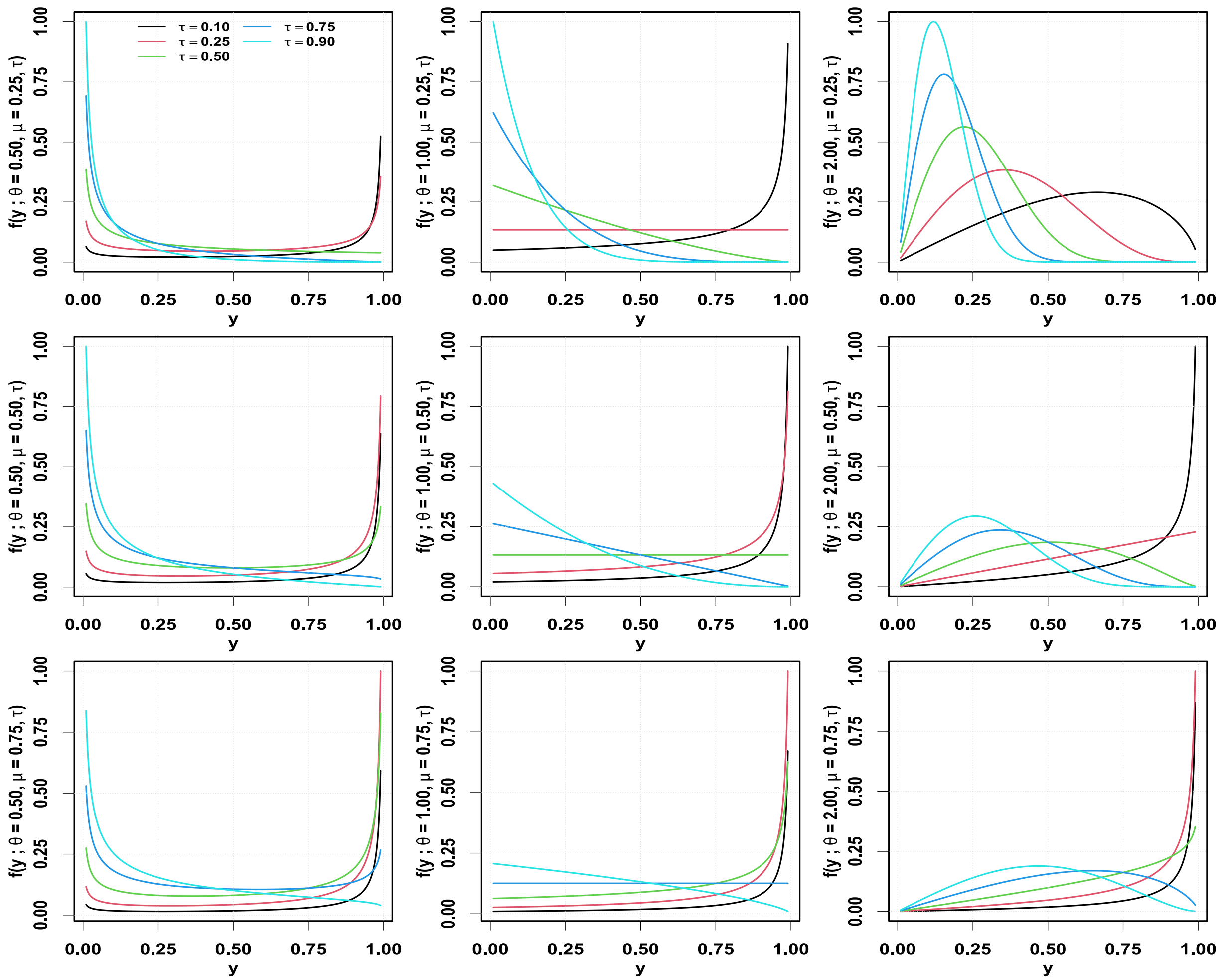

Fig. 8. Plots of the re-parameterized PDF stated from (3.23) for indicated values of μ, θ and τ.

longer a rate parameter as in the distribution of X. In contrast, α and θ can be re-parameterized according to a quantile. Note from [92] that a re-parameterization in α is more advantageous. Thus, from the expression defined in (3.24), the parameter α may be expressed as $\alpha = h^{-1}(\mu) = \log(1-\tau)/\log(1-\mu^{\theta})$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Fig. 8 shows some possible shapes of the re-parameterized Kumaraswamy PDF for selected values of μ, θ and τ. The figure illustrates well how flexible and versatile the Kumaraswamy distribution is.

In [92], a Kumaraswamy regression model was proposed considering only the median, but the model can be extended to other quantiles. The authors did not consider applications for the proposed model, but several applications may be found in the literature for comparative purposes. For example, in [77], this model was applied for two data sets related to political science and sports medicine. In [85], this model was applied to three data sets: the first one related to the stem cell recovery rate; the second one was on the access of families to piped water supply in Brazilian cities in the Southeast and Northeast regions; and the third one on the cost effectiveness of risk management. In [63], the educational level of countries of the Organization for Economic Co-operation and Development (OECD) is studied [28]. The transmuted Kumaraswamy distribution was proposed in [55]. A new quantile parametric mixed regression model for bounded response was presented in [10], whereas in [105] a Kumaraswamy regression model was introduced with an Aranda-Ordaz link function. A mode regression model for this distribution was analyzed in [91], whereas in [43] a Kumaraswamy regression to model bounded outcome scores was considered. An extension of the Kumaraswamy quantile regression model to couple extremes zero and one was presented in [9]. Considering that the continuous part follows a re-parameterized Kumaraswamy distribution, the proposed inflated model mixes the continuous and discrete parts. Therefore, their respective PDF and CDF are given by

$$g(y;\nu,\mu,\theta,\tau) = \begin{cases} \nu\,(1-c), & \text{if } y=0; \\ \nu\,c, & \text{if } y=1; \\ (1-\nu)\,f(y;\mu,\theta,\tau) & \text{if } y\in(0,1); \end{cases} \tag{3.25}$$

$$G(y;\nu,\mu,\theta,\tau) = \nu\,(1-c) + \nu\,c\,\mathbb{I}_{\{1\}}(y) + (1-\nu)\,F(y;\mu,\theta,\tau); \tag{3.26}$$

where $0 < \nu < 1$ is the mixture parameter, c is the probability of a Bernoulli distributed random variable, f is the re-parameterized Kumaraswamy PDF, and $\mathbb{I}_A$ is the indicator function that equals one if $y \in A$ and zero otherwise. The proposed model was used in [9] to analyze the impacts of several conditioning variables on the proportion of people that live in households with inadequate water supply and sewage in Brazil. Since nearly 17% of the Brazilian municipalities no one lives in households with inadequate water supply and sewage, the data display inflation at zero.

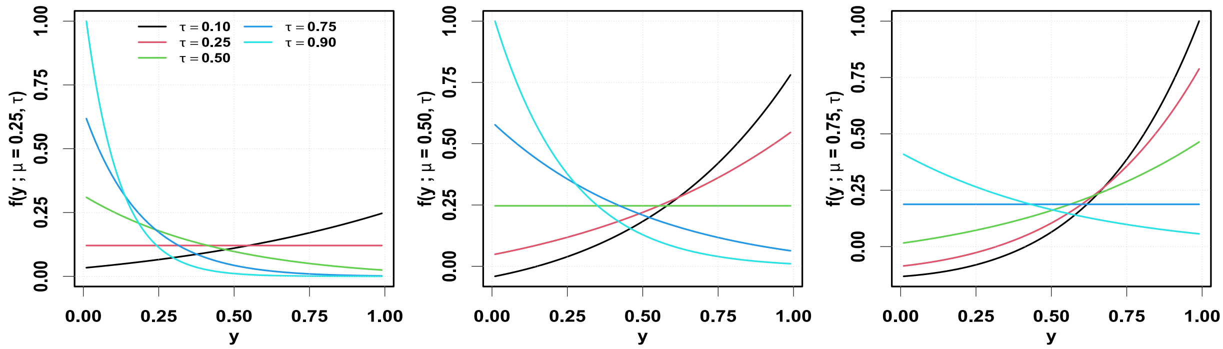

Fig. 9. Plots of the re-parameterized PDF stated from (3.27) for indicated values of μ and τ.

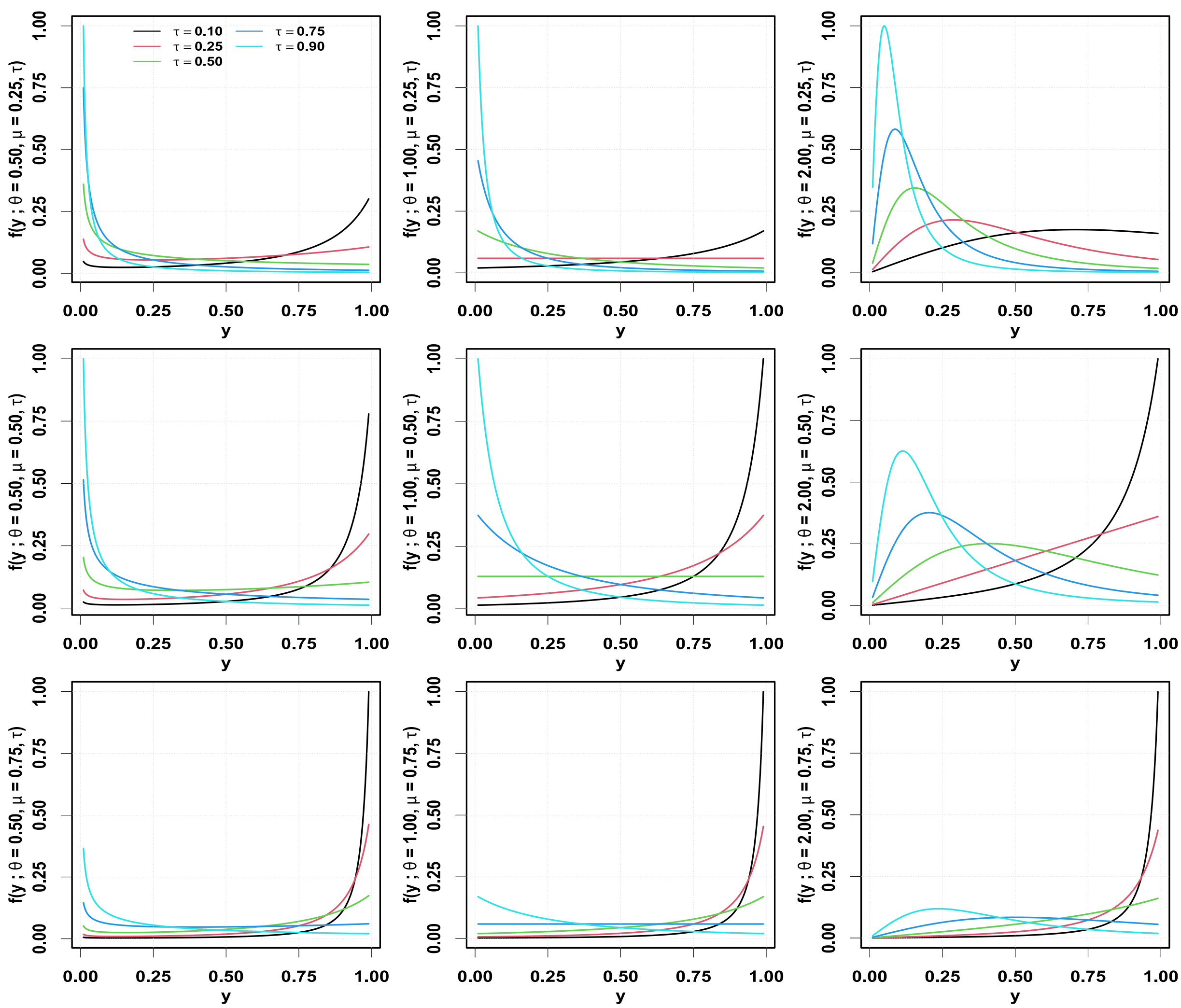

Fig. 10. Plots of the re-parameterized PDF stated from (3.30) for indicated values of μ, θ and τ.

3.3. The Lambert-uniform distribution

The Lambert-uniform model [49] has one parameter and is helpful to describel bounded data from a PDF with a monotonic (increasing or decreasing) behavior. This distribution arises directly from the Lambert-F generator [48] when considering a uniform baseline distribution. The corresponding PDF, CDF and QF of Y are established, respectively, as

$$f(y;\alpha) = \alpha^{y}[1-\log(\alpha)(1-y)], \tag{3.27}$$

$$F(y;\alpha) = 1-(1-y)\alpha^{y}$$

$$Q(\tau;\alpha) = \begin{cases} \dfrac{1}{\log(\alpha)} W_0\left[\dfrac{\log(\alpha)(\tau-1)}{\alpha}\right]+1, & \text{if}\alpha \in (0,1)\cup(1,e); \\ \tau, & \text{if } \alpha = 1; \end{cases} \tag{3.28}$$

where $0 < y < 1$, e $\simeq 2.718$ is the Euler number, $0 < \alpha < e$ is a shape parameter and W_0 is the principal branch of the Lambert-

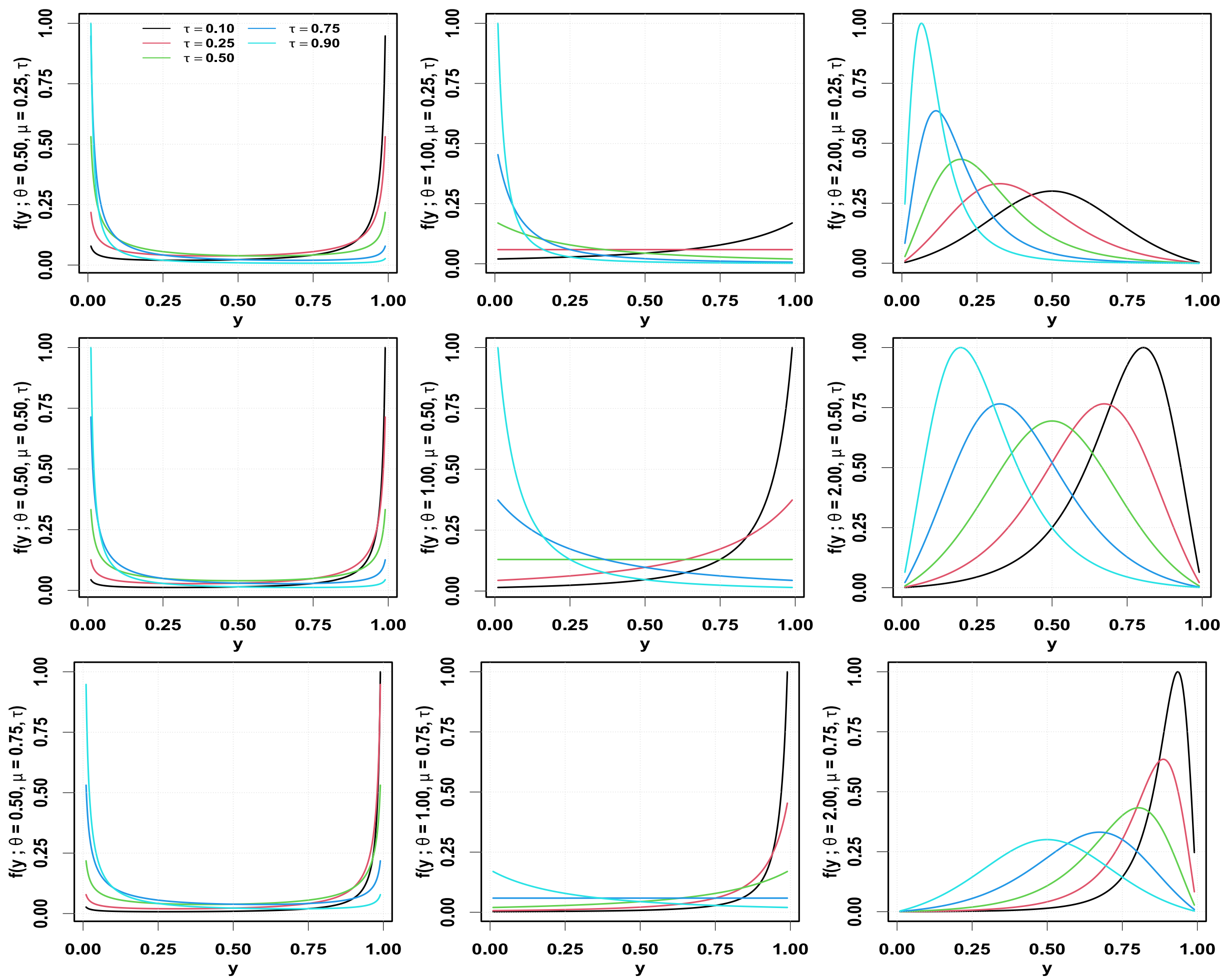

Fig. 11. Plots of the re-parameterized PDF stated from (3.32) for indicated values of μ, θ and τ.

W function [22]. The Lambert-uniform mean is given by

$$\mathrm{E}(Y) = \begin{cases} \dfrac{\alpha - 1 - \log(\alpha)}{\log(\alpha)^2}, & \text{if } \alpha \in (0, 1) \cup (1, \mathrm{e}); \\ \frac{1}{2}, & \text{if } \alpha = 1. \end{cases} \tag{3.29}$$

In the formula stated in (3.29), observe that the mean of the Lambert-uniform distribution has a closed form. However, despite this, the shape parameter α cannot be expressed explicitly as a function of the mean, which is a major drawback to formulate a regression model and quantify the effect of the covariates on the mean response. Also, α can be explicitly formulated as a function of the 100 τth quantile, which permits us to re-parameterize the Lambert-uniform distribution in terms of this quantile and, consequently, to establish a quantile regression in a simple way.

From the expression defined in (3.28), note that α may be re-parameterized as $\alpha = [(1-\tau)/(1-\mu)]^{1/\mu}$, such that μ is, for a fixed value of τ, the 100 τth quantile of the distribution of Y. Plots of the re-parameterized Lambert-uniform PDF for several values μ and τ are given in Fig. 9. A model considering the Lambert-uniform distribution was analyzed in [49] outperforming the models considering the Kumaraswamy and arcsecant-hyperbolic-normal distributions based on data on cost effectiveness of risk management.

3.4. The log-extended exponential-geometric distribution

The log-extended exponential-geometric model [52] is obtained from the transformation $Y = \exp(-X)$, where $X \sim \mathrm{EEG}(\alpha, \theta)$, which denotes an extended exponential-geometric distributed random variable [1]. The corresponding PDF, CDF and QF of Y are given, respectively, as

$$f(y; \alpha, \theta) = \frac{\theta(1+\alpha)y^{\theta-1}}{\left(1+\alpha y^{\theta}\right)^2}, \tag{3.30}$$

$$F(y; \alpha, \theta) = \frac{(1+\alpha)y^{\theta}}{1+\alpha y^{\theta}},$$
$$Q(\tau; \alpha, \theta) = \left(\frac{\tau}{1+\alpha-\alpha\tau}\right)^{\frac{1}{\theta}}, \tag{3.31}$$

where $0 < y < 1$, $\alpha > 0$ and $\theta > -1$. From the expression defined in (3.31), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = -[(1-\tau\mu^{-\theta})/(1-\tau)]$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Fig. 10 shows some possible shapes of the re-parameterized log-extended exponential-geometric PDF for selected values of μ, θ and τ. This model has as special cases the power function and uniform distributions. In [52], a log-extended exponential-geometric model was proposed and compared with the beta model in an application on the cost effectiveness of risk management. In this application, the best fit was obtained by the log-extended exponential-geometric model.

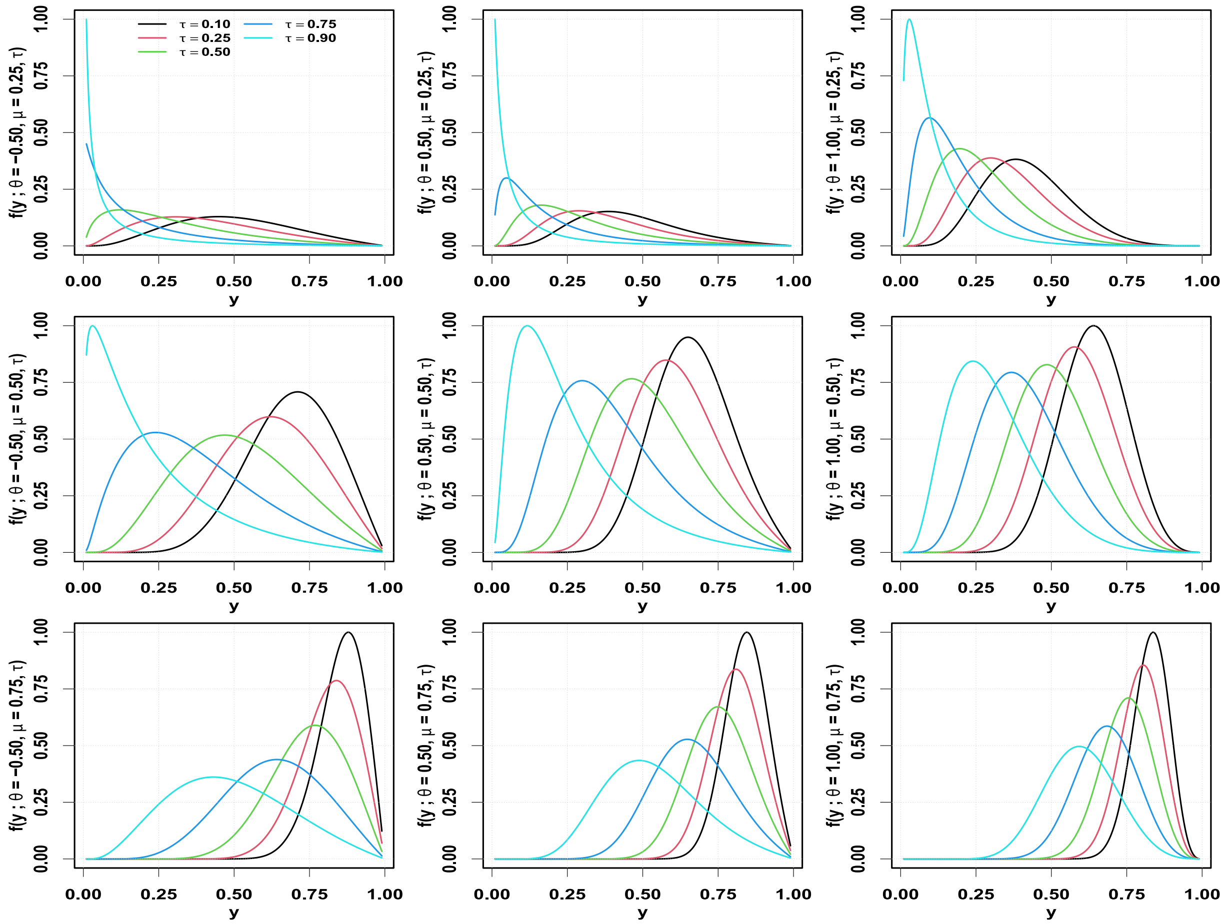

Fig. 12. Plots of the re-parameterized PDF stated from (3.34) for indicated values of μ, θ and τ.

3.5. The L-logistic distribution

The L-logistic model [121] is obtained from $Y=(1+\exp\{-[(X-\alpha)/\theta]\})^{-1}$, where $X\sim\text{Log}(0,1)$, which denotes a standard logistic distributed random variable [6]. The corresponding PDF, CDF and QF of Y are expressed, respectively, as

$$f(y;\alpha,\theta)=\frac{\theta\exp(\alpha)\left(\frac{y}{1-y}\right)^{\theta-1}}{\left[1+\exp(\alpha)\left(\frac{y}{1-y}\right)^{\theta}\right]^2}, \tag{3.32}$$

$$F(y;\alpha,\theta)=\frac{\exp(\alpha)\left(\frac{y}{1-y}\right)^{\theta}}{1+\exp(\alpha)\left(\frac{y}{1-y}\right)^{\theta}},$$

$$Q(\tau;\alpha,\theta)=\frac{\exp\left(-\frac{\alpha}{\theta}\right)\left(\frac{\tau}{1-\tau}\right)^{\frac{1}{\theta}}}{1+\exp\left(-\frac{\alpha}{\theta}\right)\left(\frac{\tau}{1-\tau}\right)^{\frac{1}{\theta}}}, \tag{3.33}$$

where $0<y<1$, $\alpha>0$ and $\theta>0$. To formulate a quantile regression, α defined in (3.33) must be re-parameterized as

$$\alpha=h^{-1}(\mu)=\log\left(\frac{\tau}{1-\tau}\right)-\theta\log\left(\frac{\mu}{1-\mu}\right),$$

such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Plots of the re-parameterized L-logistic PDF for several values μ, θ and τ are given in Fig. 11. A closed form was stated in [99] for the moments of the L-logistic distribution, which involves the multivariate Wright generalized hypergeometric function. In [99], a L-logistic quantile regression model was carried out on the relationship between vulnerability to poverty and anxiety. In this study, the beta regression model was also considered. The regression model considering the L-logistic distribution provided a better fit than the beta regression model for all the criteria stated [124].

3.6. The transmuted unit-Rayleigh distribution

The transmuted unit-Rayleigh model [63] is based on the unit-Rayleigh distribution proposed in [7] combined with the quadratic transmutation scheme used in [116,15]. The corresponding PDF, CDF and QF of Y are formulated, respectively, as

$$f(y;\alpha,\theta)=\frac{2\alpha\log(y)}{y}\exp\left[-\alpha\log(y)^2\right]\left\{-1-\theta+2\theta\exp\left[-\alpha\log(y)^2\right]\right\}, \tag{3.34}$$

$$F(y;\alpha,\theta)=\exp\left\{-\alpha[-\log(y)]^2\right\}\left(1+\theta-\theta\exp\left\{-\alpha[-\log(y)]^2\right\}\right)$$

$$Q(\tau;\alpha,\theta)=\exp\left\{-\alpha^{-\frac{1}{2}}\sqrt{-\log\left[(1+\theta-\sqrt{(1+\theta)^2-4\theta\tau})/(2\theta)\right]}\right\}, \tag{3.35}$$

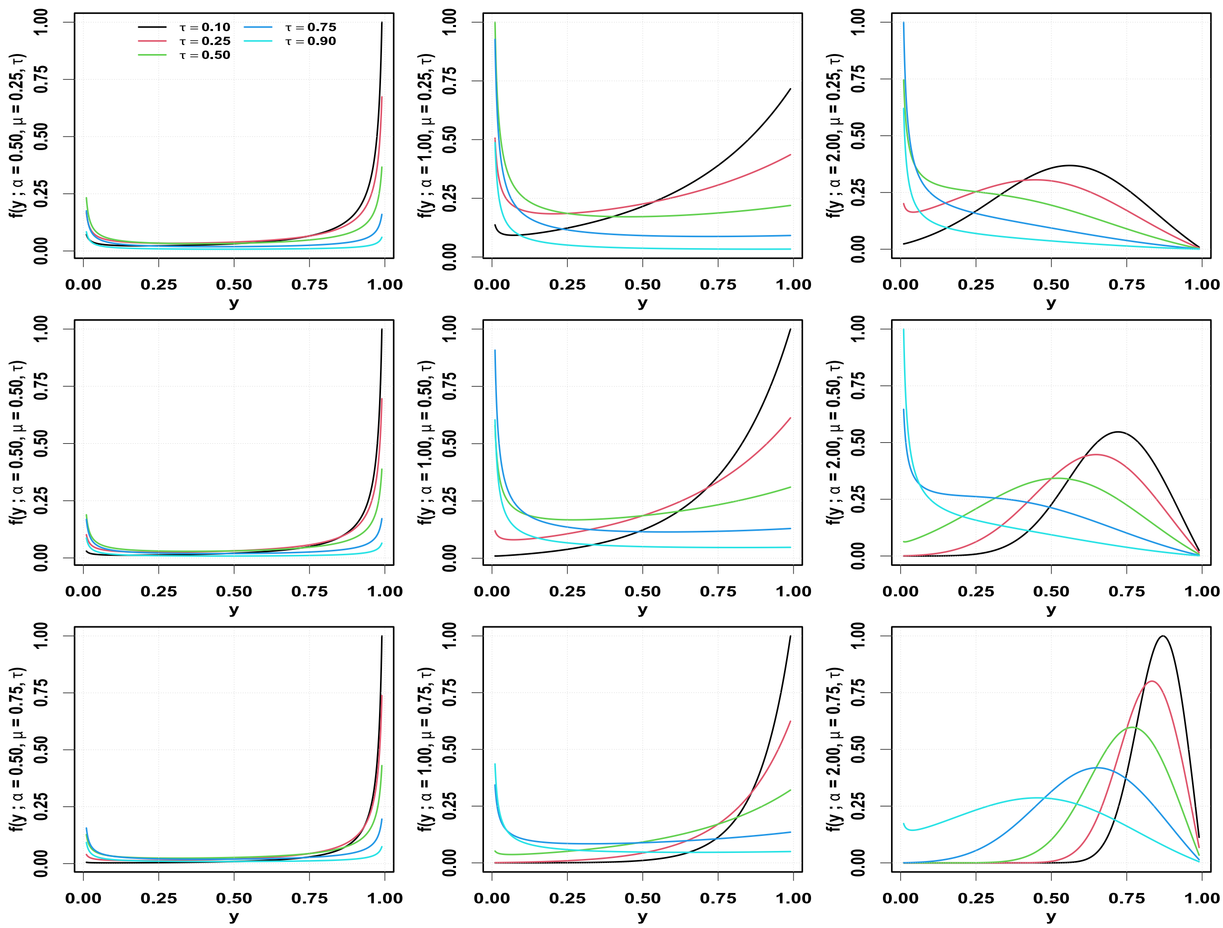

Fig. 13. Plots of the re-parameterized PDF stated from (3.36) for indicated values of μ, α and τ.

where $0 < y < 1$, $\alpha > 0$ and $\theta \in [-1, 1]$ is the shape parameter. From the expression defined in (3.35), the parameter α can be re-parameterized as $\alpha = -\log[1+\theta-((1+\theta)^2 - 4\theta\tau)^{1/2}/(2\theta)]/\log(\mu)^2$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Plots of the re-parameterized transmuted unit-Rayleigh PDF for several values μ, θ and τ are given in Fig. 12. Note that these shapes may be decreasing as well as unimodal with various skewed forms.

An application was considered in [64] to measure the educational level of OECD countries related to the covariates as life satisfaction, homicide rate, and voter turnout. The application indicated that the transmuted unit-Rayleigh quantile regression model provided a better fit than the beta and Kumaraswamy regression models[27].

3.7. The unit-Burr-XII distribution

The unit-Burr-XII model [61] is obtained from the transformation $Y = \exp(-X)$, where $X \sim \text{Burr-XII}(\alpha, \theta)$, which denotes a Burr-XII distributed random variable [16] with CDF given by $F_X(x; \alpha, \theta) = 1 - (1+x^\alpha)^{-\theta}$. The corresponding PDF, CDF and QF of Y are written, respectively, as

$$f(y; \alpha, \theta) = \frac{\alpha\theta}{y}[-\log(y)]^{\alpha-1}\left\{1+[-\log(y)]^\alpha\right\}^{(-\theta-1)}, \qquad (3.36)$$

$$F(y; \alpha, \theta) = \left\{1+[-\log(y)]^\alpha\right\}^{-\theta},$$
$$Q(\tau; \alpha, \theta) = \exp\left[-\left(\tau^{-\frac{1}{\theta}}-1\right)^{\frac{1}{\alpha}}\right], \qquad (3.37)$$

where $0 < y < 1$ and α, $\theta > 0$ are shape parameters. Note that, when $y \to 0$, $f(y; \alpha, \theta) \to +\infty$ for all the values of $\alpha > 0$ and $\theta > 0$. When $y \to 1$, if $\alpha > 1$, $f(y; \alpha, \theta) \to +\infty$; if $\alpha = 1$, $f(y; \alpha, \theta) \to \theta$; and if $\alpha < 1$, $f(y; \alpha, \theta) \to 0$.

As with other distributions mentioned, in this case, α nor θ have a direct interpretation in terms of the observed data. However, it is possible to re-parameterize both parameters as a function of the 100 τth quantile. In [61], the parameter α is expressed as

$$\alpha = h^{-1}(\mu) = \frac{\log\left(\tau^{-\frac{1}{\theta}}-1\right)}{\log[-\log(\mu)]}.$$

As shown in [61], the conditions $\tau > 2^\theta$ and $\mu > \exp(-1)$, either $\tau < 2^{-\theta}$ and $\mu < \exp(-1)$, must be satisfied. Therefore, for some values of τ, given these parameters combinations, it was not possible to display the shapes of the PDF. We can get more flexible shapes for the PDF stated in (3.36) re-parameterizing θ as

$$\theta = h^{-1}(\mu) = \frac{\log\left(\tau^{-1}\right)}{\log\left[1+\log\left(\frac{1}{\mu}\right)^\alpha\right]}. \qquad (3.38)$$

Fig. 13 shows some possible shapes of this re-parameterized PDF for selected values of μ, α and τ. In [108], it was considered the transformation $Z = 1 - Y$ and the re-parameterization defined in (3.38), where Y follows the unit-Burr-XII distribution with PDF stated in (3.36).

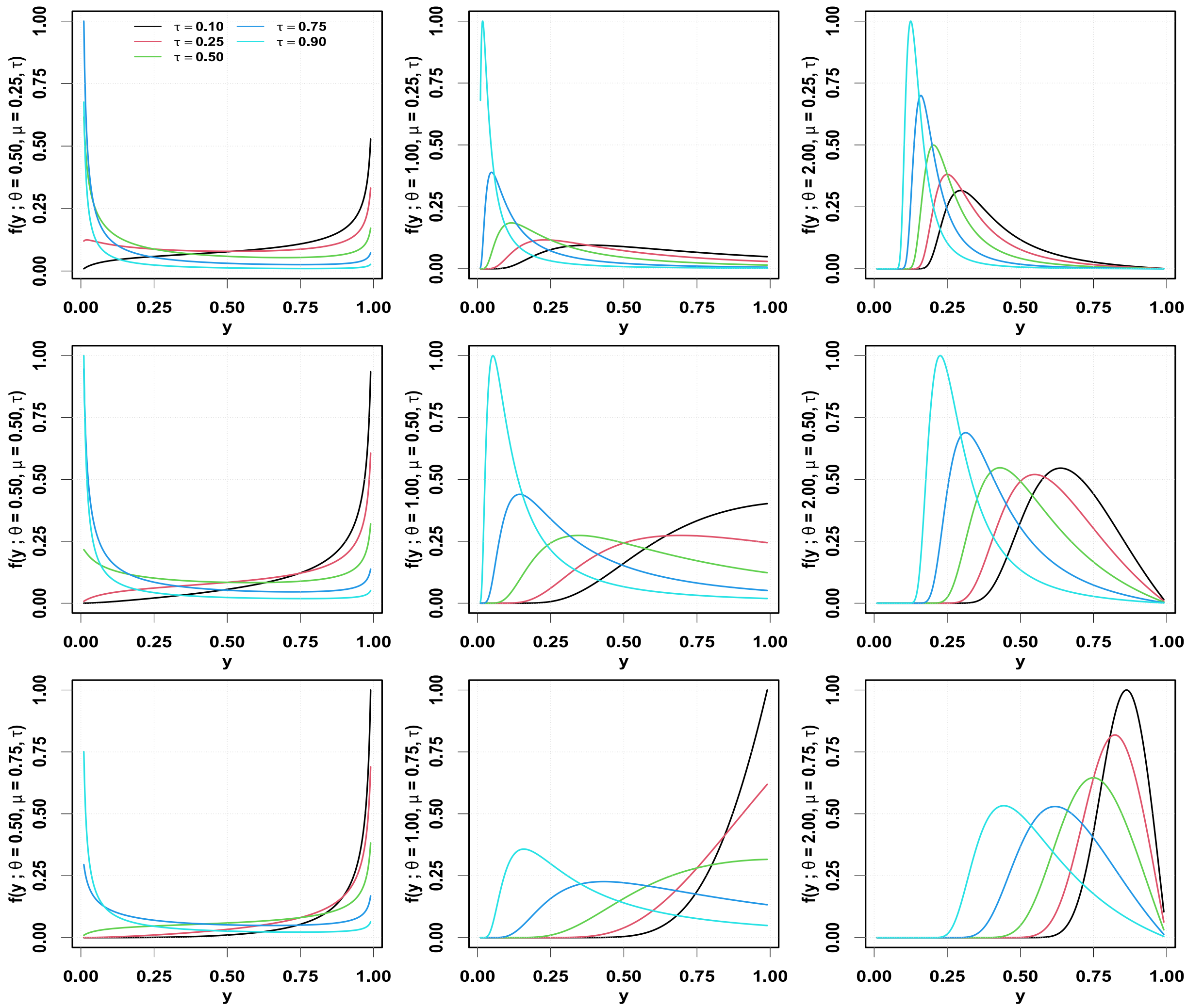

Fig. 14. Plots of the re-parameterized PDF stated from (3.39) for indicated values of μ, θ and τ.

3.8. The unit-Chen distribution

The unit-Chen model [59] is obtained from the transformation $Y = \exp(-X)$, where $X \sim \text{Chen}(\alpha, \theta)$, which denotes a Chen distributed random variable [19] with CDF given by $F_X(x; \alpha, \theta) = 1 - \exp\{\alpha[1 - \exp(x^{\theta})]\}$. The PDF, CDF and QF of Y are written, respectively, as

$$f(y; \alpha, \theta) = \frac{\alpha}{y}\theta[-\log(y)]^{\theta-1} \exp\left\{[-\log(y)]^{\theta}\right\} \exp\left[\alpha\left(1 - \exp\left\{[-\log(y)]^{\theta}\right\}\right)\right], \quad (3.39)$$

$$F(y; \alpha, \theta) = \exp[\alpha(1 - \exp\{[-\log(y)]^{\theta}\})],$$
$$Q(\tau; \alpha, \theta) = \exp\left(-\left\{\log\left[1 - \frac{\log(\tau)}{\alpha}\right]\right\}^{\frac{1}{\theta}}\right), \quad (3.40)$$

where $0 < y < 1$, and $\alpha, \theta > 0$ are shape parameters. From the expression defined in (3.40), the parameter α can be re-parameterized as

$$\alpha = h^{-1}(\mu) = \frac{\log(\tau)}{1 - \exp\left\{[-\log(\mu)]^{\theta}\right\}},$$

such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Plots of the re-parameterized unit-Chen PDF for several values μ, θ and τ are given in Fig. 14. This figure shows that the unit-Chen distribution has left and right skewed shapes as well as bathtub shape. In [59], the importance of the unit-Chen model is shown through an application with real-world data on the rate of stem cell recovery and compared with a Kumaraswamy model considering only the median. For this application, the unit-Chen model showed better performance than the Kumaraswamy model.

3.9. The unit-Gompertz distribution

The unit-Gompertz distribution [83] is obtained from the transformation $Y = \exp(-X)$, where $X \sim \text{GO}(\alpha, \theta)$, which denotes a Gompertz distributed random variable with CDF established by $F_X(x; \alpha, \theta) = 1 - \exp\{\alpha[1 - \exp(\theta x)]\}$. The corresponding PDF, CDF and QF of Y are stated as

$$f(y; \alpha, \theta) = \alpha\theta y^{-(1+\theta)} \exp\left[\alpha\left(1 - y^{-\theta}\right)\right], \quad (3.41)$$

$$F(y; \alpha, \theta) = \exp\left[\alpha\left(1 - y^{-\theta}\right)\right],$$
$$Q(\tau; \alpha, \theta) = \left[1 - \frac{\log(\tau)}{\alpha}\right]^{-\frac{1}{\theta}}, \quad (3.42)$$

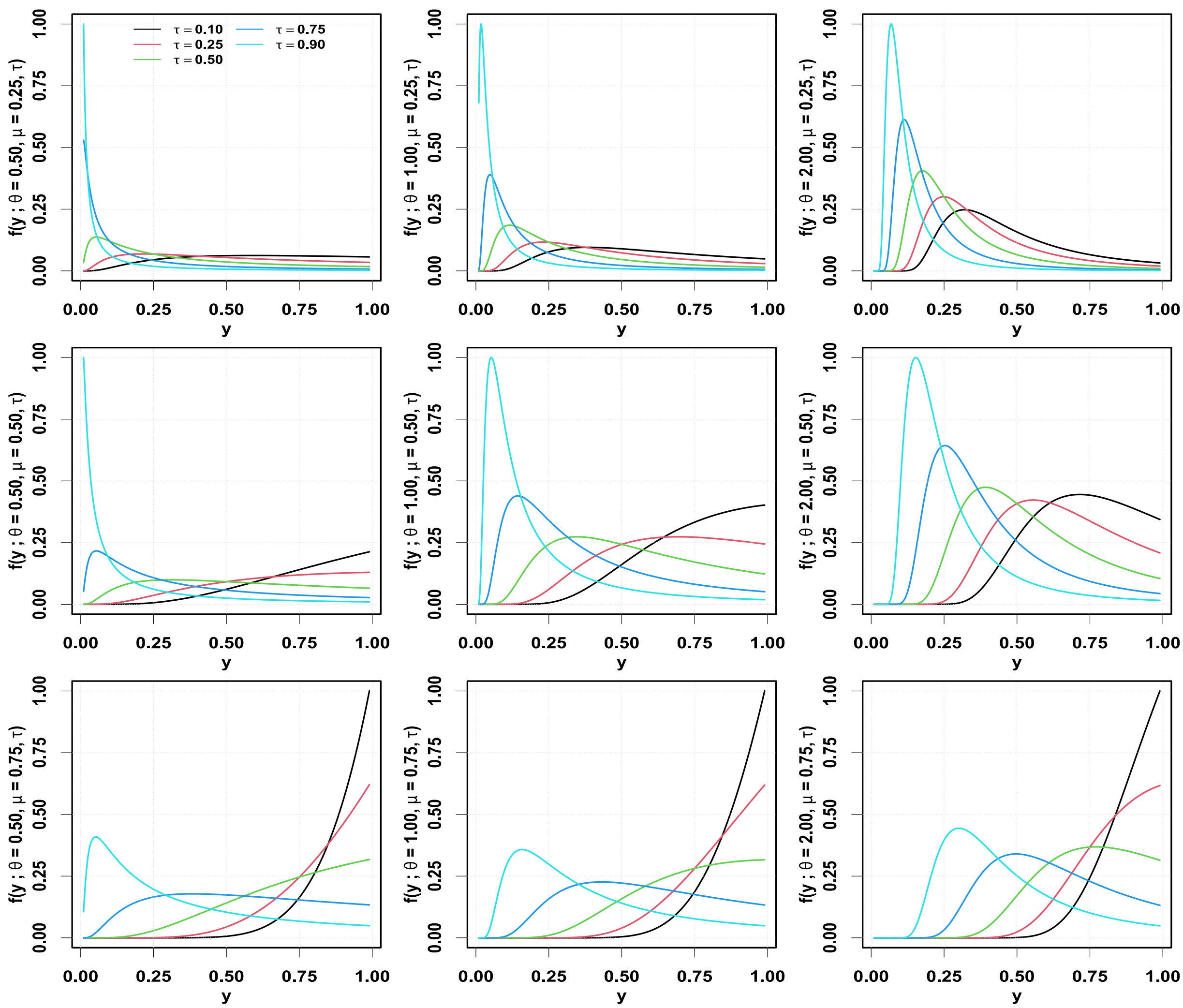

Fig. 15. Plots of the re-parameterized PDF stated from (3.41) for indicated values of μ, θ and τ.

where $0 < y < 1$, $\alpha > 0$ and $\theta > 0$. The mean of a unit-Gompertz distributed random variable Y is given as

$$\mathrm{E}(Y) = \alpha^{\frac{1}{\theta}} \exp(\alpha)\Gamma\left(\frac{\theta-1}{\theta}, \alpha\right),$$

where $\Gamma(a,b) = \int_b^\infty t^{a-1}\exp(-t)dt$ is the upper incomplete gamma function. To find the variance, consider $\mathrm{Var}(Y) = \int_0^1 [\mathrm{Q}(\tau;\alpha,\theta) - \mathrm{E}(Y)]^2 d\tau$, which depends on a term formulated as

$$\int_0^1 [\mathrm{Q}(\tau;\alpha,\theta)]^2 d\tau = 2\alpha^{\frac{3}{\theta}} \exp(2\alpha)\Gamma\left(\frac{\theta-1}{\theta}, \alpha\right)\Gamma\left(\frac{\theta-2}{\theta}, \alpha\right).$$

Note that α nor θ have a direct interpretation in terms of the observed data. For example, θ is no longer a rate parameter as in the distribution of X. However, from the expression defined in (3.42), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = \log(\tau)/(1-\mu^{-\theta})$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Similarly, we may re-parameterize θ as $\theta = h^{-1}(\mu) = -[1/\log(\mu)]\log[1-\log(\tau)/\alpha]$. We can easily verify that, by re-parameterizing α, we have a large number of shapes for the PDF [79]. Fig. 15 shows some possible shapes of the re-parameterized unit-Gompertz PDF for selected values of μ, θ and τ.

The unit-Gompertz quantile regression was used in [79] to analyze data of plants where ammonia is oxidized to nitric acid. Its fit was compared with the Kumaraswamy, Johnson SB, unit-Birnbaum-Saunders, unit-logistic, and unit-Weibull distributions. Parameter estimation, model selection, and diagnostics of these models are available on the `ugomquantreg` R package [74]. Besides the unit-Gompertz quantile regression, other works appeared in the literature considering the unit-Gompertz model. In [50,51], this model was employed for estimating the reliability of a multi-component stress-strength system. By using lower record values and inter-record times, inference procedures for estimating the parameters and predicting future record values were presented in [4,66,83] including some interesting properties.

A characterization of the unit-Gompertz distribution using truncated moments was introduced in [4], while a collection of parametric modal regression models was presented in [91], including the unit-Gompertz distribution. Furthermore, a unit-Gompertz distribution different from the one proposed in [83] was stated in [39].

3.10. The unit-Gumbel distribution

The unit-Gumbel distribution [78] is obtained from $Y = \exp[(X-\alpha)/\theta]/\{1+\exp[(X-\alpha)/\theta]\}$, where $X \sim \mathrm{SG}(0,1)$, which denotes a standard Type-I Gumbel distributed random variable with CDF given by $F_X = \exp[-x-\exp(-x)]$ [40]. The correspond-

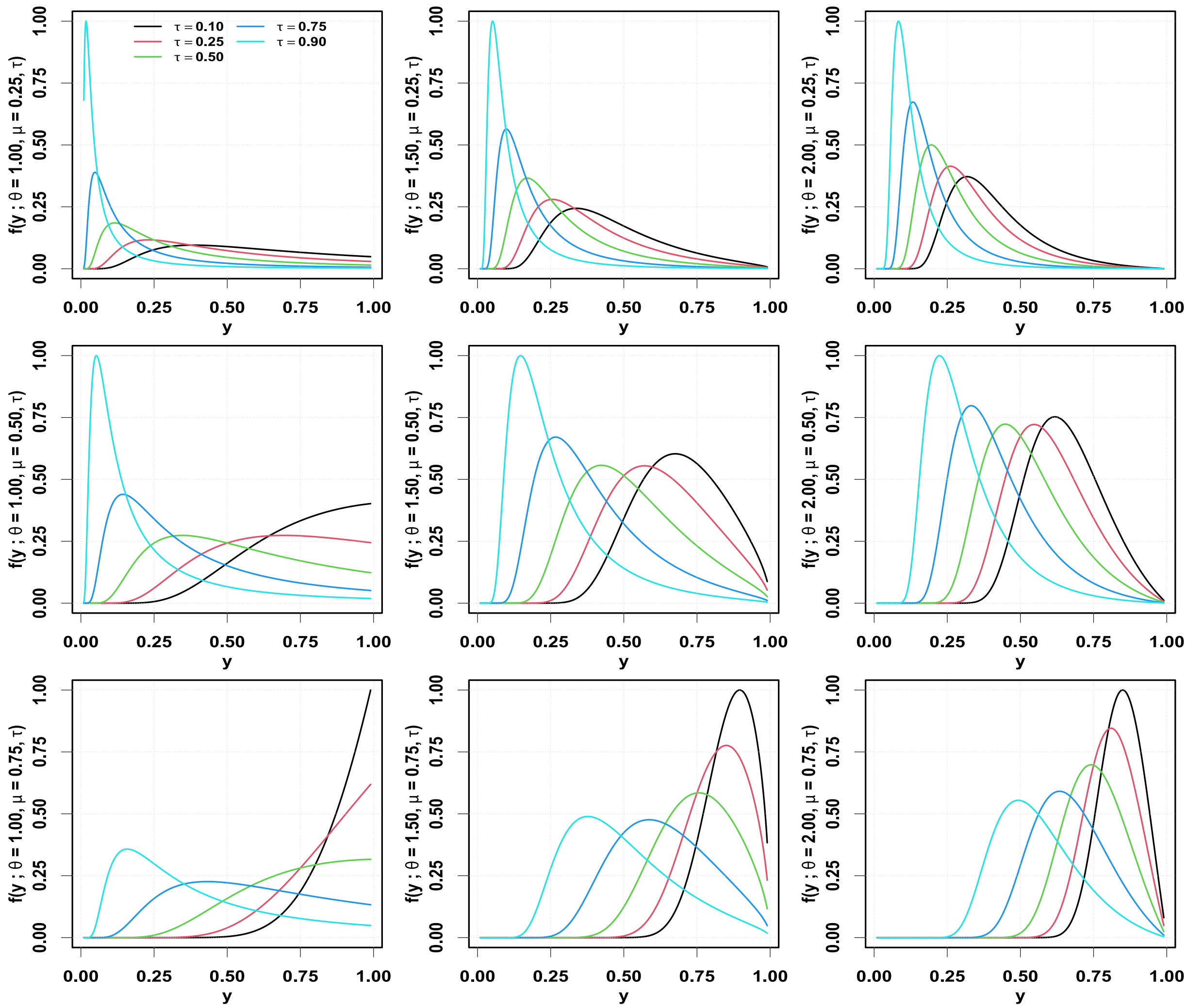

Fig. 16. Plots of the re-parameterized PDF stated from (3.43) for indicated values of μ, θ and τ.

ing PDF, CDF and QF of Y are written, respectively, as

$$f(y;\alpha,\theta)=\frac{\theta}{y(1-y)}\exp\left\{-\alpha-\theta\log\left(\frac{y}{1-y}\right)-\exp\left[-\alpha-\theta\log\left(\frac{y}{1-y}\right)\right]\right\}, \tag{3.43}$$

$$F(y;\alpha,\theta)=\exp\left[-\exp(-\alpha)\left(\frac{1-y}{y}\right)^{\theta}\right],$$

$$Q(\tau;\alpha,\theta)=\frac{\left[-\frac{1}{\log(\tau)}\right]^{\frac{1}{\theta}}}{\exp\left(\frac{\alpha}{\theta}\right)+\left[-\frac{1}{\log(\tau)}\right]^{\frac{1}{\theta}}}, \tag{3.44}$$

where $0 < y < 1$, while $\theta > 0$ and $\alpha \in \mathbb{R}$ are shape parameters. From the expression defined in (3.44), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = \theta\log[(1-\mu)/\mu] + \log[-1/\log(\tau)]$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. Fig. 16 shows some possible shapes of the re-parameterized unit-Gumbel PDF for selected values of μ, θ and τ.

3.11. The unit-Weibull distribution

The unit-Weibull distribution [84,85] is obtained from the transformation $Y=\exp(-X)$, where $X\sim\text{Weibull}(\alpha,\theta)$, which denotes a Weibull distributed random variable with CDF given by $F_X(x;\alpha,\theta)=\exp(-\alpha x^{\theta})$ [126]. The corresponding PDF, CDF and QF of Y are written, respectively, as

$$f(y;\alpha,\theta)=\frac{1}{y}\alpha\theta[-\log(y)]^{\theta-1}\exp\left\{-\alpha[-\log(y)]^{\theta}\right\}, \tag{3.45}$$

$$F(y;\alpha,\theta)=\exp\left\{-\alpha[-\log(y)]^{\theta}\right\},$$

$$Q(\tau;\alpha,\theta)=\exp\left\{-\left[-\frac{\log(\tau)}{\alpha}\right]^{\frac{1}{\theta}}\right\}, \tag{3.46}$$

where $0 < y < 1$, while $\alpha > 0$ and $\theta > 0$ are shape parameters.

Special cases of the unit-Weibull distributions include the uniform distribution over the interval (0,1) when $\alpha=\theta=1$, the power function distribution when $\theta=1$, and the unit-Rayleigh distribution when $\theta=2$.

Note that α nor θ have a direct interpretation in terms of the observed data. For example, α is no longer a scale parameter as in the distribution of X. However, from the expression

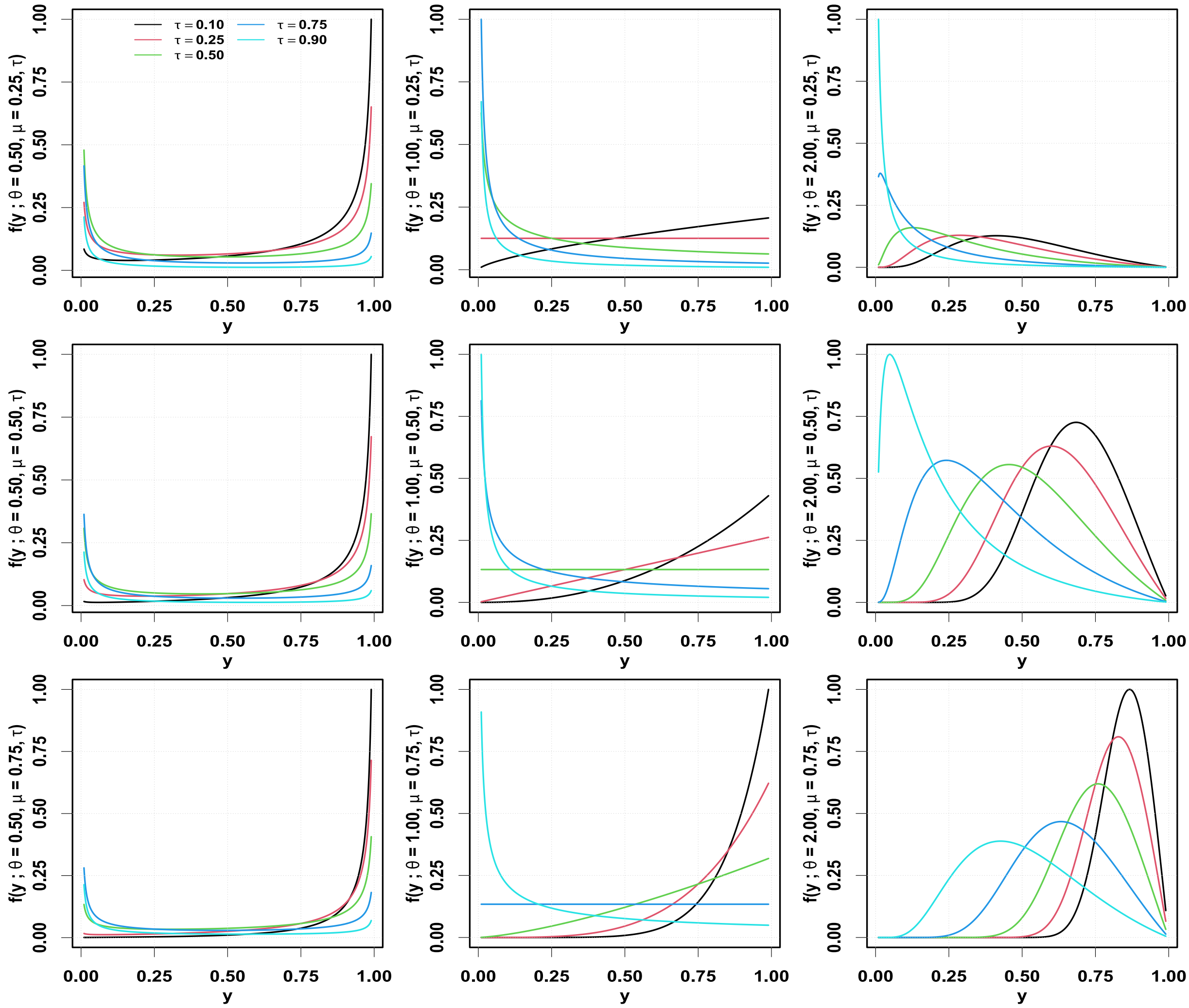

Fig. 17. Plots of the re-parameterized PDF stated from (3.45) for indicated values of μ, θ and τ.

defined in (3.46), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = -\log(\tau)/[-\log(\mu)]^{\theta}$, such that μ is, for a fixed and known value τ, the 100τth quantile of the distribution of Y. Plots of the re-parameterized unit-Weibull PDF for several values μ, θ and τ are given in Fig. 17. The PDF can assume different shapes (decreasing, increasing, unimodal, anti-unimodal) according to the values of its parameters.

In [85], three real-world data sets were analyzed for illustrative and model comparison purposes. The first application was related to the rate of recovery of stem cells; the second one was on the access of families to the supply of piped water in Brazilian cities in the Southeast and Northeast regions; and the third one was based on cost effectiveness of risk management. For these data sets, the unit-Weibull quantile regression model outperformed the Kumaraswamy and beta models according to three information criteria [124].

A Hausdorff approximation of the Heaviside step function was studied in [47] by a family of the unit-Weibull cumulative sigmoids. In [39], two new families were proposed: the unit extended Weibull and complementary unit extended Weibull distributions. In [90], three approaches were presented for bias reduction of the ML estimators of the unit-Weibull distribution parameters. The first approach is the analytical methodology suggested in [24]; the second one is based on parametric bootstrap resampling method; and the third one is the preventive approach introduced in [33]. Motivated by the presence of zeros or ones in proportion responses, an extension of the unit-Weibull quantile regression for the interval [0,1) or (0,1] was proposed in [89,91]. They assumed that the continuous mechanism is described by a re-parameterized unit-Weibull distribution, while the discrete component is a degenerate distribution in a known value c either zero or one. Under this approach, the PDF and CDF of the inflated unit-Weibull distribution in c is given by

$$m(y;\nu,\mu,\theta,\tau) = \begin{cases} \nu, & \text{if } y = c; \\ (1-\nu)f(y;\mu,\theta,\tau), & \text{if } y \in (0,1); \end{cases} \tag{3.47}$$

$$M(y;\nu,\mu,\theta,\tau) = \nu\mathbb{I}_c(y) + (1-\nu)F(y;\mu,\theta,\tau); \tag{3.48}$$

where $\mathbb{I}_A(y)$ is the indicator function above mentioned, whereas $\nu \in (0,1)$ is a mixture parameter, and $f(y;\mu,\theta,\tau), F(y;\mu,\theta,\tau)$ are the PDF and CDF of the re-parameterized unit-Weibull distribution. Notice that the random variable Y follows a unit-Weibull distribution with probability $1-\nu$ and it follows a degenerate distribution in c with probability ν.

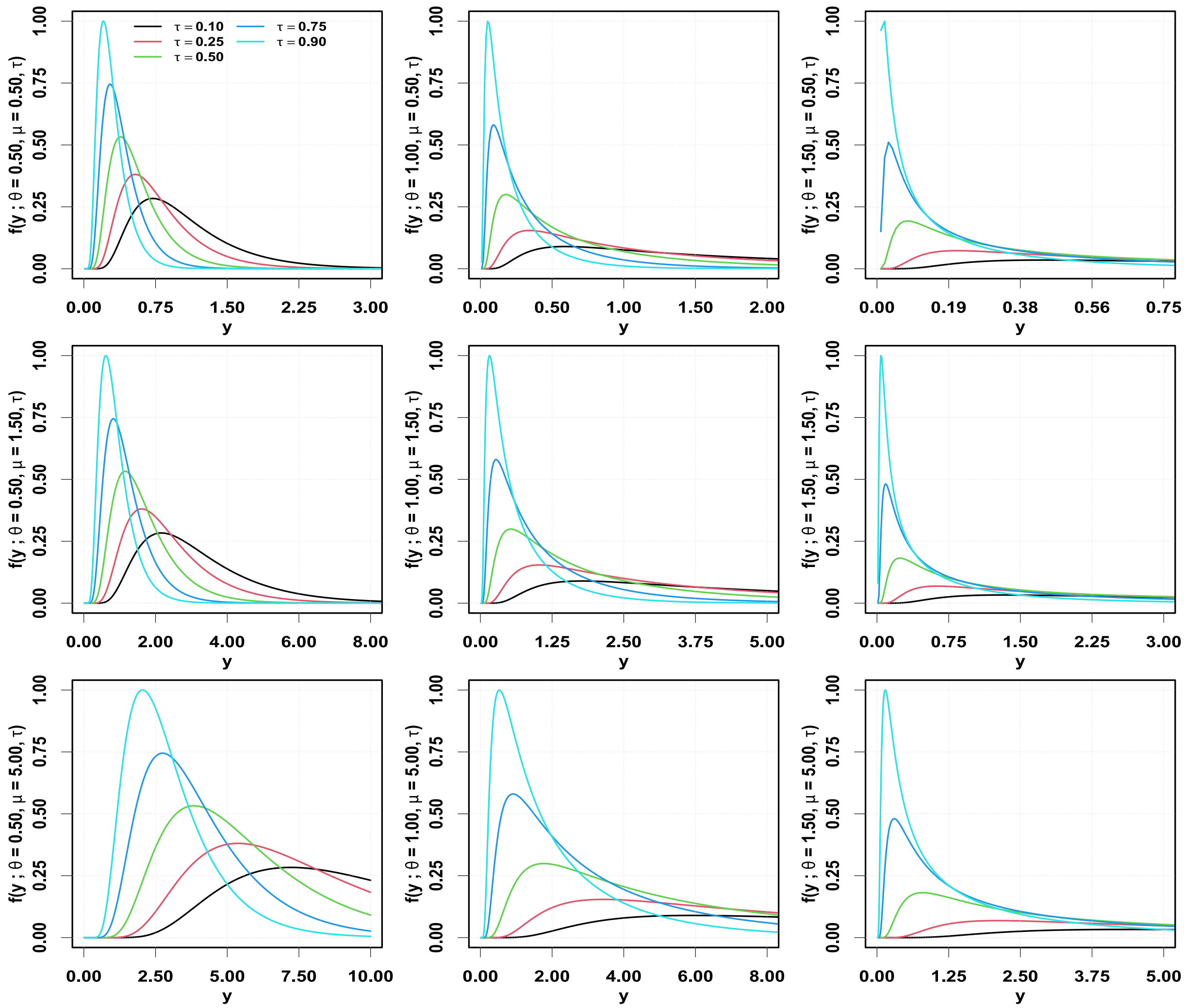

Fig. 18. Plots of the re-parameterized PDF stated from (4.49) for indicated values of μ, θ and τ.

4. Parametric quantile regressions for continuous positive and discrete responses

In this section, we discuss and present quantile regressions for continuous positive and discrete responses. For these distributions, the literature on parametric quantile regression formulated by the re-parameterization approach is scarce. We cite the models based on the continuous Birnbaum-Saunders [3,15,73,110,111], flexible Weibull [103], logistic Nadarajah-Haghighi [100], and log-symmetric [72,114] distributions, as well as the discrete generalized half-normal distribution [34,38].

The main characteristics of these distributions are presented from the next subsection. However, although not discussed in this paper, we can also mention the quantile regression model based on the gamma-sinh-Cauchy distribution [37], generalized gamma distribution [97,98] and asymmetric Laplace distribution [129]. Note that we do not present the characteristics of these distributions since their respective regression models are not formulated from re-parameterizations, as it happens with all the others distributions discussed in this paper.

4.1. The Birnbaum-Saunders distribution

Let Y be a Birnbaum-Saunders distributed random variable. Then, the corresponding PDF, CDF and QF of Y are given, respectively, by

$$f(y;\alpha,\theta) = \frac{1}{2\theta\alpha\sqrt{2\pi}}\left[\left(\frac{\alpha}{y}\right)^{\frac{1}{2}} + \left(\frac{\alpha}{y}\right)^{\frac{3}{2}}\right] \exp\left[-\frac{1}{2\theta^2}\left(\frac{y}{\alpha}+\frac{\alpha}{y}-2\right)\right], \tag{4.49}$$

$$F(y;\alpha,\theta) = \Phi\left\{\frac{1}{\theta}\left[\left(\frac{y}{\alpha}\right)^{\frac{1}{2}} - \left(\frac{\alpha}{y}\right)^{\frac{1}{2}}\right]\right\},$$

$$Q(\tau;\alpha,\theta) = \frac{\alpha}{4}\left[\theta\Phi^{-1}(\tau) + \sqrt{\theta^2\Phi^{-1}(\tau)^2+4}\right]^2, \tag{4.50}$$

where $y > 0$, $\alpha > 0$ is a scale parameter, and $\theta > 0$ is a shape parameter. The mean and variance of Y are stated, respectively, as $\mathrm{E}(Y) = \alpha(1+\theta^2/2)$ and $\mathrm{Var}(Y) = (\alpha\theta)^2(1+5/4\theta^2)$.

Note that we can easily assess the effect of covariates on the mean of the distribution of Y through some appropriate link function [68,71]. However, in many situations, modeling the effect of covariates on quantiles of the response can be also of interest. From the expression defined in (4.50), we may re-parameterize α in terms of the 100 τth quantile, $\tau \in (0,1)$ namely, as $\alpha = h^{-1}(\mu) = 4\mu/[\theta\Phi^{-1}(\tau) + (\theta^2\Phi^{-1}(\tau)^2 + 4)^{1/2}]^2$. This re-parameterization was considered in [110,111]. A geostatistical model based on a new approach to quantile regres-

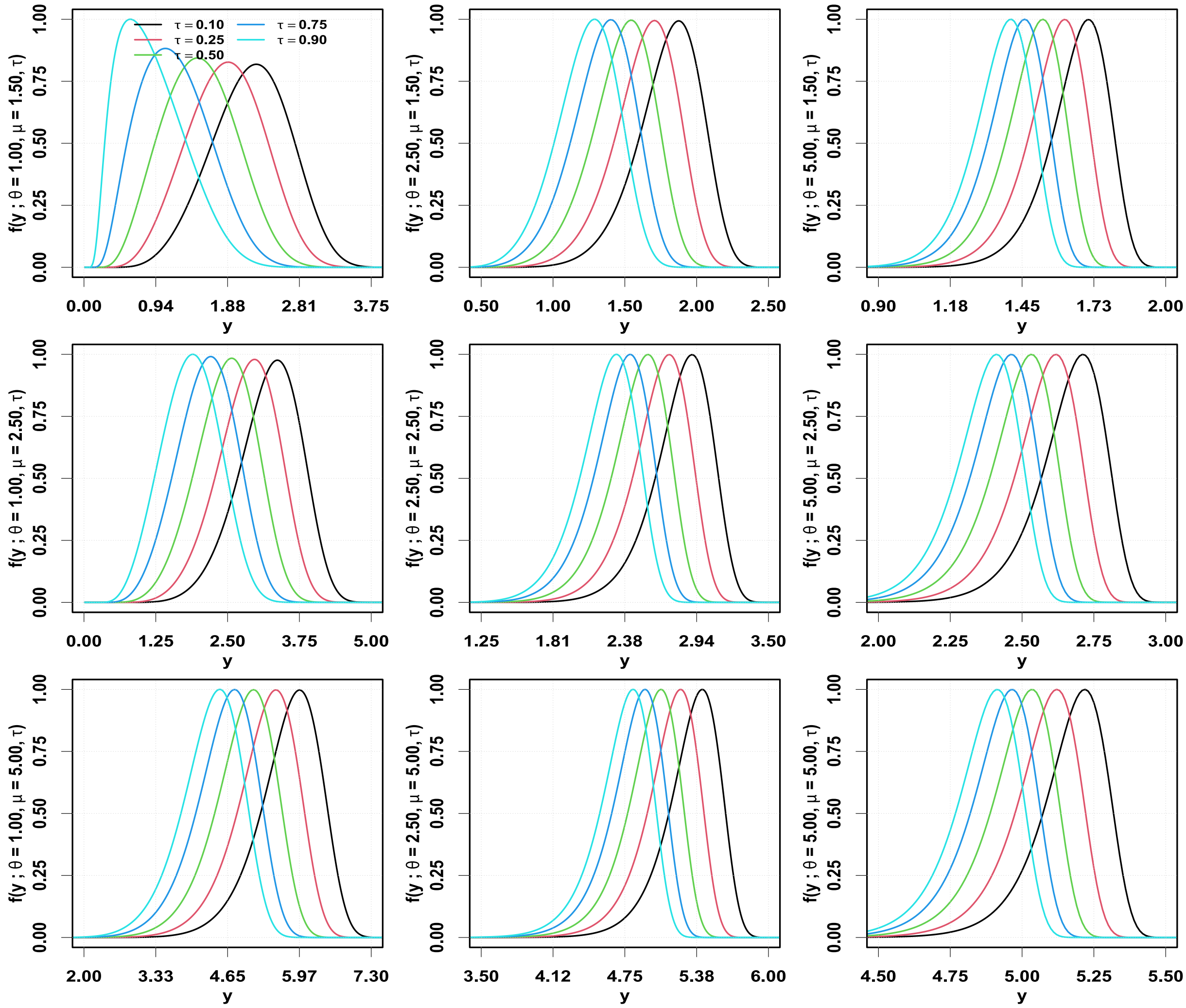

Fig. 19. Plots of the re-parameterized PDF stated from (4.51) for indicated values of μ, θ and τ.

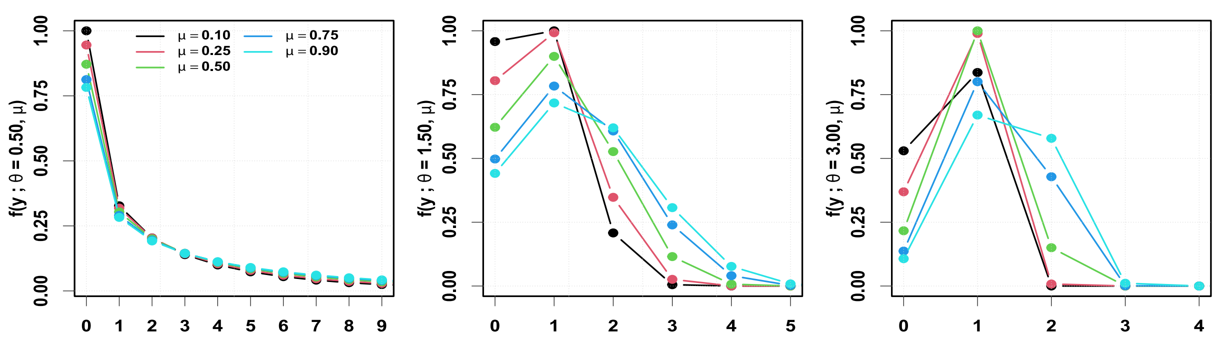

Fig. 20. Plots of the re-parameterized PDF stated from (4.57) for indicated values of μ and τ.

sion considering the Birnbaum-Saunders distribution was derived in [70,35].

4.2. The flexible Weibull distribution

The flexible Weibull distribution [11] has two parameters: $\alpha > 0$ and $\theta > 0$. A flexible Weibull distributed random variable Y has CDF, PDF and QF given by

$$f(y;\alpha,\theta) = \left(\theta + \frac{\alpha}{y^2}\right)\exp\left[\left(\theta y - \frac{\alpha}{y}\right) - \exp\left(\theta y - \frac{\alpha}{y}\right)\right], \quad (4.51)$$

$$F(y;\alpha,\theta) = 1 - \exp\left[-\exp\left(\theta y - \frac{\alpha}{y}\right)\right],$$

$$Q(\tau;\alpha,\theta) = \frac{\xi + \sqrt{\xi^2 + 4\theta\alpha}}{2\theta}, \quad (4.52)$$

where $y > 0$, $\xi = \log[-\log(1-\tau)]$ and $\theta > 0$, $\alpha > 0$ are shape parameters.

To study the relationship between the response variable and covariates in survival studies, a re-parameterization in terms of the median was introduced in [103], so that a re-parameterization in terms of all quantiles can be considered. From the expression defined in (4.52), the parameter α may be re-parameterized as

$\alpha = h^{-1}(\mu) = [(2\mu\theta - \xi)^2 - \xi^2]/(4\theta)$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. A quantile regression model and its diagnostic analytics for a Weibull distributed response with applications to engineering problems was presented in [112].

4.3. The logistic Nadarajah-Haghighi distribution

The logistic Nadarajah-Haghighi model [100] is suitable to describe random variables with positive support, such as time failures. It is derived by inserting the Nadarajah-Haghighi distribution [94] as the baseline model in the logistic-X class [121] of distributions. Let Y be a logistic Nadarajah-Haghighi distributed random variable. Then, the corresponding PDF, CDF and QF of Y are written, respectively, as

$$f(y;\alpha,\theta,\delta) = \frac{\alpha\theta\delta(1+\alpha y)^{\theta-1}\left[(1+\alpha y)^{\theta}-1\right]^{\delta-1}}{\left\{1+\left[(1+\alpha y)^{\theta}-1\right]^{\delta}\right\}^2}, \tag{4.53}$$

$$F(y;\alpha,\theta,\delta) = \frac{\left[(1+\alpha y)^{\theta}-1\right]^{\delta}}{1+\left[(1+\alpha y)^{\theta}-1\right]^{\delta}},$$
$$Q(\tau;\alpha,\theta,\delta) = \frac{1}{\alpha}\left\{\left[1+\left(\frac{\tau}{1-\tau}\right)^{\frac{1}{\delta}}\right]^{\frac{1}{\theta}}-1\right\}, \tag{4.54}$$

where $y > 0$, $\alpha > 0$ is the rate parameter and $\theta > 0$, $\delta > 0$ are shape parameters. For the parameter θ, the logistic Nadarajah-Haghighi family of models contains as one of its members to the logistic-exponential case. If $\delta = 1$, we have the Lomax distribution as particular case. If U has a standard uniform distribution, then $Q(U)$ has PDF given by (4.53).

A parametric regression model for right-censored data was constructed in [100] based on a median re-parmeterization of the logistic Nadarajah-Haghighi distribution. For that purpose, they re-parameterized the formula given in (4.53) in terms of the median, denoted by μ, which is obtained by setting $\tau = 0.5$ in the expression stated in (4.54) and then $\alpha = (1/\mu)(2^{1/\theta} - 1)$.

4.4. The class of log-symmetric distributions

The family of log-symmetric models [72,114,122] comprises several members that are generally used in the description of continuous, strictly positive and asymmetric data. This family also accommodates the possibility to model bimodal and/or light and heavy-tailed data. Log-symmetric distributions are obtained from the transformation $Y = \exp(X)$, where $X \sim \mathrm{S}(\alpha, \theta)$, which denotes a symmetric distributed random variable [114]. The corresponding PDF, CDF and QF of Y are defined, respectively, as

$$f(y;\alpha,\theta) = \frac{\xi}{\sqrt{\theta}y} f_g\left\{\frac{1}{\theta}[\log(y) - \log(\alpha)]^2\right\}, \tag{4.55}$$

$$F(y;\alpha,\theta) = F_g\left(\frac{1}{\theta}[\log(y) - \log(\alpha)]^2\right),$$
$$Q(\tau;\alpha,\theta) = \alpha\exp\left(\sqrt{\theta}z_\tau\right), \tag{4.56}$$

where $y > 0$, $\alpha > 0$ is the scale parameter, $\theta > 0$ is the power parameter, f_g is the PDF generator kernel possibly associated with an additional parameter ϑ (or parameter vector $\boldsymbol{\vartheta}$), ξ is a normalizing constant, with $F_g(w) = \xi\int_{-\infty}^{w} g(z^2)\mathrm{d}z$ and $z_\tau = F_g^{-1}(\tau)$ being the 100 τth quantile of a symmetric distribution. Some members of the log-symmetric family of distributions obtained from different f_g stated in (4.55) are the extended Birnbaum-Saunders, extended Birnbaum-Saunders-t, log-contaminated-normal, log-hyperbolic, log-normal, log-power-exponential, log-slash, and log-Student-t cases. From the expression defined in (4.56), the parameter α can be re-parameterized as $\alpha = h^{-1}(\mu) = \mu/\exp(\theta^{1/2}z_\tau)$, such that μ is, for a fixed and known value τ, the 100 τth quantile of the distribution of Y. This strategy was used in [114].

4.5. The discrete generalized half-normal distribution

The discrete generalized half-normal model [86] is obtained from the transformation $Y = P(X = k) = S_X(k) - S_X(k+1)$, where S_X denotes the survival function of a random variable X, where $X \sim \mathrm{GHN}(\alpha, \theta)$, which denotes a generalized half-normal distributed random variable [21]. The corresponding probability mass function, CDF and QF of the discrete generalized half-normal distribution, for a random variable Y, are formulated, respectively, as

$$f(y;\alpha,\theta) = 2\left\{\Phi\left[\left(\frac{y+1}{\alpha}\right)^{\theta}\right] - \Phi\left[\left(\frac{y}{\alpha}\right)^{\theta}\right]\right\}, \tag{4.57}$$

$$F(y;\alpha,\theta) = 2\Phi\left[\left(\frac{\lfloor y+1\rfloor}{\alpha}\right)^{\theta}\right] - 1,$$
$$Q(\tau;\alpha,\theta) = \left\lfloor \alpha\left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)^{\frac{1}{\theta}}\right] - 1\right\rfloor, \tag{4.58}$$

where $y \in \{0, 1, \ldots\}$ and $\lfloor a\rfloor$ denotes the floor function (integer part) of the number $a \in \mathbb{R}$. The r-th moment of Y is given by

$$\mathrm{E}(Y^r) = 2\sum_{k=0}^{\infty} y^r\Phi\left\{\Phi\left[\left(\frac{y+1}{\alpha}\right)^{\theta}\right] - \Phi\left[\left(\frac{y}{\alpha}\right)^{\theta}\right]\right\}, \quad r \in \{1, 2, \ldots\}. \tag{4.59}$$

From the expression defined in (4.59), note that the mean does not appear expressed in a closed form allowing simple re-parameterization. However, based on [34], considering $\tau = 0.5$ in (4.58), we can re-parameterize α as

$$\alpha = h^{-1}(\mu) = \frac{1+\mu}{0.6745^{1/\theta}},$$

such that μ is the median of the distribution of Y.

A median regression model of the discrete generalized half-normal was applied in [34] to the healthcare and compared with three other models: Poisson, negative binomial and generalized Poisson. A second application was considered with data on automobile insurance rate-making, in which the model was compared with the Poisson and negative binomial models. These applications showed that the discrete generalized half-normal model provided a better fit than the other models considered.

5. Regression, model fitting, computational implementation, and applications

In this section, we describe, for the two-parameters distributions presented previously, the general ML estimation method in a similar manner as for generalized linear models. In addition, we introduce some details regarding the `unitquantreg` package used for parameters estimation, as well as model selection and diagnostics. The models use a parametric regression approach where both location and shape parameters of the conditional distribution of the response are described employing covariates. The `unitquantreg` package is implemented in the framework of the `stats::lm` package. Therefore, most methods and packages that utilize this structure are also applied to it. Package and vignette are available from the GitHub at

https://github.com/AndrMenezes/unitquantreg
and may be downloaded and installed via
`devtools::install_github(''AndrMenezes/unitquantreg'')`
A full view of the package may be reached from the link
https://andrmenezes.github.io/unitquantreg/index.html
Lastly, we illustrate how the models residing in the `unitquantreg` package can be fitted.

5.1. Regression modeling

Suppose that the response Y_i, for $0 < Y_i < 1$, with $i \in \{1, \ldots, n\}$, is modeled considering the observed values $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})^\top$ and $\boldsymbol{z}_i = (1, z_{i1}, \ldots, z_{iq})^\top$ of the covariate vectors of both location and shape parameters of the conditional distribution of the response. We are interested in evaluating the effects of these covariates on both μ and θ (μ is, for a fixed τ, the 100 τth quantile of Y and θ is the shape parameter). For ML parameter estimation, we have the observed values $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ from the vector of n independent random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$. Then, let us assume the equations stated as

$$g_1(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \tag{5.60}$$

and

$$g_2(\theta_i) = \zeta_i = \delta_0 + \delta_1 z_{i1} + \cdots + \delta_q z_{iq}, \tag{5.61}$$

which relate μ_i and θ_i to the linear predictions η_i and ζ_i, respectively. Furthermore, consider that $Y_i|\boldsymbol{x}_i, \boldsymbol{z}_i \sim F(y_i; \mu_i, \theta_i)$, where F is the CDF of a two-parameter distribution. We assume that g_1 and g_2 are strictly monotonic, twice differentiable functions that map the 100 τth quantile μ_i and θ_i to the line of real numbers [29] p. 228]. Suitable choices of g_1 are the following link functions used in generalized linear models [87]: (i) the inverse CDF of the logistic (logit link); (ii) standard normal (probit link); (iii) minimum extreme-value (complementary log-log link); (iv) maximum extreme-value (log-log link); and (v) Cauchy (Cauchit link). Furthermore, the shape parameter θ_i must be positive and the link function g_2 is, for example, the logarithm or square-root links [88]. It is important to note that $\boldsymbol{x}$ and $\boldsymbol{z}$ can be identical or they could be subsets of each other.

To obtain the ML estimates of the model parameters, we need the first-order and second-order partial derivatives of the logarithm of the corresponding likelihood function. For the observed response i, y_i namely, with $i \in \{1, \ldots, n\}$, the log-likelihood function is given by $\ell_i = \ell_i(\boldsymbol{\Theta}) = \log[f(y_i; \boldsymbol{\Theta}, \boldsymbol{x}_i, \boldsymbol{z}_i, \tau)]$, such that the score equations are defined as

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}, \quad \frac{\partial \ell_i}{\partial \delta_j} = \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \zeta_i}\frac{\partial \zeta_i}{\partial \delta_j},$$

for $j \in \{1, \ldots, (p+1)\}$ and $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\delta})$, where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ is a $(p+1) \times 1$ parameter vector associated with a covariate matrix $\boldsymbol{X}_{n \times (p+1)}$ and $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_q)$ is a $(q+1) \times 1$ parameter vector associated with a covariate matrix $\boldsymbol{Z}_{n \times (q+1)}$. Considering the full log-likelihood function, we have the score vectors for $\boldsymbol{\Theta}$ written, respectively, as

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \mathrm{diag}(\boldsymbol{W}_\mu)\dot{\boldsymbol{\ell}}_\mu, \quad \frac{\partial \ell}{\partial \boldsymbol{\delta}} = \boldsymbol{Z}^\top \mathrm{diag}(\mathbf{W}_\delta)\dot{\boldsymbol{\ell}}_\delta,$$

where diag is an $n \times n$ diagonal matrix,

$$\boldsymbol{W}_\mu = \left[\frac{\partial \mu_1}{\partial \eta_1}, \ldots, \frac{\partial \mu_n}{\partial \eta_n}\right], \quad \dot{\boldsymbol{\ell}}_\mu = \left[\frac{\partial \ell}{\partial \mu_1}, \ldots, \frac{\partial \ell}{\partial \mu_n}\right]^\top, \quad \boldsymbol{W}_\delta = \left[\frac{\partial \theta_1}{\partial \zeta_1}, \ldots, \frac{\partial \theta_n}{\partial \zeta_n}\right], \quad \dot{\boldsymbol{\ell}}_\delta = \left[\frac{\partial \ell}{\partial \theta_1}, \ldots, \frac{\partial \ell}{\partial \theta_n}\right]^\top.$$

For the Hessian matrix, we have the expressions stated as

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \boldsymbol{X}^\top \mathrm{diag}(\ddot{\boldsymbol{\ell}}_{\mu\mu})\mathrm{diag}(\boldsymbol{W}_\mu^2)\boldsymbol{X},$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^\top} = \boldsymbol{Z}^\top \mathrm{diag}(\ddot{\boldsymbol{\ell}}_{\delta\delta})\mathrm{diag}(\boldsymbol{W}_\delta^2)\boldsymbol{Z}$$

and

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\delta}^\top} = \boldsymbol{X}^\top \mathrm{diag}(\ddot{\boldsymbol{\ell}}_{\mu\delta})\mathrm{diag}(\boldsymbol{W}_\mu)\mathrm{diag}(\mathbf{W}_\delta)\boldsymbol{Z}$$

where

$$\ddot{\boldsymbol{\ell}}_{\mu\mu} = \left[\frac{\partial^2 \ell}{\partial \mu_1 \partial \mu_1}, \ldots, \frac{\partial^2 \ell}{\partial \mu_n \partial \mu_n}\right], \quad \ddot{\boldsymbol{\ell}}_{\delta\delta} = \left[\frac{\partial^2 \ell}{\partial \delta_1 \partial \delta_1}, \ldots, \frac{\partial^2 \ell}{\partial \delta_n \partial \delta_n}\right], \quad \ddot{\boldsymbol{\ell}}_{\mu\delta} = \left[\frac{\partial^2 \ell}{\partial \mu_1 \partial \delta_1}, \ldots, \frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial \mu_n \partial \delta_n}\right].$$

5.2. Computational implementation

The ML estimates $(\widehat{\beta}_0, \ldots, \widehat{\beta}_p)$ and $(\widehat{\delta}_0, \ldots, \widehat{\delta}_q)$ can be obtained, for instance, through general-purpose optimization algorithms as Nelder-Mead, quasi-Newton and conjugate-gradient available in the `optim` function of the `stats` package of R [106]; see [23] for a study on ten computational algorithms in the estimation of parameters for a class of regression models. In the study performed in [23], the controlled random search, differential evolutionary, `DIRECT`, `DIRECT_L`, evolutionary, genetic, memetic, particle swarm, self-adapted evolutionary, and simulated annealing methods were evaluated by using the Monte Carlo simulation method with the R software. In that study, four algorithms presented satisfactory results (differential evolutionary, simulated annealing, stochastic ranking evolutionary, and controlled random search algorithms, with the latter one having the best performance). The `optim` function failed in most cases, but when it was successful, it is more accurate and faster than the others. The annealing algorithm obtained satisfactory estimates in viable time with few failures so that we recommend its use when the `optim` function it fails.

As mentioned, to obtain the ML estimates, we developed the `unitquantreg` package. The quantile regression models fitted by `unitquantreg` take the baseline distributions: (i) arcsecant hyperbolic Weibull, (ii) Johnson SB, (iii) Kumaraswamy, (iv) log-extended exponential-geometric, (v) unit-Birnbaum-Saunders, (vi) unit-Burr-XII, (vii) unit-Chen, (viii) unit-Gompertz, (ix) unit-Gumbel, (x) unit-logistic, (xi) unit-Weibull, and (xii-xiii) two versions of the unit generalized half-normal. These 13 distributions implemented in the `unitquantreg` package follow the standard naming convention of R, where names of CDF, PDF, QF and random generation functions follow the `d`, `p`, `q`, and `r` prefixes, as it is usual in the R software. For example, for the unit-Weibull distribution, we define:

```
duweibull(x, mu, theta, tau = 0.5, log = FALSE)
puweibull(q, mu, theta, tau = 0.5, lower.tail = TRUE, log.p = FALSE)
quweibull(p, mu, theta, tau = 0.5, lower.tail = TRUE, log.p = FALSE)
ruweibull(n, mu, theta, tau = 0.5)
```

These and all other `d`, `p`, `q`, and `r` functions are vectorized and coded in `C++` using the `Rcpp` package [30,31]. The main function of the `unitquatreg` package, `unitquantreg()` namely, works similarly to the functions `lm()`, `glm()`, `betareg()`, `simplexreg()`, `gamlss()`, `cdfquantreg()` as follows:

```
unitquantreg(formula, data, subset, na.action, tau, family,
             link = c("logit", "probit", "cloglog", "cauchit"),
             link.theta = c("log", "sqrt", "identity"), start = NULL,
             control = unitquantreg.control(), model = TRUE, x = FALSE, y = TRUE)
```

In the same way, for example, in the `simplexreg package` [132], the regression model can be specified via an `R` package named `Formula` [131]. Thus, to specify both quantile and shape equations stated as in (5.60) and (5.61), we define `formula = y ~ x1 + x2| z1 + z2`, where `y ~ x1 + x2` specifies the quantile model with covariates `z1` and `z2` being related to the shape parameter. Without the latter part, the model `formula = y ~ x1 + x2` is fitted. Note that, for g_1 given in (5.60), we have the four options:
`link = c(''logit'', ''probit'', ''cloglog'', ''cauchit'')`,
while for g_2 given in (5.61), we have the options:
`link.theta = c(''log'', ''sqrt'', ''identity'')`.

Methods for extracting information from the returned S3 class object named `unitquantreg`, such as for the generic functions `coef`, `print`, `plot` and `summary`, are available. Next, the methods within the unitquantreg class are listed:

```
> methods(class = "unitquantreg")
 [1] coef    confint   fitted   hnp    logLik    model.frame   model.matrix
 [8] plot    predict   print   residuals    summary    terms    update    vcov
```

In particular, the model parameters are estimated by the ML method using the `optimx` package [96], which is a general-purpose optimization wrapper function that calls other `R` tools for optimization, including the existing `optim` function [23]. The main advantage of `optimx` is to unify the tools allowing a number of different optimization methods and providing sanity checks. The `unitquantreg.control command`, behind the `control argument`, handles the fitting process and its default values are:

```
method     hessian gradient maxit  factr   trace dowarn  starttests fnscale
"L-BFGS-B" "FALSE" "TRUE"   "5000" "1e+07" "0"   "FALSE" "FALSE"    "1"
```

whereas its two most important arguments are: `hessian` and `gradient`, which control the `optimx` whether it should use the analytical Hessian matrices and the analytical score vectors, respectively. For all available distributions, the Hessian matrices and the score vectors are implemented in `C++` for more accurate estimates and computation performance. Starting values for the vector parameter $\boldsymbol{\beta}$ may be user-supplied, otherwise the starting values for $\boldsymbol{\beta}$ are estimated from the quantile regression model, where the response is defined as $y_i^* = g_1(y_i)$, being g_1 the link function for the quantile parameter μ.

The `rq` function from the `quantreg` package [56] is employed to obtain the starting values. For δ_j, with $j \in \{1, \ldots, q\}$, the initial values are setting as 0.1 for the logarithm link function, and 1.1 for the inverse and square roots link functions. The standard errors (SEs) are obtained employing the observed Fisher information matrix, which is computed from the inverse of the analytical Hessian matrix implemented in `C++`. For numerical stability, the inverse of Hessian matrix is calculated using a Cholesky decomposition.

The `uniquantreg` function checks if the optimization algorithm converged. If it fails, the warning message “optimization failed to converge” is printed and the user must be care about the results. Furthermore, the Karush–Kuhn–Tucker optimality conditions are checked by the `optimx::optimx` function. If some of these conditions is not satisfied, then a warning message is also printed. The package computes the Moore-Penrose inverse if the final Hessian matrix is full rank, but it has at least one negative eigen-value.

The distribution of the response is defined by the `family` argument, and the following names can used for its members:

```
families <- c("arc-secant hyperbolic Weibull" = "ashw",
              "Johnson-SB" = "johnsonsb",
              "Kumaraswamy" = "kum",
              "log-extended exponential-geometric" = "leeg",
              "unit-Birnbaum-Saunders" = "ubs",
              "unit-Burr-XII" = "uburrxii",
              "unit-Chen" = "uchen",
              "unit-generalized half-normal-E" = "ughne",
              "unit-generalized half-normal-X" = "ughnx",
              "unit-Gompertz" = "ugompertz",
              "unit-Gumbel" = "ugumbel",
              "unit-logistic" = "ulogistic",
              "unit-Weibull" = "uweibull")
```

Table 1
Summary of univariate statistics for bodyfat data set.

Variable (by level)	Mean	SD	Min	Max	Q1	Q2	Q3	CS	CK
arms	0.266	0.111	0.042	0.547	0.181	0.261	0.344	0.157	-0.772
arms (by F)	0.340	0.087	0.119	0.547	0.282	0.338	0.407	-0.096	2.414
arms (by M)	0.191	0.078	0.042	0.418	0.136	0.188	0.242	0.361	2.793
arms (by A)	0.236	0.113	0.042	0.490	0.143	0.219	0.321	0.382	2.185
arms (by I)	0.283	0.094	0.074	0.488	0.225	0.290	0.344	0.096	2.716
arms (by S)	0.324	0.100	1.132	0.547	0.244	0.329	0.405	0.139	2.152
age	46.000	19.879	18.000	87.000	25.000	47.000	65.000	0.158	-1.359
bmi	24.716	3.151	18.500	29.900	22.300	24.900	27.200	-0.101	-0.943

where F: female, M: male, A: active, I: insufficiently active, S: sedentary, SD: standard deviation, CS: coefficient of skewness, CK: coefficient of variation, Qi: ith quartile, with $i \in \{1, 2, 3\}$

Table 2
Spearman correlation coefficient (with the corresponding p-value under the null hypothesis H_0: $\rho = 0$) for the indicated variables.

Variable	bmi	age
arms	0.381 (<0.001)	0.464 (<0.001)
bmi		0.470 (<0.001)

Model diagnostics, including theoretical quantile versus empirical quantile (QQ) plots with simulated envelopes [94] for the Cox-Snell and normalized quantile residuals are available. The normalized quantile residual [29] is defined by $\widehat{r}_i = \Phi^{-1}[F(y_i; \widehat{\mu}_i, \widehat{\theta}_i)]$, for $i \in \{1, \dots, n\}$. If the model is correctly specified, then $\widehat{r}_i$ has an approximate standard normal distribution. In addition, stating $\widehat{r}_i = -\log[1 - F(y_i; \widehat{\mu}_i, \widehat{\theta}_i)]$, we have the estimated Cox-Snell residuals. The important property of the Cox-Snell residual is that if the model selected fits the data adequately, $\widehat{r}_i$ follows the standard exponential distribution.

Methods to compare two fitted models and model selection criteria are available by the functions `vuong.test()` and `likelihood_stats()`, respectively. The `vuong.test()` was designed to implement the Vuong test when comparing non-nested models [125], while `likelihood_stats()` has a variety of likelihood-based information criteria [44,115,124,2] for model diagnostics.

5.3. Biomedical application I with Brazilian data

Next, we consider a real-world data set first reported in [102] that was also analyzed in [77]. This data set contains 298 observations about body fat percentage of individuals assisted in a public hospital in Curitiba, Paraná, Brazil. The fat percentages at android, arms, gynecoid, legs, and body correspond to the five responses, and the data set is composed of two continuous and two categorical covariates. Continuous covariates refer to the age (in years) and body mass index (bmi, in kg/m^2) of the individuals, while the categorical covariates are related to gender (female or male) and ipaq (active, insufficiently active or sedentary patient). As described in [12], the ipaq is a questionnaire that allows us to obtain data about the weekly time spent on physical activities of moderate and strong intensity, in different contexts of daily life, such as: housework, leisure, transportation, and work, as well as the time spent in passive activities performed on the seating position. The bodyfat data set is stored in the dataframe `unitquantreg::bodyfat`. The four first rows of this dataframe are presented as:

```
> library("unitquantreg")
> data("bodyfat", package = "unitquantreg")
> head(bodyfat)
   arms  legs  body android gynecoid    bmi age    sex                  ipaq
1 0.163 0.234 0.238   0.295    0.314 -3.916 -28   male insufficiently active
2 0.331 0.335 0.366   0.432    0.431  0.584 -28 female                active
3 0.252 0.312 0.179   0.169    0.354 -2.616 -28 female                active
4 0.094 0.172 0.206   0.251    0.272 -2.316 -28   male                active
```

where bmi is centered at 24.72 (the average bmi), age is centered at 46.00 (the average age) and sex and ipaq are, as mentioned, categorical variables with two and three levels, respectively.

Now, we present an exploratory data analysis for all continuous variables in bodyfat data set. Tables 1 and 2, respectively, report univariate descriptive statistics for each variable and pairwise Spearman correlation coefficients for these variables. Figure 21 shows histograms and scatter-plots for the variables: arms, age and bmi. Note that the response has a symmetric empirical distribution between 0.042 and 0.547, which can be well modeled by several members of the family of models proposed in our R package. Also, observe that all the correlations are statistically significant at 1%. On the one hand, the significant correlations between the response and the continuous covariates bmi and age support the formulation of a regression model. However, on the other hand, the significant correlation between bmi and age could indicate a collinearity problem which is analyzed when the regression models are stated.

In what follows, we explore the functional relationship between the covariates and the body fat at arms through 13 quantile regression models. For a fixed $\tau = 0.5$, let us assume that the functional relationship between μ_i, θ_i and the linear predictors are given by

$$\text{logit}(\mu_i) = \beta_0 + \beta_1\, \text{bmi}_i + \beta_2\, \text{age}_i + \beta_3\, \text{sexmale}_i + \beta_4\, \text{ipaqinsufficientlyactive}_i + \beta_5 \text{ipaqactive}_i;$$
$$\log(\theta_i) = \delta_0;$$

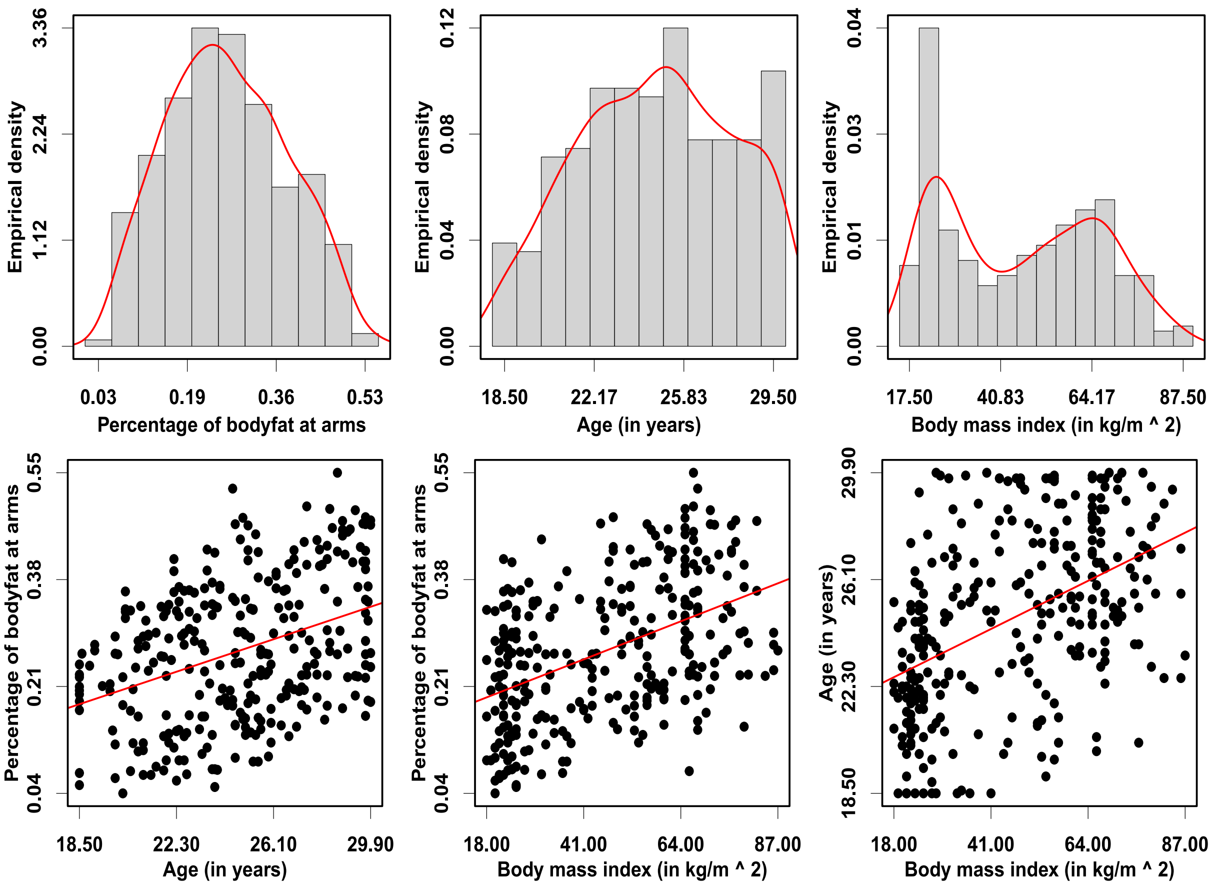

Fig. 21. Histograms and scatter-plots for bodyfat in arms data set.

where “sexmale = 1 is utilized for “sex = male and “sexmale = 0 for “sex = female; “ipaqinsufficientlyactive = 1 is employed for “ipaq = insufficiently active and “ipaqinsufficientlyactive = 0 for “ipaq = sedentary, whereas “ipaqactive = 1 is used for “ipaq = active and “ipaqactive = 0 for “ipaq = sedentary. To simultaneously fit all available models, we can use, for example,

```
> models <- c("ashw", "johnsonsb", "kum", "leeg", "ubs", "uburrxii", "uchen",
              "ughne", "ughnx", "ugompertz", "ugumbel", "ulogistic", "uweibull")
> fits   <- lapply(1:13, function(i) unitquantreg(arms ~ age + bmi + sex + ipaq,
                   family = models[i], link = "logit", link.theta = "log",
                   tau = 0.5, data = bodyfat))
```

which creates a list for each of the distributions and each of these lists contains 23 objects (`? unitquantreg::unitquantreg` for more details) named as:

```
> names(sapply(fits[[5]], mode))
 [1] "family"       "coefficients" "fitted.values"  "linear.predictors"
 [5] "link"         "tau"          "loglik"         "gradient"
 [9] "vcov"         "nobs"         "npar"           "df.residual"
[13] "theta_const"  "control"      "iterations"     "converged"
[17] "kkt"          "elapsed_time" "call"           "formula"
[21] "terms"        "model"        "y"
```

Considering all distributions available in the unitquantreg package, the ML estimates, the SEs of their corresponding estimators for the parameters, and the p-values of the associated tests, can be obtained using the following instructions:

```
> lapply(fits, function(x) round(rbind(mle = coef(x), se = sqrt(diag(vcov(x))),
  p.value = summary(x)$coeftable[,4]), 3))
$ashw
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.476 0.005 0.082  -0.898                   -0.125     -0.332      2.137
se            0.045 0.001 0.007   0.038                    0.053      0.052      0.040
p.value       0.000 0.000 0.000   0.000                    0.017      0.000      0.000

$johnsonsb
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.470 0.005 0.092  -0.938                   -0.117     -0.263      1.174
se            0.047 0.001 0.007   0.037                    0.054      0.052      0.041
p.value       0.000 0.000 0.000   0.000                    0.031      0.000      0.000

$kum
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.530 0.004 0.082  -0.820                   -0.076     -0.216      1.553
se            0.049 0.001 0.006   0.037                    0.054      0.053      0.045
p.value       0.000 0.000 0.000   0.000                    0.154      0.000      0.000

$leeg
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.450 0.005 0.093  -0.957                   -0.131     -0.257      2.032
se            0.049 0.001 0.007   0.039                    0.056      0.052      0.049
p.value       0.000 0.000 0.000   0.000                    0.019      0.000      0.000

$ubs
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.485 0.004 0.086  -0.895                   -0.115     -0.242     -1.836
se            0.040 0.001 0.006   0.035                    0.048      0.047      0.041
p.value       0.000 0.000 0.000   0.000                    0.016      0.000      0.000

$uburrxii
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.506 0.004 0.053  -0.565                   -0.083     -0.160      2.439
se            0.036 0.001 0.007   0.051                    0.042      0.043      0.070
p.value       0.000 0.008 0.000   0.000                    0.047      0.000      0.000

$uchen
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.283 0.007 0.114  -1.138                   -0.138     -0.387      0.593
se            0.075 0.002 0.010   0.056                    0.086      0.084      0.022
p.value       0.000 0.000 0.000   0.000                    0.108      0.000      0.000

$ughne
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.481 0.005 0.072  -0.833                   -0.135     -0.382      1.452
se            0.043 0.001 0.006   0.037                    0.050      0.051      0.043
p.value       0.000 0.000 0.000   0.000                    0.006      0.000      0.000

$ughnx
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.504 0.004 0.080  -0.799                   -0.098     -0.227      0.963
se            0.045 0.001 0.006   0.035                    0.049      0.050      0.044
p.value       0.000 0.000 0.000   0.000                    0.045      0.000      0.000
```

```
$ugompertz
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.362 0.006 0.104  -1.057                    -0.148     -0.385      1.308
se            0.060 0.001 0.008   0.044                     0.068      0.066      0.042
p.value       0.000 0.000 0.000   0.000                     0.029      0.000      0.000

$ugumbel
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.438 0.006 0.092   -0.97                    -0.130     -0.354      1.092
se            0.050 0.001 0.007    0.04                     0.058      0.058      0.040
p.value       0.000 0.000 0.000    0.00                     0.026      0.000      0.000

$ulogistic
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.475 0.005 0.089  -0.932                    -0.122     -0.239      1.761
se            0.043 0.001 0.007   0.036                     0.051      0.048      0.049
p.value       0.000 0.000 0.000   0.000                     0.017      0.000      0.000

$uweibull
        (Intercept)   age   bmi sexmale ipaqinsufficiently active ipaqactive log(theta)
mle          -0.494 0.005 0.077  -0.863                    -0.121      -0.32      1.789
se            0.042 0.001 0.006   0.037                     0.050       0.05      0.041
p.value       0.000 0.000 0.000   0.000                     0.014       0.00      0.000
```

From the above results reported for all the models, in general, similar values are obtained in terms of significance at 5%, only the Kumaraswamy and unit-Chen models have a slightly different value in the covariate “ipaq (insufficiently active)”.

For model selection, the function `likelihood_stats()` computes and reports various statistics to indicate how well the estimated model fits the data. We consider the following criteria: Neg2LogLike $= -2\log(L)$, AIC $= -2\log(L) + 2p$ [2], BIC $= -2\log(L) + p\log(n)$ [115] and HQIC $= -2\log(L) + 2p\log[\log(n)]$ [44], where L, n and p are, respectively, the maximized likelihood function, the sample size and the number of model parameters estimated, whereas the acronyms AIC, BIC and HQIC indicate the Akaike, Bayesian, and Hannan-Quinn information criteria, respectively; see more details in [124]. In general, when you are comparing candidate models, smaller Neg2LogLike, AIC, BIC and HQIC statistics indicate a better fitting model. For further details on likelihood-based statistics for model selection, we recommend [17,20].

If `likelihood_stats(lt = fits)` is applied to the list `fits`, then it returns:

```
 Likelihood-based statistics of fit for unit quantile regression models
Call:  likelihood_stats(lt = fits)
                                  Neg2LogLike AIC      BIC      HQIC
arc-secant hyperbolic Weibull      -844.298    -830.298 -804.418 -819.938
Johnson-SB                         -897.967    -883.967 -858.088 -873.608
Kumaraswamy                        -859.030    -845.030 -819.151 -834.671
log-extended exponential-geometric -867.814    -853.814 -827.934 -843.454
unit-Birnbaum-Saunders             -923.332    -909.332 -883.452 -898.972
unit-Burr-XII                      -661.098    -647.098 -621.218 -636.738
unit-Chen                          -705.722    -691.722 -665.843 -681.363
unit-generalized half-normal-X     -858.130    -844.130 -818.250 -833.771
unit-generalized half-normal-E     -808.135    -794.135 -768.256 -783.776
unit-Gompertz                      -771.300    -757.300 -731.420 -746.941
unit-Gumbel                        -814.947    -800.947 -775.067 -790.588
unit-logistic                      -902.401    -888.401 -862.522 -878.042
unit-Weibull                       -854.932    -840.932 -815.052 -830.573
```

According to the values of these statistics, the structure based on the unit-Birnbaum-Saunders distribution is the best model that fits the bodyfat data set.

Table 3
Coefficient of determination (R^2) for bodyfat data set.

Distribution	R^2
unit-logistic	0.7773
unit-Birnbaum-Saunders	0.7751
unit-Gumbel	0.7715
log-extended exponential-geometric	0.7678
Johnson-SB	0.7689
arc-secant hyperbolic Weibull	0.7641
unit-Gompertz	0.7558
unit-Weibull	0.7578
unit-generalized half-normal-E	0.7465
unit-generalized half-normal-X	0.7323
Kumaraswamy	0.7219
unit-Chen	0.6967
unit-Burr-XII	0.5202

Before statistically analyzing the unit-Birnbaum-Saunders quantile regression model, we must evaluate the possible collinearity problem detected in the exploratory data analysis. A more formal and often used approach to measuring correlation between covariates is the variance inflation factor (VIF). If VIF $>$ 10, then collinearity could exist [104,112]. Note that once we define `vcov`, `terms` and `model.matrix`, we can employ the `vif` function, available in the `car` package [53], for computing the VIF and detecting multicollinearity. For the unit-Birnbaum-Saunders model, we have:

```
> car::vif(fits[[5]])
         GVIF Df  GVIF^(1/(2*Df))
age  1.638593  1  1.280075
bmi  1.400078  1  1.183249
sex  1.065376  1  1.032171
ipaq 1.264615  2  1.060449
```

Therefore, as all the VIF values are less than 10, then we can continue with our quantile regression analysis based on the unit-Birnbaum-Saunders distributions; otherwise, we should refit the models consequently. In the unit-Birnbaum-Saunders case, by using `summary(fits[[5]])`, we obtain:

```
Wald-tests for unit-Birnbaum-Saunders quantile regression model
Call:  unitquantreg(formula = arms ~ age + bmi + sex + ipaq, data = bodyfat,
                    tau = 0.5, family = "ubs", link = "logit", link.theta = "log")
Mu coefficients: (quantile model with logit link and tau = 0.5):
                           Estimate   SE       z-value Pr(Z>|z|)
(intercept)               -0.485445   0.039953 -12.150  < 2e-16 ***
age                        0.004355   0.001091   3.990 6.60e-05 ***
bmi                        0.085517   0.006277  13.623  < 2e-16 ***
sexmale                   -0.895046   0.035443 -25.253  < 2e-16 ***
ipaqinsufficiently active -0.115179   0.047879  -2.406   0.0161 *
ipaqactive                -0.242332   0.046505  -5.211 1.88e-07 ***
Signif. codes:  0 '***' < 0.001  '**' < 0.01  '*' < 0.05  '.' < 0.1  'n.s.' < 1
---
Model with constant shape:
           Estimate    SE       z-value  Pr(Z>|z|)
log(theta) -1.83575    0.04096  -44.82   <2e-16 ***
---
Residual degrees of freedom: 291
Log-likelihood: 461.6659; Number of iterations: 116
```

Analogously to the coefficient of determination R^2 in ordinary regression models, we can compute it in the context of parametric quantile regression [25,32,95,112]. This coefficient is considered as a global measure of explained variation of a model for its response, given by $R^2 = 1 - [L_0/L_{p+q}]^{2/n}$, where L_0 and L_{p+q} denote the likelihood functions for models containing only the intercept and the model containing the intercept, plus a number of $p+q$ covariates, respectively. Table 3 reports the values of R^2 considering all distributions for $\tau = 0.5$. Like the AIC and BIC statistics, R^2 is most helpful for comparing competing models that are not necessarily nested, with larger values indicating better models.

It is important to point out that the `unitquantreg()` function is very flexible and allows us to use regression splines, through the `ns()` function, which is available in the `splines` package of R. In addition, one might want to look at the `gam()` function of the `mgcv` package [128], which is distributed with R. For further details, see for example [8,45,101]. To permit for non-linearity in the covariate age and the unit-Birnbaum-Saunders distribution, we can consider:

```
> model_fitted <- unitquantreg(arms ~ ns(age, df = 3) + bmi + sex + ipaq,
                               family = "ubs", link = "logit", link.theta = "log",
                               tau = 0.5, data = bodyfat)
> summary(model_fitted)

 Wald-tests for unit-Birnbaum-Saunders quantile regression model

Call:  unitquantreg(formula = arms ~ ns(age, df = 3) + bmi + sex + ipaq,
                    data = bodyfat, tau = 0.5, family = "ubs",
                    link = "logit", link.theta = "log")

Mu coefficients: (quantile model with logit link and tau = 0.5):
                           Estimate   SE         z-value Pr(Z>|z|)
(Intercept)               -0.605737   0.073065  -8.290  < 2e-16 ***
ns(age, 3)1                0.063247   0.082622   0.766 0.443972
ns(age, 3)2                0.381365   0.168609   2.262 0.023708 *
ns(age, 3)3                0.329510   0.085458   3.856 0.000115 ***
bmi                        0.086967   0.006425  13.535  < 2e-16 ***
sexmale                   -0.904574   0.036113 -25.049  < 2e-16 ***
ipaqinsufficiently active -0.110643   0.048126  -2.299 0.021504 *
ipaqactive                -0.235635   0.047189  -4.993 5.93e-07 ***
---
Signif. codes:  0 '***' < 0.001  '**' < 0.01  '*' < 0.05  '.' < 0.1  'n.s.' < 1

Model with constant shape:
           Estimate    SE       z-value  Pr(Z>|z|)
log(theta) -1.83857    0.04096  -44.88   <2e-16 ***
---
Signif. codes:  0 '***' < 0.001  '**' < 0.01  '*' < 0.05  '.' < 0.1  'n.s.' < 1

Residual degrees of freedom: 289
Log-likelihood: 462.4979
Number of iterations: 65
```

To graphically assess the adequacy of the fitted model, we can generate, as in the `hnp` package [93], the QQ (half-normal) plot with simulated envelopes using the Cox-Snell and normalized quantile residuals; see Figure 22. From this figure, we observe the good agreement between the unit-Birnbaum-Saunders quantile regression model and the bodyfat data set. Then, once again, we can continue with our quantile regression analysis based on the unit-Birnbaum-Saunders distribution; otherwise, we should refit the models consequently. The syntax to generate the half-normal plot with simulated envelopes using the Cox-Snell residual is as follows:

```
> hnp(object, nsim = 99, halfnormal = TRUE, plot = TRUE, output = TRUE,
level = 0.95, resid.type = c(''quantile'', ''cox-snell''),...)
```

Observe that the value of $\widehat{\beta}_0$ is the estimate for a female with 46 years old, body mass index equal to 24.72 kg/m^2 and sedentary. The parameter estimates for β_1 and β_2 indicate that age and bmi have a positive effect on the percentage of fat in the arms. In contrast, the parameters β_3, β_4 and β_5 are negatively estimated, indicating that this percentage is less for insufficiently active and active men, respectively. To compute 95% confidence intervals for one or more parameters in a fitted model, we use `confint(fits[[5]])`, which returns:

```
                          lower limit   upper limit
(intercept)               -0.563751859 -0.407138438
age                        0.002216026  0.006494468
bmi                        0.073213867  0.097819942
sexmale                   -0.964512508 -0.825579136
ipaqinsufficiently active -0.209020809 -0.021337875
ipaqactive                -0.333479151 -0.151184209
log(theta)                -1.916037287 -1.755470782
```

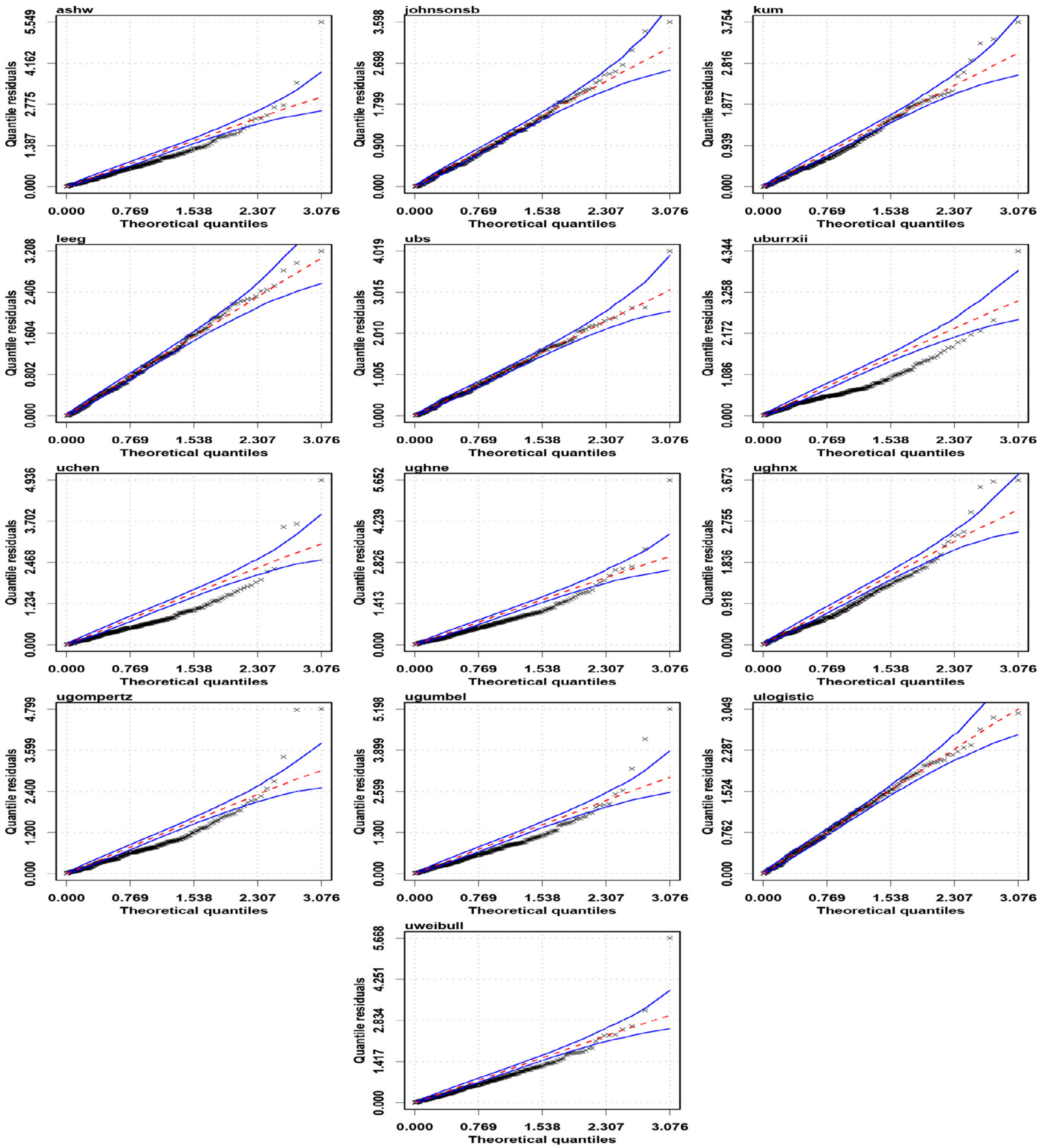

Fig. 22. QQ (half-normal) plots with a simulated envelope of quantile residuals with Brazilian data.

Point and interval estimates are more helpful if converted to the corresponding values for the odds themselves by exponentiating the values. In the case of the unit-Birnbaum-Saunders model, we have that:

```
> exp(coef(fits[[5]]))
             (intercept)                      age                      bmi
               0.6154232                1.0043647                1.0892800
                 sexmale ipaqinsufficiently active               ipaqactive
               0.4085889                0.8912063                0.7847958
              log(theta)
               0.1594932

> exp(confint(fits[[5]]))
                          lower limit upper limit
(intercept)                 0.5690700 0.6655520
age                         1.0022185 1.0065156
bmi                         1.0759606 1.1027642
sexmale                     0.3811690 0.4379813
ipaqinsufficiently active 0.8113784 0.9788882
ipaqactive                  0.7164268 0.8596893
log(theta)                  0.1471891 0.1728259
```

Notice that some methods available in the `lmtest` package [130], such as `coeftest`, `coefci`, `lrtest`, `waldtest`, `reset`, can be applied to an object created by the `unitquantreg` function. These methods show the flexibilidade of this package for: (i) testing a restricted model versus a full model and (ii) verifying whether a functional structure is reasonable or not. For example:

```
> lmtest::lrtest(fits[[5]])
Likelihood ratio test
Model 1: arms ~ age + bmi + sex + ipaq | 1
Model 2: arms ~ 1 | 1
  #Df LogLik Df Chisq Pr(>Chisq)
1   7 461.67
2   2 239.37 -5 444.6  < 2.2e-16 ***

> lmtest::resettest(fits[[5]])
RESET test
data:  fits[[5]]
RESET = 0.037096, df1 = 2, df2 = 290, p-value = 0.9636
```

returns the likelihood ratio test and Ramsey RESET test [107] for the funcional form, respectively.

Lastly, we may fit a particular distribution (the unit-Birnbaum-Saunders one, for example) for various values of τ, with $0 < \tau < 1$, just like in the `quantreg` package [56], as well as in the `SAS PROC QUANTREG` procedure [113], as follows:

```
unitquantreg(arms ~ age + bmi + sex + ipaq | bmi, family = "ubs",
             link = "logit", link.theta = "log", data = bodyfat,
             tau = c(0.25, 0.50, 0.75))

unit-Birnbaum-Saunders quantile regression model
Call:  unitquantreg(formula = arms ~ age + bmi + sex + ipaq | bmi, data = bodyfat,
                    tau = c(0.25, 0.5, 0.75), family = "ubs", link = "logit",
                    link.theta = "log")
Mu coefficients (quantile model with logit link):
                           tau = 0.2500  tau = 0.5000  tau = 0.7500
(intercept)                -0.6564       -0.4847       -0.3205
age                         0.0046        0.0044        0.0041
bmi                         0.0893        0.0855        0.0814
sexmale                    -0.9512       -0.8967       -0.8496
ipaqinsufficiently active  -0.1222       -0.1155       -0.1103
ipaqactive                 -0.2560       -0.2424       -0.2305
Theta coefficients (shape model with log link):
                     tau = 0.2500  tau = 0.5000  tau = 0.7500
(theta)_(intercept)  -1.8344       -1.8357       -1.8372
(theta)_bmi           0.0051        0.0022        -0.0001
```

where $\mathrm{logit}(\mu_i) = \beta_0 + \beta_1\,\mathrm{age}_i + \beta_2\,\mathrm{bmi}_i + \beta_3\,\mathrm{sex}_i + \beta_4\,\mathrm{ipaqinsufficientlyactive}_i + \beta_5\,\mathrm{ipaqactive}_i$ and $\log(\theta_i) = \delta_0 + \delta_1 \mathrm{bmi}_i$. For interpretation of estimated regression coefficients, we suggest see [65] Ch. 7]. Note that `plot(fitubs)` works as the `plot.rqs` function of the `quantreg` package. For an overview of all functions available in `unitquantreg`, use `ls(''package:unitquantreg'')`. The `unitquantreg` function is being restructured to fit models with responses augmented by zeros, ones, or zeros and ones, as in [88].

Table 4
Summary table of univariate statistics for COVID-19 data set.

Variable	Mean	SD	Min	Max	Q1	Q2	Q3	Skew	Kurt
RR	0.975	0.032	0.820	0.999	0.968	0.988	0.997	-2.158	5.220
RR(30)	0.997	0.004	0.981	1.000	0.997	0.998	0.999	-2.454	5.203
RR(90)	0.974	0.030	0.862	0.999	0.971	0.985	0.992	-2.087	4.064
RR(180)	0.955	0.036	0.820	0.994	0.940	0.966	0.978	-1.648	2.906
PD	203.901	265.612	1.286	1,215.198	44.809	107.784	219.942	2.211	4.485
GINI	0.452	0.018	0.419	0.499	0.440	0.453	0.466	0.134	-0.481
BEDS	2.600	0.710	1.600	4.800	2.100	2.450	3.100	0.969	0.564
SR	0.173	0.035	0.089	0.260	0.149	0.172	0.193	0.274	-0.111
PR	0.132	0.028	0.076	0.201	0.107	0.132	0.151	0.455	-0.388
LE	78.696	1.785	74.800	82.300	77.800	79.100	79.900	-0.483	-0.426

Table 5
Spearman correlation coefficient (with the corresponding p-value under the null hypothesis H_0: $\rho = 0$) for the indicated variables.

Variable	PD	GINI	BEDS	SR	PR	LE
RR	−0.364 (<0.001)	−0.310 (<0.001)	0.031 (0.709)	0.060 (0.466)	−0.002 (0.806)	−0.010 (0.902)
PD		0.549 (<0.001)	−0.242 (0.003)	−0.295 (<0.001)	−0.076 (0.357)	0.182 (0.026)
GINI			0.033 (0.686)	0.061 (0.455)	0.519 (<0.001)	−0.106 (0.198)
BEDS				0.666 (<0.001)	0.308 (<0.001)	−0.518 (<0.001)
SR					0.628 (<0.001)	−0.889 (<0.001)
PR						−0.678 (<0.001)

It is important to emphasize that the objective of this and next subsection is not to present all numerous approaches to variable selection, regression diagnostics, link function selection or parameter interpretation, but rather suggest the use of the `unitquatreg` package for quantile regression.

5.4. Biomedical application II with COVID-19 recovery rates in the United States

In this example, we consider the data set extracted from [108], available at https://github.com/tatianefribeiro/RUBXII_Regression_COVID-19/tree/master. Different from [108], we consider the recovery rate (RR) (1-mortality rates) across the 50 US states as the response variable and the following model is fitted:

$$\mathrm{logit}(\mu_i) = \beta_0 + \beta_1\,\mathrm{PD}_i + \beta_2\,\mathrm{GINI}_i + \beta_3\,\mathrm{BEDS}_i + \beta_4\,\mathrm{SR}_i + \beta_5\mathrm{PR}_i + \beta_6\,\mathrm{LE}_i + \beta_7\,\mathrm{T90} + \beta_8\,\mathrm{T180}_i, \tag{5.62}$$

where PD is the population density (p/mi^2) in 2020; GINI is the Gini coefficient in 2017; BEDS is the hospital beds per 100 thousand inhabitants in 2018; SR is the smoking rate by state in 2020; PR is the poverty rate in 2020; LE is the life expectancy in 2018; T90 is a dummy variable that is equal to one if the response corresponds to recovery rate after 90 days of the 10th confirmed case, and zero otherwise; whereas T180 is a dummy variable that is equal to one if the response corresponds to recovery rate after 180 days of the 10th confirmed case, and zero otherwise.

Table 4 reports the descriptive measures for continuous variables and for RR measured in the three periods (30, 90, and 180 days). Table 5 presents the Spearman correlation coefficients. Figure 23 shows histograms and scatter-plots for the variables RR, PD, GINI, BEDS, SR, PR, and LE. Note that the response has a leptokurtic, asymmetric empirical distribution between 0.820 and 0.999, which can be well modeled by several members of the family of models proposed in our R package. Also, on the one hand, observe that the response variable is only correlated statistically at a significant level of 1% to the covariates PD and GINI, so that the other covariates could be discarded from the model. On the other hand, some covariates present significant correlation that could indicate multicollinearity problems, which is analyzed when the regression models are stated.

We fit all available models simultaneously, for $\tau = 0.5$, as follows:

```
> models <- c("ashw", "johnsonsb", "kum", "leeg", "ubs", "uburrxii", "uchen",
              "ughne", "ughnx", "ugompertz", "ugumbel", "ulogistic", "uweibull")

> fitscovid  <- lapply(1:13, function(i) unitquantreg(RR ~ PD + GINI + BEDS +
                 SR + PR + LE + T90 + T180, family = models[i], link = "logit",
                 link.theta = "log", tau = 0.5, data = covid))
```

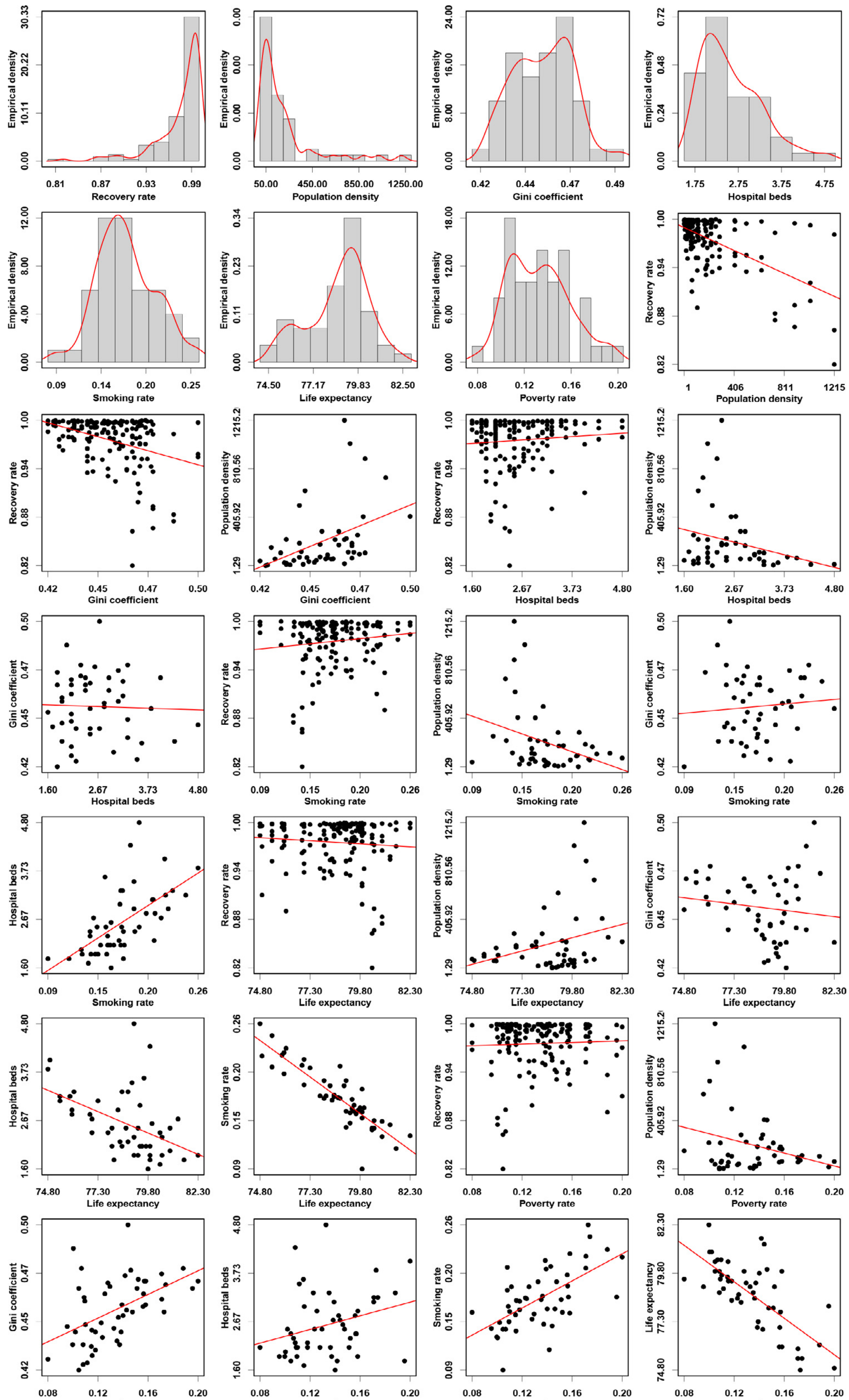

Fig. 23. Histograms and scatter-plots for COVID-19 data set.

ML estimates, SEs of their corresponding estimators for the parameters, and the p-values of the associated tests, can be obtained using the following instructions:

```
> lapply(fitscovid, function(x) round(rbind(mle = coef(x), se = sqrt(diag(vcov(x))),
  p.value = summary(x)$coeftable[,4]), 3))
$ashw
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle          12.640 -0.002 -15.385 0.195 -7.259 0.597 0.016 -1.907 -2.571      1.126
se            8.409  0.000   5.426 0.111  4.900 4.644 0.104  0.138  0.142      0.062
p.value       0.133  0.000   0.005 0.078  0.139 0.898 0.875  0.000  0.000      0.000

$johnsonsb
        (Intercept)     PD    GINI  BEDS    SR    PR    LE    T90   T180 log(theta)
mle          -2.609 -0.002 -17.179 0.040 2.229 6.124 0.199 -2.000 -2.833      0.321
se            7.993  0.000   5.654 0.108 4.919 4.251 0.096  0.148  0.148      0.058
p.value       0.744  0.000   0.002 0.715 0.650 0.150 0.039  0.000  0.000      0.000

$kum
        (Intercept)     PD    GINI   BEDS    SR     PR    LE    T90   T180 log(theta)
mle          -6.101 -0.002 -19.044 -0.182 7.253 11.745 0.234 -1.845 -2.275      4.428
se            4.185  0.000   2.284  0.092 3.325  2.292 0.052  0.194  0.187      0.092
p.value       0.145  0.000   0.000  0.046 0.029  0.000 0.000  0.000  0.000      0.000
$leeg
        (Intercept)     PD    GINI   BEDS    SR    PR    LE    T90   T180 log(theta)
mle          -3.900 -0.002 -19.499 -0.074 5.816 8.403 0.219 -2.285 -2.724      4.460
se            7.032  0.000   3.560  0.131 5.228 3.429 0.084  0.212  0.205      0.117
p.value       0.579  0.000   0.000  0.575 0.266 0.014 0.009  0.000  0.000      0.000

$ubs
        (Intercept)     PD    GINI  BEDS    SR    PR    LE    T90   T180 log(theta)
mle          -7.154 -0.002 -17.267 0.056 0.563 9.176 0.256 -2.019 -2.891     -0.223
se            8.280  0.000   5.923 0.117 5.006 4.362 0.099  0.145  0.153      0.058
p.value       0.388  0.000   0.004 0.634 0.910 0.035 0.010  0.000  0.000      0.000

$uburrxii
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle          12.591 -0.002 -15.207 0.196 -7.288 0.569 0.016 -1.899 -2.563      0.443
se            8.432  0.000   5.451 0.110  4.904 4.673 0.105  0.137  0.142      0.062
p.value       0.135  0.000   0.005 0.076  0.137 0.903 0.879  0.000  0.000      0.000

$uchen
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle          12.525 -0.002 -15.281 0.194 -7.202 0.584 0.017 -1.902 -2.566      0.434
se            8.464  0.000   5.456 0.111  4.933 4.672 0.105  0.138  0.143      0.063
p.value       0.139  0.000   0.005 0.080  0.144 0.900 0.870  0.000  0.000      0.000

$ughne
        (Intercept)     PD    GINI  BEDS      SR     PR     LE    T90   T180 log(theta)
mle          17.455 -0.002 -14.542 0.251 -10.927 -1.044 -0.042 -1.840 -2.456      0.173
se            7.976  0.000   5.250 0.112   4.465  4.942  0.100  0.133  0.142      0.066
p.value       0.029  0.000   0.006 0.025   0.014  0.833  0.675  0.000  0.000      0.009

$ughnx
        (Intercept)     PD    GINI   BEDS    SR     PR    LE    T90   T180 log(theta)
mle         -28.798 -0.002 -23.392 -0.222 8.718 25.610 0.533 -2.190 -3.208     -0.037
se            6.936  0.000   6.264  0.119 3.929  5.006 0.081  0.153  0.154      0.066
p.value       0.000  0.000   0.000  0.061 0.027  0.000 0.000  0.000  0.000      0.573

$ugompertz
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle           9.963 -0.002 -19.182 0.118 -3.551 0.283 0.070 -2.235 -2.827      3.402
se            8.678  0.000   4.520 0.129  5.401 4.135 0.103  0.183  0.164      0.160
p.value       0.251  0.000   0.000 0.363  0.511 0.945 0.495  0.000  0.000      0.000

$ugumbel
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle          12.535 -0.002 -15.336 0.194 -7.167 0.538 0.017 -1.909 -2.573      0.429
se            8.412  0.000   5.410 0.111  4.914 4.632 0.104  0.138  0.142      0.062
p.value       0.136  0.000   0.005 0.080  0.145 0.908 0.867  0.000  0.000      0.000

$ulogistic
        (Intercept)     PD    GINI  BEDS    SR    PR    LE    T90   T180 log(theta)
mle           0.468 -0.002 -16.621 0.042 2.344 5.274 0.157 -2.032 -2.825      0.919
se            7.934  0.000   5.395 0.102 5.138 3.940 0.096  0.147  0.140      0.068
p.value       0.953  0.000   0.002 0.683 0.648 0.181 0.102  0.000  0.000      0.000

$uweibull
        (Intercept)     PD    GINI  BEDS     SR    PR    LE    T90   T180 log(theta)
mle          12.563 -0.002 -15.246 0.195 -7.245 0.580 0.017 -1.899 -2.564      0.438
se            8.450  0.000   5.457 0.111  4.920 4.673 0.105  0.138  0.142      0.062
p.value       0.137  0.000   0.005 0.078  0.141 0.901 0.875  0.000  0.000      0.000
```

From the above results reported for all the models, differently to what happened with the Brazilian data set, the results obtained in terms of significance at 5% for the COVID-19 data set are reported to be somewhat distinct among the models. Then, we conduct an

Table 6
Coefficient of determination (R^2) for COVID-19 data set.

Distribution	R^2
unit-Gompertz	0.8029
unit-logistic	0.7984
Johnson-SB	0.7786
unit-generalized half-normal-X	0.7722
unit-Gumbel	0.7709
arc-secant hyperbolic Weibull	0.7702
unit-Chen	0.7702
unit-Weibull	0.7689
unit-Burr-XII	0.7680
unit-generalized half-normal-E	0.7644
unit-Birnbaum-Saunders	0.7616
log-extended exponential-geometric	0.7570
Kumaraswamy	0.7372

analysis for model selection. By using `likelihood_stats(lt = fits)`, we have the values of the likelihood-based statistics given by:

```
Likelihood-based statistics of fit for unit quantile regression models
Call:  likelihood_stats(lt = fits)
                                   Neg2LogLike AIC       BIC       HQIC
arc-secant hyperbolic Weibull      -1050.780   -1030.780 -1000.674 -1018.549
Johnson-SB                         -1055.734   -1035.734 -1005.628 -1023.503
Kumaraswamy                        -1026.769   -1006.769 -976.663  -994.5380
log-extended exponential-geometric -1044.394   -1024.394 -994.287  -1012.162
unit-Birnbaum-Saunders             -1040.872   -1020.872 -990.766  -1008.641
unit-Burr-XII                      -1049.539   -1029.539 -999.432  -1017.307
unit-Chen                          -1050.365   -1030.365 -1000.259 -1018.134
unit-generalized half-normal-E     -1040.730   -1020.730 -990.623  -1008.498
unit-generalized half-normal-X     -995.2650   -975.2650 -945.159  -963.0340
unit-Gompertz                      -1051.371   -1031.371 -1001.264 -1019.139
unit-Gumbel                        -1051.210   -1031.210 -1001.103 -1018.978
unit-logistic                      -1061.644   -1041.644 -1011.538 -1029.413
unit-Weibull                       -1049.946   -1029.946 -999.840  -1017.715
```

which indicates that the unit-logistic quantile regression is the best model according to each of the likelihood-based statistics. This model selection analysis is supported by a residual graphical study shown in Figure 24. Once again, we can consider the measure R^2 as a criterion for comparison between all models. Table 6 reports these results attributing the best fit (highest value) to the unit-Gompertz quantile regression.

To fit a full unit-logistic model, we have:

```
> fitcovidulog <- unitquantreg(RR ~  PD + GINI + BEDS + SR + PR + LE + T90 + T180,
                               family = "ulogistic",
                               link = "logit", link.theta = "log", tau = 0.5,
                               data =  covid)
```

From Table 5, we can observe that the correlation coefficient between LE and SR is high. Because correlation coefficients only show pairwise correlations, once again, we use the VIF to assess what covariates are collinear and should be dropped before starting the analyses. The resulting VIF values are given below:

```
> car::vif(fitcovidulog)
      PD      GINI     BEDS       SR       PR       LE      T90     T180
2.136667 2.551950 1.740564 6.988683 4.176044 6.791217 1.423759 1.419650
```

As mentioned, a VIF value that exceeds 10 indicates a problematic potentially amount of collinearity. In this example, the VIF score for the predictor variables SR and LE are VIF = 6.989 and VIF = 6.791, respectively; so that we must pay attention on these covariates. To find a set of covariates that does not contain collinearity, we remove one variable at a time, recalculate the VIF values, and repeat this process until all VIF values are enough small, for example, less than five to be sure no collinearity problems could be present. Then, we consider the following scenarios for restricted (reduced) unit-logistic quantile regression models:

- Reduced unit-logistic model 1 [excluding the covariate SR]

```
reduced.1 <- unitquantreg(RR ~  PD + GINI + BEDS +  PR + LE + T90 + T180,
             family = "ulogistic", link = "logit", link.theta = "log",
             tau = 0.5, data =  covid)
```

- Reduced unit-logistic model 2 [excluding the covariate LE]

```
> reduced.2 <- unitquantreg(RR ~  PD + GINI + BEDS + SR + PR + T90 + T180,
                            family = "ulogistic", link = "logit", link.theta = "log",
                            tau = 0.5, data =  covid)
```

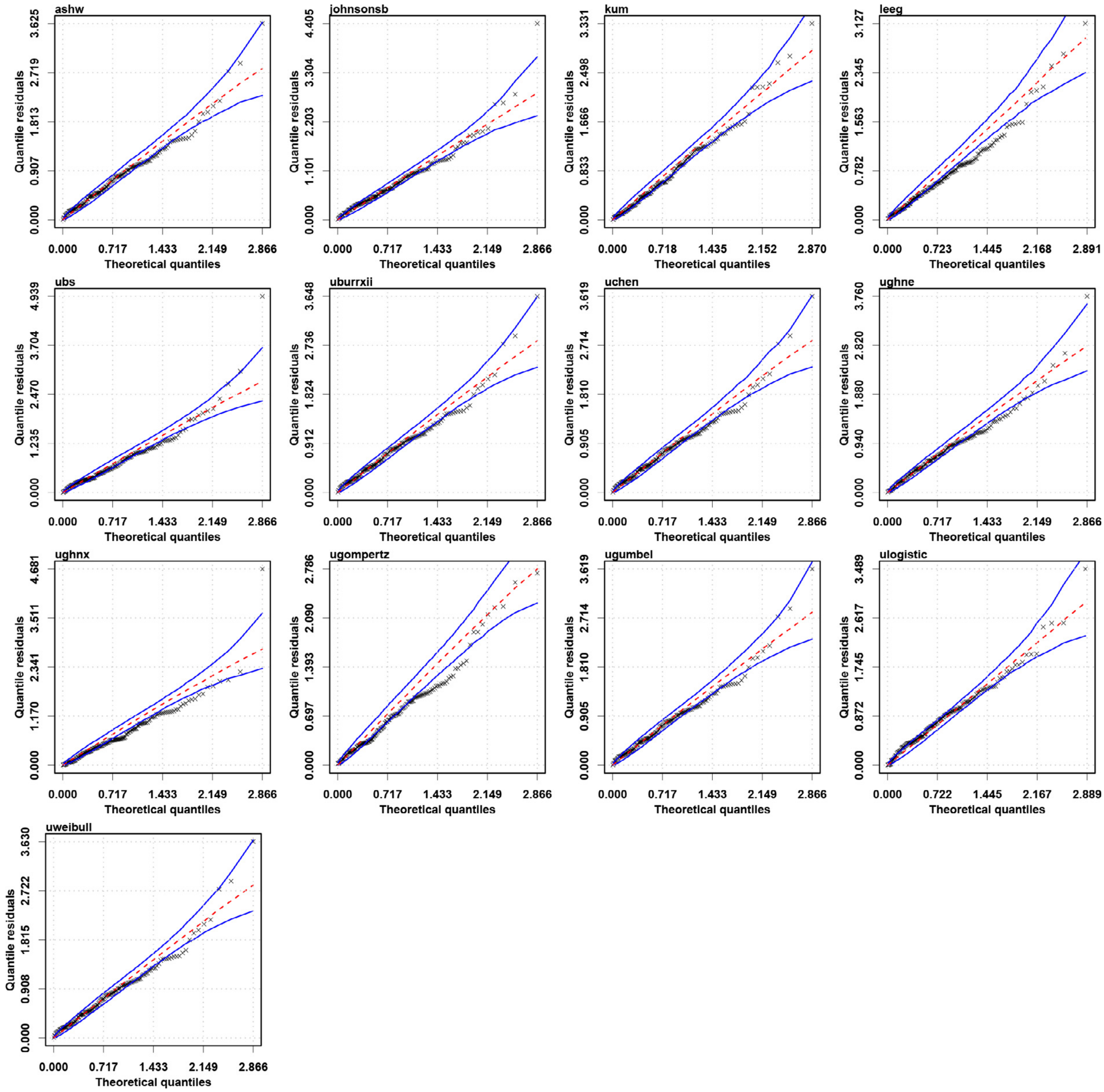

Fig. 24. QQ (half-normal) plots with a simulated envelope of quantile residuals for COVID-19 data set.

- Reduced unit-logistic model 3 [excluding the covariates SR and LE]

```
> reduced.3 <- unitquantreg(RR ~  PD + GINI + BEDS  + PR + T90 + T180,
                 family = "ulogistic", link = "logit", link.theta = "log",
                 tau = 0.5, data =  covid)
```

The VIF values for these models are given, respectively, by

```
> car::vif(reduced.1)
      PD     GINI     BEDS       PR       LE      T90     T180
2.068567 2.516064 1.183617 4.169128 2.701867 1.393607 1.374571

> car::vif(reduced.2)
      PD     GINI     BEDS       SR       PR      T90     T180
2.016583 2.253713 1.637897 2.888653 3.165383 1.433448 1.422845

> car::vif(reduced.3)
      PD     GINI     BEDS       PR      T90     T180
2.005546 2.224827 1.120871 2.155601 1.401250 1.379976
```

Note that all reduced models (1-3) have small VIF. Now, we can compare the fit of the full unit-logistic model with the reduced models through the likelihood ratio test as follows:

```
> lmtest::lrtest(fitcovidulog, reduced.1)
Likelihood ratio test
Model 1: RR ~ PD + GINI + BEDS + SR + PR + LE + T90 + T180 | 1
Model 2: RR ~ PD + GINI + BEDS + PR + LE + T90 + T180 | 1
  #DF LogLik DF  Chisq Pr(>Chisq)
1  10 530.82
2   9 530.64 -1 0.3604     0.5483

> lmtest::lrtest(fitcovidulog, reduced.2)
Likelihood ratio test
Model 1: RR ~ PD + GINI + BEDS + SR + PR + LE + T90 + T180 | 1
Model 2: RR ~ PD + GINI + BEDS + +SR + PR + T90 + T180 | 1
  #DF LogLik DF Chisq Pr(>Chisq)
1  10 530.82
2   9 528.90 -1  3.85    0.04975*

> lmtest::lrtest(full, reduced.3)
Likelihood ratio test
Model 1: RR ~ PD + GINI + BEDS + SR + PR + LE + T90 + T180 | 1
Model 2: RR ~ PD + GINI + BEDS + PR + T90 + T180 | 1
  #DF LogLik DF  Chisq Pr(>Chisq)
1  10 530.82
2   8 527.98 -2 5.6765    0.05853
```

From these results, we can conclude that there is no practically significant difference at 5% between the four models. We decide to use Model 3 due to the principle of parsimony. Then, we apply the exponential function to the 95% confidence limits obtaining:

```
> exp((confint(reduced.3)))
            lower limit  upper limit
(intercept) 3.541654e+03 1.509526e+07
PD          9.975539e-01 9.986627e-01
GINI        1.152474e-10 1.183067e-01
BEDS        8.529337e-01 1.194588e+00
PR          1.174452e-03 6.828848e+01
T90         9.837382e-02 1.748500e-01
T180        4.492931e-02 7.748723e-02
log(theta)  2.162510e+00 2.824666e+00
```

The estimates of the parameters of Model 3, for $\tau = 0.5$, are given by:

```
> summary(reduced.3)
Wald-tests for unit-logistic quantile regression model
Call:  unitquantreg(formula = RR ~ PD + GINI + BEDS + PR + T90 + T180,
    data = covid, tau = 0.5, family = "ulogistic", link = "logit",
    link.theta = "log")
Mu coefficients: (quantile model with logit link and tau = 0.5):
             Estimate  SE          z-value Pr(Z>|z|)
(intercept)  1.235e+01  2.132e+00   5.793 6.91e-09 ***
PD          -1.894e-03  2.834e-04  -6.682 2.35e-11 ***
GINI        -1.251e+01  5.293e+00  -2.363   0.0181 *
BEDS         9.364e-03  8.594e-02   0.109   0.9132
PR          -1.262e+00  2.799e+00  -0.451   0.6521
T90         -2.031e+00  1.467e-01 -13.845  < 2e-16 ***
T180        -2.830e+00  1.390e-01 -20.355  < 2e-16 ***
---
Model with constant shape:
           Estimate    SE        z-value Pr(Z>|z|)
log(theta)  0.90483    0.06814   13.28   <2e-16 ***
---
Residual degrees of freedom: 142
Log-likelihood: 527.9838
Number of iterations: 73
```

Note that the final model is consequent with the conjectures from the exploratory data analysis which indicated that, at a significant level of 1%, only the covariates PD, GINI, T90 and T180 should be present in the model. Therefore, the fit of the new model built with only the significant variables is given by:

```
 Wald-tests for unit-Logistic quantile regression model

Call:  unitquantreg(formula = RR ~ PD + GINI + T90 + T180, data = covid,
                    tau = 0.5, family = "ulogistic",
                    link = "logit", link.theta = "log")

Mu coefficients: (quantile model with logit link and tau = 0.5):
              Estimate  SE         z-value Pr(Z>|z|)
(Intercept)  1.278e+01  1.743e+00   7.336  2.2e-13 ***
PD          -1.830e-03  2.172e-04  -8.424  < 2e-16 ***
GINI        -1.381e+01  3.908e+00  -3.533  0.00041 ***
T90         -2.030e+00  1.460e-01 -13.899  < 2e-16 ***
T180        -2.828e+00  1.389e-01 -20.359  < 2e-16 ***
---
Signif. codes:  0 '***' < 0.001   '**' < 0.01   '*' < 0.05   '.' < 0.1   'n.s.' < 1

Model with constant shape:
            Estimate   SE        z-value Pr(Z>|z|)
log(theta)  0.90479    0.06814   13.28   <2e-16 ***
---
Signif. codes:  0 '***' < 0.001   '**' < 0.01   '*' < 0.05   '.' < 0.1   'n.s.' < 1

Residual degrees of freedom: 144
Log-likelihood: 527.8803
Number of iterations: 79
```

For the case with all covariates, the value $\widehat{\beta}_0$ indicates the rate of recoveries when all covariates were null. The estimate for the parameter β_3 reports that hospital beds have a positive effect on the rate of recoveries, that is, in US states with a large number of hospital beds, the rate of recoveries is higher. In contrast, the estimates for the parameters β_1, β_2 and β_4 have a negative effect on the rate of recoveries, indicating that this rate is less for US states with higher population density, Gini index and poverty rate. Likewise, $\widehat{\beta}_5$ and $\widehat{\beta}_6$ are negative, with $\widehat{\beta}_6$ being less than $\widehat{\beta}_5$, which indicates that the time after the 10th confirmed case is detrimental to the rate of recovery.

6. Concluding remarks

The quantile regression methodology provides a framework for modeling the relationship between an outcome or response variable and explanatory variables or covariates using conditional quantile functions. This methodology not only offers a more robust alternative to estimate the central tendency of the response but also allows a more detailed exploration of its conditional distribution for different quantiles. Applications of quantile regression arose in many research areas, ranging from ecology over genetics to economics [13], but only in the last decade have been works that investigate a parametric approach.

This paper presented a new computational package implemented in the R software, two biomedical applications, one of them with COVID-19 data, and an up-to-date review of the parametric quantile regression models obtained re-parameterizing a distribution in terms of a quantile. We described the main characteristics of several distributions used to model continuous variables bounded to the unit interval (based on the exponentiated arcsech-normal, generalized half-normal, generalized Johnson SB, Johnson-Student-t, Lambert-uniform, log-extended exponential-geometric, power Johnson SB, Kumaraswamy, L-logistic, transmuted unit-Rayleigh, unit-Birnbaum-Saunders, unit-Bur-XII, unit-Chen, unit-Gompertz, unit-half-normal, unit-Weibull, and Vasicek distributions), four for non-negative continuous responses (based on the Birnbaum-Saunders, flexible Weibull, logistic Nadarajah-Haghighi, and log-symmetric families), and one for discrete responses (based on the discrete generalized half-normal distribution).

For the distributions on the unit interval that are more flexible in terms of the behavior of its probability density function, an R package is available, mainly for parameter estimation and model checking. We showed how to apply the methods and functions contained in the package through two applications in biomedical data. Future versions of the package will focus on extending the support interval to include zero-inflation, one-inflation or zero-one-inflation quantile regression. Please note that that the computational implementation of zero-or-one augmented is straightforward since the likelihood function factorizes in two terms: one depending on the discrete component and another one depending on continuous component. Thus, in the R software, one can use the `stats::glm()` function to estimate the discrete component and the `uniquantreg::uniquantreg()` function to estimate the continuous component. In the next version of the `uniquantreg` package, we are planning to add a wrapper function to estimate and infer in augmented unit quantile regression models.

Author Statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, and/or revision of the article.

Statement of Ethical Approval

Not applicable.

Disclosure of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Editors and Reviewers for their constructive comments on an earlier version of this article which resulted in this improved version. The research of B. Alves was partially funded by the Coordenação de Aperfeiçoamento Pessoal de Nível Superior –Brazil (CAPES) – Finance Code 001. The research of V. Leiva was partially funded by FONDECYT, project grant number 1200525, from the National Agency for Research and Development (ANID) of the Chilean government, under the Ministry of Science and Technology, Knowledge and Innovation.

Appendix: Comparison with non-parametric quantile regressions

Although outside the scope of this paper, in this appendix, we present the parameter estimates, for both data sets considered in the applications (Brazilian and COVID-19 data sets), considering the standard quantile regression model introduced in [57]. These estimates are obtained using the `rq` function available in the `quantreg` package of R. Once the response variable is on the (0,1) interval, the fit is carried out considering its logit transformation; see, for example, [14,127]. Table A.7 reports the parameter estimates considering the Brazilian body fat data set. For the data set related to COVID-19, estimates are in Table A.8. From these results, in general, no major differences are observed between the estimates obtained by the parametric and non-parametric methodologies. Note that the coefficients of determination R^2 for the standard quantile regression model introduced in [56] were of

Table A1
Parameter estimates for Brazilian body fat data set.

	Coefficients					
Distribution	Intercept	age	bmi	sexmale	ipaqinsufficiently active	ipaqactive
ashw	-0.476	0.005	0.082	-0.898	-0.125	-0.332
johnsonsb	-0.470	0.005	0.092	-0.938	-0.117	-0.263
kum	-0.530	0.004	0.082	-0.820	-0.076	-0.216
leeg	-0.450	0.005	0.093	-0.957	-0.131	-0.257
ubs	-0.485	0.004	0.086	-0.895	-0.115	-0.242
uburrxii	-0.506	0.004	0.053	-0.565	-0.083	-0.160
uchen	-0.283	0.007	0.114	-1.138	-0.138	-0.387
ughne	-0.481	0.005	0.072	-0.833	-0.135	-0.382
ughnx	-0.504	0.004	0.080	-0.799	-0.098	-0.227
ugompertz	-0.362	0.006	0.104	-1.057	-0.148	-0.385
ugumbel	-0.438	0.006	0.092	-0.970	-0.130	-0.354
ulogistic	-0.475	0.005	0.089	-0.932	-0.122	-0.239
uweibull	-0.494	0.005	0.077	-0.863	-0.121	-0.320
non-parametric	-0.469	0.005	0.083	-0.950	-0.153	-0.220

Table A2
Parameter estimates for COVID-19 data set.

Distribution	Coefficients								
	Intercept	PD	GINI	BEDS	SR	PR	LE	T90	T180
ashw	12.640	-0.002	-15.385	0.195	-7.259	0.597	0.016	-1.907	-2.571
johnsonsb	-2.609	-0.002	-17.179	0.040	2.229	6.124	0.199	-2.000	-2.833
kum	-6.101	-0.002	-19.044	-0.182	7.253	11.745	0.234	-1.845	-2.275
leeg	-3.900	-0.002	-19.499	-0.074	5.816	8.403	0.219	-2.285	-2.724
ubs	-7.154	-0.002	-17.267	0.056	0.563	9.176	0.256	-2.019	-2.891
uburrxii	12.591	-0.002	-15.207	0.196	-7.288	0.569	0.016	-1.899	-2.563
uchen	12.525	-0.002	-15.281	0.194	-7.202	0.584	0.017	-1.902	-2.566
ughne	17.455	-0.002	-14.542	0.251	-10.927	-1.044	-0.042	-1.840	-2.456
ughnx	-28.798	-0.002	-23.392	-0.222	8.718	25.610	0.533	-2.190	-3.208
ugompertz	9.963	-0.002	-19.182	0.118	-3.551	0.283	0.070	-2.235	-2.827
ugumbel	12.535	-0.002	-15.336	0.194	-7.167	0.538	0.017	-1.909	-2.573
ulogistic	0.468	-0.002	-16.621	0.042	2.344	5.274	0.157	-2.032	-2.825
uweibull	12.563	-0.002	-15.246	0.195	-7.245	0.580	0.017	-1.899	-2.564
non-parametric	13.171	-0.002	-14.829	0.154	-3.711	3.527	-0.001	-2.229	-2.859

78.8% and 81.3%, respectively, when employing the Brazilian and COVID-19 data sets, respectively, which are slightly greater than the corresponding maximal values of the parametric quantile regressions, that is, 77.7% (unit-logistic model) and 77.5% (unit-Birnbaum-Saunders model) for the Brazilian body fat data set; and 80.3% (unit-Gompertz model) and 79.8% (unit-logistic model) for the COVID-19 data set.

References

[1] K. Adamidis, S. Loukas, A lifetime distribution with decreasing failure rate, Stat. Probab. Lett. 39 (1998) 35–42.
[2] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (1974) 716–723.
[3] C. Azevedo, V. Leiva, E. Athayde, N. Balakrishnan, Shape and change point analyses of the Birnbaum-Saunders-T hazard rate and associated estimation, Comput. Stat. Data Anal. 56 (2012) 3887–3897.
[4] M.Z. Anis, D. De, An expository note on unit-Gompertz distribution with applications, Statistica 80 (2020) 469–490.
[5] H.S. Bakouch, A.S. Nik, A. Asgharzadeh, H.S. Salinas, A flexible probability model for proportion data: Unit-half-normal distribution, Commun. Stat. Case Stud. Data Anal. Appl. 7 (2021) 271–288.
[6] N. Balakrishnan, Handbook of the Logistic Distribution, Marcel Dekker, New York, USA, 1992.
[7] R.A.R. Bantan, C. Chesneau, F. Jamal, M. Elgarhy, M.H. Tahir, A. Ali, M. Zubair, S. Anam, Some new facts about the unit-Rayleigh distribution with applications, Mathematics 8 (2020) 1954.
[8] D.M. Bates, W.N. Venables, splines: Regression spline functions and classes, 2015, R package version 4.0.4.
[9] F.M. Bayer, F. Cribari-Neto, J. Santos, Inflated Kumaraswamy regressions with application to water supply and sanitation in Brazil, Stat. Neerl. 75 (2021) 453–481.
[10] C.L. Bayes, J.L. Bazán, M. De Castro, A quantile parametric mixed regression model for bounded response variables, Stat. Interface 10 (2017) 483–493.
[11] M. Bebbington, C.D. Lai, R. Zitikis, A flexible Weibull extension, Reliab. Eng. Syst. Saf. 92 (2007) 719–726.
[12] T.R.B. Benedetti, P.d.C. Antunes, C.R. Rodriguez-Añez, G.Z. Mazo, E.L. Petroski, Reproducibility and validity of the international physical activity questionnaire (IPAQ) in elderly men, Rev. Bras. Med. Esporte 13 (2007) 11–16.
[13] D.F. Benoit, D. Van den Poel, Bayesqr: a bayesian approach to quantile regression, J. Stat. Softw. 76 (2017) 1–32.
[14] M. Bottai, B. Cai, R.E. McKeown, Logistic quantile regression for bounded outcomes, Stat. Med. 29 (2010) 309–317.
[15] M. Bourguignon, J. Leão, V. Leiva, M. Santos-Neto, The transmuted Birnbaum-Saunders distribution, REVSTAT Stat. J. 15 (2017) 601–628.
[16] I.W. Burr, Cumulative frequency functions, Ann. Math. Stat. 13 (1942) 215–232.
[17] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference, Springer, Fort Collins, CO, USA, 2002.
[18] V.G. Cancho, J.L. Bazán, D.K. Dey, A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer, Stat. Methods Med. Res. 29 (2020) 2015–2033.
[19] Z. Chen, A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, Stat. Probab. Lett. 49 (2000) 155–161.
[20] G. Claeskens, N. Hjort, Model Selection and Model Averaging, Cambridge University Press, Cambridge, UK, 2008.
[21] K. Cooray, M.M. Ananda, A generalization of the half-normal distribution with applications to lifetime data, Commun. Stat. Theory Methods 37 (2008) 1323–1337.
[22] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, On the lambertw function, Adv Comput Math 5 (1996) 329–359.
[23] L. Couri, R. Ospina, G. da Silva, V. Leiva, J. Figueroa-Zúñiga, A study on computational algorithms in the estimation of parameters for a class of beta regression models, Mathematics 10 (2022) 299.
[24] D.R. Cox, E.J. Snell, A general definition of residuals, J. R. Stat. Soc. B 30 (1968) 248–265.
[25] D.R. Cox, E.J. Snell, The Analysis of Binary Data, Chapman and Hall, London, UK, 1989.
[26] C. Daniel, Use of half-normal plots in interpreting factorial two-level experiments, Technometrics 1 (1959) 311–341.
[27] A. Dasilva, R. Dias, V. Leiva, C. Marchant, H. Saulo, Birnbaum-Saunders regression models: a comparative evaluation of three approaches, J. Stat. Comput. Simul. 90 (2020) 2552–2570.
[28] H. de la Fuente-Mella, R. Rubilar, K. Chahuan-Jimenez, V. Leiva, Modeling COVID-19 cases statistically and evaluating their effect on the economy of countries, Mathematics 9 (2021) 1558.
[29] P.K. Dunn, G.K. Smyth, Generalized Linear Models with Examples in R, Springer, New York, USA, 2018.
[30] D. Eddelbuettel, J.J. Balamuta, Extending R with C++: A brief introduction to Rcpp, Am. Stat. 72 (2018) 28–36.
[31] D. Eddelbuettel, R. Francois, Rcpp: seamless R and C++ integration, J. Stat. Softw. 40 (2011) 1–18.
[32] S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, J.Appl. Stat. 31 (2004) 799–815.
[33] D. Firth, Bias reduction of maximum likelihood estimates, Biometrika 80 (1993) 27–38.
[34] D.I. Gallardo, E. Gómez-Déniz, H.W. Gómez, Discrete generalized half-normal distribution with applications in quantile regression, Stat. Oper. Res. Trans. 44 (2020) 265–284.
[35] F. Garcia-Papani, V. Leiva, M.A. Uribe-Opazo, R.G. Aykroyd, Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data, Chemom. Intell. Lab. Syst. 177 (2018) 114–128.
[36] I. Gijbels, R. Karim, A. Verhasselt, Semiparametric quantile regression using family of quantile-based asymmetric densities, Comput. Stat. Data Anal. 157 (2021) 107–129.
[37] Y.M. Gómez, E. Gómez-Déniz, O. Venegas, D.I. Gallardo, H.W. Gómez, An asymmetric bimodal distribution with application to quantile regression, Symmetry 11 (2019) 899.
[38] E. Gómez-Déniz, F.J. Vázquez-Polo, V. García-García, A discrete version of the half-normal distribution and its generalization with applications, Stat. Pap. 55 (2014) 497–511.
[39] R.R. Guerra, F.A. Peña Ramírez, M. Bourguignon, The unit extended Weibull families of distributions and its applications, J. Appl. Stat. 48 (2021) 3174–3192.
[40] E.J. Gumbel, The return period of flood flows, Ann. Math. Stat. 12 (1941) 163–190.
[41] R.D. Gupta, R.C. Gupta, Analyzing skewed data by power normal model, Test 17 (2008) 197–210.
[42] R.D. Gupta, D. Kundu, Generalized exponential distributions, Aust. N. Z.. Stat. 41 (1999) 173–188.
[43] S. Hamedi-Shahraki, A. Rasekhi, M.S. Yekaninejad, M.R. Eshraghian, A.H. Pakpour, Kumaraswamy regression modeling for bounded outcome scores, Pak. J. Stat. Oper. Res. 17 (2021) 79–88.
[44] E.J. Hannan, B.J. Quinn, The determination of the order of an autoregression, J. R. Stat. Soc. B 41 (1979) 190–195.
[45] F.E. Harrell, Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, Springer, Nashville, USA, 2015.
[46] Q. Huang, H. Zhang, J. Chen, M. He, Quantile regression models and their applications: a review, J. Biomet. Biostat. 8 (2017) 2155–6180.

[47] A.I. Iliev, A. Rahnev, N. Kyurkchiev, S. Markov, A study on the unit-logistic, unit-Weibull and Topp-Leone cumulative sigmoids, Biomath Commun. 6 (2019) 1–15.
[48] Y.A. Iriarte, M. de Castro, H.W. Gómez, The Lambert-F distributions class: an alternative family for positive data analysis, Mathematics 8 (2020) 1398.
[49] Y.A. Iriarte, M. de Castro, H.W. Gómez, An alternative one-parameter distribution for bounded data modeling generated from the Lambert transformation, Symmetry 13 (2021) 1190.
[50] M.K. Jha, S. Dey, R.M. Alotaibi, Y.M. Tripathi, Reliability estimation of a multicomponent stress-strength model for unit-Gompertz distribution under progressive type II censoring, Qual. Reliab. Eng. Int. 36 (2020) 965–987.
[51] M.K. Jha, S. Dey, Y.M. Tripathi, Reliability estimation in a multicomponent stress-strength based on unit-Gompertz distribution, Int. J. Qual. Reliab. Manag. 37 (2019) 428–450.
[52] P. Jodrá, M.D. Jiménez-Gamero, A quantile regression model for bounded responses based on the exponential-geometric distribution, REVSTAT Stat. J. 18 (2020) 415–436.
[53] J. Fox, S. Weisberg, An R Companion to Applied Regression, Sage, Thousand Oaks, CA, USA, 2019.
[54] T. Kecojević, Bootstrap inference for parametric quantile regression, Faculty of Engineering and Physical Sciences, University of Manchester, Manchester, UK, 2011 Ph.D. thesis.
[55] M.S. Khan, R. King, I.L. Hudson, Transmuted Kumaraswamy distribution, Stat. Trans. 17 (2016) 183–210.
[56] R. Koenker, quantreg: quantile regression, 2021, https://CRAN.R-project.org/package=quantreg. R package version 5.86.
[57] R. Koenker, G. Bassett, Regression quantiles, Econometrica 46 (1978) 33–50.
[58] R. Koenker, J.A.F. Machado, Goodness of fit and related inference processes for quantile regression, J. Am. Stat. Assoc. 94 (1999) 1296–1310.
[59] M.C. Korkmaz, A. Emrah, C. Chesneau, H.M. Yousof, On the unit-Chen distribution with associated quantile regression and applications, Math. Slov. (2022) in press.
[60] M.C. Korkmaz, The unit generalized half-normal distribution: a new bounded distribution with inference and application, Sci. Bull. 82 (2022) 133–140.
[61] M.C. Korkmaz, C. Chesneau, On the unit Burr-XII distribution with the quantile regression modeling and applications, Comput. Appl. Math. 40 (2021) 29.
[62] M.C. Korkmaz, C. Chesneau, Z.S. Korkmaz, On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications, Symmetry 13 (2021) 117.
[63] M.C. Korkmaz, C. Chesneau, Z.S. Korkmaz, Transmuted unit Rayleigh quantile regression model: alternative to beta and Kumaraswamy quantile regression models, Sci. Bull. 83 (2021) 149–159.
[64] M.C. Korkmaz, C. Chesneau, Z.S. Korkmaz, A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries, J. Appl. Stat. (2021) 1–25.
[65] O. Korosteleva, Advanced Regression Models with SAS and R, CRC Press, Boca Raton, FL, USA, 2019.
[66] D. Kumar, S. Dey, E. Ormoz, S. MirMostafaee, Inference for the unit-Gompertz model based on record values and inter-record times with an application, Rend. Circ. Mat. Palermo Ser. 2 69 (2020) 1295–1319.
[67] P. Kumaraswamy, A generalized probability density function for double-bounded random processes, J. Hydrol. 46 (1980) 79–88.
[68] J. Leão, V. Leiva, H. Saulo, V. Tomazella, Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data, Stat. Med. 37 (2018) 4421–4440.
[69] V. Leiva, The Birnbaum-Saunders Distribution, Academic Press, New York, USA, 2016.
[70] V. Leiva, L. Sánchez, M. Galea, H. Saulo, Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression, Stoch. Environ. Res. Risk Assess. 34 (2020) 1457–1471.
[71] V. Leiva, M. Santos-Neto, F.J.A. Cysneiros, M. Barros, Birnbaum–Saunders statistical modelling: a new approach, Stat. Model. 14 (2014) 21–48.
[72] V. Leiva, R.A. Santos, H. Saulo, C. Marchant, Y. Lio, Bootstrap control charts for quantiles based on log-symmetric distributions with applications to monitoring of reliability data, Qual. Reliab. Eng. Int. (2022). in press
[73] C. Marchant, V. Leiva, F.J.A. Cysneiros, A multivariate log-linear model for birnbaum-saunders distributions, IEEE Trans. Reliab 65 (2016) 816–827.
[74] J. Mazucheli, B. Alves, Ugomquantreg: quantile regression modeling for unit-Gompertz responses, 2021a, R package version 1.0.0.
[75] J. Mazucheli, B. Alves, Vasicekreg: regression modeling using Vasicek distribution, 2021b, R package version 1.0.1.
[76] J. Mazucheli, S.R. Bapat, A.F.B. Menezes, A new one-parameter unit-lindley distribution, Chil. J. Stat. 11 (2020) 53–67.
[77] J. Mazucheli, V. Leiva, B. Alves, A.F.B. Menezes, A new quantile regression for modeling bounded data under a unit Birnbaum-Saunders distribution with applications in medicine and politics, Symmetry 13 (2021) 682.
[78] J. Mazucheli, B. Alves, The unit-gumbel quantile regression model for proportion data, Working Paper (2022).
[79] J. Mazucheli, B. Alves, M.C. Korkmaz, The unit-gompertz quantile regression model for bounded responses, Math. Slov. (2022) in press.
[80] J. Mazucheli, B. Alves, M.C. Korkmaz, V. Leiva, Vasicek quantile and mean regression models for bounded data: new formulation, mathematical derivations, and numerical applications, Mathematics 10 (2022) 1289.
[81] J. Mazucheli, A.F.B. Menezes, unitBSQuantReg: unit-Birnbaum-Saunders quantile regression, 2020, https://github.com/AndrMenezes/unitBSQuantReg. R package version 0.1.0.
[82] J. Mazucheli, A.F.B. Menezes, S. Dey, The unit-Birnbaum-Saunders distribution with applications, Chil. J. Stat. 9 (2018) 47–57.
[83] J. Mazucheli, A.F.B. Menezes, S. Dey, Unit Gompertz distribution with applications, Statistica 79 (2019) 25–43.
[84] J. Mazucheli, A.F.B. Menezes, L.B. Fernandes, R.P. de Oliveira, M.E. Ghitany, The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, J. Appl. Stat. 47 (2020) 954–974.
[85] J. Mazucheli, A.F.B. Menezes, M.E. Ghitany, The unit-Weibull distribution and associated inference, J. Appl. Probab. Stat. 13 (2018) 1–22.
[86] J. Mazucheli, R.P. Oliveira, J.C. Cardoso, A distribuição half-normal generalizada discreta: uma distribuição alternativa para a análise de dados de contagem, Ciên. e Nat. 41 (2019) 1–11.
[87] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman and Hall, London, UK, 1989.
[88] A.F.B. Menezes, Uwquantreg: unit-Weibull quantile regression, 2020, https://github.com/AndrMenezes/uwquantreg. R package version 0.1.0.
[89] A.F.B. Menezes, J. Mazucheli, M. Bourguignon, A parametric quantile regression approach for modelling zero-or-one inflated double bounded data, Biomet. J. 63 (2021) 841–858.
[90] A.F.B. Menezes, J. Mazucheli, F. Alqallaf, M.E. Ghitany, Bias-corrected maximum likelihood estimators of the parameters of the unit-weibull distribution, Austrian J. Stat. 50 (2021) 41–53.
[91] A.F.B. Menezes, J. Mazucheli, S. Chakraborty, A collection of parametric modal regression models for bounded data, J. Biopharm. Stat. 31 (2021) 490–506.
[92] P.A. Mitnik, S. Baek, The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation, Stat. Pap. 54 (2013) 177–192.
[93] R.A. Moral, J. Hinde, C.G.B. Demétrio, Half-normal plots and overdispersed models in R: the hnp package, J. Stat. Softw. 81 (2017) 1–23.
[94] S. Nadarajah, F. Haghighi, An extension of the exponential distribution, Statistics 45 (2011) 543–558.
[95] N.J.D. Nagelkerke, A note on a general definition of the coefficient of determination, Biometrika 78 (1991) 691–692.
[96] J.C. Nash, R. Varadhan, Unifying optimization algorithms to aid software system users: optimx for R, J. Stat. Softw. 43 (2011) 1–14.
[97] A. Noufaily, Parametric Quantile Regression Based on the Generalised Gamma Distribution, The Open University, 2011. Ph.D. thesis
[98] A. Noufaily, M.C. Jones, Parametric quantile regression based on the generalized gamma distribution, J. R. Stat. Soc. C 62 (2013) 723–740.
[99] R.F. Paz, N. Balakrishnan, J.L. Bazán, L-logistic regression models: prior sensitivity analysis, robustness to outliers and applications, Braz. J. Probab. Stat. 33 (2019) 455–479.
[100] F.A. Peña Ramírez, R.R. Guerra, D.R. Canterle, G.M. Cordeiro, The logistic Nadarajah–Haghighi distribution and its associated regression model for reliability applications, Reliab. Eng. Syst. Saf. 204 (2020) 1–13.
[101] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, M. Schmid, A review of spline function procedures in R, BMC Med. Res. Methodol. 19 (2019) 1–16.
[102] R.R. Petterle, W.H. Bonat, C.T. Scarpin, T. Jonasson, V.Z.C. Borba, Multivariate quasi-Beta regression models for continuous bounded data, Int. J. Biostat. 17 (2020) 39–53.
[103] F. Prataviera, Reparameterized flexible Weibull distribution with some applications, Am. J. Math. Manag. Sci. (2021). in press
[104] R. Puentes, C. Marchant, V. Leiva, J. Figueroa-Zúñiga, F. Ruggeri, Predicting PM2.5 and PM10 levels during critical episodes management in Santiago, Chile, with a bivariate Birnbaum-Saunders log-linear model, Mathematics 9 (2021) 645.
[105] G. Pumi, C. Rauber, F.M. Bayer, Kumaraswamy regression model with Aranda-Ordaz link function, Test 29 (2020) 1051–1071.
[106] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. https://www.R-project.org/.
[107] J.B. Ramsey, Tests for specification errors in classical linear least-squares regression analysis, J. R. Stat. Soc. B 31 (1969) 350–371.
[108] T.F. Ribeiro, G.M. Cordeiro, F.A. Peña Ramírez, R.R. Guerra, A new quantile regression for the COVID-19 mortality rates in the United States, Comput. Appl. Math. 40 (2021) 255.
[109] B.M. Robert, G.R. Brindha, B. Santhi, G. Kanimozhi, N.R. Prasad, Computational models for predicting anticancer drug efficacy: a multi linear regression analysis based on molecular, cellular and clinical data of oral squamous cell carcinoma cohort, Comput. Methods Programs Biomed. 178 (2019) 105–112.
[110] L. Sánchez, V. Leiva, M. Galea, H. Saulo, Birnbaum-Saunders quantile regression and its diagnostics with application to economic data, Appl. Stoch. Models Bus. Ind. 37 (2020) 53–73.
[111] L. Sánchez, V. Leiva, M. Galea, H. Saulo, Birnbaum-Saunders quantile regression models with application to spatial data, Mathematics 8 (2020) 1000.
[112] L. Sánchez, V. Leiva, C. Marchant, H. Saulo, J.M. Sarabia, A new quantile regression model and its diagnostic analytics for a Weibull distributed response with applications, Mathematics 9 (2022) 2768.
[113] SAS, SAS/STAT® 14.1 Users Guide, SAS Institute, Cary, NC, 2015.
[114] H. Saulo, A. Dasilva, V. Leiva, L. Sánchez, H. de la Fuente-Mella, Log-symmetric quantile regression models, Stat. Neerl. 76 (2022) 124–163.

[115] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[116] W.T. Shaw, I.R.C. Buckley, The alchemy of probability distributions: Beyond Gram-Charlier expansions, and a Skew-kurtotic normal distribution from a rank transmutation map, 2009, ArXiv:0901.0434.
[117] M. Smithson, Y. Shou, CDF-quantile distributions for modelling random variables on the unit interval, Br. J. Math. Stat. Psychol. 70 (2017) 412–438.
[118] M. Smithson, J. Verkuilen, A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, Psychol. Methods 11 (2006) 54–71.
[119] P.X.K. Song, M. Tan, Marginal models for longitudinal continuous proportional data, Biometrics 56 (2000) 496–502.
[120] P.R. Tadikamalla, N.L. Johnson, Systems of frequency curves generated by transformations of logistic variables, Biometrika 69 (1982) 461–465.
[121] M.H. Tahir, G.M. Cordeiro, A. Alzaatreh, M. Mansoor, M. Zubair, The logistic-X family of distributions and its applications, Commun. Stat. Theory Methods 45 (2016) 7326–7349.
[122] L.H. Vanegas, G.A. Paula, Log-symmetric distributions: Statistical properties and parameter estimation, Br. J. Probab. Stat. 30 (2016) 196–220.
[123] O.A. Vasicek, The distribution of loan portfolio value, Risk 15 (2002) 160–162.
[124] M. Ventura, H. Saulo, V. Leiva, S. Monsueto, Log-symmetric regression models: Information criteria, application to movie business and industry data with economic implications, Appl. Stoch. Models Bus. Ind. 35 (2019) 963–977.
[125] Q.H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57 (1989) 307–333.
[126] W. Weibull, A statistical distribution function of wide applicability, J. Appl. Mech. 18 (1951) 293–297.
[127] V. Wong, Logistic Quantile Regression to Evaluate Bounded Outcomes, Stockholm University, 2018. Bachelor Thesis, Mathematical Statistics
[128] S. Wood, mgcv: Mixed GAM computation vehicle with automatic smoothness estimation, 2021, R package version 4.0.5.
[129] K. Yu, R.A. Moyeed, Bayesian quantile regression, Stat. Probab. Lett. 54 (2001) 437–447.
[130] Z. Zeileis, T. Hothorn, Diagnostic checking in regression relationships, R J. 2 (2002) 7–10.
[131] A. Zeileis, Y. Croissant, Extended model formulas in R: multiple parts and multiple responses, J. Stat. Softw. 34 (2010) 1–13.
[132] P. Zhang, Z. Qiu, C. Shi, Simplexreg: An R package for regression analysis of proportional data using the simplex distribution, J. Stat. Softw. 71 (2016) 1–21.