Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (2) (2025), 633 – 655
DOI : 10.15672/hujms.1533205

Research Article

The unit-Cauchy quantile regression model with variates observed on (0, 1): percentages, proportions, and fractions

Talha Arslan[*,1,2], Keming Yu[2]

[1] *Department of Econometrics, Van Van Yüzüncü Yıl University, 65080 Van, Türkiye.*
[2] *Department of Mathematics, Brunel University London, UB83PH London, United Kingdom.*

Abstract

In this study, a new parametric quantile regression model is introduced as an alternative to the beta regression and Kumaraswamy quantile regression model. The proposed quantile regression model is obtained by reparametrization of the unit-Cauchy distribution in terms of its quantiles. The model parameters are estimated using the maximum likelihood method. A Monte-Carlo simulation study is conducted to show the efficiency of the maximum likelihood estimation of the model parameters. The implementation of the proposed quantile regression model is shown by using real datasets. Quantile regression models based on unit-Weibull, unit generalized half normal, and unit Burr XII are also considered in the applications. The application results show that the proposed quantile regression model is preferable over its rivals when several comparison criteria are taken into account. In addition, the fitting plots indicate that the proposed quantile regression model fits extreme observations on the right tail better than its strong rivals, which is important in quantile regression modeling.

Mathematics Subject Classification (2020). 60E05, 60E10, 62F10, 62J12, 62J20

Keywords. Maximum likelihood, Monte-Carlo simulation, parametric model, quantile regression, unit-Cauchy

1. Introduction

Researchers from different scientific areas usually aim to explore phenomena by using the data extracted from them. In this context, statistical distributions are widely used in modeling data from different fields. The normal distribution is traditionally the most famous one used for modeling real-valued data. In addition to the wide usage of the normal distribution, it has a deficiency in modeling data having atypical observations. In this regard, distributions having heavy tail(s) are used for modeling purposes, and Cauchy distribution is one of the well-known ones. Cauchy distribution looks similar to a normal distribution, i.e., its probability density function (pdf) is symmetric around the center. However, it has much heavier tails than the normal distribution. The basic theory and

*Corresponding Author.
Email addresses: mstalhaarslan@yyu.edu.tr (T. Arslan), keming.yu@brunel.co.uk (K. Yu)
Received: 14.08.2024; Accepted: 15.02.2025

conditions of the Cauchy distribution are not given here for brevity; see [14, 15, 17] for details.

Real-valued distributions, such as normal, Cauchy, Laplace, and some skew extensions, cannot model all kinds of data. For example, specific experiments (different indices, rates, etc.) or observations from real life (infant mortality rate, human development index, etc.) may have a bounded range on the unit interval $(0,1)$ and should be modeled unit distribution. The power, beta, Johnson [9], Topp-Leone [37], Kumaraswamy (Kum) [16] distributions are the well-known early examples of the unit distributions. Later, new unit distributions have been introduced; for example, Mazucheli et al. [19] and Korkmaz [13] proposed unit-Weibull (UW) and unit-generalized half normal (UGHN) distributions, respectively. Recently, Arslan [1] proposed a general definition for a family of unit distributions which includes the unit-Cauchy distribution (UC) having the pdf

$$f_Z(z;\alpha,\sigma) = \frac{\alpha}{\pi\sigma}\left(z\left(1-z^{\alpha}\right)\right)^{-1}\left(1+\left(-\frac{1}{\sigma}\log\left(z^{-\alpha}-1\right)\right)^2\right)^{-1} \tag{1.1}$$

and the cumulative distribution function (cdf)

$$F_Z(z;\alpha,\sigma) = 0.5 + \frac{1}{\pi}\arctan\left(-\frac{1}{\sigma}\log\left(z^{-\alpha}-1\right)\right). \tag{1.2}$$

Here, $z \in (0,1)$, $\alpha > 0$, and $\sigma > 0$. The cdf of the UC distribution, given in (1.2), can be readily inverted to yield the quantile function (qf) of the UC distribution

$$\kappa(p) = F_Z^{-1}(p;\alpha,\sigma) = \left(1+\exp\left(-\sigma\tan\left(\pi(p-0.5)\right)\right)\right)^{-\frac{1}{\alpha}}, \quad p \in (0,1). \tag{1.3}$$

Note that the UC distribution given in (1.1) is an extended version of the unit distribution given in [30]; see also [31].

Unit distributions are essential in the construction of regression models when the response variable takes a value between 0 and 1, that is, the response variable follows a unit support distribution. In this context, the beta regression model, proposed by [10] and [6], is one of the first attempts in which the mean response is considered as a function of covariates. It is known that the mean of the data is not robust against outliers; therefore, it cannot represent the measure of central tendency efficiently when the response variable includes anomalies. In such a case, the beta regression model, which is a mean conditional model, cannot perform well as expected. In this regard, Ribeiro and Ferrari [35] also pointed out that the maximum likelihood (ML) estimate of the beta regression model is affected by outliers since scores equations of the corresponding parameters are not bounded; therefore, they derived a robust estimation method to rehabilitate the beta regression model in the context of robustness. Also note that the qf of the beta distribution cannot explicitly be formulated; thus, quantile-based reparametrization, e.g., median-based parametrization of it becomes intractable, which is another deficiency of the beta regression model. Therefore, the UC distribution, given in (1.1), has an advantage over the beta distribution, since its qf, given in (1.3), has a simple form that allows a quantile-based analysis of the UC distribution. In addition, a median conditional regression model based on the UC distribution can be readily obtained, and the resulting quantile regression model will be more robust against outliers than beta regression.

In the literature, there are also different regression models alternative to beta regression. In this context, Mitnik and Baek [24] proposed the Kum quantile regression model for a bounded response to eliminate the deficiency of the beta regression model. In addition, new quantile regression models based on the odd log-logistic unit omega, unit Lindley, UW, UGHN, exponential-geometric and unit Burr XII (UBXII) distributions were introduced by [3, 20, 21, 23, 28, 34] respectively. See [22] for a comprehensive overview of parametric quantile regression models.

Based on the conclusion given above, the UC distribution can be a strong alternative to the beta distribution in not only modeling univariate data but also the context of regression model. To the best of authors' knowledge, beside the robustness property of Cauchy distribution, its unit case has not been considered in the context of quantile regression model. In this regard, the motivation for this study comes from introducing a new quantile regression model due to the attractive properties of the UC distribution and the manageable form of its qf, which allows the reparametrization of the pdf given in (1.1) in terms of quantiles. Therefore, in this study, the UC distribution is reparametrized by its quantiles, and the resulting distribution is called RUC. Then, a new quantile regression model is introduced based on the RUC distribution in which the model parameters are estimated using the ML method. In addition, a Monte Carlo simulation study is conducted to show the efficiencies of the ML estimation of the model parameters. Not least of all, three real datasets are used to show the implementation of the proposed quantile regression model.

The paper is structured as follows. In Section 2, the RUC distribution is defined, its parameters are named, characteristic measures of it are obtained based on its qf, and an estimation of its parameters is provided. In Section 3, the new quantile regression model is derived, say the RUC quantile regression model, and ML inference is investigated. Section 4 is reserved for model diagnostic criteria. The application of the RUC quantile regression model is provided in Section 5. The paper ends with some concluding remarks.

2. The RUC distribution

Let $Z \sim UC(\alpha, \sigma)$ and $\kappa(p; \alpha, \sigma)$, the p -th quantile of Z, be denoted as κ to simplify the notation. Then, the following equation

$$\alpha = \frac{\log\left(1 + \exp\left(-\sigma \tan\left(\pi\left(p - 0.5\right)\right)\right)\right)}{\log\left(\frac{1}{\kappa}\right)} \tag{2.1}$$

is obtained after some straightforward algebraic manipulation. By using the parametrization in (2.1), the pdf of the RUC distribution is expressed as

$$\begin{aligned} f_Z(z;\kappa,\sigma) =& \frac{1}{\pi\sigma}\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\log\left(1+\exp\left(-\sigma\tan\left(\pi\left(p-0.5\right)\right)\right)\right) \\ & \left(z\left(1-z^{\left(\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\log(1+\exp(-\sigma\tan(\pi(p-0.5))))\right)}\right)\right)^{-1} \\ & \times\left(1+\left(-\frac{1}{\sigma}\log\left(z^{-\left(\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\log(1+\exp(-\sigma\tan(\pi(p-0.5))))\right)}-1\right)\right)^2\right)^{-1}, \end{aligned} \tag{2.2}$$

where $\kappa \in (0, 1)$ denotes the quantile parameter (also can be called location parameter), $\sigma > 0$, and $p \in (0, 1)$ is quantile order and assumed to be known.

When $p = 0.5$, the pdf given in (2.2) is reduced to be a median-parametrized RUC distribution, and its pdf and cdf turn out to be, respectively,

$$\begin{aligned} f_Z(z;\kappa,\sigma) =& \frac{1}{\pi\sigma}\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\left(z\left(1-z^{\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}}\right)\right)^{-1} \\ & \left(1+\left(-\frac{1}{\sigma}\log\left(z^{-\left(\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\right)}-1\right)\right)^2\right)^{-1} \end{aligned} \tag{2.3}$$

and

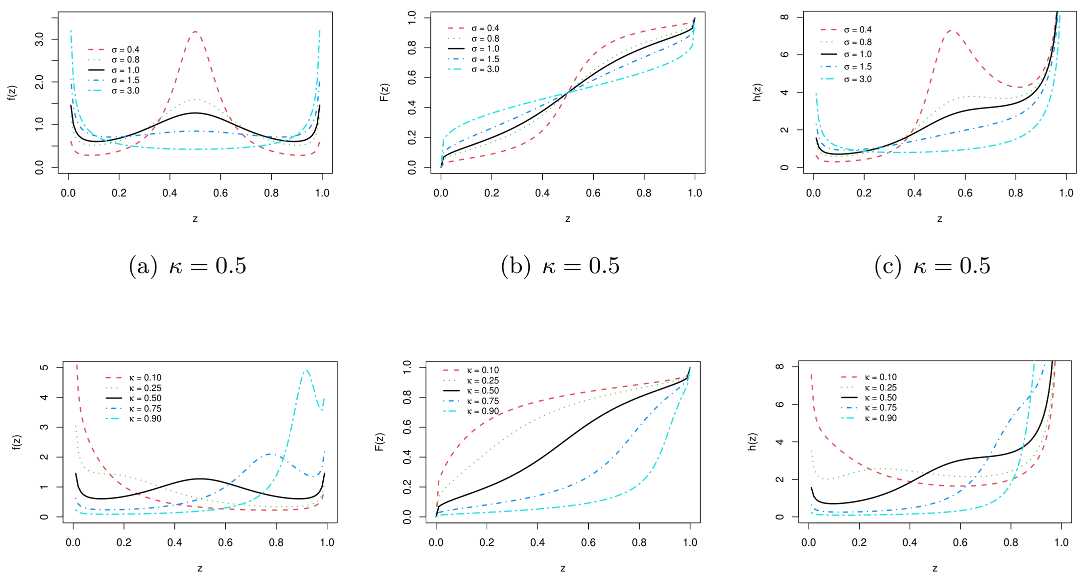
$$F_Z(z;\kappa,\sigma) = 0.5 + \frac{1}{\pi}\arctan\left(-\frac{1}{\sigma}\log\left(z^{-\left(\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\right)}-1\right)\right). \tag{2.4}$$

Also, by using the well-known definition of hazard rate function (hrf), the hrf of the RUC distribution is

$$h_Z(z;\kappa,\sigma) = \frac{f_Z(z;\kappa,\sigma)}{1-F_Z(z;\kappa,\sigma)},$$

where $f_Z(\cdot)$ and $F_Z(\cdot)$ are the pdf and cdf of the RUC distribution, respectively. In the following, a random variable Z having pdf in (2.3) is denoted by $Z \sim \text{RUC}(\kappa,\sigma)$.

The plots for the pdf, cdf, and hrf of the RUC distribution are provided in column 1, column 2, and column 3, respectively, in Figure 1 for illustration.



(a) $\kappa = 0.5$ (b) $\kappa = 0.5$ (c) $\kappa = 0.5$

(d) $\sigma = 1.0$ (e) $\sigma = 1.0$ (f) $\sigma = 1.0$

Figure 1. The plots for the pdf, cdf, and hrf of the RUC distribution for different parameter settings.

It could be seen from Figure 1(a) and Figure 1(d) that the density plots of the RUC distribution can be symmetric, skewed left or right based on the different values of the parameters σ and κ. The RUC density plots have J, reversed J, U and bell shapes as well. Also, the plots for the hrf of the RUC distribution have the forms J, monotonically increasing, decreasing-increasing(bathtub), increasing-decreasing-increasing (tilde) for the particular values of the parameters; see Figure 1(c) and Figure 1(f).

Remark 2.1. The support space of the pdf in (2.3) can be extended from $(0,1)$ to (u,l) using the transformation $Y = l + (u-l)Z$, where l and u denote the lower and upper limits which come from the nature of the random variable Y; that is, l and u are assumed to be known. Thus, the pdf of the Y is

$$\begin{aligned} f_Y(y;l,u,\kappa,\sigma) =& \frac{1}{\pi\sigma}\frac{\log(2)}{(u-l)}\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\left(\left(\frac{y-l}{u-l}\right)\left(1-\left(\frac{y-l}{u-l}\right)^{\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}}\right)\right)^{-1} \\ & \left(1+\left(-\frac{1}{\sigma}\log\left(\left(\frac{y-l}{u-l}\right)^{-\left(\log(2)\left(\log\left(\frac{1}{\kappa}\right)\right)^{-1}\right)}-1\right)\right)^2\right)^{-1}. \end{aligned}$$

Proposition 2.2. *The pdf of RUC distribution in* (2.3) *is symmetric when* $\kappa = 0.5$.

Proof. To show that the corresponding pdf is symmetric when $\kappa = 0.5$ whatever the value of parameter σ, is sufficient to show that $f_Z(0.5 - z; 0.5, \sigma) = f_Z(0.5 + z; 0.5, \sigma)$. Using the pdf given in (2.3), it is easy to put forth that

$$\begin{aligned} f_Z(0.5 - z; 0.5, \sigma) =& \frac{1}{\pi\sigma}\left((0.5 - z)(0.5 + z)\right)^{-1}\left(1 + \left(-\frac{1}{\sigma}\left(\log(0.5 + z) - \log(0.5 - z)\right)\right)^2\right)^{-1} \\ =& f_Z(0.5 + z; 0.5, \sigma). \end{aligned}$$

□

2.1. Quantile-based analysis

The quantile-based analysis of the RUC distribution can be carried out by using its qf

$$\kappa(p) = F_Z^{-1}(p; \kappa, \sigma) = \left(1 + \exp\left(-\sigma \tan\left(\pi(p - 0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}}, \quad p \in (0, 1). \tag{2.5}$$

Remark 2.3. Let κ_p represent the p-th quantile of the sample, e.g., $\kappa_{0.5}$ denotes the sample median. Then, if $p = 0.5$ in (2.5), $2^{\frac{\log(\kappa)}{\log(2)}} = \kappa_{0.5}$, which means that the median of the RUC distribution depends only on the parameter κ as expected; therefore, the parameter κ can be estimated by using the sample's median via straightforward calculation, that is, $\tilde{\kappa} = \kappa_{0.5}$. Also, the parameter σ can be estimated by using $\kappa(p)$ and its conterparts in the sample κ_p; one of the following equations $\tilde{\sigma} = -\log\left((\kappa_{0.75})^{\frac{\log(2)}{\log(\tilde{\kappa})}} - 1\right)$, $\tilde{\sigma} = \log\left((\kappa_{0.25})^{\frac{\log(2)}{\log(\tilde{\kappa})}} - 1\right)$ or $\tilde{\sigma} = \frac{\log(2)}{\log(\tilde{\kappa})}\log\left(\kappa_{0.75}/\kappa_{0.25}\right)$ can be used to estimate the parameter σ.

The qf of the RUC distribution, given in (2.5), has a simple form; therefore, a quantile-based analysis of the RUC distribution can be easily performed. For example, the inter-quantile range (IQR) of the RUC distribution

$$IQR = \kappa(0.75) - \kappa(0.25) = \left(\exp\left(\sigma\frac{\log(1/\kappa)}{\log(2)}\right) - 1\right)\left(1 + \exp(\sigma)\right)^{\frac{\log(\kappa)}{\log(2)}}$$

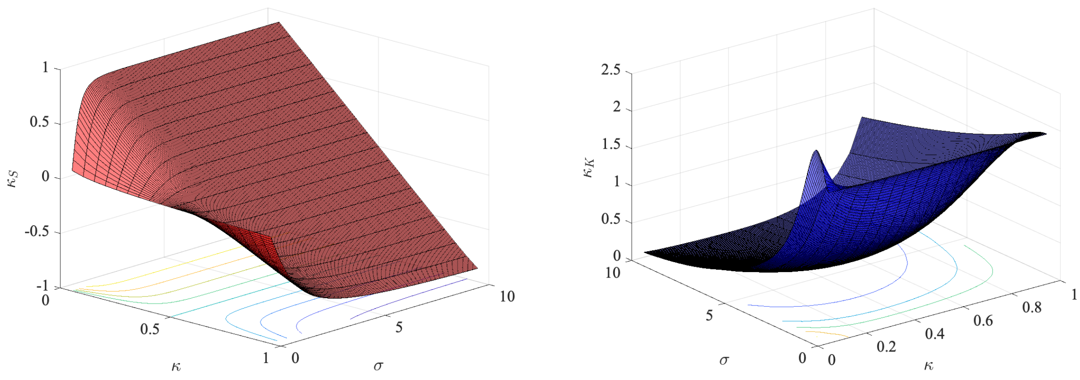
can be used to identify outliers in the data. Octile skewness measure ($-1 \leq \kappa_S \leq 1$) proposed by [2], which is a member of the class of skewness measures introduced by [8], of the RUC distribution

$$\kappa_S = \frac{(\kappa(0.875) - \kappa(0.5)) - (\kappa(0.5) - \kappa(0.125))}{\kappa(0.875) - \kappa(0.125)}$$

can be used to detect asymmetry in the data. Also, Moors's [25] kurtosis measure (κ_K) of the RUC distribution is

$$\kappa_K = \frac{\kappa(0.875) - \kappa(0.625) + \kappa(0.375) - \kappa(0.125)}{\kappa(0.75) - \kappa(0.25)}.$$

The surface plots for the κ_S and κ_K of the RUC distribution are given, for an illustrative purpose, in Figure 2(a)-2(b), respectively.



(a) surface plot for the κ_S (b) surface plot for the κ_K

Figure 2. The surface plots for the κ_S and κ_K of the RUC distribution.

In Figure 2, it is clear that the pdf of the RUC distribution can be right skewed, left skewed and symmetric when $\kappa < 0.5$, $\kappa > 0.5$, and $\kappa = 0.5$, respectively. The RUC distribution can be used to model data that have various values of skewness and kurtosis, which motivates practitioners to use it for statistical modeling.

Definition 2.4. Let X be a random variable having qf F_X^{-1}. Then the quantile spread, introduced by Townsend and Colonius [38], of X is

$$QS_X(p) = F_X^{-1}(1-p) - F_X^{-1}(p),$$

for $0 < p < 0.5$. If X and Y are random variables with quantile spreads $QS_X(p)$ and $QS_Y(p)$, respectively, and $QS_X(p) \leq QS_Y(p)$, for all $0 < p < 0.5$, then X is called smaller than Y in quantile spread order and it is denoted by $X \leq_{QS} Y$.

Proposition 2.5. *If $X \sim RUC(\kappa, \sigma_1)$ and $Y \sim RUC(\kappa, \sigma_2)$ then X is smaller than Y in the quantile spread order ($X \leq_{QS} Y$) if and only if $\sigma_2 \leq \sigma_1$. See Appendix-A for the proof.*

Corollary 2.6. *Mitnik and Baek [24] used quantile spread order to identify the dispersion parameter of a distribution; therefore, by using the following arguments in study of [24] and the result given in Proposition 2.5, it can be said that the σ is a dispersion parameter of the RUC distribution.*

Remark 2.7. Random variates of the RUC distribution, having the pdf given in (2.3), can be generated via the equation

$$z = \left(1 + \exp\left(-\sigma \tan\left(\pi(u - 0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \tag{2.6}$$

where u is generated from uniform$(0, 1)$.

3. Quantile regression

Quantile regression, which has been popular since the paper of [12], is utilized to determine the conditional distribution of a response variable given the values of the covariates. See [3] for the characteristics and main advantages of quantile regression. As stated in [21], there exist several approaches, such as the nonparametric approach, based on a pseudo-likelihood through an asymmetric Laplace distribution, and the parametric approach based on ML inference, in modeling quantiles conditional on covariates. In this context, Noufaily and Jones [27] considered the generalized gamma distribution to obtain a fully parametric approach to quantile regression for the response variable in $\mathbb{R}^+$. See [11, 40–42] in which detailed information about quantile regression is presented.

In this study, the new quantile regression model is formulated by using the pdf of RUC distribution, given in (2.3), for bounded response variable. Therefore, this paper can be classified into papers that use a fully parametric approach.

3.1. The RUC quantile regression model

Let $\boldsymbol{z} = (z_1, z_2, \cdots, z_n)^\top$ be a vector of n independent observations of the variables $Z_i \sim RUC(\kappa_i, \sigma)$ for $(i = 1, 2, \cdots, n)$. Then, the proposed quantile regression model is defined as

$$g(\kappa_i) = \boldsymbol{x_i}^\top \boldsymbol{\xi} = \eta_i$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3, \cdots, \xi_k)^\top$, $\boldsymbol{\xi} \in \mathbb{R}^k$, is the parameter vector associated with the covariates $\boldsymbol{x_i}^\top = (x_{i1}, x_{i2}, x_{i3}, \cdots, x_{ik})$ which are assumed to be fixed and known; $(k < n)$. Here, $g : (0, 1) \to \mathbb{R}$ is a strictly monotonic and twice differentiable link function that is used the covariates to conditional of the response variable. In the literature, several link functions are used for $g(\cdot)$. In the rest of this paper, the logit link function

$$g(\kappa_i) = \log\left(\frac{\kappa_i}{1-\kappa_i}\right) = \boldsymbol{x_i}^\top \boldsymbol{\xi}$$

is considered due to its useful interpretation of the regression coefficients as an odds ratio. Therefore, the quantities κ_i are

$$\kappa_i = \frac{\exp\left(\boldsymbol{x_i}^\top \boldsymbol{\xi}\right)}{1 + \exp\left(\boldsymbol{x_i}^\top \boldsymbol{\xi}\right)}. \tag{3.1}$$

Note that data may have an asymmetric form and may include unusual observations. In such cases, the median, a more robust measure against the presence of the anomalies in the data than the mean, is usually used as a measure of the location of the data rather than the mean; therefore, median-based regressions are preferable to mean-based ones, e.g., beta regression. Therefore, in this study, the covariates are linked to a conditional median of the response variable, i.e., the pdf given in (2.3) is considered in constructing the proposed quantile regression model.

Remark 3.1. In this study, median conditional quantile regression model is considered. However, quantile regression models with conditional to the other quantiles, such as $0.1, 0.25, 0.75$ or 0.90, can definitely be obtained by using the pdf given in (2.2) and estimated quantile curves associated with the considered quantiles cannot cross as a yield of the parametric approach. This issue and its importance were also emphasized by [27].

3.2. The ML estimation for the RUC quantile regression model

From (2.3), the $\log L$ for $(\boldsymbol{\xi}, \sigma)$ is

$$\log L(\boldsymbol{\xi}, \sigma) = \sum_{i=1}^{n} \log L_i(\kappa_i, \sigma),$$

where

$$\begin{aligned}\log L_i(\kappa_i, \sigma) =& \log(2) - \log \pi - \log\left(\log\left(\frac{1}{\kappa_i}\right)\right) - n\log\sigma - \log(z_i) - \log\left(1 - z_i^{\frac{\log(2)}{\log\left(\frac{1}{\kappa_i}\right)}}\right) \\ & - \log\left(1 + \left(-\frac{1}{\sigma}\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}} - 1\right)\right)^2\right).\end{aligned}$$

After differentiating $\log L_i(\kappa_i,\sigma)$ with respect to the parameters of interest, i.e., κ_i and σ,

$$\begin{aligned}\frac{\partial \log L_i(\kappa_i,\sigma)}{\partial \kappa_i} &= \frac{1}{\kappa_i \log(\kappa_i)} + \frac{\log(2)\log(z_i)}{\kappa_i(\log(\kappa_i))^2\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)} \\ &+ 2\frac{\log(2)\log(z_i) z_i^{\frac{\log(2)}{\log(\kappa_i)}}\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)}{\kappa_i(\log(\kappa_i))^2\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\left(\sigma^2+\left(\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)^2\right)} \\ &= a_i + b_i \dot{z}_i(1+\ddot{z}_i)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log L_i(\kappa_i,\sigma)}{\partial \sigma} &= -\frac{1}{\sigma}+\frac{2}{\sigma}\frac{\left(\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)^2}{\sigma^2+\left(\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)^2} \\ &= -\frac{1}{\sigma}+\frac{2}{\sigma}\left(\ddot{z}_i\left(z_i^{-\frac{\log(2)}{\log(\kappa_i)}}\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)\right)\end{aligned}$$

are obtained. Here,

$$a_i=\frac{1}{\kappa_i\log(\kappa_i)},\quad b_i=\frac{\log(2)}{\kappa_i(\log(\kappa_i))^2},\quad \dot{z}_i=\frac{\log(z_i)}{z^{\frac{\log(2)}{\log(\kappa_i)}}-1},\quad \text{and}\quad \ddot{z}_i=2\frac{z_i^{\frac{\log(2)}{\log(\kappa_i)}}\log\left(z^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)}{\sigma^2+\left(\log\left(z^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)^2}$$

To sum up, the differential total is given by

$$\frac{\partial \log L(\boldsymbol{\xi},\sigma)}{\partial \xi_j}=\sum_{i=1}^{n}\frac{\partial \log L_i(\kappa_i,\sigma)}{\partial\kappa_i}\frac{d\kappa_i}{d\eta_i}\frac{\partial\eta_i}{\partial\xi_j},\quad (j=1,2,3,\cdots,k)$$

and

$$\frac{\partial \log L(\boldsymbol{\xi},\sigma)}{\partial \sigma}=\sum_{i=1}^{n}\frac{\partial \log L_i(\kappa_i,\sigma)}{\partial\sigma}.$$

Note that $\frac{d\kappa_i}{d\eta_i}=\frac{1}{g'(\kappa_i)}=\kappa_i(1-\kappa_i)$ and $\frac{\partial\eta_i}{\partial\xi_j}=x_{ij}$; therefore, the score equations for the parameters of the RUC quantile regression model are

$$\frac{\partial \log L(\boldsymbol{\xi},\sigma)}{\partial \xi_j}=\sum_{i=1}^{n}(a_i+b_i\dot{z}_i(1+\ddot{z}_i))(\kappa_i(1-\kappa_i))x_{ji}=0,\quad (j=1,2,3,\cdots,k)$$

and

$$\frac{\partial \log L(\boldsymbol{\xi},\sigma)}{\partial \sigma}=\sum_{i=1}^{n}s_i=0,$$

where $s_i=\left(\ddot{z}_i\left(z_i^{-\frac{\log(2)}{\log(\kappa_i)}}\log\left(z_i^{\frac{\log(2)}{\log(\kappa_i)}}-1\right)\right)\right)-1$.

As a result, the score vector's components can be expressed in matrix form as

$$U_{\boldsymbol{\xi}}(\boldsymbol{\xi},\sigma)=\boldsymbol{X}^{\top}\boldsymbol{K}\boldsymbol{c}\quad\text{and}\quad U_{\sigma}(\boldsymbol{\xi},\sigma)=\boldsymbol{s}^{\top}\mathbf{1}_n,$$

where $\boldsymbol{X}$ is an $(n \times k)$ matrix whose i-th row is $\boldsymbol{x}_{\boldsymbol{i}}^{\top}$, $\boldsymbol{s} = (s_1, s_2, s_3, \cdots, s_n)^{\top}$, and $\mathbf{1}_n$ is n-dimensional vector of 1's, $\boldsymbol{K} = diag\{\kappa_1(1-\kappa_1), \kappa_2(1-\kappa_2), \kappa_3(1-\kappa_3), \cdots, \kappa_n(1-\kappa_n)\}$, $\boldsymbol{c} = (a_1 + b_1\dot{z}_1(1+\ddot{z}_1), a_2 + b_2\dot{z}_2(1+\ddot{z}_2), a_3 + b_3\dot{z}_3(1+\ddot{z}_3), \cdots, a_n + b_n\dot{z}_n(1+\ddot{z}_n))^{\top}$.

The ML estimates of the model parameters vector, say $\left(\hat{\boldsymbol{\Omega}} = \left(\hat{\boldsymbol{\xi}}^{\top}, \hat{\sigma}\right)\right)$, are simultaneous solutions of the nonlinear system $U_{\boldsymbol{\xi}}(\boldsymbol{\xi}, \sigma) = U_{\boldsymbol{\sigma}}(\boldsymbol{\xi}, \sigma) = 0$ which cannot be obtained explicitly; thus, $\hat{\boldsymbol{\Omega}}$ should be obtained by using numerical optimization methods.

The asymptotic distribution of $(\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega})$ is the multivariate normal distribution with zero mean vector and variance-covariance matrix $\boldsymbol{H}^{-1}$, that is, $(\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}) \sim N_{k+1}(\mathbf{0}, \boldsymbol{H}^{-1})$, where $\boldsymbol{H}^{-1}$ is the inverse of the expected information matrix. Therefore, hypothesis testing for the parameters can be conducted under the normal distribution. In an application, the observed information matrix, obtained numerically by the software, is usually used instead of the expected information matrix for easy calculation.

3.3. Monte-Carlo simulation

In this section, the performance of the ML method in estimating the parameters of the RUC quantile regression model is investigated via a Monte-Carlo simulation study. The simulations are conducted for $\lfloor 100,000/n \rfloor$ Monte-Carlo runs, where $\lfloor \cdot \rfloor$ denotes the integer value function, and the sample size, n, is considered 30 (small), 50 (moderate), and 100 (large). The parameter vector $\boldsymbol{\Omega} = (\xi_0, \xi_1, \sigma)$ of the model is taken to be $(-0.5, -1.5, 1.0), (1.2, 2.5, 1.0), (-1.5, 2.8, 1.0)$, and $(1.5, -1.2, 1.0)$. These values were arbitrarily chosen, that is, no specific reason for choosing them, and simulations are performed in software `MATLAB2015b`.

The simulation structure is given below:

i. Set the values for the sample size n and the parameters ξ_0, ξ_1, and σ.
ii. Generate random variates for $x_{i1} \sim \text{uniform}(-3, 3), \quad (i = 1, 2, 3, \cdots, n)$.
iii. Set $i = 0$
iv. Generate $z_i \sim \text{RUC}(\kappa_i, \sigma)$ from (2.6), where $\kappa_i = \exp(\xi_0 + \xi_1 x_{i1})/(1 + \exp(\xi_0 + \xi_1 x_{i1}))$.
v. Obtain the estimates for $\hat{\xi}_0, \hat{\xi}_1$, and $\hat{\sigma}$.
vi. If the $\log L\left(\hat{\xi}_0, \hat{\xi}_1, \hat{\sigma}\right)$ attains its global maximum, then $i = i + 1$ and

$$\hat{\xi}_0^i = \hat{\xi}_0, \quad \hat{\xi}_1^i = \hat{\xi}_1, \quad \hat{\sigma}^i = \hat{\sigma}.$$

vii. Repeat steps (iv) - (vi) for $\lfloor 100,000/n \rfloor$.

For each generated sample, the ML estimates of the parameters are obtained using the optimization tool "fminsearch", which is available in software `MATLAB2015b`. Then, simulated bias, variance, and MSE values of the ML estimators of the parameters are calculated. The results of the Monte-Carlo simulation study for the RUC quantile regression model are reported in Table 1.

Table 1. The simulated bias, variance, and MSE values of the ML estimators of the parameters of the RUC quantile regression model.

(ξ_0, ξ_1, σ)	n	Bias			Variance			MSE		
		$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\sigma}$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\sigma}$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\sigma}$
(-0.5, -1.5, 1.0)	30	-0.0064	0.0847	0.0185	0.0384	0.2180	0.1014	0.0384	0.2251	0.1017
	50	0.0077	0.0658	0.0226	0.0204	0.1397	0.0432	0.0205	0.1439	0.0437
	100	0.0088	0.0638	0.0186	0.0093	0.0821	0.0204	0.0094	0.0860	0.0207
(1.2, 2.5, 1.0)	30	-0.0534	-0.2325	0.0246	0.0979	0.9305	0.0745	0.1007	0.9842	0.0750
	50	-0.0276	-0.1755	0.0170	0.0633	0.5803	0.0406	0.0640	0.6108	0.0409
	100	-0.0203	-0.1713	0.0201	0.0394	0.4516	0.0199	0.0398	0.4805	0.0203
(-1.5, 2.8, 1.0)	30	-0.0664	-0.2467	0.0201	0.2636	0.7512	0.0776	0.2679	0.8118	0.0780
	50	-0.0525	-0.2123	0.0249	0.1740	0.4836	0.0407	0.1766	0.5284	0.0413
	100	-0.0240	-0.2221	0.0211	0.1205	0.3241	0.0193	0.1210	0.3731	0.0197
(1.5, -1.2, 1.0)	30	-0.1482	-0.0275	0.0124	0.0959	0.1296	0.0797	0.1178	0.1303	0.0799
	50	-0.1267	-0.0014	0.0181	0.0622	0.0903	0.0423	0.0782	0.0902	0.0426
	100	-0.0935	-0.0032	0.0149	0.0351	0.0516	0.0212	0.0438	0.0516	0.0214

It can be seen from Table 1 that the simulated bias, variance and MSE values for the ML estimators of the model parameters are slight for the moderate and large sample sizes. Also note that the ML estimators of the model parameters are asymptotically unbiased, and the MSE values of them decrease when the sample size increases, as the theory says.

4. Model diagnostic

After the estimation of the model parameters, residual analysis should be performed to verify the suitability of the regression model fitted. In real-life applications, different models may also be considered, and their modeling performances are compared to ensure that the selected model performs better than its possible rivals. In this step, the models considered in the comparisons should already pass the model adequacy check; otherwise, comparisons may not make sense.

4.1. Model adequacy

The randomized quantile residuals proposed by [5]

$$\hat{e}_i = \Phi^{-1}\left(F\left(z_i; \hat{\mathbf{\Omega}}\right)\right)$$

and Cox-Snell residuals introduced by [4]

$$\hat{r}_i = -\log\left(1 - F\left(z_i; \hat{\mathbf{\Omega}}\right)\right)$$

are analyzed to check if the regression model suits the data set. Here, $F(\cdot)$ is the cdf of the RUC distribution, given in (2.4), evaluated in $\hat{\mathbf{\Omega}}$ and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.See [29, 43] for using randomized quantile residuals and Cox-Snell residuals in the context of beta regression.

Note that if $\hat{e}_i$ and $\hat{r}_i$ follow the standard normal distribution and the exponential distribution with the scale parameter one (standard exponential distribution), respectively, the corresponding regression model is well-adjusted; otherwise, the corresponding model may not give reliable results. The Shapiro-Wilk (SW) and Kolmogorov-Simirnov (KS) tests can be used to check the goodness of fit of $\hat{e}_i$ and $\hat{r}_i$ to the standard normal and standard exponential distribution, respectively.

4.2. Model comparison

In the literature, several information criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are used to compare models. The formulas for AIC and BIC are $-2\log L(\hat{\boldsymbol{\Omega}}) + 2k$ and $-2\log L(\hat{\boldsymbol{\Omega}}) + k\log(n)$, respectively, where k is the number of parameters in the corresponding model and n is the number of observations in the sample. It is clear from their formulas that when the number of parameters in the models is equal, the value of $\log L(\hat{\boldsymbol{\Omega}})$ can only be used to determine which model reflects more information than others from the corresponding samples. Also, note that the AIC and BIC criteria are generally preferred to compare nested models, and Raftery [33] stated that the significant difference in the BIC of the models should be greater than 2.0.

In addition, the proportion of explained variation in the response variable by the regression model can be considered a model comparison criterion. In this context, the generalized pseudo-R^2 (R_G^2), defined by [26],

$$R_G^2 = 1 - \exp\left(-\frac{2}{n}\left(\log L_{fit} - \log L_{null}\right)\right)$$

can be used for comparing not only nested but also non-nested models. Here, $\log L_{fit}$ and $\log L_{null}$ represent the log-likelihood of the fitted regression model and the log-likelihood for the null model, without covariates, respectively. A higher value of R_G^2 means a better explanation of the variation in the response variable.

Likelihood ratio tests can also be used to compare models. Vuong [39] proposed to use the following likelihood ratio statistics

$$Vuong = \frac{1}{\hat{\omega}\sqrt{n}}\sum_{i=1}^{n}\log\frac{f\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\delta}}\right)}{g\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\theta}}\right)}$$

for testing whether there is any significant difference in two non-nested models. Here,

$$\hat{\omega} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\delta}}\right)}{g\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\theta}}\right)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{f\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\delta}}\right)}{g\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\theta}}\right)}\right)^2\right]^{1/2}$$

is the standard deviation of the $\frac{1}{n}\sum_{i=1}^{n}\log\frac{f\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\delta}}\right)}{g\left(z_i;\boldsymbol{x_i},\hat{\boldsymbol{\theta}}\right)}$. Note that when n goes to infinity, the distribution of Vuong statistics converges to the standard normal distribution.

5. Data modeling

In this section, real datasets from the related literature are modeled using the RUC quantile regression model. Also, the beta regression model proposed by [6] and the UW, UGHN, Kum, and UBXII quantile regression models introduced by [21,23,24,34], respectively, were included in the study to strengthen the application section. See Appendix B for the pdf of the model considered. The ML method is used to obtain estimates of the parameters of the corresponding regression models.

The modeling performances of the beta, Kum, UW, UGHN, UBXII, and RUC quantile regression models are compared to each other. In the comparisons, adequacy of the corresponding model is assessed by using the SW and KS tests, which show whether the randomized residuals and Cox-Snell residuals follow the standard normal and standard exponential distributions, respectively.

After the model adequency check, the R_G^2 and Vuong statistics are used to determine which model explains the variation in the response variable better than the others, and Vuong statistics are used to determine whether the compared models are equivalent, i.e., to check if there exist statistical differences between the compared models. Note that the Vuong statistics are obtained for the RUC quantile regression model against others. Furthermore, values of $\log L$ and BIC of the beta, Kum, UW, UGHN, UBXII and RUC quantile regression models are provided.

Note that all computations were performed via the software `R` [32]. The ML estimates of the beta, Kum, UW, UGHN, UBXII, and RUC regression models along with the corresponding p-values were computed by using the ``‘‘maxBFGS’’`` function which is available in the ``‘‘maxLik’’`` package proposed by [7].

5.1. Application-I: Modeling of tuna data

In this section, the tuna dataset taken from [35] is considered. The data set includes 77 observations of longliner catches of the tropical tuna percentage (TTP) and the sea surface temperature (SST) in the corresponding points, where the observations were obtained, of the southern Indian Ocean in the year 2000. Here, the TTP is considered a response variable, and the SST is taken to be a covariate; in other words, the TTP is assumed to be a function of the SST. Ribeiro and Ferrari [35] replaced the value of observation 46 from 1 to 0.99 since observations at the boundary of the unit interval make it problematic in beta regression. Here, the following regression structure

$$\text{logit}(\kappa_i) = \xi_0 + \xi_1 x_{i1}, \quad (i = 1, 2, 3, \cdots, 77)$$

is considered for κ_i.

The ML estimates of the parameters, with an associated value p in parentheses, of the beta, Kum, UW, UGHN, UBXII and RUC quantile regression models along with the $\log L$, BIC, SW (p-value), KS (p-value), Vuong statistics (p value) and R_G^2 values are tabulated in Table 2 and Table 3.

Table 2. The ML estimates of the parameters of the employed regression models in modeling tuna data (the associated p-values are given in parentheses).

	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\sigma}$
Beta	-5.039 (<0.01)	0.145 (<0.01)	7.781 (<0.01)
Kum	-7.529 (<0.01)	0.235 (<0.01)	1.131 (<0.01)
UW	-7.360 (<0.01)	0.223 (<0.01)	3.246 (<0.01)
UGHN	-7.728 (<0.01)	0.238 (<0.01)	2.291 (<0.01)
UBXII	-7.138 (<0.01)	0.216 (<0.01)	1.313 (<0.01)
RUC	-7.058 (<0.01)	0.209 (<0.01)	0.217 (<0.01)

Table 3. Model diagnostic results of the employed models in modeling tuna data (the associated p-values are given in parentheses).

	Model Adequency		Model Comparison			
	SW	KS	Vuong	R_G^2	$\log L$	BIC
Beta	0.595 (<0.01)	0.253 (<0.01)	1.448 (0.074)	0.324	69.808	-126.584
Kum	0.682 (<0.01)	0.179 (0.013)	1.857 (0.032)	0.341	71.071	-129.111
UW	0.810 (<0.01)	0.144 (0.072)	0.820 (0.206)	0.526	89.134	-165.237
UGHN	0.809 (<0.01)	0.167 (0.024)	1.370 (0.085)	0.519	86.365	-159.699
UBXII	0.811 (<0.01)	0.142 (0.081)	1.546 (0.061)	0.388	82.706	-152.381
RUC	0.962 (0.021)	0.111 (0.278)	—	0.639	98.722	-184.413

From Table 2, it is clear that the covariate “SST” has a statistically significant positive effect on the response variable “TTP” in the beta, Kum, UW, UGH, UBXII and RUC regression models. Furthermore, the results of the adequacy of the model, given in Table 3, show that only the randomized quantile residuals of the RUC regression model follow the standard normal distribution at a 1% significant level; see also Figure 3 in which Q-Q plots are provided for the randomized quantile residuals of the models considered. This conclusion can also be extended for the Cox-Snell residuals in which only the RUC regression model meets the criterion at a significant level 10%. Also, Ribeiro and Ferrari [35] spotted that observation 46 is an outlier and stated that the beta regression model with ML estimates has a lack of fit because of it; therefore, they proposed robust estimation approaches which are not affected by the outliers.

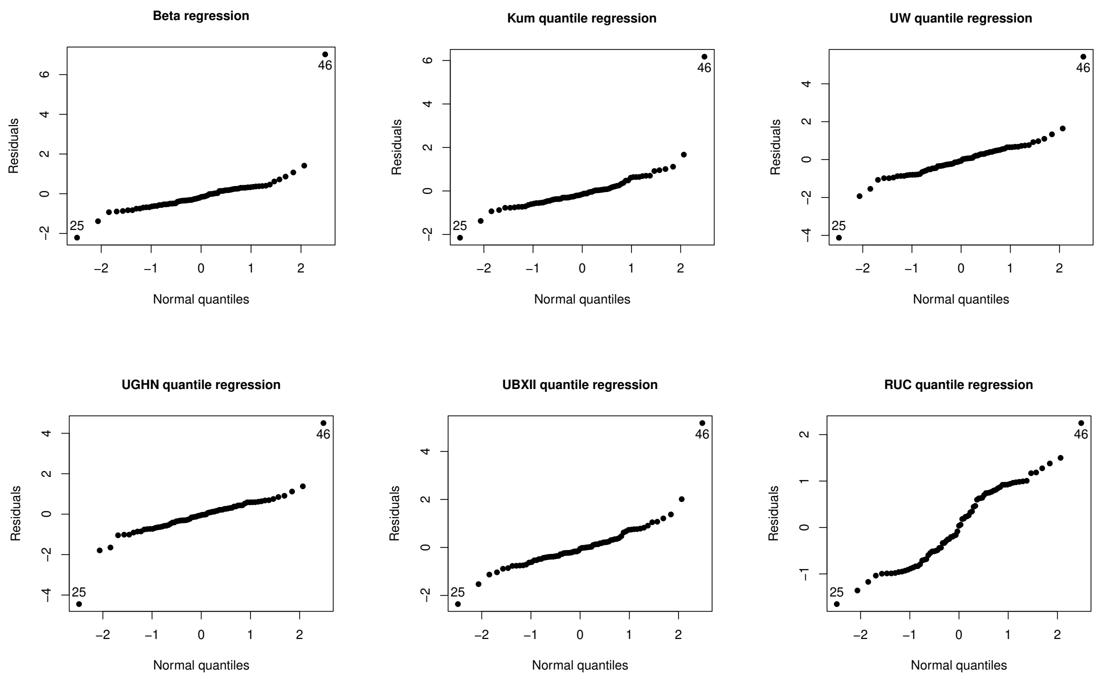


Figure 3. Normal Q-Q plot for the quantile residuals of the corresponding fitted models to the TTP.

The RUC quantile regression model explains the variation in the TTP better than its rivals since it has the highest R_G^2 value; see also Figure 4, in which the fitting performances of the considered regression models are illustrated. Furthermore, the RUC quantile regression model is not equal to the beta, Kum, UGHN, and UBXII regression models (at a 10% significant level) when the corresponding Vuong statistics are considered. The RUC quantile regression model also has the highest $\log L$ and the lowest BIC values.

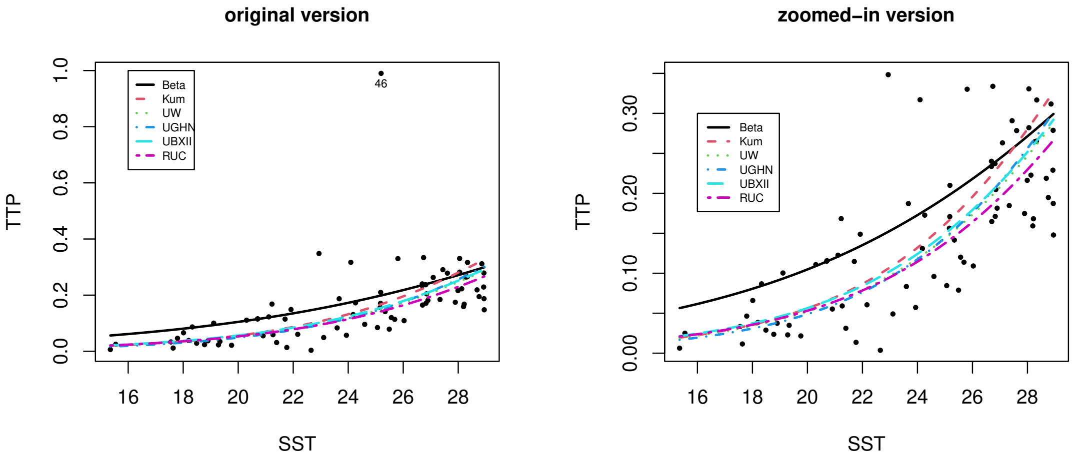


Figure 4. Scatter plot of the pairs (SST, TTP) along with the fitted regression lines of the corresponding models.

In summary, from Table 2, Table 3, Figure 3 and Figure 4, it can be said that the RUC quantile regression model performs better modeling performance than its rivals when not only goodness-of-fit statistics are taken into account, but also information criteria results are taken into account.

5.2. Application-II: Modeling of health insurance coverage data

In this section, the health insurance coverage (HIC) data includes information on 80 cities in the state of São Paulo, Brazil, in 2010, from Maluf et al. [18], is modeled using beta, Kum, UW, UGH, UBXII, and RUC regression models. The HIC index is assumed to be a function of the percentage of the total population who lives in the city's urban zone (URB) and the per capita gross domestic product (GDP), i.e., the HIC index is considered a response variable, and the URB and GDP are stand for covariates.

In this study, the following regression structure

$$\mathrm{logit}(\kappa_i) = \xi_0 + \xi_1 x_{i1} + \xi_2 x_{i2}, \quad (i = 1, 2, 3, \cdots, 80)$$

is considered for κ_i.

The ML estimates of the parameters, with associated p-value in the parentheses, of the beta, Kum, UW, UGHN, UBXII, and RUC quantile regression models along with the $\log L$, BIC, SW (p-value), KS (p-value), Vuong statistics (p-value), and R_G^2 values are tabulated in Table 4 and Table 5.

Table 4. The ML estimates of the parameters of the employed regression models in modeling the HIC data (the associated p-values are given in parentheses).

	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\sigma}$
Beta	-4.281 (<0.01)	3.269 (<0.01)	0.010 (<0.01)	7.398 (<0.01)
Kum	-4.595 (<0.01)	3.344 (<0.01)	0.011 (<0.01)	1.246 (<0.01)
UW	-5.853 (<0.01)	4.908 (<0.01)	0.006 (<0.01)	2.944 (<0.01)
UGHN	-5.838 (<0.01)	4.874 (<0.01)	0.006 (<0.01)	2.265 (<0.01)
UBXII	-4.554 (<0.01)	3.203 (<0.01)	0.011 (<0.01)	1.421 (<0.01)
RUC	-6.922 (<0.01)	5.999 (<0.01)	0.011 (<0.01)	0.245 (<0.01)

Table 5. Model diagnostic results of the employed models in modeling the HIC data (the associated p-values are given in parentheses).

	Model Adequacy		Model Comparison			
	SW	KS	Vuong	R_G^2	$\log L$	BIC
Beta	0.775 (<0.01)	0.138 (0.085)	0.999 (0.159)	0.289	53.384	-89.241
Kum	0.735 (<0.01)	0.149 (0.052)	1.125 (0.106)	0.244	50.358	-83.187
UW	0.920 (<0.01)	0.078 (0.690)	0.358 (0.360)	0.377	66.012	-114.495
UGHN	0.933 (<0.01)	0.089 (0.523)	0.429 (0.334)	0.361	66.057	-114.586
UBXII	0.836 (<0.01)	0.120 (0.183)	0.832 (0.203)	0.304	59.352	-101.176
RUC	0.978 (0.185)	0.062 (0.904)	—	0.594	69.083	-120.639

The results of the analysis show that the covariates "URB" and "GDP" statistically significantly affect the response variable "HIC" in the beta, Kum, UW, UGH, UBXII and RUC regression models. Note that the estimates of parameters $\hat{\xi}_0$ and $\hat{\xi}_1$ differ, while $\hat{\xi}_2$ remain more or less the same for each of the models considered. From Table 5, the randomized residuals from only the RUC quantile regression model follow the standard normal distribution, and it shows that only the RUC quantile regression model meets the model adequacy criterion. See also Figure 5 in which Q-Q plots are provided for the randomized quantile residuals of the models considered. In contrast to the SW test, the results of the KS test show that only the beta regression and Kum quantile regression models perform a lack of fit (at a 10% significant level) to the Cox-Snell residuals.

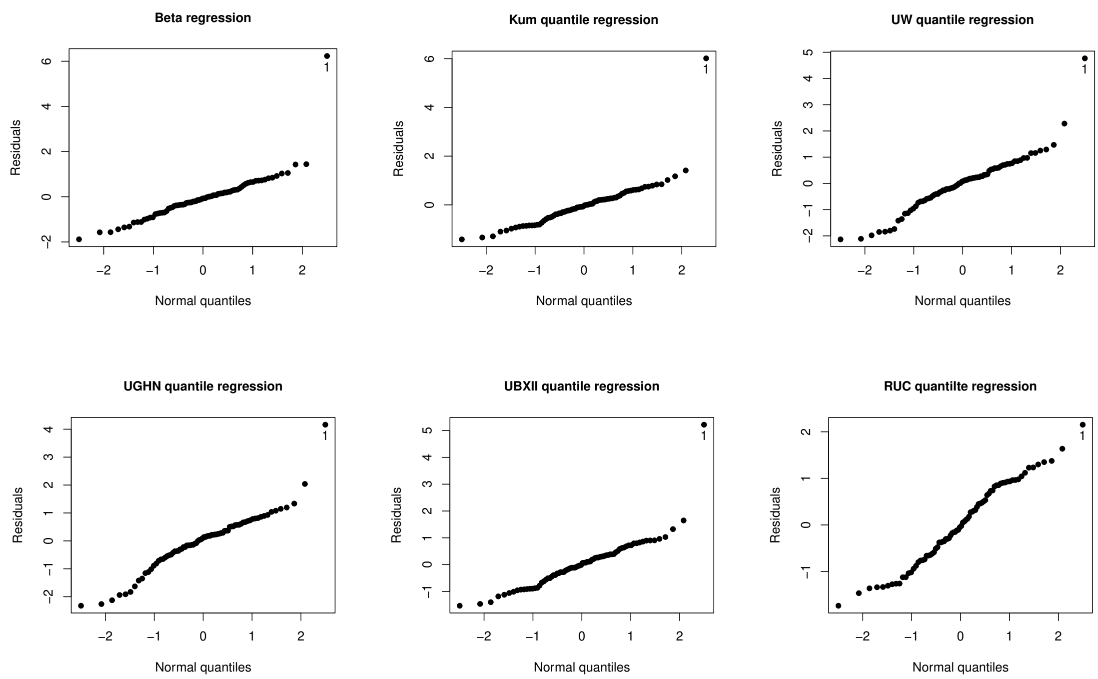


Figure 5. Normal Q-Q plot for the quantile residuals of the corresponding models fitted to the HIC.

Maluf et al. [18] stated that the city with an atypical value for HIC, around 0.98, could be an outlier when the beta regression model is taken into account. This conclusion can be extended when the Kum, UW, UGHN, and UBXII quantile regression models are considered; see Figure 5. In addition, Maluf et al. [18] show that the ML estimates of the beta regression model are not robust against outliers; therefore, they proposed robust estimation approaches.

The quantile regression of RUC has the highest value R_G^2, which means that it explains the variation in the HIC better than the other considered models; see Figure 6 in which the fitting performances of the beta, UW, and RUC regression models are illustrated. In addition, it can be said that the Vuong statistics are in favor of the quantile regression model RUC at significant levels 16%, 11%, 0.36%, 0.33%, and 0.20% when the beta, Kum, UW, UGHN, and UBXII regression models are taken as rivals, respectively. Furthermore, the RUC quantile regression model reflects more information from the corresponding samples than the beta, Kum, UW, UGHN and UBXII regression models, since it has the highest $\log L$ and lowest BIC values.

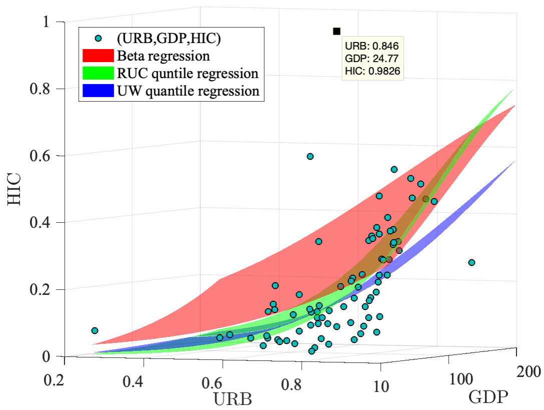


Figure 6. Scatter plot of the pairs (URB, GDP, HIC) along with the fitted surface plots of the beta, UW, and RUC regression models.

As a result, from Table 4, Table 5, Figure 5 and Figure 6, it can be said that the RUC quantile regression model performs better modeling performance than its rivals when not only goodness-of-fit statistics are taken into account, but also information criteria results are taken into account.

5.3. Application-III: Modeling of risk survey data

In this section, the data of the Schmitand and Roth risk survey [36] are analyzed. In their study, unit risk management costs are assumed to be a function of the firm's retention strategy, whether or not the firm uses a captive insurer; the log of the firm's size and the firm's industry risk; the firm's centralization strategy; the degree to which the firm employs analytical tools in performing the risk management function. Therefore, the response variable (z) and the covariates (x_1, x_2, x_3, x_4, x_5, and x_6) associated with z are defined, respectively, as follows:

- Firm-specific ratio of premiums plus uninsured losses divided by total assets (FIRM-COST)

and

- Firm-specific ratio of the summation of per occurrence retention levels (ASSUM),
- Whether or not the firm uses a captive insurer (CAP),
- Log of the firm's total asset value (SIZELOG),
- Industry average of premiums plus uninsured losses divided by total assets (IND-COST),
- Importance of the local manager in choosing local retention levels (CENTRAL),

- Importance of analytical tools in making risk management decisions (SOPH).

Note that the covariates x_1 and x_5 were measured by the corporate risk manager, and the covariate x_4 was obtained from the 1985 Cost of Risk Survey; see Schmitand and Roth [36] and Mazucheli et al. [21] for details. Recently, Mazucheli et al. [21] utilized the regression structure

$$\operatorname{logit}(\kappa_i) = \xi_0 + \xi_1 x_{i1} + \xi_2 x_{i2} + \xi_3 x_{i3} + \xi_4 x_{i4} + \xi_5 x_{i5} + \xi_6 x_{i6}, \quad (i = 1, 2, 3, \cdots, 73)$$

for κ_i and showed that the unit-Weibull (UW) quantile regression model is preferable over the beta and Kumaraswamy (Kum) regression models.

The data set from the risk survey is analyzed by embedding the same regression structure into the RUC quantile regression model. The ML estimates of the parameters (p-value) of the beta, Kum, UW, UGHN, UBXII and RUC quantile regression models along with the values $\log L$, BIC, SW (p-value), KS (p-value), Vuong statistics (p-value) and R_G^2 are tabulated in Table 6 and Table 7.

Table 6. The ML estimates of the parameters of the employed regression models in modeling risk survey data (the associated p-values are given in parentheses).

	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\xi}_3$
Beta	1.888 (0.107)	-0.012 (0.383)	0.178 (0.443)	-0.511 (<0.01)
Kum	2.539 (0.089)	-0.036 (0.038)	0.596 (0.118)	-0.798 (<0.01)
UW	3.471 (<0.01)	-0.008 (0.595)	0.128 (0.610)	-0.804 (<0.01)
UGHN	3.142 (<0.01)	-0.001 (0.921)	0.045 (0.848)	-0.757 (<0.01)
UBXII	2.274 (0.179)	-0.027 (0.142)	0.297 (0.408)	-0.782 (<0.01)
RUC	4.000 (<0.01)	-0.016 (<0.01)	0.053 (0.797)	-0.894 (<0.01)
	$\hat{\xi}_4$	$\hat{\xi}_5$	$\hat{\xi}_6$	$\hat{\sigma}$
Beta	1.236 (<0.01)	-0.012 (0.889)	-0.004 (0.861)	6.331 (<0.01)
Kum	5.256 (<0.01)	-0.028 (0.816)	-0.027 (0.389)	0.978 (<0.01)
UW	1.439 (<0.01)	-0.024 (0.777)	-0.002 (0.916)	3.353 (<0.01)
UGHN	1.248 (<0.01)	-0.017 (0.824)	-0.003 (0.875)	2.680 (<0.01)
UBXII	4.870 (<0.01)	-0.051 (0.647)	-0.012 (0.682)	1.058 (<0.01)
RUC	2.035 (<0.01)	-0.238 (<0.01)	0.019 (0.298)	0.204 (<0.01)

Table 7. Model diagnostic results of the employed models in modeling risk survey data (the associated p-values are given in parentheses).

	Model Adequency		Model Comparison			
	SW	KS	Vuong	R_G^2	$\log L$	BIC
Beta	0.677 (<0.01)	0.217 (<0.01)	1.914 (0.028)	0.272	87.723	-141.122
Kum	0.901 (<0.01)	0.120 (0.227)	1.981 (0.024)	0.425	98.827	-163.330
UW	0.902 (<0.01)	0.078 (0.737)	0.513 (0.304)	0.468	111.111	-187.899
UGHN	0.932 (<0.01)	0.087 (0.613)	0.333 (0.369)	0.500	113.679	-193.033
UBXII	0.918 (<0.01)	0.099 (0.449)	1.305 (0.096)	0.399	106.517	-178.711
RUC	0.981 (0.320)	0.076 (0.771)	—	0.532	116.729	-199.134

From the results of the SW test given in Table 7, the randomized quantile residuals only from the RUC quantile regression model follow the standard normal distribution while the other models fail. Therefore, it can be said that only the fitting of the RUC quantile regression model is acceptable. This conclusion is also supported by the Q-Q plots, shown in Figure 7, for the randomized quantile residuals of the models considered.

However, when the KS test for the Cox-Snell residuals is taken into account, only the beta regression model shows a lack of fit to the corresponding data set.

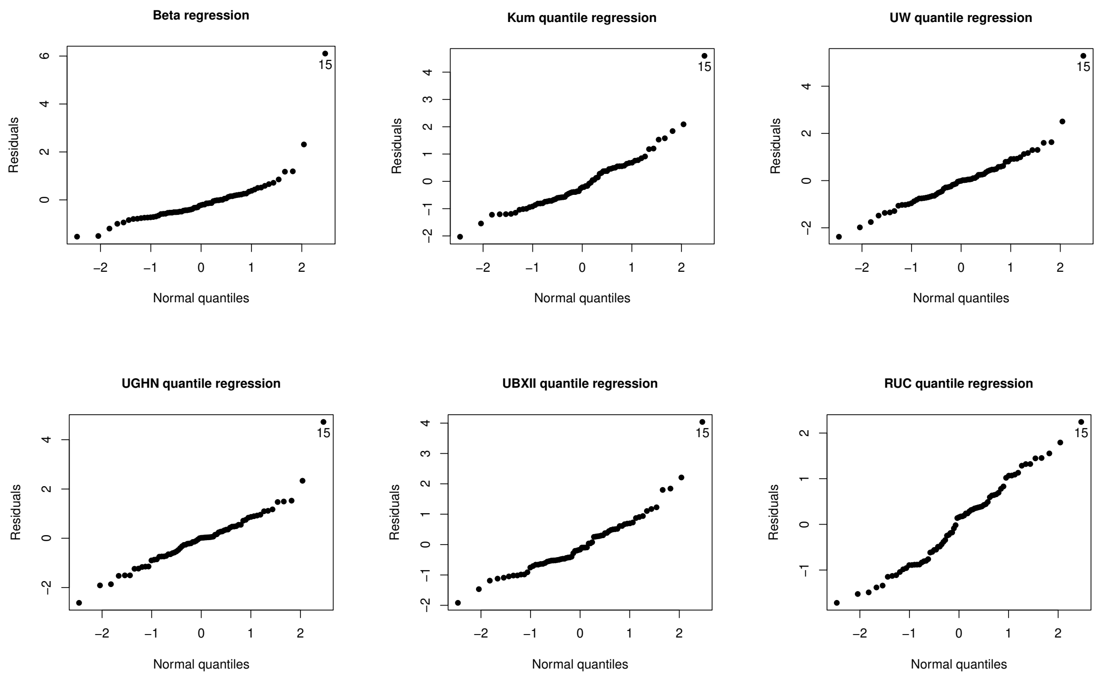


Figure 7. Normal Q-Q plot for the quantile residuals of the corresponding models fitted to the FIRMCOST.

The RUC quantile regression model explains the variation in "FIRMCOST" better than its rivals; therefore, the RUC quantile regression model is one step ahead of its rivals based on the R_G^2 criterion. Moreover, the Vuong test shows that the RUC quantile regression model is not equivalent to the beta, KUM, and UBXII regression models at a 10% significant level. In addition, if the UW and UGHN quantile regression models are taken as rivals, the Vuong statistics favor the RUC quantile regression model at significant levels 30% and 37%, respectively. Furthermore, the RUC quantile regression model has higher $\log L$ and smaller BIC values than the beta, Kum, UW, UGHN, and UBXII regression models; therefore, it can be said that the RUC quantile regression model is also preferable to its rivals when information criteria are considered.

Not least of all, the results of the analysis show that the covariates "SIZELOG" and "INDICOST" have statistically significant negative and positive impacts, respectively, on the response variable "FIRMCOST". In contrast, the remaining covariates in the beta, Kum, UW, UGH, and UBXII quantile regression models do not statistically significantly affect the "FIRMCOST". In contrast, the covariates "ASSUM" and "CENTRAL" are also statistically significant in the RUC quantile regression model.

Ribeiro and Ferrari [35] conducted a comprehensive analysis of the data from the risk survey and spotted atypical observations in the data set. They showed that the ML estimates of the parameters of the beta regression model are highly affected by anomalies in the data set. Similarly to Ribeiro and Ferrari [35], the scatter plots for the pairs (x_3, z) and (x_4, z) along with the fitted regression lines of the beta, Kum, UW, UGHN, UBXII and RUC regression model are produced – by setting values of the remaining covariates at their sample median value – to show the effects of the outliers on the regression lines; see Figure 8.

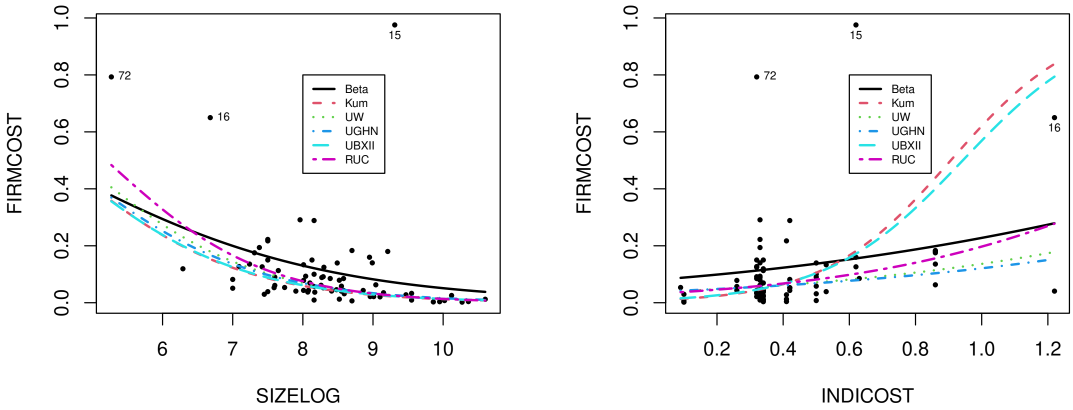


Figure 8. Scatter plots of the pairs (SIZELOG, FIRMCOST) and (INDICOST, FIRMCOST) along with the fitted regression lines of the corresponding models.

Note that the RUC quantile regression model considered in this study performs more or less the same fitting performance as the robust beta regression counterparts given in Ribeiro and Ferrari [35]; see Figure 8 in this study and Figure 13 in Ribeiro and Ferrari [35].

6. Conclusion

In this study, the RUC distribution is obtained by reparametrizing the UC distribution and quantile-based analysis of the RUC distribution is performed to show some of its characteristic measures. Then, a new quantile regression model is derived based on the RUC distribution, and the ML method is used in estimating the parameters of the RUC quantile regression model. In addition, a Monte Carlo simulation study is conducted to show the efficiencies of the ML estimation of the model parameters, and simulation results show that the corresponding estimates are asymptotically unbiased with small variances. Furthermore, implementation of the RUC quantile regression model is shown by modeling the three datasets, which were previously modeled via the competitive regression model by the several authors. Last but not least, modeling performance of the RUC quantile regression model is compared with its strong rivals' by means of the several criteria, and comparison results show that the RUC quantile regression model stands one step ahead from not only the beta regression model but also the other competetive models which were recently introduced.

Acknowledgements

The editor and reviewers are thanked for comments that led to presentational improvements. Talha Arslan would like to thank Brunel University London (UK) for a visiting position in 2023 to collaborate with Prof. Keming Yu and providing a peaceful environment to conduct this study.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. 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.

Funding. There is no funding regarding this research work

Data availability. Enquiries about data availability should be directed to the authors.

References

[1] T. Arslan, *A new family of unit-distribution: Definition, properties and applications*, TWMS J. App. Eng. Math. **13** (2), 782–791, 2023.

[2] G. Brys, M. Hubert and A. Struyf, *A comparison of some new measures of skewness*, in: Developments in Robust Statistics (R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw, eds.), 98–113, Physica, Heidelberg, 2003.

[3] G.M. Cordeiro, G.M. Rodrigues, F. Prataviera and E.M.M. Ortega, *A new quantile regression model with application to human development index*, Comput. Stat. **39**, 2925–2948 2024.

[4] D.R. Cox and E.J. Snell, *A general definition of residuals*, J. R. Stat. Soc. B **30** (2), 248–275, 1968.

[5] P.K. Dunn and G.K. Smyth, *Randomized quantile residuals*, J. Comput. Graph. Stat. **5** (3), 236–244, 1996.

[6] S. Ferrari and F. Cribari-Neto, *Beta regression for modelling rates and proportions*, J. Appl. Stat. **31** (7), 799–815, 2004.

[7] A. Henningsen and O.Toomet, *maxlik: A package for maximum likelihood estimation in R*, Comput. Stat **26**, 443–458, 2011.

[8] D.V. Hinkley, *On power transformations to symmetry*, Biometrika **62** (1), 101–111, 1975.

[9] N.L. Johnson, *Systems of frequency curves generated by methods of translation*, Biometrika **36** (1), 149–176, 1949.

[10] R. Kieschnick and B.D. McCullough, *Regression analysis of variates observed on (0, 1): percentages, proportions and fractions*, Stat. Model. **3** (3), 193–213, 2003.

[11] R. Koenker, *Quantile Regression*, Cambridge University Press, Cambridge, 2005.

[12] R. Koenker and G. Bassett, *Regression quantiles*, Econometrica **46** (1), 33–50, 1978.

[13] M.C. Korkmaz, *The unit generalized half-normal distribution: A new bounded distribution with inference and application*, U.P.B. Sci. Bull. Ser. A **82** (2), 133–140, 2020.

[14] S. Kotz, N.L. Johnson and N. Balakrishnan, *Univariate continuous distributions-Volume I*, John Wiley & Sons, 1994.

[15] K. Krishnamoorthy, *Handbook of statistical distributions with applications*, Chapman and Hall/CRC, 2016.

[16] P. Kumaraswamy, *Generalized probability density function for double-bounded random processes*, J. Hydrol. **46** (1), 79–88, 1980.

[17] H.Y. Lee, H.J. Park and H.M. Kim, *A clarification of the Cauchy distribution*, Commun. Stat. Appl. Methods **21** (2), 183–191, 2014.

[18] Y.S. Maluf, S.L.P. Ferrari and F.F. Queiroz, *Robust beta regression through the logit transformation*, Metrika **88**, 61–81, 2025.

[19] J. Mazucheli, A.F.B. Menezes and M.E. Ghitany, *The unit-Weibull distribution and associated inference*, J. Appl. Probab. Stat. **13** (2), 1–22, 2018.

[20] J. Mazucheli, A.F.B. Menezes and S. Chakraborty, *On the one parameter unit-Lindley distribution and its associated regression model for proportion data*, J. Appl. Stat. **46** (4), 700–714, 2019.

[21] J. Mazucheli, A.F.B. Menezes, L.B. Fernandes, R.P. de Oliveira and 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** (6), 954–974, 2020.

[22] J. Mazucheli, B. Alves, A.F.B. Menezes and V. Leiva, *An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data*, Comput. Methods Programs Biomed. **221**, 106816, 2022.

[23] J. Mazucheli, M.C. Korkmaz, A.F.B. Menezes and V. Leiva, *The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics, and numerical applications*, Soft Comput. **27**, 279–295, 2023.

[24] P.A. Mitnik and S. Baek, *The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation*, Stat. Pap. **54**, 177–192, 2013.

[25] J.A.A. Moors, *A quantile alternative for kurtosis*, J. R. Stat. Soc. D **37** (1), 25–32, 1988.

[26] N.J.D. Nagelkerke, *A note on a general definition of the coefficient of determination*, Biometrika **78** (3), 691–692, 1991.

[27] A. Noufaily and M.C. Jones, *Parametric quantile regression based on the generalized gamma distribution*, J. R. Stat. Soc. C **62** (5), 723–740, 2013.

[28] J. Pedro and M.D. Jimenez-Gamero, *A quantile regression model for bounded response based on the exponential-geometric distribution*, Revstat Stat. J. **18** (4), 415–436, 2020.

[29] G.H.A. Pereira, *On quantile residuals in beta regression*, Commun. Stat. Simul. Comput. **48** (1), 302–316, 2019.

[30] A.R. Qader, A unit-Cachy distribution and unit-Cauchy-generated family of distributions, MSc Thesis, Van Yüzüncü Yıl University, 2021.

[31] A.R. Qader and T. Arslan, *A unit-Cauchy distribution: Definition, properties and application*, In: 2nd Int. Appl. Stat. Conf., p. 225, 2021.

[32] R Core Team, R: A language and environment for statistical computing, R Found. Stat. Comput., Vienna, 2023.

[33] A.E. Raftery, *Bayesian model selection in social research*, Sociol. Methodol. **25**, 111-163, 1995.

[34] T.F. Ribeiro, G.M. Cordeiro, F.A. Peña-Ramìrez and R.R. Guerra, *A new quantile regression for the COVID-19 mortality rates in the United States*, Comput. Appl. Math. **40**, 255, 2021.

[35] T.K.A. Ribeiro and S. Ferrari, *Robust estimation in beta regression via maximum Lq-likelihood*, Stat. Pap. **64**, 321–353, 2023.

[36] J.T. Schmit and K. Roth, *Cost effectiveness of risk management practices*, J. Risk Insur. **57** (3), 455–470, 1990.

[37] C.W. Topp and F.C. Leone, *A family of J-shaped frequency functions*, J. Am. Stat. Assoc. **50** (269), 209–219, 1955.

[38] J.T. Townsend and H. Colonius, *Variability of the max and min statistic: a theory of the quantile spread as a function of sample size*, Psychometrika **70** (4), 759–772, 2005.

[39] Q.H. Vuong, *Likelihood ratio tests for model selection and non-nested hypotheses*, Econometrica **57** (2), 307–333, 1989.

[40] K. Yu and M.C. Jones, *Local linear quantile regression*, J. Am. Stat. Assoc. **93** (441), 228–237, 1998.

[41] K. Yu and R. Moyeed, *Bayesian quantile regression*, Stat. Probab. Lett. **54** (4), 437–447, 2001.

[42] K. Yu, Z. Lu and J. Stander, *Quantile regression: applications and current research areas*, J. R. Stat. Soc. D **52** (3), 331–350, 2003.
[43] H. Zhou and X. Huang, *Bayesian beta regression for bounded responses with unknown supports*, Comput. Stat. Data Anal. **167**, 107345, 2022.

Appendix A. Quantile spread order

The proof of Proposition 2.5 is provided below:

Proof. By using the qf in (2.5), quantile spread of X and Y are

$$\begin{aligned} QS_X(p) =& \kappa^{-1}(1-p) - \kappa^{-1}(p) \\ &= \left(1+\exp\left(-\sigma_1 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_1 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \end{aligned}$$

and

$$\begin{aligned} QS_Y(p) =& \kappa^{-1}(1-p) - \kappa^{-1}(p) \\ &= \left(1+\exp\left(-\sigma_2 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_2 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}}, \end{aligned}$$

for $0 < p < 0.5$, respectively. Therefore,

$$\begin{aligned} & QS_X(p) \leq QS_Y(p) \Leftrightarrow \\ & \left(1+\exp\left(-\sigma_1 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_1 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \\ & \leq \left(1+\exp\left(-\sigma_2 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_2 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}}; \\ & \Leftrightarrow \left(1+\exp\left(-\sigma_1 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_2 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq 0; \\ & \Leftrightarrow \left(1+\exp\left(-\sigma_2 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_1 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq 0. \end{aligned}$$

Then,

$$\begin{aligned} & \left(1+\exp\left(-\sigma_1 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_2 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq 0 \\ & \Leftrightarrow \left(1+\exp\left(-\sigma_1 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq \left(1+\exp\left(-\sigma_2 \tan\left(\pi(0.5-p)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \\ & \Leftrightarrow -\sigma_1 \tan\left(\pi(0.5-p)\right) \leq -\sigma_2 \tan\left(\pi(0.5-p)\right) \Leftrightarrow -\sigma_1 \leq -\sigma_2 \Leftrightarrow \sigma_2 \leq \sigma_1. \end{aligned}$$

Additionally,

$$\begin{aligned} & \left(1+\exp\left(-\sigma_2 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} - \left(1+\exp\left(-\sigma_1 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq 0 \\ & \left(1+\exp\left(-\sigma_2 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \leq \left(1+\exp\left(-\sigma_1 \tan\left(\pi(p-0.5)\right)\right)\right)^{\frac{\log(\kappa)}{\log(2)}} \\ & \Leftrightarrow -\sigma_2 \tan\left(\pi(p-0.5)\right) \leq -\sigma_1 \tan\left(\pi(p-0.5)\right) \\ & \Leftrightarrow \sigma_2 \tan\left(\pi(0.5-p)\right) \leq \sigma_1 \tan\left(\pi(0.5-p)\right) \Leftrightarrow \sigma_2 \leq \sigma_1. \end{aligned}$$

As stated above, $QS_X(p) \leq QS_Y(p)$ if and only if $\sigma_2 \leq \sigma_1$. Therefore, the proof for the quantile spread order $X \leq_{QS} Y$ is complete immediately if and only if $\sigma_2 \leq \sigma_1$. □

Appendix B. The unit distributions considered in the application section

In the application section, the beta, Kum, UW, UGHN, UBXII distributions having the pdf

$$f_{Beta}(z;\kappa,\sigma)=\frac{\Gamma(\sigma)}{\Gamma(\kappa\sigma)\Gamma\left((1-\kappa)\sigma\right)}z^{\kappa\sigma-1}\left(1-z\right)^{(1-\kappa)\sigma-1},$$

$$f_{Kum}(z;\kappa,\sigma)=\frac{\sigma\log(0.5)}{\log\left(1-\kappa^{\sigma}\right)}z^{\sigma-1}\left(1-z^{\sigma}\right)^{\frac{\log(0.5)}{\log(1-\kappa^{\sigma})}-1},$$

$$f_{UW}(z;\kappa,\sigma)=\frac{\sigma}{z}0.5^{\left(\frac{\log(z)}{\log(\kappa)}\right)^{\sigma}}\left(\frac{\log(0.5)}{\log(\kappa)}\right)\left(\frac{\log(z)}{\log(\kappa)}\right)^{\sigma-1},$$

$$f_{UGHN}(z;\kappa,\sigma)=\sqrt{\frac{2}{\pi}}\frac{\sigma\Phi^{-1}(0.25)}{\log(z)}\left(\frac{\log(z)}{\log(\kappa)}\right)^{\sigma}\exp\left(-\frac{1}{2}\left(\Phi^{-1}(0.25)\right)^{2}\left(\frac{\log(z)}{\log(\kappa)}\right)^{2\sigma}\right),$$

and

$$f_{UBXII}(z;\kappa,\sigma)=\frac{\left(\log\left(0.5\right)^{-\sigma}\right)\right)\left(\log\left((1-z)^{-1}\right)\right)^{\sigma-1}}{(1-z)\log\left(1+\left(\log\left((1-\kappa)^{-1}\right)\right)^{\sigma}\right)}\times\left(1+\left(\log\left((1-z)^{-1}\right)\right)^{\sigma}\right)^{\frac{\log(0.5)}{\log\left(1+\left(\log\left((1-\kappa)^{-1}\right)\right)^{\sigma}\right)}-1},$$

respectively, are considered. Here, $z\in(0,1)$, $\kappa\in(0,1)$, $\sigma>0$, $\Gamma(\cdot)$ represents the gamma function, and $\Phi^{-1}(\cdot)$ represents the quantile function of the standard normal distribution.