Soft Computing (2018)
https://doi.org/10.1007/s00500-022-07278-3

MATHEMATICAL METHODS IN DATA SCIENCE

The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics, and numerical applications

Josmar Mazucheli[1] · Mustafa Ç. Korkmaz[2] · André F. B. Menezes[3] · Víctor Leiva[4]

Accepted: 9 May 2022 / Published online: 27 July 2022

Abstract
In this paper, we propose and derive a new regression model for response variables defined on the open unit interval. By reparameterizing the unit generalized half-normal distribution, we get the interpretation of its location parameter as being a quantile of the distribution. In addition, we can evaluate effects of the explanatory variables in the conditional quantiles of the response variable as an alternative to the Kumaraswamy quantile regression model. The suitability of our proposal is demonstrated with two simulated examples and two real applications. For such data sets, the obtained fits of the proposed regression model are compared with that provided by a Kumaraswamy regression model.

Keywords Kumaraswamy distribution · Likelihood methods · Monte Carlo simulation · R software · Residual analysis · Unit generalized half-normal distribution.

1 Introduction

Interest in distributions with support on the unit interval has been increasingly coming into prominence in the literature of probability and statistics because of their usefulness in almost all disciplines to model indexes, percentage, proportions, and rates. The recent COVID-19 pandemic has intensified this type of modeling (Ribeiro et al. 2021; Mazucheli et al. 2022b). Such a kind of models has been employed in applications related to diverse areas linked to computer vision, deep learning, economics, education, finance, health, hydrology, mining, psychology, reliability, and risk management; see, for example, Kumaraswamy (1980); Van Dorp and Kotz (2002a, b); Smithson and Verkuilen (2006); Huerta et al. (2018); Kizilaslan and Nadar (2018); Kohansal (2019); Martinez-Florez et al. (2020); Mazucheli et al. (2020b); and Korkmaz et al. (2021b). Generally, the new unit models have been proposed via a transformation of well-known models; see Korkmaz and Chesneau (2021). Therefore, studies on unit modeling are increasing day by day.

It is well known that standard regressions aim to model the conditional mean of a dependent (response) variable by utilizing its relationship with independent (explanatory) variables (covariates or regressors). This relationship may be linear or nonlinear functionally and it generally relates the mean of the response variable with given values of the covariates. Thus, employing a standard regression is suitable if the response has a non-skewed distribution (symmetric).

If the response is defined on unit interval, the beta regression (Ferrari and Cribari Neto 2004) is the most used model to relate the unit response with covariates based on the mean; see also Figueroa-Zuniga et al. (2022a). Since the beta distribution is set according to its mean, its unit response regression was introduced by Ferrari and Cribari Neto (2004). However, if the unit response is skew distributed, or it has some outliers, using a beta regression may be unsuitable since it affects the mean. Similarly, unit mean response regressions were introduced based on the beta rectangular (Bayes et al. 2012), Birnbaum-Saunders (Mazucheli et al. 2018, 2021), log-Bilal (Altun et al. 2021), log-Lindley (Gómez-Déniz et al. 2014), log-weighted exponential (Altun 2021), unit-Burr XII (Korkmaz and Chesneau 2021), unit Lindley (Mazucheli et al. 2019a, b), unit second degree improved

✉ Mustafa Ç. Korkmaz
mustafacagataykorkmaz@gmail.com

1 Department of Statistics, Universidade Estadual de Maringá, Maringá, Brazil

2 Department of Measurement and Evaluation, Artvin Çoruh University, Artvin, Turkey

3 Department of Statistics, Universidade Estadual de Campinas, Campinas, Brazil

4 School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Lindley (Altun and Cordeiro 2020), and Vasicek (Mazucheli et al. 2022a) distributions. For other exponentiated distributions, see, for example, Akgul (2021).

If the response follows a skewed distribution, or outliers are present in the data, the usage of robust regression models is more suitable than a standard regression model. This is because the responses can be affected by these two situations making that the inference on the mean is affected as well, and the results from the model can provide possible wrong interpretations. As an alternative to the mean response regression, the quantile regression, introduced by Koenker and Bassett Jr (1978), has been popularly and recently considered as robust alternative to the mean regressions. The quantile regression relates the conditional quantiles of the response to given values of the covariates, instead of explaining its conditional mean. Indeed, a quantile regression is a more robust model than the ordinary regression and does not impose any distributional assumption on the error term, so the quantile regression was derived as a model that is essentially nonparametric. In addition, quantile regressions may be considered as competitors of standard regressions when the moments of a distribution cannot be obtained in a closed analytical form. Furthermore, the main advantage of the quantile regression is its flexibility for modeling data with heterogeneous conditional distributions (Bayes et al. 2017).

Note that, if we wish to employ a parametric quantile regression, the quantile function (QF) of the basis distribution, from where this regression is obtained, must have a closed form. To introduce a parametric quantile regression model based on a probability distribution, we must parameterize it based on its QF. As mentioned, this parameterization can be applied to any distribution that has its QF expressed in closed form and it is settable according to the model parameters, even if its mean does not have a closed form. Some examples of the quantile modeling are based on the exponentiated arcsecant hyperbolic normal (Korkmaz et al. 2021a), Kumaraswamy (Mitnik and Baek 2013), L-logistic (Paz et al. 2019), log-extended exponential-geometric (Jodrá and Jiménez-Gamero 2020), log-symmetric (Saulo et al. 2022), power Johnson SB, (Johnson 1949; Cancho et al. 2020), transmuted unit Rayleigh (Korkmaz et al. 2021b), unit-Birnbaum–Saunders (Mazucheli et al. 2021; Sanchez et al. 2021), unit Burr-XII, (Korkmaz and Chesneau 2021), unit-Chen (Korkmaz et al. 2021c), unit-Weibull (Mazucheli et al. 2020b; Sanchez et al. 2022), unit-log exponential power (Korkmaz et al. 2021d), unit-folded normal (Korkmaz et al. 2022b), unit-log-log (Korkmaz and Korkmaz 2022), unit arcsecant hyperbolic Weibull (Korkmaz et al. 2022a) and Vasicek (Mazucheli et al. 2022a) distributions. For a complete overview on parametric quantile regression models and their computational implementation with diverse applications, see Mazucheli et al. (2022b).

Despite the different proposals related to unit quantile regressions and the wide use of the normal distribution and its extensions, to the best of our knowledge, quantile regression based on the unit generalized half-normal (UGHN) distribution (Korkmaz 2020) has been no studied until now. Note that the moments of the UGHN distribution cannot be obtained in a closed form, but its QF is given in a very manageable form. In this paper, we parameterize the UGHN distribution in terms of its QF to evaluate the relationship of covariates on a quantile using the distribution of the response variable.

The main objective of the paper is to propose, derive, and apply a UGHN quantile regression model as an alternative to the existing quantile regression models. Our secondary objectives are: (i) to estimate the model parameters with the maximum likelihood (ML) method; (ii) to provide a diagnostic tool based on residuals for model checking; (iii) to conduct Monte Carlo simulation studies to assess the performance of our methodology; and (iv) to apply the obtained results to real-world data.

This paper consists of six sections. In Sect. 2, we introduce and characterize the UGHN distribution in terms of its QF. The estimation method of parameters for the UGHN distribution, based on the ML method, is discussed also here. The UGHN quantile regression model is formulated in Sect. 3. Monte Carlo simulations are conducted in Section 4 for evaluating the statistical performance of the ML estimators as well as the coverage probability of the asymptotic confidence intervals for the parameters in two regression frameworks. Two applications using the UGHN and Kumaraswamy quantile regression models are proposed in Sect. 5. Finally, Sect. 6 gives the conclusions of this paper and ideas for further investigation.

2 The unit generalized half-normal distribution

In this section, we provide some characteristics of the UGHN distribution and its quantile parameterization.

2.1 Cumulative distribution, probability density, and quantile functions

The UGHN distribution is obtained from the transformation $Y = \exp(-X)$, where $X \sim \text{GHN}(\alpha, \theta)$ denotes a generalized half-normal distributed random variable (Cooray and Ananda 2008). The corresponding cumulative distribution (CDF) and probability density (PDF) functions are written, respectively, as

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

$$\times \exp\left\{-\frac{1}{2}\left[-\frac{\log(y)}{\alpha}\right]^{2\theta}\right\},$$

$$F(y;\alpha,\theta) = 2\Phi\left[-\left(-\frac{\log(y)}{\alpha}\right)^{\theta}\right], \quad 0 < y < 1, \qquad (1)$$

where $\alpha > 0, \theta > 0$ are the model parameters, and Φ is the standard normal CDF. Note that a random variable Y is UGHN distributed on the unit interval, that is, on (0,1), if its log transformation, $-\log(Y)$ namely, is GHN(α,θ) distributed. The distribution with CDF stated in (1) is a strong competitor of well-known unit distributions such as the beta, Johnson and Kumaraswamy distributions. In addition, the quantile function of the UGHN distribution has a simple expression given by $y_\tau = Q(\tau;\alpha,\theta) = \exp\{-\alpha\left[-\Phi^{-1}(\tau/2)\right]^{1/\theta}\}$, for $0 < \tau < 1$, where Φ^{-1} is the inverse standard normal CDF, that is, the standard normal QF. Hence, if τ is a uniformly distributed random variable on (0,1), then Y_τ is a UGHN distributed random variable.

2.2 Shapes of the UGHN distribution

The first derivatives of the UGHN PDF and hazard rate function (HRF), defined by $h(y;\alpha,\beta) = f(y;\alpha,\beta)/[1 - F(y;\alpha,\beta)]$, are

$$\frac{\mathrm{d}}{\mathrm{d}y} f(y;\alpha,\theta) = -\frac{f(y;\alpha,\theta)}{y\log(y)} \zeta_1(y;\alpha,\theta),$$
$$\frac{\mathrm{d}}{\mathrm{d}y} h(y;\alpha,\theta) = -\frac{h(y;\alpha,\theta)}{\pi y\log(y)\left[1 - F(y;\alpha,\theta)\right]} \zeta_2(y;\alpha,\theta),$$

respectively, where $\zeta_2(y;\alpha,\theta) = -(\pi y\log(y)/\sqrt{2})\, f(y;\alpha,\theta) + (\pi\sqrt{2}/2)\left[1 - F(y;\alpha,\theta)\right]\zeta_1(y;\alpha,\theta)$ and $\zeta_1(y;\alpha,\theta) = 1 - \alpha + \alpha\left[-\theta^{-1}\log(y)\right]^{2\alpha} + \log(y)$.

Clearly, $\mathrm{d}f(y;\alpha,\theta)/\mathrm{d}y$ and $\mathrm{d}h(y;\alpha,\theta)/\mathrm{d}y$ have the same signs than $\zeta_1(y;\alpha,\theta)$ and $\zeta_2(y;\alpha,\theta)$, respectively. Note that, if $\zeta_1(y;\alpha,\theta)$ is increasing on y, then $\zeta_2(y;\alpha,\theta)$ also is increasing on y, that is, when $\zeta_1(y;\alpha,\theta)$ has positive sign. Observe that, analytically, it may be difficult to identify other signs of these equations. Thus, we sketch the plots of the UGHN PDF and HRF to see their possible shapes. Hence, we give some examples of $\zeta_1(y;\alpha,\theta)$ and $\zeta_2(y;\alpha,\theta)$ for fixed values of the parameters α and θ in Fig. 1. From this figure, we see that the UGHN PDF has U-shaped (bathtub), n-shaped (upside-down bathtub), unimodal, increasing, and decreasing shapes. Furthermore, the UGHN HRF has U-shaped, n-shaped, and increasing shapes. Figure 2 verifies these shapes of the model also and deals with the results of Fig. 1. Such shapes can be seen as distinguishing feature for the data modeling on the unit interval. Consequently, we can say that this new model can be more helpful for various data sets than other bounded models.

2.3 Maximum likelihood estimation

Next, we use the ML method to estimate the UGHN parameters. Let $Y_1,\ldots,Y_n$ be a random sample from the UGHN distribution with observed values $y_1,\ldots,y_n$, and the vector $\boldsymbol{\Theta} = (\theta,\alpha)^\top$ be the model parameters. Then, the corresponding log-likelihood function for $\boldsymbol{\Theta}$ is stated as

$$\begin{aligned}\ell &= \ell(\boldsymbol{\Theta}) \\ &= \frac{n}{2}\log\left(\frac{2}{\pi}\right) + n\log(\theta) - n\theta\log(\alpha) \\ &\quad - \sum_{i=1}^{n}\log(y_i) + (\theta-1)\sum_{i=1}^{n}\log\left[-\log(y_i)\right] \\ &\quad - \frac{1}{2\alpha^{2\theta}}\sum_{i=1}^{n}\log\left[-\log(y_i)\right]^{2\theta}. \end{aligned} \qquad (2)$$

Differentiating the function defined in (2) with respect to the corresponding parameter, the normal equations of the score vector are reached, with elements expressed in general as $U_\theta = \partial\ell/\partial\theta$ and $U_\alpha = \partial\ell/\partial\alpha = 0$ and specified as

$$\begin{aligned} U_\theta &= \frac{n}{\theta} - n\log(\alpha) + \sum_{i=1}^{n}\log\left[-\log(y_i)\right] \\ &\quad - \frac{1}{\alpha^{2\theta}}\sum_{i=1}^{n}\left[-\log(y_i)\right]^{2\theta}\log\left[-\frac{\log(y_i)}{\alpha}\right] = 0, \\ U_\alpha &= -\frac{n\theta}{\alpha} + \frac{\theta}{\alpha^{2\theta+1}}\sum_{i=1}^{n}\left[-\log(y_i)\right]^{2\theta} = 0. \end{aligned} \qquad (3)$$

The ML estimators, $\widehat{\theta}$ and $\widehat{\alpha}$ namely, are obtained from above the equations stated in (3) as

$$\widehat{\alpha} = \widehat{\alpha}(\theta) = \left(\frac{1}{n}\sum_{i=1}^{n}\left[-\log(y_i)\right]^{2\theta}\right)^{\frac{1}{2\theta}}.$$

Thus, $\widehat{\alpha}$ can be calculated as a function of $\widehat{\theta}$, say $\widehat{\alpha}(\theta)$. Substituting $\widehat{\alpha}$ in the expression given in (2), the corresponding profile log-likelihood function based on θ is generated as

$$\begin{aligned}\ell(\theta) &= \frac{n}{2}\log\left(\frac{2}{\pi}\right) + n\log(\theta) \\ &\quad - \frac{n}{2}\log\left[\frac{1}{n}\sum_{i=1}^{n}\left[-\log(y_i)\right]^{2\theta}\right] - \sum_{i=1}^{n}\log(y_i) \\ &\quad + (\theta-1)\sum_{i=1}^{n}\log\left[-\log(y_i)\right] - \frac{n}{2}. \end{aligned} \qquad (4)$$

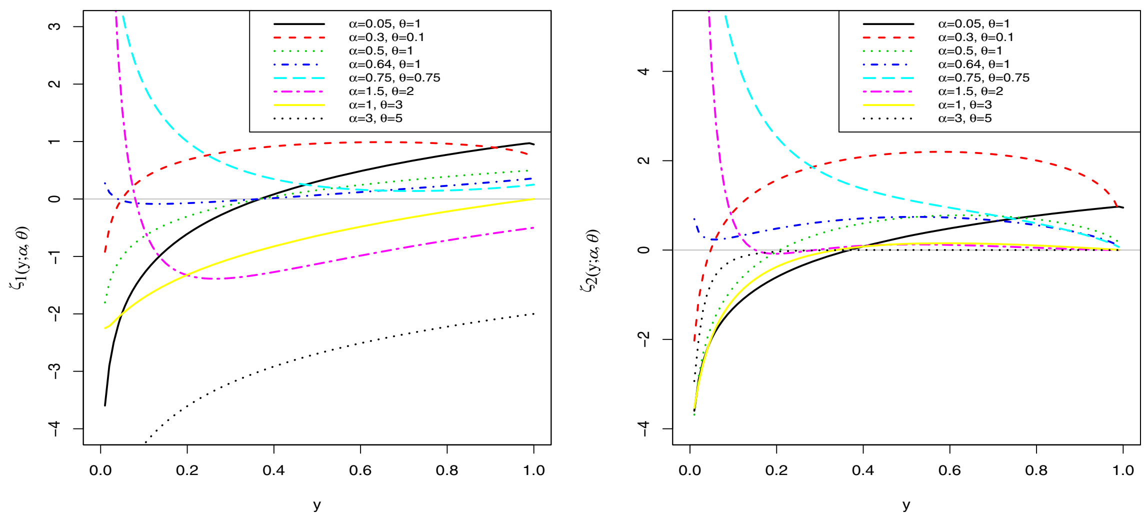

Fig. 1 Plots of the possible signs of the functions $\zeta_1(x, \alpha, \beta)$ and $\zeta_2(x, \alpha, \beta)$ for the indicated values of parameters

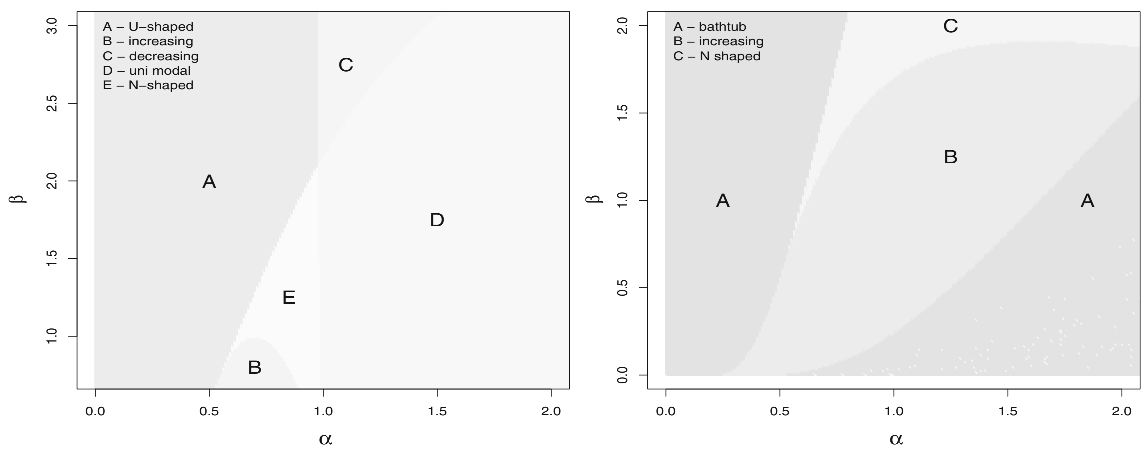

Fig. 2 Plots of the possible regions of the PDF (left) and HRF (right) for the indicated values of parameters

Hence, equating the formula given in (4) to zero, we obtain

$$\frac{\partial \ell(\theta)}{\partial \theta}=\frac{n}{\theta}+\sum_{i=1}^{n}\log\left[-\log\left(y_i\right)\right]-\frac{n\left(\sum_{i=1}^{n}\left[-\log\left(y_i\right)\right]^{2\theta}\log\left[-\log\left(y_i\right)\right]\right)}{\sum_{i=1}^{n}\left[-\log\left(y_i\right)\right]^{2\theta}}=0. \tag{5}$$

Note that numerical methods, such as the Newton–Raphson algorithm are required to find the solution of the nonlinear system defined in (5).

The corresponding inverse Fisher expected information matrix is established as

$$\begin{aligned}\mathcal{I}^{-1}(\theta,\alpha) &= \begin{bmatrix} \mathrm{E}\left(-\frac{\partial^2\ell}{\partial\theta^2}\right) & \mathrm{E}\left(-\frac{\partial^2\ell}{\partial\theta\partial\alpha}\right) \\ \mathrm{E}\left(-\frac{\partial^2\ell}{\partial\alpha\partial\theta}\right) & \mathrm{E}\left(-\frac{\partial^2\ell}{\partial\alpha^2}\right)\end{bmatrix}^{-1}_{|\theta=\widehat{\theta},\alpha=\widehat{\alpha}} \\ &= \begin{bmatrix} \mathrm{Var}\left(\widehat{\theta}\right) & \mathrm{Cov}\left(\widehat{\theta},\widehat{\alpha}\right) \\ \mathrm{Cov}\left(\widehat{\alpha},\widehat{\theta}\right) & \mathrm{Var}\left(\widehat{\alpha}\right)\end{bmatrix}. \end{aligned} \tag{6}$$

The elements of the matrix given in (6) are available from authors upon request. Notice that the most of the statistical properties such as moments, stochastic orderings, order statistics, and estimation procedures of the UGHN distribution have been obtained by Korkmaz (2020).

Assuming usual regularity conditions (Davidson 2003, pp. 118–119), observe that

$$\widehat{\boldsymbol{\Theta}} \mathrel{\dot\sim} \mathrm{N}_2(\boldsymbol{\Theta}, (\mathcal{I}(\boldsymbol{\Theta}))^{-1}), \tag{7}$$

where $\mathcal{I}(\boldsymbol{\Theta})$ is the expected information matrix defined in (6). Notice that approximate confidence intervals may be reached employing the expression given in (7), while for obtaining the information matrix stated in (6), we can utilize the observed Fisher information matrix formulated as

$$\mathcal{J}(\boldsymbol{\Theta}) = -\partial^2\ell(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}\,\partial\boldsymbol{\Theta}^\top, \tag{8}$$

with the elements given in (8) calculated from results above presented, evaluated at $\mathbf{\Theta} = \widehat{\mathbf{\Theta}}$. To test the hypotheses $\mathrm{H}_0\colon \mathbf{\Theta} = \mathbf{\Theta}_0$ versus $\mathrm{H}_1\colon \mathbf{\Theta} \neq \mathbf{\Theta}_0$, we may utilize the Wald and likelihood ratio tests. Their test statistics based on the observed information matrix are, respectively, expressed as

$$\begin{aligned} W &= (\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0)^\top \mathcal{J}(\widehat{\mathbf{\Theta}})(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0), \\ L &= -2\big(\ell(\mathbf{\Theta}_0) - \ell(\widehat{\mathbf{\Theta}})\big). \end{aligned} \tag{9}$$

If $n \to \infty$, such statistics converge to a random variable following a χ^2 distribution, with r being its degrees of freedom, denoted by χ^2_r, and r is the number of parameters under H_0. This hypothesis is rejected, at a significance level κ, if any of the statistics calculated using (9) is greater than $\chi^2_{r,1-\kappa}$, which denotes the $100(1-\kappa)$th quantile of the χ^2_r distribution.

2.4 The quantile UGHN distribution

Since the UGHN QF contains the standard normal QF, we can calculate this last one employing any mathematical program. Hence, the UGHN PDF and CDF may be parameterized using its QF.

To propose a quantile regression model based on the UGHN distribution, it must be parameterized in terms of the $100\,\tau$th quantile, $\mu = Q(\tau;\alpha,\theta)$ namely, such that α is written as

$$\alpha = k^{-1}(\mu) = -\log(\mu)\left[-\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^{-\frac{1}{\theta}}.$$

Then, the PDF and CDF of the reparameterized UGHN distribution are, respectively, given by

$$\begin{aligned} f(y;\theta,\mu,\tau) = {} & \sqrt{\frac{2}{\pi}}\frac{\theta\Phi^{-1}\left(\frac{\tau}{2}\right)}{\log(y)}\left[\frac{\log(y)}{\log(\mu)}\right]^{\theta} \\ & \times \exp\left\{-\frac{1}{2}\left[\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^2\left[\frac{\log(y)}{\log(\mu)}\right]^{2\theta}\right\}, \end{aligned} \tag{10}$$

$$\begin{aligned} F(y;\theta,\mu,\tau) = {} & 2\Phi\left[\Phi^{-1}\left(\frac{\tau}{2}\right)\left[\frac{\log(y)}{\log(\mu)}\right]^{\theta}\right], \\ & 0 < y < 1, \theta > 0, \mu \in (0,1), \tau \in (0,1), \end{aligned} \tag{11}$$

where θ, μ are the shape and quantile parameters, and τ is known, which is denoted by $Y \sim \mathrm{UGHN}(\alpha,\mu,\tau)$.

Figure 3 displays graphical plots of the UGHN PDF for different values of μ, α and τ. We consider the first decile ($\tau = 0.10$), the first quartile ($\tau = 0.25$), the median ($\tau = 0.50$), the third quartile ($\tau = 0.75$), and the ninth decile ($\tau = 0.90$). From this figure, we see that this PDF has bathtub and unimodal shapes. Other possible shapes of the PDF can be obtained with different parameters settings. Now, we are ready to introduce a quantile regression based on the UGHN distribution.

3 The UGHN quantile regression model

In this section, we provide a quantile regression model based on the UGHN distribution.

3.1 Model formulation

Since the moments of the UGHN distribution have no analytic expressions, it is difficult to relate the mean response and covariates. After defining the quantile UGHN distribution, we can formulate a quantile regression based on the UGHN distribution with its PDF given in (10).

Let $y_1,\ldots,y_n$ be observations of $Y \sim \mathrm{UGHN}(\theta,\mu_i,\tau)$, for $i \in \{1,\ldots,n\}$, where μ_i and θ are unknown parameters, and τ is known. Hence, the proposed quantile regression model is defined by means of

$$g(\mu_i) = \boldsymbol{x}_i\boldsymbol{\delta}^\top,$$

where $\boldsymbol{\delta} = \big(\delta_0,\delta_1,\ldots,\delta_p\big)^\top$ and $\boldsymbol{x}_i = (1,x_{i1},\ldots,x_{ip})^\top$ are the regression coefficient vector and the ith vector of values of the covariates. Note that g is the link function which connects the covariates and response variable. We obtain a median quantile regression for unit response variable, that is, $\tau = 0.5$. Since the UGHN distribution is a probability model supported on (0,1), the logit-link function is used and defined as

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right), \quad i \in \{1,\ldots,n\}.$$

Observe that the probit and log-log link functions can be employed to connect conditional quantiles of the response variable as well.

3.2 ML method for the regression coefficients

Now, we estimate the unknown parameters of the UGHN quantile regression model via the ML method. With this aim, we consider the logit link function stated as

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{x}_i\boldsymbol{\delta}^\top. \tag{12}$$

From the expression defined in (12), we obtain the relationship given by

$$\mu_i = \frac{\exp\left(\boldsymbol{x}_i\boldsymbol{\delta}^\top\right)}{1+\exp\left(\boldsymbol{x}_i\boldsymbol{\delta}^\top\right)}. \tag{13}$$

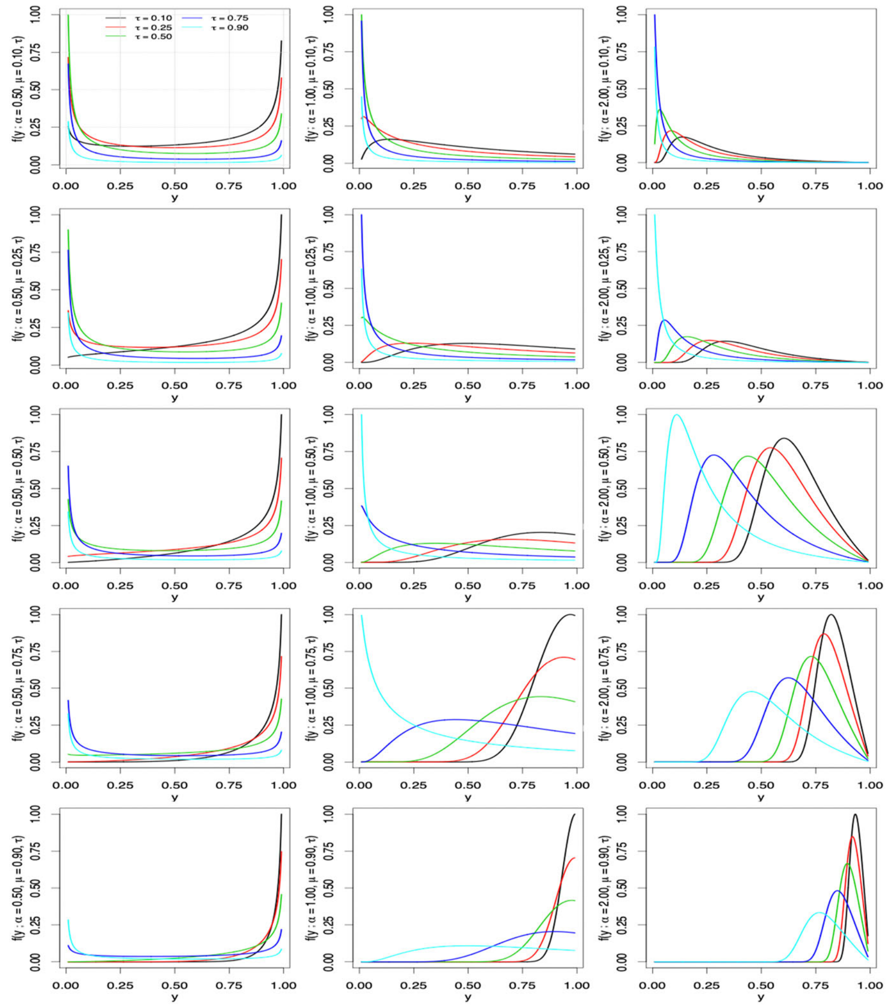

Fig. 3 Plots of the PDF of the quantile UGHN distribution for the indicated values of μ, α and τ

Let $Y_1, \ldots, Y_n$ be a sample of size n from $Y \sim \text{UGHN}(\theta, \mu_i, \tau)$, with observed values denoted by $y_1, \ldots, y_n$, where μ_i is given in (13), for $i \in \{1, \ldots, n\}$. Then, using the PDF stated in (10), the associated log-likelihood function is established as

$$\ell(\boldsymbol{\Theta}) = n \log\left(\frac{2}{\pi}\right) + n \log(\theta) + n \log\left[-\Phi^{-1}\left(\frac{\tau}{2}\right)\right] \\ + \theta \sum_{i=1}^{n} \log\left[\frac{\log(y_i)}{\log(\mu_i)}\right] - \sum_{i=1}^{n} \log\left[-\log(y_i)\right] \\ - \frac{1}{2}\left[\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^2 \sum_{i=1}^{n}\left[\frac{\log(y_i)}{\log(\mu_i)}\right]^{2\theta}, \tag{14}$$

where now $\boldsymbol{\Theta} = (\theta, \boldsymbol{\delta})^\top$ is the unknown parameter vector. The ML estimators of $\boldsymbol{\Theta}$, say $\widehat{\boldsymbol{\Theta}} = (\widehat{\theta}, \widehat{\boldsymbol{\delta}})^\top$, is obtained by maximizing $\ell(\boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta}$. Following usual routine, the first derivatives of $\ell(\boldsymbol{\Theta})$ for estimating the model parameters are given by

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log\left[\frac{\log(y_i)}{\log(\mu_i)}\right] - \left[\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^2 \\ \times \sum_{i=1}^{n}\left[\frac{\log(y_i)}{\log(\mu_i)}\right]^{2\theta} \log\left[\frac{\log(y_i)}{\log(\mu_i)}\right], \tag{15}$$

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \delta_r} = -\theta \sum_{i=1}^{n} \frac{1}{\mu_i \log \mu_i} \frac{\partial \mu_i}{\partial \delta_r} + \theta \left[\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^2 \\ \times \sum_{i=1}^{n}\left[\frac{\log(y_i)}{\log(\mu_i)}\right]^{2\theta} \frac{1}{\mu_i \log \mu_i} \frac{\partial \mu_i}{\partial \delta_r}, \tag{16}$$

where $\partial \mu_i / \partial \delta_r = \mu_i (1 - \mu_i) x_{ir}$, with $i \in \{1, \ldots, n\}$ and $r \in \{1, \ldots, p\}$. Since the equations stated in (15) and (16) contain nonlinear functions according to model parameters, they must be solved by numerical methods. For example, we may employ the `optim` function of the `R` software to maximize the function defined in (14) directly. We note that, when $\tau = 0.5$, the corresponding solutions to the equations are equivalent to modeling the conditional median.

Similarly to the case of no covariates, based on usual regularity conditions, the asymptotic distribution of $(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$ is multivariate normal, that is, $(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \dot{\sim} \mathrm{N}_{p+1}(\mathbf{0}, \mathcal{I}^{-1})$, where, as mentioned, $\mathcal{I}^{-1}$ is the inverse expected information matrix. In practice, as also mentioned, one utilizes the $(p+1) \times (p+1)$ observed information matrix $\mathcal{J}$ instead of $\mathcal{I}$. The elements of this observed information matrix are calculated numerically by any mathematical software. The `optim` function of `R` also provides asymptotic standard errors (SEs) of the estimators numerically. Then, approximate $100(1-\kappa)\%$ confidence intervals for $\boldsymbol{\Theta}$ may be determined by $\widehat{\theta}_r \pm z_{\kappa/2} s(\widehat{\theta}_r)$, where $z_{\kappa/2}$ is the upper $100(\kappa/2)$th percentile of the standard normal distribution and $s(\widehat{\theta}_r)$ is the estimated SE of the estimator of θ_r, which may be obtained from the rth diagonal element of $\mathcal{J}^{-1}$.

3.3 Model checking

After model fitting, we must evaluate whether the regression model is suitable to data or not. In this context, the residual analysis plays a relevant role for the model checking or validation. In order to make this, we focus on the Cox-Snell residual (Cox and Snell 1968) that is defined for observation i by

$$\widehat{\eta}_i = -\log\left[1 - F(y_i; \widehat{\theta}, \widehat{\mu}_i, \tau)\right], \quad i \in \{1, \ldots, n\},$$

where $F(y_i; \widehat{\theta}, \widehat{\mu}_i, \tau)$ is the estimated CDF of the quantile UGHN distribution given in (11) and $\widehat{\mu}_i$ is defined in (13). If the model fits to the data accordingly, the residual $\widehat{\eta}_i$ is an observation of an exponentially distributed random variable with scale parameter equal to one. Then, the Cox-Snell residual may be checked by the empirical quantile versus theoretical quantile (QQ) plot with simulated envelopes proposed by Atkinson (1981). If the model is fitted properly to the data, then no more than $\upsilon \times 100\%$ of the observations are expected to appear outside the envelope bands.

4 Simulation study

In this section, we present the results of the Monte Carlo simulations which are helpful to assess some statistical properties of the ML estimators, as the bias and root mean-squared error (RMSE), of the parameters of the UGHN quantile regression model. Also, we study the coverage probability (CP) of the 95% confidence interval ($\mathrm{CP}_{95\%}$) based on asymptotic normality of the ML estimators.

4.1 Setting for the simulation

We consider sample sizes $n \in \{20, 50, 100, 200, 300\}$; $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$; and $\theta \in \{0.5, 1.0, 2.0\}$ on two regression frameworks formulated as:

(i) $\mathrm{logit}(\mu_i) = \delta_0 + \delta_1 z_{i1}$ for $\delta_0 = 1.0$, $\delta_1 = 2.0$ and $z_{i1} \sim \mathrm{N}(0, 1)$.

(ii) $\mathrm{logit}(\mu_i) = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2}$ for $\delta_0 = 1.0$, $\delta_1 = 1.0$, $\delta_2 = 2.0$, $z_{i1} \sim \mathrm{N}(0, 1)$ and $z_{i2} \sim \mathrm{N}(0, 1)$.

For each (n, τ, θ) and the two regression frameworks stated in (i) and (ii) above, with the covariates remaining constant throughout the simulations, $M = 5{,}000$ Monte Carlo replicates were simulated employing the `SAS Data-Step`, while parameter estimates were obtained by the quasi-Newton method in `PROC SAS/NLMIXED` (SAS Institute Inc. 2018). The values of the response variable, given the values of n, τ, θ and the covariates, are generated as

$$y_i = \exp\left\{-\alpha_i \left[-\Phi^{-1}\left(\frac{u_i}{2}\right)\right]^{1/\theta}\right\},$$

where u_i is an observation of $U \sim \mathrm{U}(0, 1)$ and

$$\alpha_i = k^{-1}(\mu_i) = -\log(\mu_i)\left[-\Phi^{-1}\left(\frac{\tau}{2}\right)\right]^{-\frac{1}{\theta}}.$$

The empirical bias, RMSE, and CP are calculated, respectively, by

$$\mathrm{Bias}(\widehat{\varrho}) = \frac{1}{M}\sum_{i=1}^{M}(\widehat{\varrho}_i - \varrho),$$

$$\mathrm{RMSE}(\widehat{\varrho}) = \left[\frac{1}{M}\sum_{i=1}^{B}(\widehat{\varrho}_i - \varrho)^2]\right]^{\frac{1}{2}},$$

$$\mathrm{CP}_{95\%}(\widehat{\varrho}) = \frac{1}{M}\sum_{i=1}^{M}\mathbb{1}\left[\widehat{\varrho} \pm 1.96 \times \mathrm{SE}(\widehat{\varrho})\right],$$

where $\varrho = \alpha, \delta_0, \delta_1$ or δ_2, $\mathbb{1}$ is the indicator function, and $\mathrm{SE}(\widehat{\varrho})$ is the corresponding estimated SE.

4.2 Results of the simulation

Tables 1, 2, and 3, as well as Tables 4, 5 and 6, present the results for the first and second regression models, respectively. Such tables report a small bias when estimating θ and $\boldsymbol{\delta}$ for all settings considered in this study. In addition, the estimated RMSE is small and quickly approaches to zero as n, the sample size, increases. Larger values of the bias and RMSE are detected as the quantiles are distant from $\tau = 0.5$, either from the left or right, that is, the values $\tau \in \{0.1, 0.25, 0.75, 0.9\}$. Moreover, for all scenarios, the CPs tend to the nominal confidence coefficient, that is, 95%, as n increases. Observe that the empirical results based on the Monte Carlo simulations for large sample are coherent with the asymptotic theoretical results presented at the end of Subsect. 3.2 and just before the diagnostic analysis.

Table 1 Empirical bias, RMSE, and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 2.0$, and $\alpha = 0.5$) with simulated data

τ	n	Bias			RMSE			$CP_{95\%}$		
		δ_0	δ_1	α	δ_0	δ_1	α	δ_0	δ_1	α
0.10	20	0.1356	0.0645	0.0632	0.4922	0.5945	0.1348	0.8685	0.8969	0.9423
	50	0.0540	0.0247	0.0229	0.2854	0.3217	0.0685	0.9211	0.9302	0.9463
	100	0.0261	0.0113	0.0109	0.1990	0.2177	0.0447	0.9329	0.9390	0.9488
	200	0.0130	0.0069	0.0050	0.1391	0.1499	0.0302	0.9415	0.9442	0.9501
	300	0.0082	0.0051	0.0033	0.1128	0.1200	0.0242	0.9456	0.9459	0.9534
0.25	20	0.0655	0.0812	0.0639	0.4844	0.6018	0.1353	0.8974	0.8956	0.9412
	50	0.0262	0.0308	0.0230	0.2870	0.3233	0.0683	0.9333	0.9311	0.9464
	100	0.0125	0.0142	0.0110	0.2012	0.2182	0.0447	0.9419	0.9394	0.9494
	200	0.0068	0.0082	0.0050	0.1417	0.1502	0.0302	0.9452	0.9428	0.9510
	300	0.0041	0.0060	0.0033	0.1148	0.1201	0.0242	0.9471	0.9475	0.9539
0.50	20	−0.0538	0.1191	0.0643	0.6050	0.6348	0.1353	0.9183	0.8928	0.9403
	50	−0.0199	0.0436	0.0230	0.3623	0.3315	0.0676	0.9386	0.9313	0.9457
	100	−0.0099	0.0202	0.0110	0.2536	0.2220	0.0442	0.9470	0.9403	0.9497
	200	−0.0033	0.0109	0.0050	0.1790	0.1525	0.0299	0.9453	0.9422	0.9512
	300	−0.0027	0.0078	0.0033	0.1448	0.1217	0.0239	0.9495	0.9463	0.9527
0.75	20	−0.2473	0.1988	0.0634	0.9477	0.7371	0.1336	0.9223	0.8914	0.9334
	50	−0.0910	0.0680	0.0224	0.5423	0.3557	0.0657	0.9390	0.9310	0.9443
	100	−0.0438	0.0312	0.0107	0.3741	0.2330	0.0430	0.9466	0.9402	0.9482
	200	−0.0187	0.0159	0.0049	0.2616	0.1590	0.0290	0.9501	0.9425	0.9503
	300	−0.0132	0.0112	0.0033	0.2117	0.1265	0.0233	0.9506	0.9474	0.9523
0.90	20	−0.5733	0.3282	0.0667	1.5580	0.9942	0.1313	0.9233	0.8888	0.9038
	50	−0.2482	0.0921	0.0286	0.8230	0.4139	0.0659	0.9358	0.9263	0.9307
	100	−0.1567	0.0291	0.0179	0.5572	0.2536	0.0440	0.9385	0.9432	0.9361
	200	−0.1061	0.0047	0.0125	0.3890	0.1713	0.0304	0.9408	0.9428	0.9344
	300	−0.0999	−0.0001	0.0114	0.3211	0.1377	0.0251	0.9330	0.9414	0.9279

5 Applications with real-world data

In this section, we present two real analyses to show potential applications of the proposed quantile regression model. We compare this model with the Kumaraswamy quantile regression. The data sets are provided in Appendix.

5.1 Kumaraswamy regression model

For the purpose of comparison, in addition to the UGHN quantile regression model, we also consider the Kumaraswamy quantile regression model. The Kumaraswamy regression model introduced has PDF and CDF given, respectively, by

$$f(y;\theta,\mu,\tau) = \frac{\theta \log(1-\tau)}{\log\left(1-\mu^{\theta}\right)} y^{\theta-1}\left(1-y^{\theta}\right)^{\frac{\log(1-\tau)}{\log(1-\mu^{\theta})}-1}, \quad y \in (0,1),$$

$$F(y;\theta,\mu,\tau) = 1-\left(1-y^{\theta}\right)^{\frac{\log(1-\tau)}{\log(1-\mu^{\theta})}}, \quad y \in (0,1),$$

where $\mu \in (0,1)$ is the quantile parameter and $\theta > 0$, with $\tau \in (0,1)$ being a known value.

5.2 The stack loss data

This data set, extracted from https://support.sas.com/rnd/app/stat/examples/, corresponds to the operation of a plant where ammonia is oxidized to nitric acid (Brownlee 1965; Yu and Moyeed 2001). The oxidation is measured on 21 consecutive days and the response variable, `StackLoss` namely, is the percentage of ammonia lost (times 10), and there are three explanatory variables: `AirFlow`, which measures the air flow to the plant; `WaterTemp`, which measures the cooling water inlet temperature; and `AcidConc`, which measures the acid concentration. The regression structure assumed for μ_i is given by

$$\begin{aligned}\operatorname{logit}(\mu_i) &= \delta_0 + \delta_1 \texttt{AirFlow}_i + \delta_2 \texttt{WaterTemp}_i \\ &\quad + \delta_3 \texttt{AcidConc}_i, \quad i \in \{1,\ldots,21\}.\end{aligned}$$

Tables 7 and 8, respectively, display the ML estimates and their standard errors (SEs), and model selection criteria for the UGHN and Kumaraswamy quantile regressions models. Each of the likelihood-based statistics for the model selection

Table 2 Empirical bias, RMSE, and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 2.0$, and $\alpha = 1.0$) with simulated data

τ	n	Bias			RMSE			$CP_{95\%}$		
		δ_0	δ_1	α	δ_0	δ_1	α	δ_0	δ_1	α
0.10	20	0.0767	0.0220	0.1256	0.2472	0.2859	0.2691	0.8694	0.9017	0.9422
	50	0.0300	0.0093	0.0456	0.1431	0.1590	0.1369	0.9217	0.9327	0.9464
	100	0.0145	0.0043	0.0218	0.0996	0.1082	0.0895	0.9349	0.9396	0.9490
	200	0.0072	0.0028	0.0100	0.0696	0.0747	0.0604	0.9415	0.9442	0.9501
	300	0.0046	0.0021	0.0066	0.0564	0.0599	0.0483	0.9455	0.9464	0.9535
0.25	20	0.0421	0.0300	0.1271	0.2409	0.2882	0.2700	0.8974	0.9003	0.9414
	50	0.0162	0.0124	0.0460	0.1433	0.1595	0.1366	0.9310	0.9327	0.9463
	100	0.0077	0.0057	0.0220	0.1005	0.1084	0.0894	0.9404	0.9402	0.9492
	200	0.0041	0.0034	0.0101	0.0708	0.0748	0.0603	0.9449	0.9432	0.9511
	300	0.0025	0.0026	0.0067	0.0574	0.0599	0.0483	0.9468	0.9477	0.9539
0.50	20	−0.0135	0.0449	0.1279	0.2931	0.2971	0.2698	0.9131	0.8964	0.9408
	50	−0.0058	0.0178	0.0461	0.1793	0.1624	0.1352	0.9364	0.9323	0.9458
	100	−0.0030	0.0083	0.0221	0.1262	0.1099	0.0885	0.9455	0.9404	0.9496
	200	−0.0008	0.0046	0.0101	0.0893	0.0758	0.0598	0.9449	0.9422	0.9512
	300	−0.0008	0.0033	0.0067	0.0723	0.0606	0.0479	0.9489	0.9465	0.9527
0.75	20	−0.0965	0.0710	0.1265	0.4383	0.3214	0.2662	0.9078	0.8913	0.9384
	50	−0.0383	0.0268	0.0453	0.2644	0.1707	0.1316	0.9325	0.9296	0.9458
	100	−0.0188	0.0125	0.0217	0.1849	0.1144	0.0862	0.9416	0.9396	0.9488
	200	−0.0079	0.0065	0.0099	0.1301	0.0787	0.0581	0.9485	0.9420	0.9501
	300	−0.0056	0.0046	0.0066	0.1054	0.0628	0.0466	0.9491	0.9471	0.9527
0.90	20	−0.2003	0.1099	0.1206	0.6537	0.3713	0.2555	0.9005	0.8920	0.9258
	50	−0.0778	0.0394	0.0430	0.3799	0.1861	0.1250	0.9294	0.9286	0.9436
	100	−0.0378	0.0183	0.0205	0.2626	0.1225	0.0817	0.9393	0.9412	0.9452
	200	−0.0166	0.0091	0.0094	0.1832	0.0838	0.0550	0.9479	0.9443	0.9511
	300	−0.0116	0.0064	0.0063	0.1485	0.0667	0.0442	0.9500	0.9479	0.9525

criteria is given as follows. The negative value of the log-likelihood is defined as

$$\text{Neg2LogLike} = -2\log(L).$$

The Akaike information criterion (Akaike 1974) is stated as

$$\text{AIC} = -2\log(L) + 2p.$$

The Schwarz Bayesian information criterion (Schwarz et al. 1978) is expressed as

$$\text{BIC} = -2\log(L) + p\log(n).$$

The Hannan and Quinn information criterion (Hannan and Quinn 1979) is formulated by

$$\text{HQIC} = -2\log(L) + 2p\log[\log(n)],$$

where n is the sample size, p is the number of parameters, and L is the maximized likelihood function.

For all quantile levels, the log-likelihood values of the UGHN quantile regression model are greater than those of the Kumaraswamy quantile regression model with all the smallest likelihood-based statistics. The largest log-likelihood value of the UGHN quantile regression is found when the quantile regression model is based on $\tau = 0.10$. For all τ, the `AirFlow` and `WaterTemp` covariates are obtained as statistically significant at the usual significance levels whereas the `AcidConc` covariate is statistically not significant. In addition, the coefficients δ_1 and δ_2 are obtained as with positive sign for all τ. Hence, when the measurements of the `AirFlow` and `WaterTemp` covariates increase, the percentage of the ammonia lost (`Stackloss`) increases also. The `AcidConc` is not significant for the related percentage.

To support these inferences, the model's fit plays an important role. With this aim, we construct QQ plots with simulated envelopes for the Cox-Snell residuals in Figure 4. This figure indicates a good fit of the UGHN quantile regression model to the `Stackloss` data set.

Table 3 Empirical bias, RMSE and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 2.0$ and $\alpha = 2.0$) with simulated data

τ	n	Bias			RMSE			$CP_{95\%}$		
		δ_0	δ_1	α	δ_0	δ_1	α	δ_0	δ_1	α
0.10	20	0.0405	0.0086	0.2505	0.1247	0.1411	0.5375	0.8723	0.9041	0.9426
	50	0.0158	0.0039	0.0910	0.0718	0.0791	0.2738	0.9224	0.9326	0.9465
	100	0.0076	0.0018	0.0436	0.0499	0.0540	0.1789	0.9364	0.9398	0.9489
	200	0.0038	0.0012	0.0200	0.0348	0.0373	0.1208	0.9417	0.9443	0.9500
	300	0.0024	0.0010	0.0132	0.0282	0.0299	0.0967	0.9467	0.9469	0.9536
0.25	20	0.0233	0.0125	0.2535	0.1209	0.1419	0.5396	0.8972	0.9040	0.9416
	50	0.0089	0.0054	0.0920	0.0718	0.0793	0.2733	0.9307	0.9335	0.9463
	100	0.0042	0.0025	0.0440	0.0503	0.0541	0.1788	0.9404	0.9398	0.9494
	200	0.0022	0.0015	0.0202	0.0354	0.0373	0.1207	0.9441	0.9435	0.9511
	300	0.0014	0.0012	0.0134	0.0287	0.0299	0.0966	0.9466	0.9477	0.9539
0.50	20	−0.0038	0.0192	0.2555	0.1457	0.1452	0.5394	0.9116	0.8990	0.9407
	50	−0.0019	0.0079	0.0924	0.0894	0.0806	0.2706	0.9348	0.9329	0.9458
	100	−0.0011	0.0037	0.0443	0.0630	0.0548	0.1772	0.9453	0.9396	0.9492
	200	−0.0002	0.0021	0.0203	0.0446	0.0378	0.1196	0.9450	0.9421	0.9510
	300	−0.0002	0.0015	0.0135	0.0361	0.0302	0.0958	0.9478	0.9468	0.9525
0.75	20	−0.0430	0.0297	0.2534	0.2147	0.1538	0.5325	0.9025	0.8917	0.9392
	50	−0.0175	0.0117	0.0912	0.1312	0.0841	0.2635	0.9296	0.9290	0.9463
	100	−0.0087	0.0055	0.0437	0.0921	0.0568	0.1725	0.9397	0.9397	0.9486
	200	−0.0036	0.0029	0.0200	0.0650	0.0392	0.1163	0.9482	0.9417	0.9503
	300	−0.0026	0.0021	0.0134	0.0526	0.0313	0.0932	0.9495	0.9471	0.9530
0.90	20	−0.0905	0.0437	0.2438	0.3144	0.1697	0.5139	0.8937	0.8877	0.9328
	50	−0.0364	0.0167	0.0875	0.1874	0.0904	0.2509	0.9247	0.9262	0.9451
	100	−0.0178	0.0079	0.0418	0.1304	0.0605	0.1638	0.9368	0.9390	0.9459
	200	−0.0078	0.0040	0.0192	0.0913	0.0416	0.1101	0.9463	0.9443	0.9512
	300	−0.0054	0.0028	0.0129	0.0741	0.0332	0.0884	0.9495	0.9479	0.9526

5.3 Educational attainment of the OECD countries data set

We consider a second data set related to education values of OECD countries. The measurement unit of the data set is calculated as a percentage. We relate linearly the educational attainment values (EAV) of the OECD countries and better life indexes of such countries as covariates employing a quantile regression structure. These data have been analyzed by Altun (2021) and the covariates are related to labor market insecurity (LMI) and homicide rate (HR). The regression structure assumed for μ_i is formulated as

$$\text{logit}(\mu_i) = \delta_0 + \delta_1 \texttt{LMI}_i + \delta_2 \texttt{HR}_i, \quad i \in \{1, \ldots, 37\}.$$

Tables 9 and 10, respectively, report the ML estimates and their associated SEs, as well as model selection criteria for the UGHN and Kumaraswamy quantile regressions. For all quantile levels, all regression coefficients are statistically significant. Therefore, the LMI and HR covariates have affected the response variable EAV in a statistically significant form at the usual levels. This affection has been seen as opposite sign, that is, as LMI and HR increase, the EAV decreases.

The log-likelihood values of the UGHN quantile regression are once again greater than those of the Kumaraswamy quantile regression with all the smallest likelihood-based statistics. The largest log-likelihood value of the UGHN quantile regression model was detected when the regression is based on the quantile $\tau = 0.10$. This inference is also supported by the QQ plots for the Cox-Snell residuals with simulated envelopes shown in Figure 5.

6 Concluding remarks

In this paper, we have proposed and derived a new quantile regression model based on the unit generalized half-normal distribution. In the proposed regression, the covariates were related to the quantile of the response variable by the logit link function. Different aspects of estimation, inference, based on the maximum likelihood method, and diagnostics, using the Cox-Snell residual, were considered.

Table 4 Empirical bias, RMSE and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 1.0$, $\delta_2 = 2.0$ and $\alpha = 0.5$) with simulated data

τ	n	Bias				RMSE				$CP_{95\%}$			
		δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α
0.10	20	0.1866	0.0322	0.0875	0.0874	0.5605	0.5910	0.6659	0.1570	0.8368	0.8882	0.8754	0.9276
	50	0.0727	0.0159	0.0353	0.0308	0.3052	0.3029	0.3345	0.0730	0.9090	0.9218	0.9254	0.9428
	100	0.0368	0.0101	0.0166	0.0151	0.2081	0.1994	0.2223	0.0472	0.9266	0.9350	0.9342	0.9442
	200	0.0162	0.0051	0.0115	0.0075	0.1441	0.1332	0.1501	0.0311	0.9370	0.9464	0.9410	0.9482
	300	0.0091	0.0028	0.0098	0.0049	0.1174	0.1078	0.1234	0.0249	0.9404	0.9472	0.9430	0.9476
0.25	20	0.0829	0.0444	0.1122	0.0883	0.5349	0.5967	0.6766	0.1579	0.8782	0.8846	0.8764	0.9264
	50	0.0309	0.0205	0.0446	0.0311	0.3026	0.3046	0.3375	0.0733	0.9268	0.9224	0.9214	0.9420
	100	0.0158	0.0123	0.0210	0.0152	0.2088	0.2002	0.2231	0.0474	0.9380	0.9336	0.9330	0.9438
	200	0.0056	0.0061	0.0136	0.0075	0.1451	0.1335	0.1502	0.0312	0.9442	0.9456	0.9424	0.9486
	300	0.0020	0.0035	0.0112	0.0050	0.1184	0.1080	0.1234	0.0249	0.9458	0.9468	0.9436	0.9484
0.50	20	−0.0889	0.0703	0.1633	0.0884	0.6516	0.6185	0.7155	0.1575	0.9074	0.8850	0.8722	0.9228
	50	−0.0363	0.0291	0.0622	0.0310	0.3780	0.3093	0.3486	0.0729	0.9346	0.9216	0.9220	0.9380
	100	−0.0176	0.0163	0.0293	0.0151	0.2634	0.2025	0.2277	0.0471	0.9398	0.9334	0.9330	0.9416
	200	−0.0112	0.0081	0.0176	0.0074	0.1825	0.1344	0.1523	0.0309	0.9468	0.9454	0.9434	0.9450
	300	−0.0092	0.0048	0.0139	0.0049	0.1486	0.1087	0.1248	0.0247	0.9472	0.9454	0.9432	0.9472
0.75	20	−0.3577	0.1197	0.2648	0.0857	1.0317	0.6773	0.8348	0.1531	0.9064	0.8852	0.8686	0.9120
	50	−0.1357	0.0445	0.0937	0.0300	0.5711	0.3211	0.3783	0.0709	0.9322	0.9278	0.9202	0.9316
	100	−0.0659	0.0233	0.0435	0.0146	0.3918	0.2079	0.2411	0.0457	0.9350	0.9332	0.9326	0.9382
	200	−0.0346	0.0113	0.0241	0.0071	0.2680	0.1368	0.1587	0.0299	0.9456	0.9456	0.9450	0.9428
	300	−0.0249	0.0069	0.0181	0.0047	0.2167	0.1104	0.1292	0.0239	0.9490	0.9428	0.9444	0.9454
0.90	20	−0.8381	0.2128	0.4365	0.0926	1.7858	0.8510	1.1579	0.1547	0.8946	0.8816	0.8638	0.8587
	50	−0.3661	0.0623	0.1311	0.0400	0.9019	0.3515	0.4496	0.0742	0.9159	0.9234	0.9148	0.9035
	100	−0.2274	0.0267	0.0550	0.0252	0.6011	0.2212	0.2700	0.0485	0.9218	0.9310	0.9306	0.9080
	200	−0.1637	0.0046	0.0204	0.0182	0.4096	0.1441	0.1705	0.0332	0.9297	0.9400	0.9502	0.9114
	300	−0.1391	−0.0053	0.0142	0.0156	0.3311	0.1162	0.1346	0.0273	0.9333	0.9360	0.9480	0.9027

Table 5 Empirical bias, RMSE and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 1.0$, $\delta_2 = 2.0$ and $\alpha = 1.0$) with simulated data

τ	n	Bias				RMSE				$CP_{95\%}$			
		δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α
0.10	20	0.1082	0.0100	0.0309	0.1726	0.2794	0.2797	0.3106	0.3121	0.8404	0.8910	0.8832	0.9294
	50	0.0409	0.0060	0.0139	0.0613	0.1531	0.1492	0.1642	0.1457	0.9100	0.9238	0.9272	0.9422
	100	0.0205	0.0041	0.0066	0.0300	0.1043	0.0989	0.1102	0.0945	0.9268	0.9348	0.9354	0.9442
	200	0.0091	0.0021	0.0049	0.0149	0.0721	0.0663	0.0747	0.0623	0.9370	0.9466	0.9424	0.9480
	300	0.0052	0.0011	0.0044	0.0098	0.0587	0.0537	0.0615	0.0497	0.9404	0.9468	0.9426	0.9478
0.25	20	0.0573	0.0159	0.0424	0.1742	0.2623	0.2815	0.3138	0.3137	0.8778	0.8896	0.8836	0.9266
	50	0.0203	0.0082	0.0183	0.0619	0.1507	0.1497	0.1651	0.1465	0.9260	0.9238	0.9254	0.9422
	100	0.0101	0.0052	0.0088	0.0304	0.1042	0.0992	0.1104	0.0947	0.9378	0.9344	0.9330	0.9438
	200	0.0038	0.0027	0.0060	0.0150	0.0725	0.0664	0.0747	0.0624	0.9446	0.9468	0.9428	0.9488
	300	0.0017	0.0015	0.0051	0.0099	0.0591	0.0538	0.0615	0.0498	0.9460	0.9464	0.9428	0.9484
0.50	20	−0.0223	0.0262	0.0628	0.1741	0.3075	0.2864	0.3246	0.3128	0.8998	0.8880	0.8814	0.9244
	50	−0.0119	0.0119	0.0259	0.0620	0.1855	0.1510	0.1689	0.1458	0.9296	0.9224	0.9224	0.9400
	100	−0.0061	0.0070	0.0124	0.0303	0.1305	0.1000	0.1121	0.0941	0.9376	0.9338	0.9336	0.9418
	200	−0.0044	0.0036	0.0078	0.0149	0.0908	0.0668	0.0756	0.0618	0.9438	0.9458	0.9430	0.9452
	300	−0.0038	0.0021	0.0063	0.0099	0.0740	0.0541	0.0621	0.0493	0.9478	0.9450	0.9422	0.9476

Table 5 continued

τ	n	Bias				RMSE				$CP_{95\%}$			
		δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α
0.75	20	−0.1358	0.0428	0.0963	0.1688	0.4613	0.2975	0.3524	0.3044	0.8946	0.8880	0.8716	0.9182
	50	−0.0572	0.0177	0.0378	0.0602	0.2745	0.1543	0.1785	0.1417	0.9240	0.9260	0.9204	0.9346
	100	−0.0286	0.0097	0.0180	0.0294	0.1923	0.1019	0.1172	0.0913	0.9298	0.9348	0.9342	0.9394
	200	−0.0155	0.0049	0.0105	0.0144	0.1328	0.0678	0.0783	0.0599	0.9440	0.9460	0.9458	0.9434
	300	−0.0113	0.0030	0.0081	0.0095	0.1078	0.0548	0.0641	0.0478	0.9474	0.9440	0.9430	0.9456
0.90	20	−0.2678	0.0657	0.1421	0.1555	0.6911	0.3187	0.4038	0.2838	0.8800	0.8924	0.8684	0.9034
	50	−0.1085	0.0251	0.0530	0.0558	0.3958	0.1601	0.1958	0.1335	0.9206	0.9280	0.9198	0.9276
	100	−0.0536	0.0131	0.0250	0.0270	0.2723	0.1052	0.1263	0.0854	0.9298	0.9364	0.9318	0.9360
	200	−0.0275	0.0065	0.0137	0.0131	0.1864	0.0694	0.0833	0.0560	0.9440	0.9474	0.9466	0.9448
	300	−0.0194	0.0040	0.0102	0.0087	0.1508	0.0560	0.0678	0.0447	0.9456	0.9440	0.9434	0.9436

Table 6 Empirical bias, RMSE and 95% CP (true values: $\delta_0 = 1.0$, $\delta_1 = 1.0$, $\delta_2 = 2.0$ and $\alpha = 2.0$) with simulated data

τ	n	Bias				RMSE				$CP_{95\%}$			
		δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α	δ_0	δ_1	δ_2	α
0.10	20	0.0575	0.0036	0.0124	0.3431	0.1411	0.1377	0.1520	0.6220	0.8448	0.8932	0.8882	0.9294
	50	0.0216	0.0025	0.0060	0.1222	0.0769	0.0742	0.0815	0.2912	0.9110	0.9244	0.9278	0.9424
	100	0.0108	0.0018	0.0029	0.0600	0.0523	0.0493	0.0549	0.1889	0.9274	0.9352	0.9368	0.9442
	200	0.0048	0.0010	0.0023	0.0299	0.0361	0.0331	0.0373	0.1245	0.9372	0.9470	0.9430	0.9478
	300	0.0028	0.0005	0.0021	0.0197	0.0294	0.0268	0.0307	0.0994	0.9410	0.9464	0.9430	0.9480
0.25	20	0.0321	0.0065	0.0181	0.3465	0.1316	0.1382	0.1530	0.6257	0.8796	0.8926	0.8870	0.9268
	50	0.0113	0.0036	0.0082	0.1237	0.0754	0.0744	0.0818	0.2929	0.9258	0.9244	0.9258	0.9428
	100	0.0056	0.0024	0.0040	0.0607	0.0522	0.0494	0.0549	0.1895	0.9388	0.9352	0.9340	0.9440
	200	0.0021	0.0012	0.0028	0.0301	0.0362	0.0332	0.0373	0.1247	0.9448	0.9464	0.9430	0.9486
	300	0.0010	0.0007	0.0024	0.0199	0.0296	0.0269	0.0307	0.0995	0.9458	0.9458	0.9434	0.9488
0.50	20	−0.0066	0.0112	0.0273	0.3467	0.1519	0.1397	0.1568	0.6242	0.8974	0.8910	0.8828	0.9246
	50	−0.0045	0.0053	0.0117	0.1239	0.0923	0.0749	0.0834	0.2917	0.9264	0.9228	0.9230	0.9408
	100	−0.0024	0.0032	0.0057	0.0607	0.0651	0.0498	0.0557	0.1883	0.9370	0.9350	0.9344	0.9418
	200	−0.0019	0.0017	0.0037	0.0299	0.0453	0.0333	0.0377	0.1237	0.9438	0.9466	0.9434	0.9452
	300	−0.0017	0.0010	0.0030	0.0198	0.0370	0.0270	0.0310	0.0987	0.9478	0.9452	0.9416	0.9474
0.75	20	−0.0603	0.0180	0.0410	0.3375	0.2240	0.1431	0.1663	0.6096	0.8890	0.8908	0.8734	0.9208
	50	−0.0264	0.0078	0.0169	0.1208	0.1356	0.0761	0.0874	0.2839	0.9194	0.9242	0.9196	0.9360
	100	−0.0133	0.0044	0.0082	0.0590	0.0956	0.0506	0.0580	0.1827	0.9280	0.9360	0.9344	0.9392
	200	−0.0073	0.0022	0.0049	0.0289	0.0662	0.0338	0.0390	0.1198	0.9442	0.9462	0.9454	0.9442
	300	−0.0054	0.0014	0.0038	0.0192	0.0538	0.0273	0.0319	0.0956	0.9470	0.9448	0.9434	0.9458
0.90	20	−0.1207	0.0264	0.0579	0.3133	0.3295	0.1488	0.1830	0.5711	0.8692	0.8894	0.8662	0.9118
	50	−0.0508	0.0108	0.0230	0.1128	0.1940	0.0782	0.0942	0.2677	0.9132	0.9274	0.9174	0.9312
	100	−0.0254	0.0058	0.0110	0.0548	0.1349	0.0520	0.0619	0.1711	0.9282	0.9368	0.9308	0.9378
	200	−0.0132	0.0029	0.0062	0.0266	0.0927	0.0345	0.0413	0.1122	0.9438	0.9466	0.9450	0.9454
	300	−0.0093	0.0018	0.0047	0.0178	0.0752	0.0279	0.0337	0.0896	0.9452	0.9434	0.9420	0.9438

Table 7 ML estimates with SEs in parentheses for the indicated distribution and parameter with Stackloss data

τ	Distribution	$\widehat{\delta}_0$	$\widehat{\delta}_1$	$\widehat{\delta}_2$	$\widehat{\delta}_3$	$\widehat{\theta}$
0.10	UGHN	−6.4533	0.0333	0.1169	0.0010	12.3569
		(0.5090)	(0.0044)	(0.0135)	(0.0064)	(2.4559)
	Kumaraswamy	−6.2346	0.0467	0.0726	−0.0013	6.5158
		(0.8529)	(0.0118)	(0.0183)	(0.0109)	(1.0431)
0.25	UGHN	−6.2927	0.0328	0.1148	0.0008	12.3503
		(0.4978)	(0.0043)	(0.0133)	(0.0063)	(2.4538)
	Kumaraswamy	−6.1955	0.0495	0.0735	−0.0018	6.5809
		(0.8582)	(0.0121)	(0.0186)	(0.0110)	(1.0547)
0.50	UGHN	−6.0612	0.0320	0.1119	0.0005	12.3404
		(0.4853)	(0.0041)	(0.0130)	(0.0061)	(2.4506)
	Kumaraswamy	−6.1865	0.0523	0.0744	−0.0022	6.6447
		(0.8638)	(0.0123)	(0.0189)	(0.0110)	(1.0648)
0.75	UGHN	−5.7531	0.0309	0.1080	0.0002	12.3266
		(0.4749)	(0.0039)	(0.0127)	(0.0059)	(2.4456)
	Kumaraswamy	−6.1990	0.0550	0.0753	−0.0026	6.7000
		(0.8684)	(0.0124)	(0.0191)	(0.0110)	(1.0721)
0.90	UGHN	−5.3944	0.0298	0.1036	−0.0002	12.3106
		(0.4707)	(0.0038)	(0.0123)	(0.0056)	(2.4383)
	Kumaraswamy	−6.2206	0.0572	0.0760	−0.0030	6.7437
		(0.8719)	(0.0125)	(0.0194)	(0.0110)	(1.0764)

Table 8 Likelihood-based statistics for the indicated distribution and criteria with Stackloss data

τ	Distribution	Neg2LogLike	AIC	BIC	HQIC
0.10	UGHN	−100.4086	−90.4086	−85.1860	−89.2751
	Kumaraswamy	−93.9218	−83.9218	−78.6992	−82.7884
0.25	UGHN	−100.3941	−90.3941	−85.1715	−89.2607
	Kumaraswamy	−94.4329	−84.4329	−79.2103	−83.2995
0.50	UGHN	−100.3678	−90.3678	−85.1451	−89.2343
	Kumaraswamy	−94.9782	−84.9782	−79.7556	−83.8447
0.75	UGHN	−100.3212	−90.3212	−85.0986	−89.1878
	Kumaraswamy	−95.4819	−85.4819	−80.2593	−84.3485
0.90	UGHN	−100.2487	−90.2487	−85.0261	−89.1153
	Kumaraswamy	−95.8952	−85.8952	−80.6726	−84.7618

We have conducted Monte Carlo simulations for the proposed quantile regression model to evaluate the performance of the estimators under some settings. We have found that the model parameters are well estimated in terms of bias and root mean-squared error. Also, observe that the coverage probabilities tend to the nominal confidence coefficient as the sample size increases.

Two real data sets based on a chemical problem and the better life index of the countries have been analyzed with the proposed and Kumaraswamy quantile regression models. The results have indicated that the proposed quantile regression showed an excellent agreement with the two real-world data sets, being it an alternative to other competing models available in the literature on the topic. A comparison was performed according to information criteria and goodness-of-fit, showing their adequacy.

Our analyses have revealed that the unit generalized half-normal distribution has great potential for analyzing data restricted to the unit interval in the presence of covariates. The data analysis is facilitated by means of `R` codes, which were implemented by the authors of this paper and are available upon request.

Some aspects of future research are related to the incorporation of multivariate, functional, temporal, and spatial structures, as well as errors-in-variables and partial least squares, in the quantile regression framework and its influence diagnostics (Huerta et al. 2018; Figueroa-Zuniga et al. 2022b; Liu et al. 2022). In addition, Tobit and Cobb–Douglas

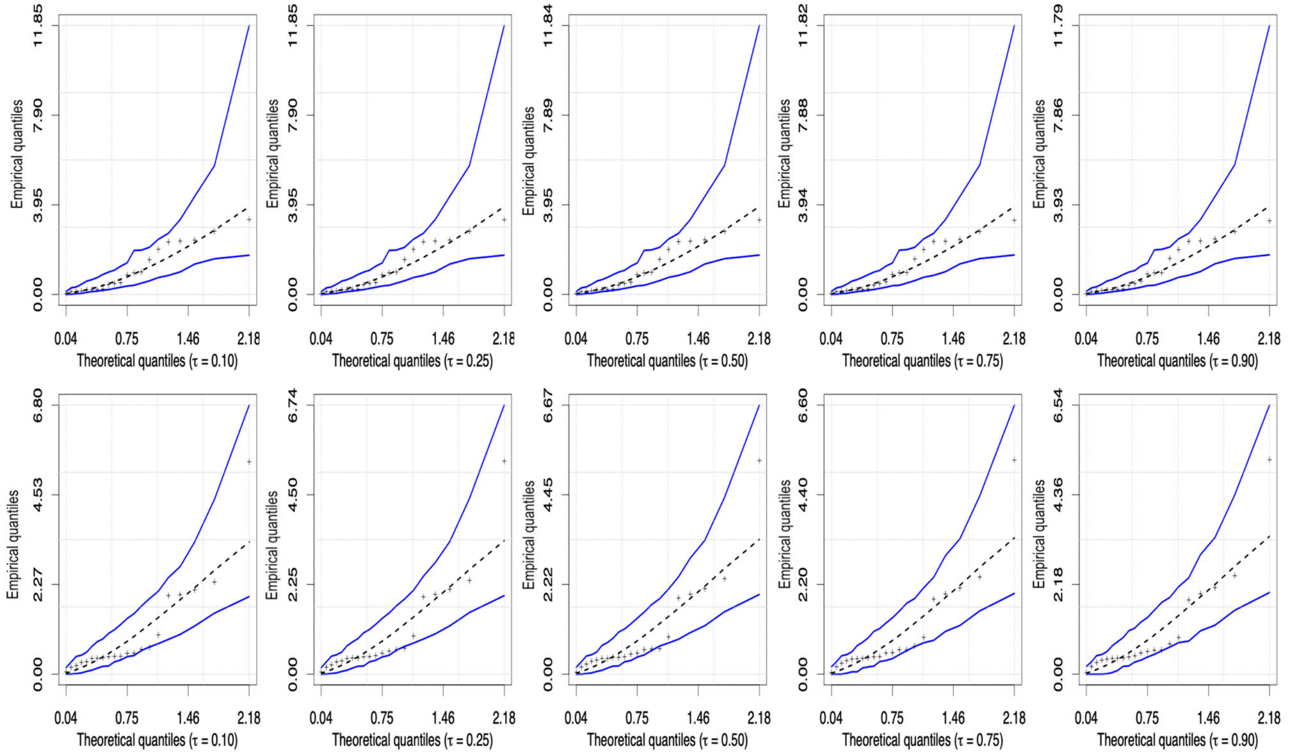

Fig. 4 QQ plots with envelopes of Cox-Snell residuals for the indicated 100 τth quantile level of the (first row) UGHN model and (second row) Kumaraswamy model with Stackloss data

Table 9 ML estimates with SEs in parentheses for the indicated distribution and parameter with education data

τ	Distribution	$\widehat{\delta}_0$	$\widehat{\delta}_1$	$\widehat{\delta}_2$	$\widehat{\theta}$
0.10	UGHN	1.6575	−0.1729	−0.0693	1.4840
		(0.1934)	(0.0351)	(0.0209)	(0.1910)
	Kumaraswamy	0.7753	−0.0520	−0.0191	6.8157
		(0.1714)	(0.0088)	(0.0158)	(1.1916)
0.25	UGHN	1.8881	−0.1613	−0.0656	1.4771
		(0.1874)	(0.0338)	(0.0192)	(0.1887)
	Kumaraswamy	1.2681	−0.0611	−0.0251	6.8511
		(0.1573)	(0.0100)	(0.0193)	(1.2010)
0.50	UGHN	2.2355	−0.1474	−0.0614	1.4662
		(0.1937)	(0.0325)	(0.0176)	(0.1851)
	Kumaraswamy	1.8452	−0.0726	−0.0346	6.8693
		(0.1605)	(0.0116)	(0.0242)	(1.2030)
0.75	UGHN	2.7329	−0.1331	−0.0574	1.4505
		(0.2258)	(0.0320)	(0.0161)	(0.1798)
	Kumaraswamy	2.4766	−0.0851	−0.0495	6.8545
		(0.2001)	(0.0134)	(0.0291)	(1.1882)
0.90	UGHN	3.3689	−0.1214	−0.0545	1.4317
		(0.2855)	(0.0331)	(0.0153)	(0.1744)
	Kumaraswamy	3.1228	−0.0969	−0.0702	6.8278
		(0.2721)	(0.0148)	(0.0301)	(1.1660)

Table 10 Likelihood-based statistics for the indicated distribution and criteria with education data

τ	Distribution	Neg2LogLike	AIC	BIC	HQIC
0.10	UGHN	−70.8264	−62.8264	−56.2761	−60.4959
	Kumaraswamy	−63.8678	−55.8678	−49.3175	−53.5372
0.25	UGHN	−70.3243	−62.3243	−55.7739	−59.9937
	Kumaraswamy	−63.9700	−55.9700	−49.4197	−53.6394
0.50	UGHN	−69.6207	−61.6207	−55.0704	−59.2902
	Kumaraswamy	−63.9677	−55.9677	−49.4174	−53.6371
0.75	UGHN	−68.7867	−60.7867	−54.2364	−58.4562
	Kumaraswamy	−63.8325	−55.8325	−49.2821	−53.5019
0.90	UGHN	−68.0432	−60.0432	−53.4929	−57.7126
	Kumaraswamy	−63.6619	−55.6619	−49.1116	−53.3313

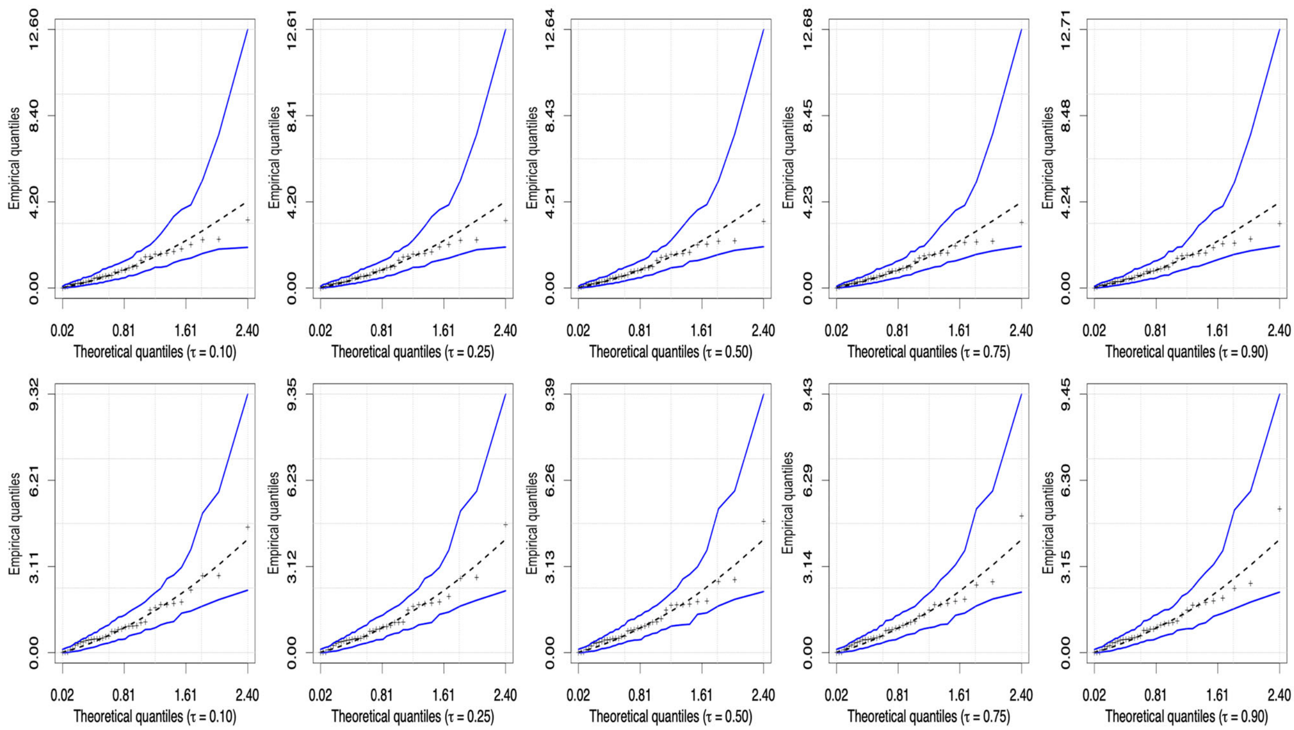

Fig. 5 QQ plots with envelopes of Cox-Snell residuals for the indicated 100 τth quantile level of the (first row) UGHN model and (second row) Kumaraswamy model) with education data

type frameworks can be considered in the topic of this study (de la Fuente et al. 2019; Martinez-Florez et al. 2020). Moreover, censored observations and frailty models could be investigated in the present context (Leao et al. 2018). The authors are analyzing these aspects associated with the present investigation and their findings are expected to be proposed promptly.

Acknowledgements The authors would like to thank the Editors and Reviewers for their constructive comments on an earlier version of this article which led to an improved presentation.

Author Contributions All authors have contributed equally to this work.

Funding This research was funded partially by FONDECYT grant number 1200525 (V. Leiva) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation.

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

Declarations

Conflict of interest All the authors in the paper have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent All authors have read and agreed to the published version of this article.

Appendix: Data sets

First data set

```
stack loss<-c(42,37,37,28,18,18,19,20,
15,14,14,13,11,12,8,7,8,8,9,15,15)
#percentage

airflow<-c(80,80,75,62,62,62,62,62,58,
58,58,58,58,58,50,50,50,50,50,56,70)

watertemp<-c(27,27,25,24,22,23,24,24,23,
18,18,17,18,19,18,18,19,19,20,20,20)

acidconc<-c(89,88,90,87,87,87,93,93,87,
80,89,88,82,93,89,86,72,79,80,82,91)
```

Second data set

```
y <-c(080,85,75,91,65,93,81,89,88,78,86,
72,83,78,80,87,60,94,87,89,79,37,77,77,
82,91,47,92,87,58,83,87,39,81,90,49,95,
43)
#EAV Percentage

x1 <-c(4.3,2.7,4.8,3.9,8.1,1.8,2.3,4,2.
7,5,2,17.4,4.8,2.6,2.1,2.6,8.1,1.5,2.4,
6.8,3.2,4.6,2.1,4.9,2.7,4.3,6.5,6.7,
4,17.3,5.7,1.8,13,2.6,3.8,4.9,3.6,26.5)
#LMI

x2 <-c(1,0.4,1,1.4,4.5,0.8,0.7,3.1,1.4,
0.6,0.4,1,1.2,0.9,0.6,1.7,0.8,0.3,1.1,
6.6,0.6,17.9,0.6,1.3,0.6,0.8,1,0.8,0.6,
0.6,1,0.5,1.7,0.2,4.9, 27.6,11.3,10)
#HR
```

References

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

Akgul FG (2021) Classical and Bayesian estimation of multicomponent stress-strength reliability for exponentiated Pareto distribution. Soft Comput 25:9185–9197

Altun E (2021) The log-weighted exponential regression model: alternative to the beta regression model. Commun in Stat-Theory Methods. https://doi.org/10.1080/03610926.2019.1664586

Altun E, Cordeiro GM (2020) The unit-improved second-degree Lindley distribution: inference and regression modeling. Comput Sta 35:259–279

Altun E, El-Morshedy M, Eliwa M (2021) A new regression model for bounded response variable: An alternative to the beta and unit-Lindley regression models. Plos One 16:e0245627

Atkinson AC (1981) Two graphical displays for outlying and influential observations in regression. Biometrika 68:13–20

Bayes CL, Bazán JL, García C (2012) A new robust regression model for proportions. Bayesian Analy 7:841–866

Bayes CL, Bazán JL, Castro M (2017) A quantile parametric mixed regression model for bounded response variables. Stat Interface 10:483–493

Brownlee KA (1965) Stat Theory Methodol Sci Eng. Wiley, New York

Cancho VG, Bazán JL, Dey DK (2020) A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer. Stat Method Med Res 29:2015–2033

Cooray K, Ananda MMA (2008) A generalization of the half-normal distribution with applications to lifetime data. Commun Stat - Theory Methods 37:1323–1337

Cox DR, Snell EJ (1968) A general definition of residuals. J Royal Stat Soc: Series B (Methodol) 30:248–265

Davison A (2003) Statl Models. Cambridge University Press, Cambrigde, UK

Ferrari S, Cribari Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31:799–815

Figueroa-Zúñiga J, Bayes CL, Leiva V, Liu S (2022) Robust beta regression modeling with errors-in-variables: a Bayesian approach and numerical applications. Stat Papers 63:919–942

Figueroa-Zúñiga J, Niklitschek S, Leiva V, Liu S (2022) Modeling heavy-tailed bounded data by the trapezoidal beta distribution with applications. REVSTAT - Stat J. 20:387–404

de la Fuente-Mella H, Rojas Fuentes JL, Leiva V (2020) Econometric modeling of productivity and technical efficiency in the Chilean manufacturing industry. Comput Industrial Eng 139:105793

Gómez-Déniz E, Sordo MA, Calderín-Ojeda E (2014) The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. Insurance: Mathematics and Economics, 54, 49–57

Hannan EJ, Quinn BJ (1979) The determination of the order of an autoregression. J Royal Stat Soc: Series B (Methodol) 41:190–195

Huerta M, Leiva V, Lillo C, Rodriguez M (2018) Beta partial least squares regression model: Diagnostics and application to mining industry data. Appl Stochastic Models Business Industry 34:305–321

Jodrá P, Jiménez-Gamero MD (2020) A quantile regression model for bounded responses based on the exponential-geometric distribution. REVSTAT - Stat J 4:415–436

Johnson NL (1949) Systems of frequency curves generated by methods of translation. Biometrika 36:149–176

Kizilaslan F, Nadar M (2018) Estimation of reliability in a multicomponent stress-strength model based on a bivariate Kumaraswamy distribution. Stat Papers 59:307–340

Koenker R, Bassett G Jr (1978) Regression quantiles. Econometrica 46:33–50

Kohansal A (2019) On estimation of reliability in a multicomponent stress-strength model for a kumaraswamy distribution based on progressively censored sample. Stat Papers 60:2185–2224

Korkmaz MÇ (2020) The unit generalized half normal distribution: A new bounded distribution with inference and application. UPB Sc Bul Series A: Appl Math Phys 82:133–140

Korkmaz MÇ, Chesneau C (2021) On the unit Burr-XII distribution with the quantile regression modeling and applications. Comput Appl Math 40:1–26

Korkmaz MÇ, Korkmaz ZS (2022) The unit log-log distribution: a new unit distribution with alternative quantile regression modeling and educational measurements applications. J Appl Stat. https://doi.org/10.1080/02664763.2021.2001442

Korkmaz, MÇ, Chesneau C, Korkmaz ZS (2021a) On the arcsecant hyperbolic normal distribution. properties, quantile regression modeling and applications. Symmetry 13, 117

Korkmaz MÇ, Chesneau C, Korkmaz ZS (2021b) Transmuted unit Rayleigh quantile regression model: alternative to beta and kumaraswamy quantile regression models. UPB Sc Bul Series A: Appl Math Phys 83:149–158

Korkmaz MÇ, Emrah A, Chesneau C, Yousof HM (2021c) On the unit-Chen distribution with associated quantile regression and applications. Math Slovaca 72:765–786

Korkmaz MÇ, Emrah A, Alizadeh M, El-Morshedy M (2021d) The log exponential-power distribution: Properties, estimations and quantile regression model. Mathematics 9:2634

Korkmaz MÇ, Chesneau C, Korkmaz ZS (2022a) A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries. J Appl Stat. https://doi.org/10.1080/02664763.2021.1981834

Korkmaz MÇ, Chesneau C, Korkmaz ZS (2022b) The Unit Folded Normal Distribution: A New Unit Probability Distribution with the Estimation Procedures, Quantile Regression Modeling and Educational Attainment Applications. J Reliability Stat Studies 15:261–298

Kumaraswamy P (1980) A generalized probability density function for double-bounded random processes. J Hydrol 46:79–88

Leao J, Leiva V, Saulo H, Tomazella V (2018) Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. Stat Med 37:4421–4440

Liu S, Leiva V, Zhuang D, Ma T, Figueroa-Zuniga J (2022) Matrix differential calculus with applications in the multivariate linear model and its diagnostics. J Multivar Anal 188:104849

Martinez-Florez G, Leiva V, Gomez-Deniz E, Marchant C (2020) A family of skew-normal distributions for modeling proportions and rates with zeros/ones excess. Symmetry 12:1439

Mazucheli J, Menezes AFB, Dey S (2018) The unit-Birnbaum-Saunders distribution with applications. Chilean J Stat 9:47–57

Mazucheli J, Bapat SR, Menezes AFB (2019a) A new one-parameter unit-Lindley distribution. Chilean J Stat 11:53–67

Mazucheli J, Menezes AFB, Chakraborty S (2019b) On the one parameter unit-Lindley distribution and its associated regression model for proportion data. J Appl Stat 46:700–714

Mazucheli J, Menezes AFB, Fernandes LB, de Oliveira RP, Ghitany ME (2020) The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. J Appl Stat 47:954–974

Mazucheli J, Leiva V, Alves B, Menezes AFB (2021) A new quantile regression for modeling bounded data under a unit Birnbaum-Saunders distribution with applications in medicine and politics. Symmetry 13:682

Mazucheli J, Alves B, Korkmaz MC, Leiva V (2022a) Vasicek quantile and mean regression models for bounded data: New formulation, mathematical derivations, and numerical applications. Mathematics 10:1389

Mazucheli M, Alves B, Menezes AFB, Leiva V (2022b) 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

Mitnik PA, Baek S (2013) The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation. Stat Papers 54:177–192

Paz RF, Balakrishnan N, Bazán JL (2019) L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. Brazilian J Probab Stat 33:455–479

Ribeiro TF, Cordeiro GM, Peña-Ramírez FA, Guerra RR (2021) A new quantile regression for the COVID-19 mortality rates in the United States. Comput Appl Math 40:255

Sanchez L, Leiva V, Galea M, Saulo H (2021) Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. Appl Stochastic Models Bus Industry 37:53–73

Sanchez L, Leiva V, Marchant C, Saulo H, Sarabia JM (2021) A new quantile regression model and its diagnostic analytics for a Weibull distributed response with applications. Mathematics 9:2768

Saulo H, Dasilva A, Leiva V, Sanchez L, de la Fuente-Mella H (2022) Log-symmetric quantile regression models. Stat Neerl 76:124–163

SAS Institute Inc. (2018) SAS/STAT® 15.1 User's Guide. The NLMIXED Procedure. Cary, NC: SAS Institute

Schwarz G (1978) Estimating the dimension of a model. Annals Stat 6:461–464

Smithson M, Verkuilen J (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods 11:54

Van Dorp JR, Kotz S (2002a) The standard two-sided power distribution and its properties: with applications in financial engineering. Am Stat 56:90–99

Van Dorp JR, Kotz S (2002b) A novel extension of the triangular distribution and its parameter estimation. J Royal Stat Soc D 51:63–79

Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54:437–447

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.