



On the unit Burr-XII distribution with the quantile regression modeling and applications

Mustafa Ç. Korkmaz¹ · Christophe Chesneau²

Received: 12 October 2020 / Revised: 31 December 2020 / Accepted: 9 January 2021
© SBMAC - Sociedade Brasileira de Matemática Aplicada e Computacional 2021

Abstract

In this paper, we modify the Burr-XII distribution through the inverse exponential scheme to obtain a new two-parameter distribution on the unit interval called the unit Burr-XII distribution. The basic statistical properties of the newly defined distribution are studied. Parameters estimation is dealt and different estimation methods are assessed through two simulation studies. A new quantile regression model based on the proposed distribution is introduced. Applications of the proposed distribution and its regression model to real data sets show that the proposed models have better modeling capabilities than competing models.

Keywords Probability weighted moments · Order statistics · Stochastic ordering · Data analysis · Regression model · Unit distribution · Burr-XII distribution · Recovery rate · Viable CD34+ cells

Mathematics Subject Classification 60E05 · 62E15 · 62F10

1 Introduction

Several unit distributions have been used for modelling data for percentage and proportions in many areas such as biological studies, mortality and recovery rates, economics, health, risks, and measurements sciences. No doubt that the beta, Johnson S_B (see Johnson 1949) and Kumaraswamy (see Kumaraswamy 1980) distributions quickly come to the mind, both to model and to obtain inferences based on data sets from the above areas. However, these classical models may be inadequate, which pose many significant problems for accurate data

Communicated by Clémentine Prieur.

✉ Christophe Chesneau
christophe.chesneau@gmail.com

Mustafa Ç. Korkmaz
mustafacagataykorkmaz@gmail.com

¹ Department of Measurement and Evaluation, Artvin Çoruh University, City Campus, Artvin 08000, Turkey

² LMNO, University of Caen-Normandie, Caen 14032, France

analysis. For this reason, the number of studies on unit modeling increases in the literature. The newly proposed unit distributions have usually been introduced by the transformation of the well-known continuous distributions. The advantage of these unit distributions is to give more flexibility to the basic distribution over the unit interval without adding new parameters. For example, the unit gamma (see Consul and Jain 1971), log-Lindley (see Gómez-Déniz et al. 2014), unit Weibull (see Mazucheli et al. 2018), unit Gompertz (see Mazucheli et al. 2019b), unit Birnbaum-Saunders (see Mazucheli et al. 2018), log-xgamma (see Altun and Hamedani 2018), unit inverse Gaussian (see Ghitany et al. 2019), unit generalized half normal (see Korkmaz 2020b) and log-weighted exponential (see Altun 2021) distributions have been obtained via the negative exponential function transformation of the gamma, Lindley (see Lindley 1958), Weibull, Gompertz, Birnbaum-Saunders, xgamma (see Sen et al. 2016), inverse Gaussian, generalized half normal (see Cooray and Ananda 2008) and log-weighted exponential (see Gupta and Kundu 2009) distributions, respectively. One may refer to Mazucheli et al. (2019a), Altun and Cordeiro (2020), Korkmaz (2020a), Gündüz and Korkmaz (2020), Korkmaz et al. (2021a) and Korkmaz et al. (2021b) for the other unit models which have been obtained with other transformation methods.

On the other hand, Burr (see Burr 1942) has pioneered a system of continuous distributions, which has 12 distributions. This system has been obtained with a differential equation of the form $dF(x)/dx = F(x)(1 - F(x))g(x)$, where $g(x)$ is a non-negative function and $F(x)$ is the satisfying cumulative distribution function (cdf) to this equation. In this way, the author has generalized the Pearson equation. Among these 12 distributions, the Burr III, Burr XII (BXII) and Burr X distributions are commonly known, and have received much more attention in the literature. The BXII distribution has the following cdf and probability density function (pdf): $\Pi(z, \alpha, \beta) = 1 - (1 + z^\beta)^{-\alpha}$ and $\pi(z, \alpha, \beta) = \alpha\beta z^{\beta-1} (1 + z^\beta)^{-\alpha-1}$, respectively, where $z > 0$ and $\alpha, \beta > 0$ are the shape parameters. Its pdf shapes are unimodal or decreasing. It has found many applications in the literature such as reliability analysis (see Zimmer et al. 1998), portfolio segmentation (see Beirlant et al. 1998), and regression modeling (see Afify et al. 2018; Altun et al. 2018a, b; Cordeiro et al. 2018; Lanjoni et al. 2016; Silva et al. 2008 and Yousof et al. 2019). The Weibull distribution is the limiting distribution of the BXII distribution when α tends to $+\infty$. The inverse BXII distribution is known as the Burr type III distribution. For $\alpha = 1$ and $\beta = 1$, the BXII model is reduced to the Lomax and log-logistic (Fisk) distributions, respectively. One may see Rodriguez (1977) and Tadikamalla (1980) for its relations to other distributions in detail.

On the other hand, the classic regression models relate the mean response by given certain values of the covariates. If the response variable follows a skew distribution or has outliers, then the classic regression model is not suitable for the inferences based on the relation between the response variable and covariates. Since the mean is affected by these specific situations, the median is a more informative and robust estimate for these situations. As a solution, quantile regression models were proposed by Koenker and Bassett (1978). Mazucheli et al. (2020) have introduced the unit Weibull quantile regression modeling and have compared its performance with the beta regression (see Ferrari and Cribari-Neto 2004) and Kumaraswamy quantile regression (see Mitnik and Baek 2013) modelings.

The scope of this study is to propose a new unit alternative distribution and to introduce its alternative quantile regression modeling to the beta regression (see Ferrari and Cribari-Neto 2004) and Kumaraswamy quantile regression modeling (see Mitnik and Baek 2013) in order to model the percentages and proportions when data has outliers. To obtain a new unit distribution, we use the negative exponential transformation of the BXII distribution. In this way, we will transport its applicability and work-ability to the unit interval. Hence, we will

bring into the different pdf and hazard rate function (hrf) properties that the BXII distribution has not on the unit interval.

The paper is organized as follows. The proposed distribution is defined in Sect. 2. Its basic distributional properties are described in Sect. 3. Section 4 is devoted to procedures of the different estimation methods to estimate its unknown parameters. Two different simulation studies are given to see the performance of the different estimates of the model parameters in Sect. 5. The new quantile regression model based on the newly defined distribution is discussed in Sect. 6. Two real data illustrations, which one is related to the univariate data modeling and other is the quantile regression modeling, are illustrated in Sect. 7. Finally, the paper ends in Sect. 8.

2 The new unit distribution and its properties

The new unit distribution is defined as follows: Let Z be a random variable having the BXII distribution with parameters α and β and $X = e^{-Z}$. Then, the cdf and pdf of X are presented by

$$F(x, \alpha, \beta) = (1 + (-\log x)^\beta)^{-\alpha} \tag{1}$$

and

$$f(x, \alpha, \beta) = \alpha\beta x^{-1} (-\log x)^{\beta-1} (1 + (-\log x)^\beta)^{-\alpha-1}, \tag{2}$$

where $x \in (0, 1)$ and $\alpha, \beta > 0$ are the shape parameters. We denote the newly defined distribution as $UBXII$ or $UBXII(\alpha, \beta)$ when the parameters need to be indicated. The hrf of the $UBXII(\alpha, \beta)$ distribution is given by

$$h(x, \alpha, \beta) = \alpha\beta x^{-1} (-\log x)^{\beta-1} \frac{(1 + (-\log x)^\beta)^{-\alpha-1}}{1 - (1 + (-\log x)^\beta)^{-\alpha}}. \tag{3}$$

The analytical behavior of $f(x, \alpha, \beta)$ and $h(x, \alpha, \beta)$ is discussed below. When x tends to 0, the following equivalences hold:

$$f(x, \alpha, \beta) \sim h(x, \alpha, \beta) \sim \alpha\beta x^{-1} (-\log x)^{-\alpha\beta-1},$$

which tends to $+\infty$ for all the values of $\beta > 0$ and $\alpha > 0$.

When x tends to 1, we have

$$f(x, \alpha, \beta) \sim \alpha\beta (1 - x)^{\beta-1}, \quad h(x, \alpha, \beta) \sim \beta (1 - x)^{-1}.$$

Therefore, in this case, if $\beta < 1$, $f(x, \alpha, \beta)$ tends to $+\infty$, if $\beta = 1$, $f(x, \alpha, \beta)$ tends to α , and if $\beta > 1$, $f(x, \alpha, \beta)$ tends to 0. Moreover, for all the values of $\beta > 0$ and $\alpha > 0$, $h(x, \alpha, \beta)$ tends to $+\infty$. The critical points of $f(x, \alpha, \beta)$ can be determined by solving in a numerical way the following equation with respect to x :

$$\alpha\beta x^{-2} (-\log x)^{\beta-2} ((-\log x)^\beta + 1)^{-\alpha-2} ((-\log x)^\beta (\alpha\beta + \log x + 1) - \beta + \log x + 1) = 0.$$

If we excluded the bounds 0 and 1, this equation is equivalent to $(-\log x)^\beta (\alpha\beta + \log x + 1) - \beta + \log x + 1 = 0$, which remains of a high complexity from a mathematical point of view. After numerical tests, 0, 1 or 2 critical points into $(0, 1)$ can be found, depending on

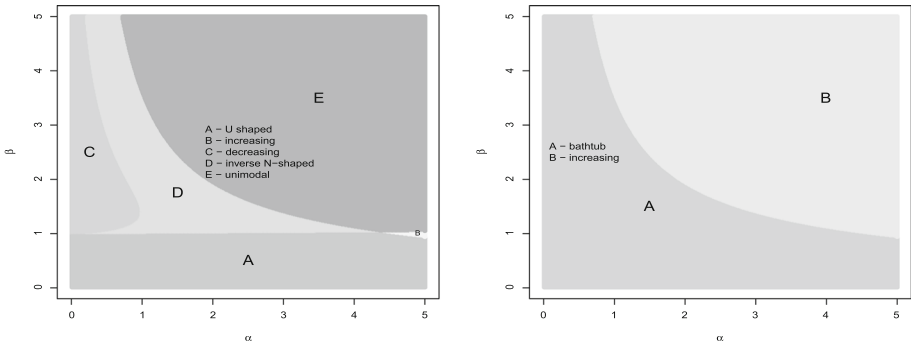


Fig. 1 The possible shapes regions of the pdf (left) and hrf (right)

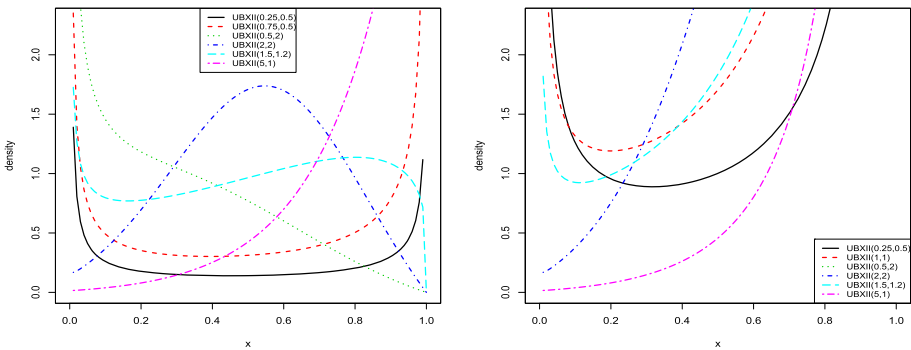


Fig. 2 The possible shapes of the pdf (left) and hrf (right)

the values α and β . For instance, in the simple case where $\beta = 1$, there is only one critical point into $(0, 1)$, it is given by $x = e^{-\alpha}$ and is a maximum point.

The critical points of $h(x, \alpha, \beta)$ can be discussed in a similar manner. More elegantly, the possible shapes of $f(x, \alpha, \beta)$ and $h(x, \alpha, \beta)$ can be represented graphically. Figure 1 determines the regions of the values of α and β where $f(x, \alpha, \beta)$ is U-shaped, increasing, decreasing, inverse N-shaped and unimodal, and $h(x, \alpha, \beta)$ is bathtub and increasing (the other kinds of shapes being excluded). Figure 2 completes the previous analysis by plotting diverse curves of $f(x, \alpha, \beta)$ and $h(x, \alpha, \beta)$ to visualize their possible shapes.

From Figs. 1 and 2, it is clear that the *UB XII* distribution enjoys a high level of flexibility, and we can take advantage of it for various statistical applications.

The *UB XII* distribution is also completely determined by its quantile function (qf) given as the inverse function of $F(x, \alpha, \beta)$. After some development, it is defined by

$$Q(u, \alpha, \beta) = e^{-(u^{-1/\alpha}-1)^{1/\beta}}, \tag{4}$$

where $u \in (0, 1)$. The median of the *UB XII* distribution is derived by taking $u = 0.5$ in the above equation. Similarly, the other quartiles and octiles can be determined. Also, some important qfs can be simply expressed, such as the quantile density and hazard quantile functions obtained as

$$q(u, \alpha, \beta) = \frac{1}{\alpha\beta} u^{-1/\alpha-1} (u^{-1/\alpha} - 1)^{1/\beta-1} e^{-(u^{-1/\alpha}-1)^{1/\beta}}$$

and

$$h(u, \alpha, \beta) = \frac{\alpha\beta}{1-u} u^{1/\alpha+1} (u^{-1/\alpha} - 1)^{1-1/\beta} e^{(u^{-1/\alpha}-1)^{1/\beta}},$$

where $u \in (0, 1)$, respectively. For the general interpretation of these functions, we refer the reader to Nair and Sankaran (2009).

3 Some properties

This section presents several theoretical results of interest involving the *UBXII* distribution.

3.1 First-order stochastic dominance

The first-order stochastic dominance is a simple concept allowing to compare distributions through their respective cdfs. We say that a distribution *A* first-order stochastically dominates another distribution *B* if their respective cdfs $F_A(x)$ and $F_B(x)$ satisfy the following inequalities: $F_A(x) \leq F_B(x)$ for all $x \in \mathbb{R}$. This concept is particularly developed in biometrics, reliability, econometrics and actuarial sciences. See, for instance, Shaked and Shanthikumar (2007) and Muller and Stoyan (2002).

We now expose some results on first-order stochastic dominance satisfied by the *UBXII* distribution. They are contained in the following proposition.

Proposition 1 *The two following results hold.*

- *If $\alpha_1 \geq \alpha_2$, then the $UBXII(\alpha_1, \beta)$ distribution first-order stochastically dominates the $UBXII(\alpha_2, \beta)$ distribution.*
- *The $UBXII(\alpha, \beta)$ distribution is first-order stochastically dominated by the unit Weibull distribution with parameters α and β introduced by Mazucheli et al. (2020), and defined by the following cdf:*

$$F_{UW}(x, \alpha, \beta) = e^{-\alpha(-\log x)^\beta},$$

where $x \in (0, 1)$.

Proposition 1 shows that the *UBXII* distribution offers a different alternative to the unit Weibull distribution, in the first-order stochastic sense, and in statistical modelling as well.

3.2 Probability weighted moments with applications

The probability weighted moments of a random variable generalize the ordinary moments, and naturally appear when we deal with the ordinary moments of order statistics. They are also involved in some parametric estimation methods. A detailed description of these moments can be found in Greenwood et al. (1979). In the context of the Burr-XII distribution, we may also refer to Usta (2013).

The following result is about an expression of the probability weighted moments of a random variable having the *UBXII* distribution.

Proposition 2 *Let s and u be two integers, and X be a random variable having the $UBXII(\alpha, \beta)$ distribution. Then, the $(s, u)^{th}$ probability weighted moment of X can be*

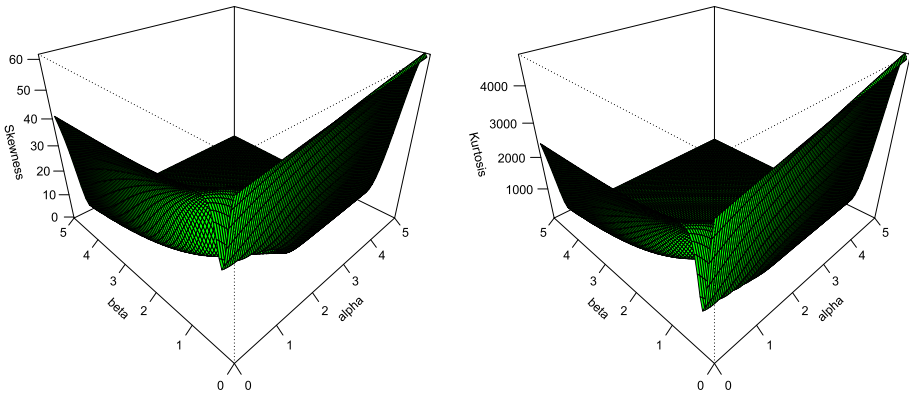


Fig. 3 The skewness and kurtosis of the $UBXII$ distribution for the parameter values into $(0, 5)$

determined as

$$\mu'_{s,u} = E [X^s F(X, \alpha, \beta)^u] = \frac{1}{u + 1} M_{BXII}(-s, \alpha(u + 1), \beta),$$

where, in full generality, $M_{BXII}(t, \alpha, \beta)$ denotes the well-known moment generating function (mgf) of the Burr-XII distribution with parameters α and β taken at the point t .

An analytical expression of $M_{BXII}(t, \alpha, \beta)$ exists through the use of the Meijer G-function. See, for instance, Paranaíba et al. (2011) and Silva and Cordeiro (2015). It can also be evaluated numerically with an immediate implementation through any mathematical software, such as R, Python, Matlab, etc.

From Proposition 2, we easily derive the s^{th} moment of X by taking $u = 0$ which gives $\mu'_s = E(X^s) = \mu'_{s,0}$. Also, the mean of X is given by $\mu = \mu'_1$, the standard deviation of X is obtained as $\sigma = (\mu'_2 - \mu^2)^{1/2}$ and the s^{th} general coefficient of X is specified by

$$\Upsilon_s = E \left[\left(\frac{X - \mu}{\sigma} \right)^s \right] = \sigma^{-s} \sum_{k=0}^s \binom{s}{k} (-1)^{s-k} \mu^{s-k} \mu'_k.$$

From this general coefficient, the coefficients of skewness and kurtosis of X are defined by Υ_3 and Υ_4 , respectively. The numerical behavior of these coefficients are shown in Fig. 3.

From Fig. 3, we see that Υ_3 and Υ_4 have versatile values with non-monotonic shapes. Also, Υ_3 can be close to 0 and large, showing that the $UBXII$ distribution is mainly right-skewed.

3.3 Entropy

The entropy of the $UBXII$ distribution can be evaluated through numerous ways. Here, we propose the Tsallis entropy, revealing a manageable expression through the use of the mgf of the Burr-XII distribution. Generalities on the Tsallis entropy can be found in Amigo et al. (2018). The following result concerns a series expansion of this entropy measure in the context of the $UBXII$ distribution.

Proposition 3 *Let $\tau < 1$. Then, the Tsallis entropy of the $UBXII$ distribution exists and it is given as*

$$\begin{aligned}
 T(\tau, \alpha, \beta) &= \frac{1}{\tau - 1} \left[1 - \int_0^1 f(x, \alpha, \beta)^\tau dx \right] \\
 &= \frac{1}{\tau - 1} \left[1 - \alpha^\tau \beta^\tau \int_0^{+\infty} e^{(\tau-1)y} y^{\tau(\beta-1)} (1 + y^\beta)^{-\tau(\alpha+1)} dy \right].
 \end{aligned}$$

For $\tau \in (0, 1)$, the following upper bound holds:

$$T(\tau, \alpha, \beta) \leq \frac{1}{\tau - 1} \left[1 - \frac{\alpha^\tau \beta^\tau}{(1 - \tau)^{\tau(\beta-1)+1}} \Gamma(\tau(\beta - 1) + 1) \right],$$

where $\Gamma(x)$ denotes the standard Euler gamma function.

Proof The proof is centered around the expression of the integral term. From Equation (2), we get

$$\int_0^1 f(x, \alpha, \beta)^\tau dx = \alpha^\tau \beta^\tau \int_0^1 x^{-\tau} (-\log x)^{\tau(\beta-1)} (1 + (-\log x)^\beta)^{-\tau(\alpha+1)} dx.$$

By performing the change of variables $x = e^{-y}$, we obtain

$$\int_0^1 f(x, \alpha, \beta)^\tau dx = \alpha^\tau \beta^\tau \int_0^{+\infty} e^{(\tau-1)y} y^{\tau(\beta-1)} (1 + y^\beta)^{-\tau(\alpha+1)} dy,$$

which exists if $\tau < 1$. Now, for $\tau \in (0, 1)$, by using the following inequalities: $(1 + y^\beta)^{-\tau(\alpha+1)} \leq 1$ and applying the change of variables $z = (1 - \tau)y$, we have

$$\begin{aligned}
 \int_0^1 f(x, \alpha, \beta)^\tau dx &\leq \alpha^\tau \beta^\tau \int_0^{+\infty} e^{-(1-\tau)y} y^{\tau(\beta-1)} dy \\
 &= \frac{\alpha^\tau \beta^\tau}{(1 - \tau)^{\tau(\beta-1)+1}} \int_0^{+\infty} z^{\tau(\beta-1)+1-1} e^{-z} dz \\
 &= \frac{\alpha^\tau \beta^\tau}{(1 - \tau)^{\tau(\beta-1)+1}} \Gamma(\tau(\beta - 1) + 1).
 \end{aligned}$$

By substituting this inequality in the definition of $T(\tau, \alpha, \beta)$ and taking into account that the factor term $(\tau - 1)^{-1}$ is negative, we obtain the stated result. □

The integral term in the Tsallis entropy can not be developed. We can however evaluate it numerically, and the Tsallis entropy as well. In a similar way, we can express the Rényi entropy which also depends on the same integral term.

3.4 Stress–strength parameter

In lifetime testing, various resistance measures exist to measure the lifetime of a system. Among them, there is the stress-strength parameter defined by the probability that a component of the system will function satisfactorily if the applied stress is lower than its strength. More details are given in Surles and Padgett (2001). The following proposition exhibits a simple form for this parameter in the setting of the $UBXII$ distribution.

Proposition 4 Let X_1 and X_2 be two independent random variables, with X_1 having the $UBXII(\alpha_1, \beta)$ distribution and X_2 having the $UBXII(\alpha_2, \beta)$ distribution. Then, if we define the stress-strength parameter as $R(\alpha_1, \alpha_2, \beta) = P(X_2 \leq X_1)$, then we have

$$R(\alpha_1, \alpha_2, \beta) = \frac{\alpha_1}{\alpha_1 + \alpha_2}.$$

The 1/2 value is obtained for $\alpha_1 = \alpha_2$. In this case, X_1 and X_2 are identically distributed. The closed-form expression of $R(\alpha_1, \alpha_2, \beta)$ making it interesting for statistical estimation purposes.

3.5 Order statistics

The modeling of various lifetime systems having some component structures needs the consideration of ordered random variables, called order statistics. The fundamental of this notion can be found in David and Nagaraja (2003). Here, standard distributional properties of the order statistics of the $UBXII$ distribution are presented.

Let X_1, X_2, \dots, X_n be a random sample from the $UBXII$ distribution with sample size n , and $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be their ordered statistics, such that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. For any $j = 1, \dots, n$, from Equations (1) and (2), the pdf of $X_{(j)}$ is defined as

$$\begin{aligned} f_{X_{(j)}}(x, \alpha, \beta) &= \frac{n!}{(j-1)!(n-j)!} f(x, \alpha, \beta) F(x, \alpha, \beta)^{j-1} [1 - F(x, \alpha, \beta)]^{n-j} \\ &= \frac{n!}{(j-1)!(n-j)!} \alpha \beta x^{-1} (-\log x)^{\beta-1} (1 + (-\log x)^\beta)^{-\alpha j-1} \\ &\quad \left[1 - (1 + (-\log x)^\beta)^{-\alpha} \right]^{n-j}, \end{aligned}$$

where $x \in (0, 1)$. In particular, the pdf of the infimum of X_1, X_2, \dots, X_n , that is $X_{(1)}$, can be expressed as

$$\begin{aligned} f_{X_{(1)}}(x, \alpha, \beta) &= n \alpha \beta x^{-1} (-\log x)^{\beta-1} (1 + (-\log x)^\beta)^{-\alpha-1} \\ &\quad \left[1 - (1 + (-\log x)^\beta)^{-\alpha} \right]^{n-1}, \end{aligned}$$

where $x \in (0, 1)$. Also, the pdf of the supremum of X_1, X_2, \dots, X_n , that is $X_{(n)}$, is

$$f_{X_{(n)}}(x, \alpha, \beta) = n \alpha \beta x^{-1} (-\log x)^{\beta-1} (1 + (-\log x)^\beta)^{-\alpha n-1},$$

where $x \in (0, 1)$. We recognize the pdf of the $UBXII(\alpha n, \beta)$ distribution. Therefore, the theory developed in the previous sections for the $UBXII$ distribution can be directly applied for $X_{(n)}$.

The moments of $X_{(j)}$ are discussed in the next proposition.

Proposition 5 Let s be an integer. Then, with the above notations, the s^{th} moment of $X_{(j)}$ can be expressed as a finite linear combination of mgfs of the $BXII$ distribution. More precisely, we have

$$\mu_{s,(j)}^* = E(X_{(j)}^s) = \sum_{k=0}^{n-j} a_{(j),k} M_{BXII}(-s, \alpha(k+j), \beta),$$

where

$$a_{(j),k} = \frac{n!}{(j-1)!(n-j)!} \binom{n-j}{k} (-1)^k \frac{1}{k+j}.$$

Thanks to Proposition 5, one can express diverse parameters of $X_{(j)}$, such as its mean, variance, standard deviation, central moments, etc. Also, the order statistics of the $UBXII$ distribution will have important roles in the estimation procedures described in the next section.

4 Estimation procedures for the model parameters

In this section, six different estimation methods have been pointed out to estimate the parameters of the $UBXII$ distribution. The details are given below.

4.1 Maximum likelihood estimation

In this subsection, we derive estimations of the parameters α and β via the method of the maximum likelihood (ML) estimation. Let X_1, X_2, \dots, X_n be a random sample of size n from the $UBXII(\alpha, \beta)$ distribution with observed values x_1, x_2, \dots, x_n . Let $\Upsilon = (\alpha, \beta)^T$ be the vector of the model parameters. Then, the log-likelihood function is given by

$$\begin{aligned} \ell(\Upsilon) &= n \log \alpha + n \log \beta - \sum_{i=1}^n \log x_i + (\beta - 1) \sum_{i=1}^n \log(-\log x_i) - (\alpha + 1) \\ &\quad \times \sum_{i=1}^n \log(1 + (-\log x_i)^\beta). \end{aligned} \tag{5}$$

Then, the ML estimates (MLEs) of α and β , say $\hat{\alpha}$ and $\hat{\beta}$, are obtained by maximizing $\ell(\Upsilon)$ with respect to Υ . Mathematically, this is equivalent to solve the following non-linear equation with respect to the parameters:

$$\frac{\partial}{\partial \alpha} \ell(\Upsilon) = \frac{n}{\alpha} - \sum_{i=1}^n \log(1 + (-\log x_i)^\beta) = 0 \tag{6}$$

and

$$\frac{\partial}{\partial \beta} \ell(\Upsilon) = \frac{n}{\beta} + \sum_{i=1}^n \log(-\log x_i) - (\alpha + 1) \sum_{i=1}^n \frac{\log(-\log x_i)(-\log x_i)^\beta}{1 + (-\log x_i)^\beta} = 0.$$

From Equation (6), these solutions are governed by the following relation:

$$\alpha = n \left[\sum_{i=1}^n \log(1 + (-\log x_i)^\beta) \right]^{-1}. \tag{7}$$

Substituting Equation (7) in Equation (5), the profile log-likelihood (PLL) based on the parameter β is given by

$$\ell(\beta) = n \log \left(n \left[\sum_{i=1}^n \log(1 + (-\log x_i)^\beta) \right]^{-1} \right) - \sum_{i=1}^n \log x_i + (\beta - 1) \sum_{i=1}^n \log(-\log x_i)$$

$$-n - \sum_{i=1}^n \log(1 + (-\log x_i)^\beta) + n \log \beta. \tag{8}$$

Therefore, the MLE $\hat{\beta}$ is obtained by maximizing Equation (8) based on the parameter β , which can be obtained through the solving of the following non-linear equation:

$$\begin{aligned} \frac{\partial}{\partial \beta} \ell(\beta) &= \frac{n}{\beta} - n \sum_{i=1}^n \frac{(-\log x_i)^\beta \log(-\log x_i)}{1 + (-\log x_i)^\beta} \left[\sum_{i=1}^n \log(1 + (-\log x_i)^\beta) \right]^{-1} \\ &+ \sum_{i=1}^n \log(-\log x_i) - \sum_{i=1}^n \frac{(-\log x_i)^\beta \log(-\log x_i)}{1 + (-\log x_i)^\beta} = 0. \end{aligned}$$

Hence, the numerical methods are needed to obtain the MLE $\hat{\beta}$. After the MLE $\hat{\beta}$ is obtained, the $\hat{\alpha}$ is derived from the relation Equation (7).

Under mild regularity conditions, one can use the bivariate normal distribution with the mean $\mu = (\alpha, \beta)$ and variance-covariance matrix I^{-1} , where I denotes the 2×2 Fisher information matrix, to construct confidence intervals or likelihood ratio test on the parameters. The asymptotic variance-covariance matrix of the maximum likelihood estimators of the parameters is given by the inverse of the Fisher information matrix. If the considered pdf is smooth, the related Fisher information matrix is the matrix whose elements are negative of the expected values of the second partial derivatives of the log-likelihood function with respect to the parameters. In applications, the observed information matrix denoted by J , which is the negative of the Hessian matrix of the log-likelihood function with the parameters replaced by their estimates and is a sample-based version of the Fisher information, is usually used instead of I . The components of the observed information matrix can be requested from the authors when it is needed. It is noticed that it is also obtained numerically by computer packages such as R. Then, approximate $100(1 - \vartheta)\%$ confidence intervals for α and β can be determined by the following bounds:

$\hat{\alpha} \pm z_{\vartheta/2} s_{\hat{\alpha}}$ and $\hat{\beta} \pm z_{\vartheta/2} s_{\hat{\beta}}$, where $z_{\vartheta/2}$ is the upper $(\vartheta/2)^{th}$ percentile of the standard normal distribution, $s_{\hat{\alpha}} = \sqrt{J_{\alpha\alpha}^{-1}}$, $s_{\hat{\beta}} = \sqrt{J_{\beta\beta}^{-1}}$, and $J_{\epsilon\epsilon}^{-1}$ are diagonal elements of J^{-1} corresponding to $\epsilon = \alpha$ and β .

4.2 Maximum product spacing estimation

The maximum product spacing (MPS) method has been introduced in Cheng and Amin (1979). It is based on the idea that differences (spacings) between the values of the cdf at consecutive data points should be identically distributed. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the ordered statistics from the $UBXII$ distribution with sample size n , and $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be the ordered observed values. Then, we define the MPS function by

$$MPS(\Upsilon) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log [F(x_{(i)}, \alpha, \beta) - F(x_{(i-1)}, \alpha, \beta)]. \tag{9}$$

The MPS estimates (MPSEs) can be obtained by maximizing $MPS(\Upsilon)$ with respect to Υ . They are also given as the simultaneous solutions of the following non-linear equations:

$$\frac{\partial MPS(\Upsilon)}{\partial \alpha} = \frac{1}{n+1} \sum_{i=1}^{n+1} \left[\frac{F'_\alpha(x_{(i)}, \alpha, \beta) - F'_\alpha(x_{(i-1)}, \alpha, \beta)}{F(x_{(i)}, \alpha, \beta) - F(x_{(i-1)}, \alpha, \beta)} \right] = 0$$

and

$$\frac{\partial MPS(\Upsilon)}{\partial \beta} = \frac{1}{n+1} \sum_{i=1}^{n+1} \left[\frac{F'_\beta(x_{(i)}, \alpha, \beta) - F'_\beta(x_{(i-1)}, \alpha, \beta)}{F(x_{(i)}, \alpha, \beta) - F(x_{(i-1)}, \alpha, \beta)} \right] = 0,$$

where

$$F'_\alpha(x, \alpha, \beta) = -(1 + (-\log x)^\beta)^{-\alpha} \log(1 + (-\log x)^\beta)$$

and

$$F'_\beta(x, \alpha, \beta) = -\alpha (1 + (-\log x)^\beta)^{-\alpha-1} (-\log x)^\beta \log(-\log x) \log(1 + (-\log x)^\beta).$$

4.3 Least squares estimation

The least square estimates (LSEs) $\hat{\alpha}_{LSE}$ and $\hat{\beta}_{LSE}$ of α and β , respectively, are obtained by minimizing the following function:

$$LSE(\Upsilon) = \sum_{i=1}^n (F(x_{(i)}, \alpha, \beta) - E[F(X_{(i)}, \alpha, \beta)])^2, \tag{10}$$

with respect to Υ , where $E[F(X_{(i)}, \alpha, \beta)] = i/(n+1)$ for $i = 1, 2, \dots, n$. Then, $\hat{\alpha}_{LSE}$ and $\hat{\beta}_{LSE}$ are solutions of the following equations:

$$\frac{\partial LSE(\Upsilon)}{\partial \alpha} = 2 \sum_{i=1}^n F'_\alpha(x_{(i)}, \alpha, \beta) \left(F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1} \right) = 0$$

and

$$\frac{\partial LSE(\Upsilon)}{\partial \beta} = 2 \sum_{i=1}^n F'_\beta(x_{(i)}, \alpha, \beta) \left(F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1} \right) = 0,$$

where $F'_\alpha(x, \alpha, \beta)$ and $F'_\beta(x, \alpha, \beta)$ are mentioned before.

4.4 Weighted least squares estimation

Similarly to LSEs, the weighted least square estimates (WLSEs) $\hat{\alpha}_{WLSE}$ and $\hat{\beta}_{WLSE}$ of α and β , respectively, are obtained by minimizing the following function:

$$WLSE(\Upsilon) = \sum_{i=1}^n \frac{1}{V[F(X_{(i)}, \alpha, \beta)]} (F(x_{(i)}, \alpha, \beta) - E[F(X_{(i)}, \alpha, \beta)])^2, \tag{11}$$

with respect to Υ , where $E[F(X_{(i)}, \alpha, \beta)] = i/(n+1)$ and $V[F(X_{(i)}, \alpha, \beta)] = i(n-i+1)/[(n+2)(n+1)^2]$ for $i = 1, 2, \dots, n$. Then, $\hat{\alpha}_{WLSE}$ and $\hat{\beta}_{WLSE}$ are solutions of the following equations:

$$\frac{\partial WLSE(\Upsilon)}{\partial \alpha} = 2 \sum_{i=1}^n \frac{(n+2)(n+1)^2}{i(n-i+1)} \left(F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1} \right) F'_\alpha(x_{(i)}, \alpha, \beta) = 0$$

and

$$\frac{\partial WLSE(\mathbf{Y})}{\partial \beta} = 2 \sum_{i=1}^n \frac{(n+2)(n+1)^2}{i(n-i+1)} \left(F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1} \right) F'_{\beta}(x_{(i)}, \alpha, \beta) = 0.$$

4.5 Anderson–Darling estimation

The Anderson–Darling minimum distance estimates (ADEs) $\hat{\alpha}_{AD}$, $\hat{\beta}_{AD}$ of α and β , respectively, are obtained by minimizing the following function:

$$AD(\mathbf{Y}) = -n - \sum_{i=1}^n \frac{2i-1}{n} [\log F(x_{(i)}, \alpha, \beta) + \log \{1 - F(x_{(n+1-i)}, \alpha, \beta)\}], \quad (12)$$

with respect to \mathbf{Y} . Therefore, $\hat{\alpha}_{AD}$ and $\hat{\beta}_{AD}$ can be obtained as the solutions of the following system of equations:

$$\frac{\partial AD(\mathbf{Y})}{\partial \alpha} = - \sum_{i=1}^n \frac{2i-1}{n} \left[\frac{F'_{\alpha}(x_{(i)}, \alpha, \beta)}{F(x_{(i)}, \alpha, \beta)} - \frac{F'_{\alpha}(x_{(n+1-i)}, \alpha, \beta)}{1 - F(x_{(n+1-i)}, \alpha, \beta)} \right] = 0$$

and

$$\frac{\partial AD(\mathbf{Y})}{\partial \beta} = - \sum_{i=1}^n \frac{2i-1}{n} \left[\frac{F'_{\beta}(x_{(i)}, \alpha, \beta)}{F(x_{(i)}, \alpha, \beta)} - \frac{F'_{\beta}(x_{(n+1-i)}, \alpha, \beta)}{1 - F(x_{(n+1-i)}, \alpha, \beta)} \right] = 0.$$

4.6 The Cramér–von Mises estimation

The Cramér–von Mises minimum distance estimates (CVMs) $\hat{\alpha}_{CVM}$ and $\hat{\beta}_{CVM}$ of α and β , respectively, are obtained by minimizing the following function:

$$CVM(\mathbf{Y}) = \frac{1}{12n} + \sum_{i=1}^n \left[F(x_{(i)}, \alpha, \beta) - \frac{2i-1}{2n} \right]^2, \quad (13)$$

with respect to \mathbf{Y} . Therefore, the desired estimates can be obtained as the solutions of the following system of equations:

$$\frac{\partial CVM(\mathbf{Y})}{\partial \alpha} = 2 \sum_{i=1}^n \left(F(x_{(i)}, \alpha, \beta) - \frac{2i-1}{2n} \right) F'_{\alpha}(x_{(i)}, \alpha, \beta) = 0$$

and

$$\frac{\partial CVM(\mathbf{Y})}{\partial \beta} = 2 \sum_{i=1}^n \left(F(x_{(i)}, \alpha, \beta) - \frac{2i-1}{2n} \right) F'_{\beta}(x_{(i)}, \alpha, \beta) = 0.$$

Since all the equations above contain non-linear functions, it is not possible to obtain explicit forms of all estimates directly. Therefore, they have to be solved by using numerical methods such as the Newton–Raphson and quasi-Newton algorithms. In addition, Equations (5), (9), (10), (11), (12) and (13) can be also optimized directly by using the software such as R (constrOptim, optim and maxLik functions), S-Plus and Mathematica to numerically optimize $\ell(\mathbf{Y})$ and $MPS(\mathbf{Y})$, $LSE(\mathbf{Y})$, $WLSE(\mathbf{Y})$, $AD(\mathbf{Y})$ and $CVM(\mathbf{Y})$ functions.

5 Simulations

In this section, we perform two graphical simulation studies to see the performance of the above estimates of the *UBXII* distribution with respect to varying sample size n . The first step is the generation of $N = 1000$ samples of size $n = 20, 25, \dots, 1000$ from the *UBXII* distribution based on the actual parameter values. We take them as $\alpha = 5, \beta = 2$ and $\alpha = 2, \beta = 5$ for the first and second simulation studies, respectively. The random numbers generation is obtained by the use of the `qf` of the model. All the estimations based on the estimation methods have been obtained by employing the `constrOptim` function in the R program. Further, we calculate the empirical mean, bias and mean square error (MSE) of the estimations for comparisons between estimation methods. For $\epsilon = \alpha$ and β , the bias and MSE are calculated from all the samples as

$$Bias_{\epsilon}(n) = \frac{1}{N} \sum_{i=1}^N (\epsilon - \hat{\epsilon}_i), \quad MSE_{\epsilon}(n) = \frac{1}{N} \sum_{i=1}^N (\epsilon - \hat{\epsilon}_i)^2,$$

respectively. We expect that the empirical means are close to true values when the MSEs and biases are near zero. The results of this simulation study are shown in Figs. 4 and 5.

Figures 4 and 5 show that all estimates are consistent since the MSE and biasedness decrease to zero with increasing sample size as expected. All estimates are asymptotic unbiased also. According to the simulation studies, the amount of the biases and MSEs of the CVM, MLE and MPS methods are bigger than other methods initially for both parameters. However, the performances of all estimation methods come closer to each other while the sample size increases. It can be said that the LSE, WLSE and AD methods may be used to obtain inferences from small samples size. Therefore, all methods can be chosen as alternative results for the newly defined model according to large sample sizes. It is noticed that since the MLE and MPS estimation methods have very important properties such as consistency, efficiency and asymptotic normality, they offer many advantages in terms of inferences. Similar results can be also obtained for different parameter settings.

Moreover, we also give a simulation study based on above results of the MLEs for the efficiency of the 95% confidence intervals. We use the coverage length (CL) criteria to see the performance of the MLEs. For $\epsilon = \alpha$ or β , the estimated CLs are given by

$$CL_{\epsilon}(n) = \frac{3.919928}{N} \sum_{i=1}^N s_{\hat{\epsilon}_i},$$

where $s_{\hat{\epsilon}_i}$ is the standard error of the MLE ϵ at the i^{th} sample (which is evaluated by inverting the observed information matrix).

Figure 6 displays the simulation results for CLs. As seen from Fig. 6, as expected, when the sample size increases the CL decreases for each parameter.

6 An alternative quantile regression model

The quantile regression has been originally proposed by Koenker and Bassett (1978) as a way to model the conditional quantiles of an outcome variable as a function of explanatory variables without any distributional assumptions on the error term. When the response variable has a skewed distribution or outliers in the measurements, robust estimation results based on the regression model are needed for the model inference. For this reason, the quantile

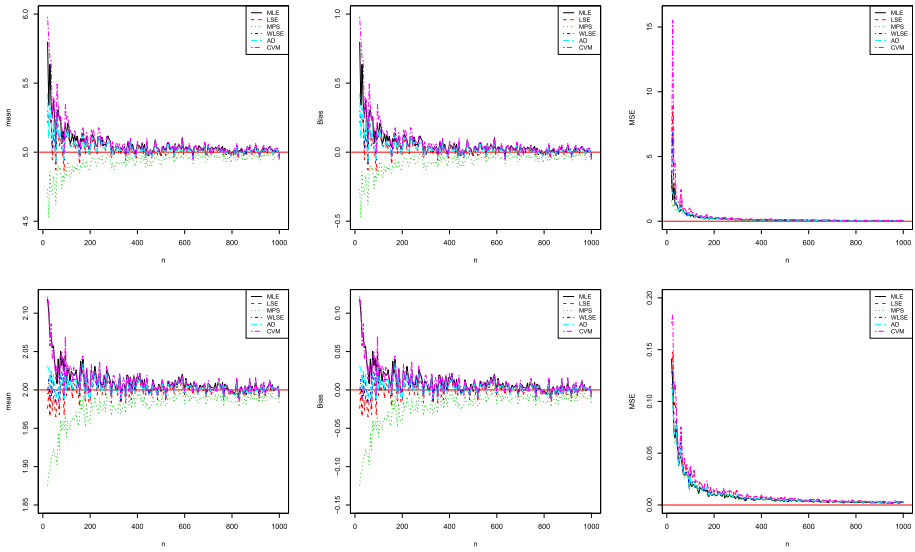


Fig. 4 The results of on the parameters α (top) and β (bottom) for the first simulation study

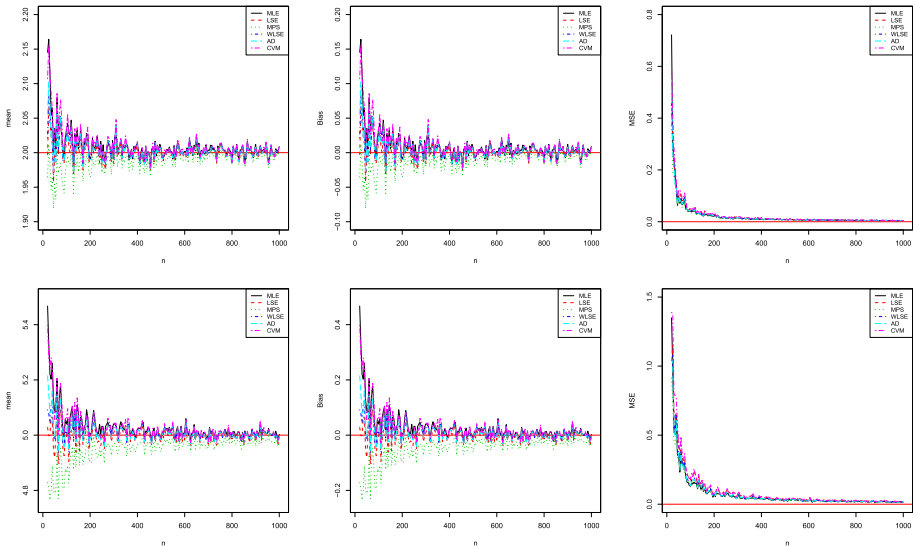


Fig. 5 The results of the parameters α (top) and β (bottom) for the second simulation study

regression is a good robust alternative model to the ordinary LSE model, which estimates the conditional mean of the response variable. In other words, quantile regression can be thought of as an extension of linear regression used when the conditions of linear regression are not met. The quantile regression has also been applied to the analysis of continuous bounded outcomes (see, e.g., Jung 1996; Geraci and Bottai 2007; Liu and Bottai 2009; Bottai et al. 2010). Since the moments of the *UBXII* distribution are not obtained with a closed-form, its mean is not obtained with a simple term. Although it has no moments with closed-form,

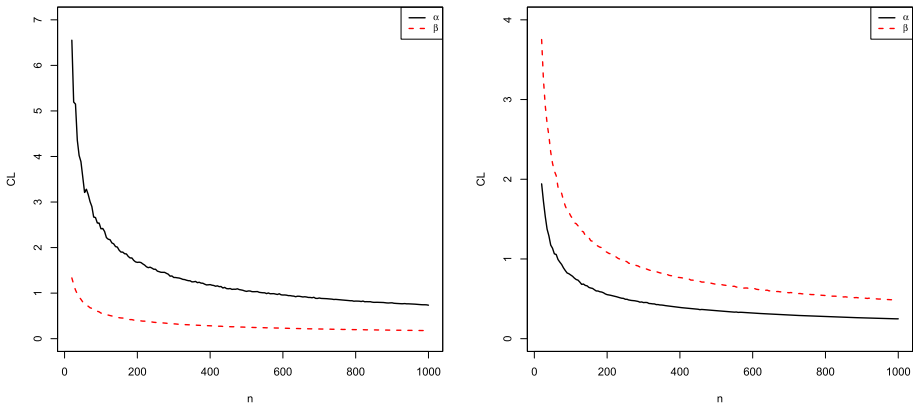


Fig. 6 The estimated ALs for the first (left) and second (right) simulation studies

its qf given is very manageable. So, we are also motivated with quantile regression modeling thanks to its nice statistical property.

On the other hand, if the support of the response variable is defined on the unit interval, a unit regression model based on the unit distribution can be used for modeling the conditional mean or quantiles of the response variable via independent variables (covariates). No doubt that the beta regression Ferrari and Cribari-Neto (2004) model comes to mind firstly to relate with continuous mean response variables in the standard unit interval with covariates. However, using the beta regression model may not be suitable when the unit response variable has a skewed distribution because the mean is affected by the skewing of the distribution precisely. If the conditional dependent variable is skewed, especially, the median may be more appropriate when compared with the mean Mazucheli et al. (2020).

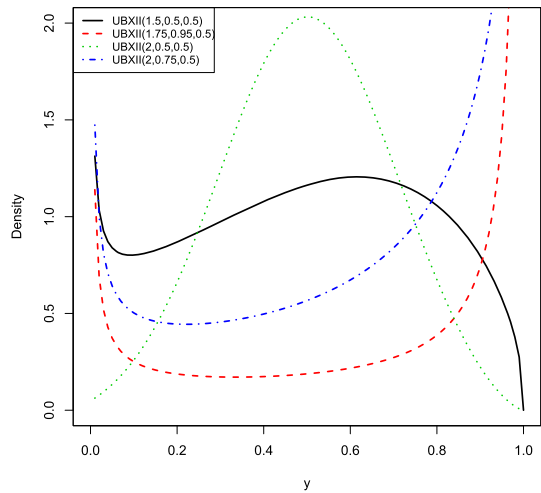
With the re-parameterizing the probability distribution as a function of the quantile approach, the Kumaraswamy Mitnik and Baek (2013), Bayes et al. (2017) and unit Weibull Mazucheli et al. (2020) quantile regression models have been proposed for modeling the conditional quantiles of the unit response. In light of these references, this Section aims to introduce an alternative quantile regression model considering a parametrization of the *UBXII* distribution in terms of any quantile. The re-parameterizing process has been applied via a shape parameter as being a quantile of the *UBXII* distribution. Now, we introduce an alternative quantile regression based on the *UBXII* distribution.

6.1 The *UBXII* distribution based on the its quantiles

Since the qf of the *UBXII* distribution is a very tractable analytic equation, its pdf and cdf can be re-parameterized in terms of its quantiles easily. Firstly, the pdf of the *UBXII* distribution can be given with a re-parameterization based on its qf given in Equation (4). Let $\mu = Q(u, \alpha, \beta)$ and $\beta = \log(u^{-1/\alpha} - 1) / \log(-\log \mu)$. Then, the pdf and cdf of the re-parameterized distribution are given by

$$g(y, \alpha, \mu) = \frac{\log(u^{-1/\alpha} - 1)^\alpha}{\log(-\log \mu)^y} (-\log y)^{\frac{\log(u^{-1/\alpha} - 1)}{\log(-\log \mu)} - 1} \left(1 + (-\log y)^{\frac{\log(u^{-1/\alpha} - 1)}{\log(-\log \mu)}} \right)^{-\alpha - 1} \tag{14}$$

Fig. 7 The pdf shapes of the *UBXII* quantile distribution



and

$$G(y, \alpha, \mu) = \left(1 + (-\log y) \frac{\log(u^{-1/\alpha} - 1)}{\log(-\log \mu)} \right)^{-\alpha}, \tag{15}$$

respectively, where $\alpha > 0$ is the shape parameter, the parameter $\mu \in (0, 1)$ represents the quantile parameter, and u is known. A random variable Y having this pdf is denoted by $Y \sim UBXII(\alpha, \mu, u)$. Some possible shapes of the quantile distribution are shown in Fig. 7. We see that the possible pdf shapes the distribution are bathtub shaped, inverse N-shaped and unimodal. It is noticed that we must have either $u > 2^{-\alpha}$ and $\mu > e^{-1}$, either $u < 2^{-\alpha}$ and $\mu < e^{-1}$.

6.2 The *UBXII* quantile regression model

Now, we focus on the quantile regression model based on the *UBXII* distribution with pdf in Eq. (14). Let y_1, y_2, \dots, y_n such that y_i is an observation of $Y_i \sim UBXII(\alpha, \mu_i, u)$ for $i = 1, \dots, n$, with unknown parameters μ_i and α . Note that the parameter u is known. Then, the new quantile regression model is defined as

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

where $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2, \dots, \delta_p)^T$ and $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ are the unknown regression parameter vector and known i^{th} vector of the covariates. The function $g(x)$ is the link function which is used to relate the covariates to conditional quantile of the response variable. For instance, when the parameter $u = 0.5$, the covariates are linked to the conditional median of the response variable. Since the *UBXII* distribution is defined on the unit-interval, we use the logit-link function such that

$$g(\mu_i) = \text{logit}(\mu_i) = \log \left(\frac{\mu_i}{1 - \mu_i} \right), \quad i = 1, 2, \dots, n.$$

Alternatively, the probit and log-log link functions can be used for linking to conditional quantiles of the response variable.

6.3 Parameter estimation

We point out the unknown parameters of the *UBXII* quantile regression model to obtain them via the MLE method. Thus, we consider

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{x}_i \boldsymbol{\delta}^T. \tag{16}$$

From Equation (16), the following relation is obtained

$$\mu_i = \frac{\exp(\mathbf{x}_i \boldsymbol{\delta}^T)}{1 + \exp(\mathbf{x}_i \boldsymbol{\delta}^T)}. \tag{17}$$

Let Y_1, Y_2, \dots, Y_n be a random sample of size n with $Y_i \sim \text{UBXII}(\alpha, \mu_i, u)$ and observed values y_1, y_2, \dots, y_n , where the μ_i is given by (17) for $i = 1, \dots, n$. Then using Equation (14), the associated log-likelihood function is given by

$$\begin{aligned} \ell(\boldsymbol{\Omega}) = & n \log \alpha + n \log [\log(u^{-1/\alpha} - 1)] - \sum_{i=1}^n \log y_i - \sum_{i=1}^n \log [\log(-\log \mu_i)] \\ & - \sum_{i=1}^n \left(\frac{\log(u^{-1/\alpha} - 1)}{\log(-\log \mu_i)} - 1 \right) \log(-\log y_i) - (\alpha + 1) \sum_{i=1}^n \log \left(1 + (-\log y_i)^{\frac{\log(u^{-1/\alpha} - 1)}{\log(-\log \mu_i)}} \right), \end{aligned} \tag{18}$$

where $\boldsymbol{\Omega} = (\alpha, \boldsymbol{\delta})^T$ is the unknown parameter vector. The MLEs of $\boldsymbol{\Omega}$, say $\hat{\boldsymbol{\Omega}} = (\hat{\alpha}, \hat{\boldsymbol{\delta}})^T$, is obtained by maximizing $\ell(\boldsymbol{\Omega})$ with respect to $\boldsymbol{\Omega}$. Since Equation (18) includes nonlinear function according to model parameters, it can be maximized directly by software such as R, S-Plus, and Mathematica.

It can be noticed that, when $u = 0.5$, it is equivalent to model the conditional median. Under mild regularity conditions, the asymptotic distribution of $(\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega})$ is multivariate normal $N_{p+1}(0, I)$, where I is the inverse of the expected information matrix. One may use the $(p + 1) \times (p + 1)$ observed information matrix instead of I . The elements of this observed information matrix are evaluated numerically by the software. We use the **maxLik** function (see Henningsen and Toomet 2011) of R software to maximize Equation (18). This function also gives asymptotic standard errors numerically, which are obtained by the observed information matrix.

6.4 Residual analysis

The residual analysis can be needed to check whether the regression model is suitable. In order to see this, we will point out the randomized quantile residuals (see Dunn and Smyth 1996) and the Cox-Snell residuals (see Cox and Snell 1968).

For $i = 1, \dots, n$, the i^{th} randomized quantile residual is defined by

$$\hat{r}_i = \Phi^{-1} [G(y_i, \hat{\alpha}, \hat{\mu}_i)],$$

where the $G(y, \alpha, \mu)$ is the cdf of the re-parameterized *UBXII* distribution given by Equation (15), $\Phi^{-1}(x)$ is the qf of the standard normal distribution, and $\hat{\mu}_i$ is defined by Equation (17) with $\hat{\boldsymbol{\delta}}$ instead of $\boldsymbol{\delta}$. If the fitted model successfully deals with the data set, the distribution of the randomized quantile residuals will correspond to the standard normal distribution.

Table 1 Some summary statistics of the data set

Minimum	Mean	Median	Maximum	Variance	Skewness	Kurtosis	<i>n</i>
0.4000	0.7907	0.8000	0.9900	0.0130	-0.7851	3.8237	239

Alternatively, for $i = 1, \dots, n$, the i^{th} Cox and Snell residual is given by

$$\hat{e}_i = -\log [1 - G(y_i, \hat{\alpha}, \hat{\mu}_i)].$$

If the model fits to data accordingly, the distribution of these residuals will distribute the exponential distribution with scale parameter 1.

7 Data analysis

This section provides two real data sets applications for illustrating the modeling ability of the *UBXII* distribution. The first data set is about univariate real data modeling and the other is about quantile regression modeling. The used data set for both applications consists of some measurements about 239 patients who consented to autologous peripheral blood stem cell (PBSC) transplant after myeloablative doses of chemotherapy between the years 2003 and 2008 at the Edmonton Hematopoietic Stem Cell Lab in Cross Cancer Institute - Alberta Health Services. Autologous PBSC transplants have been widely used for rapid hematologic recovery following myeloablative therapy for various malignant hematological disorders. The data set informs about the patients age, gender, as well as their clinical characteristics such as recovery rates for viable CD34+ cells and chemotherapy receiving of the patients. The data set can be easily found in the `simplexreg` package, proposed by Zhang et al. (2016), of the R software. We use variable recovery rates of the viable CD34+ cells for univariate data modeling and relate this recovery rate with some covariates in the data set for the second application. The calculations have been obtained by the `maxLik` (see Henningsen and Toomet 2011) and `gofTest` functions of the R software. The details are the following.

7.1 Univariate modeling for the recovery rates of the viable CD34+ cells data

Here, used data is recovery rates of the viable CD34+ cells of the 239 patients, who consented to autologous PBSC transplant after myeloablative doses of chemotherapy. Some summary statistics and box plot of the data set have been given by Table 1 and Fig. 8. As it can be seen that the data is left skewed and some observations can be considered as outliers.

Based on this data set, we also compare fitting performances of the proposed distribution under the MLE method with well known unit distributions in the literature. These comparing distributions have the following pdfs:

- Beta distribution:

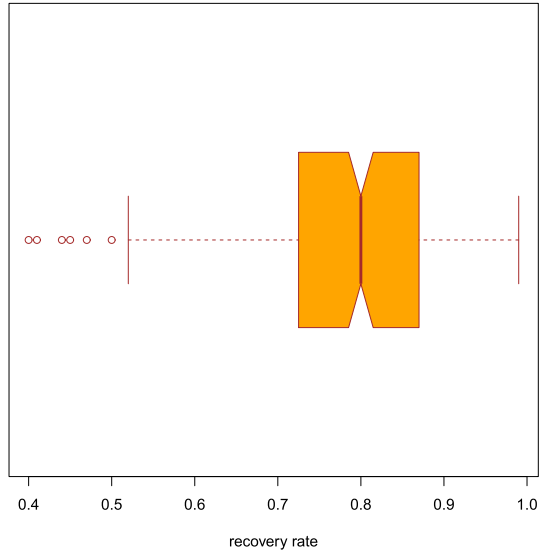
$$f_{Beta}(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

where $x \in (0, 1)$, $\alpha > 0$, $\beta > 0$ and $B(\alpha, \beta)$ denotes the classical beta function.

- Johnson S_B distribution:

$$f_{S_B}(x, \alpha, \beta) = \frac{\beta}{x(1-x)} \phi \left[\beta \log \left(\frac{x}{1-x} \right) + \alpha \right]$$

Fig. 8 The box plot of the data set



where $x \in (0, 1)$, $\alpha \in \mathbb{R}$, $\beta > 0$ and $\phi(x)$ denotes the pdf of the standard normal distribution.

- Kumaraswamy (Kw) distribution:

$$f_{Kw}(x, \alpha, \beta) = \alpha\beta x^{\alpha-1} (1 - x^\alpha)^{\beta-1},$$

where $x \in (0, 1)$, $\alpha > 0$ and $\beta > 0$.

- Standard two-sided power (STSP) distribution (see van Dorp and Kotz 2002):

$$f_{STSP}(x, \alpha, \beta) = \begin{cases} \beta \left(\frac{x}{\alpha}\right)^{\beta-1}, & x \in (0, \alpha) \\ \beta \left(\frac{1-x}{1-\alpha}\right)^{\beta-1}, & x \in (\alpha, 1), \end{cases}$$

where $\alpha \in (0, 1)$ and $\beta > 0$.

We use the estimated log-likelihood values $\hat{\ell}$, Akaike information criteria (AIC), Bayesian information criterion (BIC), Kolmogorov–Smirnov (KS), Cramér–von–Mises (W^*) and Anderson–Darling (A^*) goodness of-fit statistics criteria to determine the optimum model. The optimum model will have the smaller the values of the AIC, BIC, KS , W^* and A^* statistics and the larger the values of $\hat{\ell}$ and p-value of the goodness-of-statistics.

We give the estimates and the values of goodness-of-fits statistics in Table 2. In particular, Table 2 indicates that the $UBXII$ distribution has the lowest values of AIC and BIC statistics as well as it has the lowest values of A^* and W^* and K-S statistics with higher p-value. These results show that the $UBXII$ distribution is the best model for the considered data set.

Figure 9 presents the estimated pdfs, cdfs, and, the quantile–quantile (QQ) plot of the $UBXII$ model to see how they are the suitability for data set graphically. From this figure, it can be said that the $UBXII$ model has fitted by successfully capturing the skewness and kurtosis of the data set. The QQ plot indicates that the fitting performance of the $UBXII$ distribution is suitable for the data.

Table 2 MLEs, standard errors (SE) of the estimates (in parentheses), $\hat{\ell}$ and goodness-of-fits statistics for the data set (p value is given in [·])

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\ell}$	AIC	BIC	A*	W*	K S
UBXII	10.0754 (1.0526)	1.7321 (0.0811)	193.5027	-383.0054	-376.0525	0.5807	0.0887	0.0522 [0.5321]
Johnson S_B	-1.7834 (0.1041)	1.1823 (0.0541)	179.2383	-354.4766	-347.5236	2.8124	0.4337	0.0989 [0.0185]
Beta	8.6671 (0.9929)	2.2859 (0.2419)	191.8672	-379.7345	-372.7816	0.8732	0.1402	0.0650 [0.2647]
Kw	6.6943 (0.4499)	2.4353 0.2391	190.7640	-377.5280	-370.5751	1.1448	0.1919	0.0723 [0.1641]
STSP	0.9900 (0.0006)	4.2288 (0.2740)	162.2770	-320.2555	-313.3025	14.2260	2.7684	0.1966 [< 0.0001]

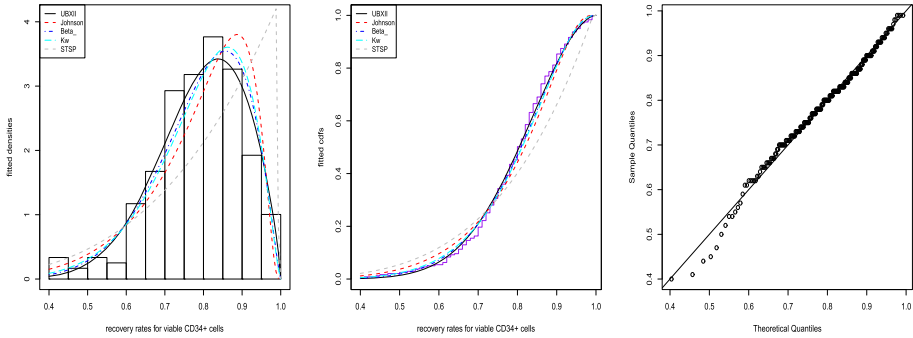


Fig. 9 The fitted pdfs (left), cdfs (center) and QQ plot (right) of the *UB XII* model

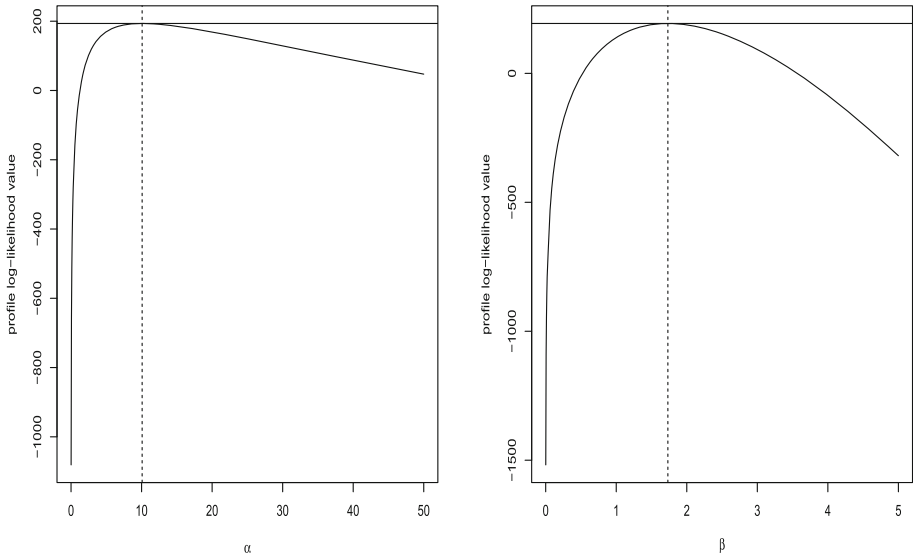


Fig. 10 The plots of the PLL functions for the data set

Further, to show the likelihood equations have a unique solution, we plot the PLL functions of the parameters α and β for the data set in Fig. 10. From this figure, we see that the likelihood equations have a unique solution for the MLEs.

7.2 Quantile regression modeling for the recovery rates of the viable CD34+ cells data

To see the applicability of the *UB XII* regression model, this section presents a real application based on the above data set. The aim is to associate these response values (y) with covariates. The response variable y and covariates associated with this response variable are:

- y : recovery rate of CD34+cells;
- x_1 (Gender): 0 for female, 1 for male;
- x_2 (Chemotherapy): 0 for receiving chemotherapy on a one-day protocol, 1 for a 3-day protocol;

- x_3 (Age): adjusted patient’s age, i.e. the current age minus 40.

This data set has been analyzed by Mazucheli et al. (2020). Since the response variable has some observed outliers, the use of the quantile regression will be better for its inferences. Two competitor regression models, which are well-known literature, are considered and compared to the *UBXII* quantile regression model. They are the beta regression (see Ferrari and Cribari-Neto 2004) and Kumaraswamy quantile regression (see Mitnik and Baek 2013) models. Their pdfs are

$$f(y, \alpha, \mu) = \frac{\Gamma(\alpha)}{\Gamma(\alpha\mu)\Gamma((1-\mu)\alpha)} y^{\alpha\mu-1} (1-y)^{(1-\mu)\alpha-1}, \quad y \in (0, 1),$$

where $\mu \in (0, 1)$ is the mean and $\alpha > 0$ and

$$f(y, \alpha, \mu) = \frac{\alpha \log(0.5)}{\log(1-\mu^\alpha)} y^{\alpha-1} (1-y^\alpha)^{\log(0.5)/(\alpha(1-\mu)-1)}, \quad y \in (0, 1),$$

where $\mu \in (0, 1)$ is the median and $\alpha > 0$, for the beta and Kumaraswamy quantile regression models, respectively.

The regression model based on μ_i is given by

$$\text{logit}(\mu_i) = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3} \quad i = 1, 2, \dots, 239.$$

It is noticed that the quantile parameter u is taken 0.5 for the *UBXII* and Kumaraswamy quantile regression models. We give the results of the regression analysis in Table 3. From this table, parameter estimations of all models are greater than zero. It means that if there is any relation statistically significant between the recovery rate of CD34+cells and covariates, they will affect the recovery rate of CD34+cells positively. The parameter δ_1 has not been seen statistically significant at usual level for all regression models. So, the gender variable has no effect on the response variable recovery rate. However, the parameters δ_2 and δ_3 have been seen statistically significant at 7% level for the *UBXII* regression model. Hence, the recovery rate of the old patients is higher than the young patients as well as the recovery rate of the patients who are receiving chemotherapy on a 3-day protocol is higher than those of the patients who are receiving chemotherapy on a one-day protocol.

Further, the *UBXII* regression model has lower values of the AIC and BIC statistics have an upper log-likelihood value than those of other regression models. So, it can be concluded that the proposed regression model is the best model among application models in terms of better modeling ability than other regression models.

Figures 11 and 12 display the QQ plots of the randomized quantile residuals and PP plots of the Cox-Snell residuals for all regression models, respectively. These figures indicate that the fitting of the *UBXII* regression model is better than those of the beta and Kumaraswamy models.

Since the distribution behind the randomized quantile residuals are theoretically in adequateness with a standard normal distribution, one may see whether they fit this corresponding distribution. The KS, A^* and W^* results are given in Table 4. From this table, it is clear that the results based on the *UBXII* quantile regression model of the randomized quantile residuals are more suitable than those of the beta and Kumaraswamy regression models.

8 Conclusion

We define a new unit model, called unit Burr-XII distribution, in order to model percentage, proportion and rate measurements. We investigate general structural properties of the new

Table 3 The results of fitted regression models with model selection criteria

Parameters	Beta Estimate	SE	<i>p</i> -value	Kumaraswamy Estimate	SE	<i>p</i> -value	UBXII Estimate	SE	<i>p</i> -value
δ_0	0.9990	0.1278	< 0.0001	1.1996	0.1397	< 0.0001	1.0428	0.1484	< 0.0001
δ_1	0.0659	0.0939	0.4829	0.0418	0.0955	0.6620	0.0584	0.1060	0.5818
δ_2	0.2116	0.1017	0.0374	0.1835	0.1150	0.1106	0.2248	0.1218	0.0649
δ_3	0.0142	0.0053	0.0072	0.0107	0.0059	0.0679	0.0165	0.0063	0.0088
α	11.3450	1.0180	< 0.0001	6.7271	0.4556	< 0.0001	10.4135	1.0432	< 0.0001
$\hat{\ell}$	195.8956			192.8299			197.1325		
AIC	-381.7912			-375.6599			-384.2649		
BIC	-364.4089			-358.2775			-366.8826		

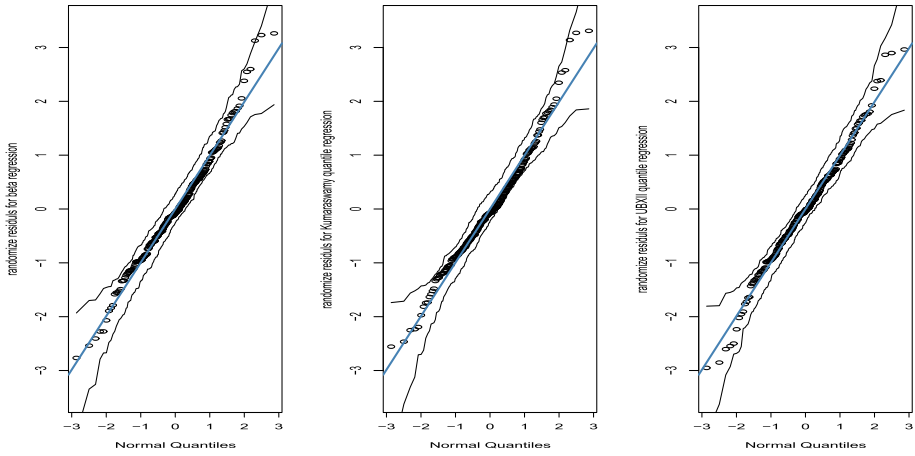


Fig. 11 The QQ plot of the randomized quantile residuals based on the regression application

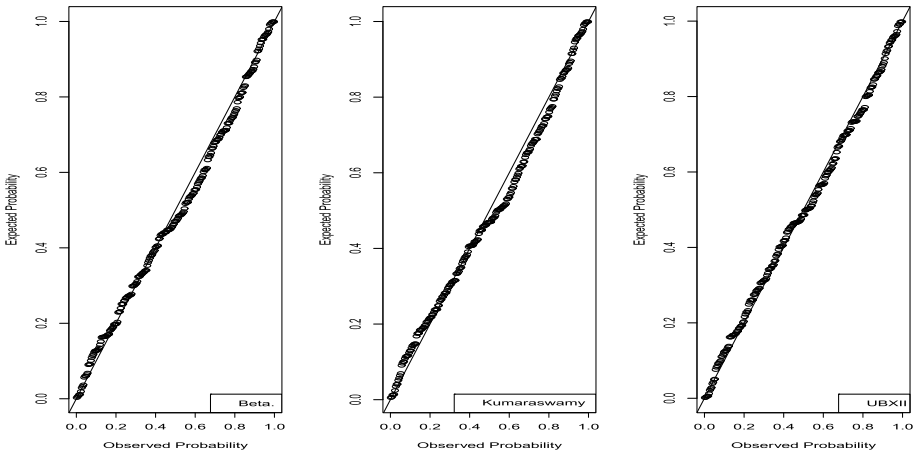


Fig. 12 The PP plots of the Cox-Snell residuals based on the regression application

Table 4 The goodness-of-fit results of the randomized quantile residuals for the regression models

Models	KS	<i>p</i> -value	<i>A</i> *	<i>p</i> -value	<i>W</i> *	<i>p</i> -value
UBXII	0.0422	0.7893	0.5652	0.6813	0.0816	0.6829
Beta	0.0518	0.5425	0.9217	0.4010	0.1470	0.3991
Kumaraswamy	0.0673	0.2285	1.2699	0.2423	0.2147	0.2409

distribution. The model parameters are estimated by the six different methods. The simulation studies are performed to see the performances of these estimates. The empirical findings indicate that the proposed model provides better fits than the well-known unit probability distributions in the literature for both its univariate data modeling and its regression modeling. It is hoped that the new distribution will attract attention in the other disciplines.

References

- Afify AZ, Cordeiro GM, Ortega EM, Yousof HM, Butt NS (2018) The four-parameter Burr XII distribution: properties, regression model, and applications. *Commun Stat Theory Methods* 47(11):2605–2624
- Altun E (2021) The log-weighted exponential regression model: alternative to the beta regression model. *Commun 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 Stat* 35(1):259–279
- Altun E, Hamedani GG (2018) The log-xgamma distribution with inference and application. *Journal de la Société Française de Statistique* 159(3):40–55
- Altun E, Yousof HM, Chakraborty S, Handique L (2018a) Zografos-Balakrishnan Burr XII distribution: regression modeling and applications. *Int J Math Stat* 19(3):46–70
- Altun E, Yousof HM, Hamedani GG (2018b) A new log-location regression model with influence diagnostics and residual analysis. *Facta Universit Ser Math Inf* 33(3):417–449
- Amigo JM, Balogh SG, Hernandez S (2018) A brief review of generalized entropies. *Entropy* 20:813
- Bayes CL, Bazán JL, De Castro M (2017) A quantile parametric mixed regression model for bounded response variables. *Stat Interface* 10(3):483–493
- Beirlant J, Goegebeur Y, Verlaak R, Vynckier P (1998) Burr regression and portfolio segmentation. *Insur Math Econ* 23(3):231–250
- Bottai M, Cai B, McKeown RE (2010) Logistic quantile regression for bounded outcomes. *Stat Med* 29(2):309–317
- Burr IW (1942) Cumulative frequency functions. *Ann Math Stat* 13(2):215–232
- Cheng RCH, Amin NAK (1979) Maximum product of spacings estimation with application to the lognormal distribution. *Math Report*, 791
- Consul PC, Jain GC (1971) On the log-gamma distribution and its properties. *Stat Pap* 12:100–106
- Cox DR, Snell EJ (1968) A general definition of residuals. *J R Stat Soc Ser B (Methodol)* 30(2):248–265
- Cooray K, Ananda MM (2008) A generalization of the half-normal distribution with applications to lifetime data. *Commun Stat Theory Methods* 37(9):1323–1337
- Cordeiro GM, Yousof HM, Ramires TG, Ortega EM (2018) The Burr XII system of densities: properties, regression model and applications. *J Stat Comput Simul* 88(3):432–456
- David HA, Nagaraja H (2003) *Order statistics*, 3rd edn. Wiley, New York
- Dunn PK, Smyth GK (1996) Randomized quantile residuals. *J Comput Graph Stat* 5(3):236–244
- Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. *J Appl Stat* 31(7):799–815
- Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 8(1):140–154
- Ghitany ME, Mazucheli J, Menezes AFB, Alqallaf F (2019) The unit-inverse Gaussian distribution: a new alternative to two-parameter distributions on the unit interval. *Commun Stat Theory Methods* 48(14):3423–3438
- 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. *Insur Math Econ* 54:49–57
- Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour Res* 15:1049–1054
- Gupta RD, Kundu D (2009) A new class of weighted exponential distributions. *Statistics* 43(6):621–634
- Gündüz S, Korkmaz MÇ (2020) A new unit distribution based on the unbounded Johnson distribution rule: the unit Johnson SU distribution. *Pak J Stat Oper Res* 16(3):471–490
- Henningesen A, Toomet O (2011) maxLik: a package for maximum likelihood estimation in R. *Comput Stat* 26(3):443–458
- Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36(1/2):149–176
- Jung SH (1996) Quasi-likelihood for median regression models. *J Am Stat Assoc* 91(433):251–257
- Koenker R, Bassett G Jr (1978) Regression quantiles. *Econometrica J Econ Soc* 33–50
- Korkmaz MÇ (2020a) A new heavy-tailed distribution defined on the bounded interval: the logit slash distribution and its application. *J Appl Stat* 47(12):2097–2119
- Korkmaz MÇ (2020b) The unit generalized half normal distribution: a new bounded distribution with inference and application. *Univ Politeh Bucharest Sci Bull Ser A Appl Math Phys* 82(2):133–140
- Korkmaz MÇ, Chesneau C, Korkmaz Z S (2021a) Transmuted unit Rayleigh quantile regression model: alternative to beta and Kumaraswamy quantile regression models. *Univ Politeh Bucharest Sci Bull Ser A Appl Math Phys* to appear

- Korkmaz MÇ, Chesneau C, Korkmaz Z S (2021b) On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications. *Symmetry* (to appear)
- Kumaraswamy P (1980) A generalized probability density function for double-bounded random processes. *J Hydrol* 46(1–2):79–88
- Lanjoni BR, Ortega EM, Cordeiro GM (2016) Extended Burr XII regression models: theory and applications. *J Agric Biol Environ Stat* 21(1):203–224
- Lindley DV (1958) Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)* :102–107
- Liu Y, Bottai M (2009) Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostat* 5(1)
- Mazucheli J, Menezes AFB, Chakraborty S (2019a) On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *J Appl Stat* 46(4):700–714
- Mazucheli J, Menezes AF, Dey S (2019b) Unit-Gompertz distribution with applications. *Statistica* 79(1):25–43
- Mazucheli J, Menezes AF, Dey S (2018) The unit-Birnbaum–Saunders distribution with applications. *Chile J Stat* 9(1):47–57
- Mazucheli J, Menezes AFB, Ghitany ME (2018) The unit-Weibull distribution and associated inference. *J Appl Probab Stat* 13:1–22
- 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(6):954–974
- Mitnik PA, Baek S (2013) The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Stat Pap* 54(1):177–192
- Muller A, Stoyan D (2002) Comparison methods for stochastic models and risks. Wiley, Chichester
- Nair NU, Sankaran PG (2009) Quantile based reliability analysis. *Commun Stat Theory Methods* 38:222–232
- Paranaíba PF, Ortega EMM, Cordeiro GM, Pescim RR (2011) The beta Burr XII distribution with application to lifetime data. *Comput Stat Data Anal* 55:1118–1136
- Pourdarvish A, Mirmostafae SMTK, Naderi K (2015) The exponentiated Topp-Leone distribution: Properties and application. *J Appl Environ Biol Sci* 5(7):251–6
- Rodriguez RN (1977) A guide to the Burr type XII distributions. *Biometrika* 64(1):129–134
- Sen S, Maiti SS, Chandra N (2016) The xgamma distribution: statistical properties and application. *J Modern Appl Stat Methods* 15(1):38
- Silva RB, Cordeiro GM (2015) The Burr XII power series distributions: a new compounding family. *Br J Probab Stat* 29(3):565–589
- Silva GO, Ortega EM, Cancho VG, Barreto ML (2008) Log-Burr XII regression models with censored data. *Comput Stat Data Anal* 52(7):3820–3842
- Shaked M, Shanthikumar JG (2007) Stochastic orders. Wiley, New York
- Surles JG, Padgett WJ (2001) Inference for reliability and stress–strength for a scaled Burr-type X distribution. *Lifetime Data Anal* 7:187–200
- Tadikamalla PR (1980) A look at the Burr and related distributions. *Int Stat Rev/Revue Internationale de Statistique* 337–344
- Usta I (2013) Different estimation methods for the parameters of the extended Burr XII distribution. *Journal of Applied Statistics* 40(2):397–414
- van Dorp JR, Kotz S (2002) The standard two-sided power distribution and its properties: with applications in financial engineering. *Am Stat* 56(2):90–99
- Zhang P, Qiu Z, Shi C (2016) simplexreg: an R package for regression analysis of proportional data using the simplex distribution. *J Stat Softw* 71(11)
- Yousof HM, Altun E, Rasekhi M, Alizadeh M, Hamedani GG, Ali MM (2019) A new lifetime model with regression models, characterizations and applications. *Commun Stat Simul Comput* 48(1):264–286
- Zimmer WJ, Keats JB, Wang FK (1998) The Burr XII distribution in reliability analysis. *Journal of quality technology* 30(4):386–394