Heliyon 9 (2023) e22260

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Unit upper truncated Weibull distribution with extension to 0 and 1 inflated model – Theory and applications

Idika E. Okorie [a], Emmanuel Afuecheta [b,d,*], Hassan S. Bakouch [c,e]

[a] *Department of Mathematics, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates*
[b] *Department of Mathematics, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia*
[c] *Department of Mathematics, College of Science, Qassim University, Buraydah, Saudi Arabia*
[d] *Interdisciplinary Research Center for Finance and Digital Economy, KFUPM, Dhahran, Saudi Arabia*
[e] *Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt*

ARTICLE INFO

MSC:
62E05
62G08
97K80
62J05

Keywords:
Inequality measures
Cox-Snell residuals
Regression
Household income and expenditure
Statistical model
Simulation

ABSTRACT

A two-parameter unit distribution and its regression model plus its extension to 0 and 1 inflation is introduced and studied. The distribution is called the unit upper truncated Weibull (UUTW) distribution, while the inflated variant is called the $0-1$ inflated unit upper truncated Weibull (ZOIUUTW) distribution. The UUTW distribution has an increasing and a J-shaped hazard rate function. The parameters of the proposed models are estimated by the method of maximum likelihood estimation. For the UUTW distribution, two practical examples involving household expenditure and maximum flood level data are used to show its flexibility and the proposed distribution demonstrates better fit tendencies than some of the competing unit distributions. Application of the proposed regression model demonstrates adequate capability in describing the real data set with better modeling proficiency than the existing competing models. Then, for the ZOIUUTW distribution, the CD34+ data involving cancer patients are analyzed to show the flexibility of the model in characterizing inflation at both endpoints of the unit interval.

1. Introduction

In practice, phenomena that can give rise to data on the restricted domain of [0,1) are quite ubiquitous, where such data could represent proportions or even some sort of standardized or transformed values, for example in demography we have the sex, dependency, abortion and child-woman ratios; in economics and finance we have the proportion of household income spent on food, as well as the following famous ratios: savings to income, capital to output, labor's share income, income velocity of circulation and capital to labor ratios; in engineering and related fields, we have the crawl ratio in automotive engineering, aspect ratio in aeronautics, fineness ratio in naval architecture and aerospace engineering, extinction ratio (r_e) in telecommunication, common mode rejection ratio (CMRR) in electronics and lighting ratio in photography. Other applications can be found in the beta distribution's guide [1]. To model this kind of data, one distribution that first come to mind is the beta distribution. However, in most application problems, we are constrained with the much we could do with the beta distribution because, its distribution function (cdf) does not appear in any analytical closed form. A similar and popular alternative to the beta distribution is the Kumaraswamy distribution [2]. Unlike the cdf of the beta distribution, the cdf of the Kumaraswamy distribution come in a nice close form therefore, making

* Corresponding author.
E-mail addresses: idika.okorie@ku.ac.ae (I.E. Okorie), emmanuel.afuecheta@kfupm.edu.sa (E. Afuecheta), h.bakouch@qu.edu.sa (H.S. Bakouch).

https://doi.org/10.1016/j.heliyon.2023.e22260
Received 25 November 2022; Received in revised form 24 October 2023; Accepted 8 November 2023

Available online 13 November 2023

it easier to work with compare to the beta distribution. Jones [3] presents a list of similarities and dissimilarities between the beta distribution and the Kumaraswamy distribution as well as the advantages and the disadvantages of one over the other.

Apart from the Kumaraswamy distribution, many other unit distributions have emerged for modeling the ever growing number of data-sets on the unit interval which could emanate from different complex phenomena, with the ultimate goal of providing better fit capabilities for this kind of data. Some of these distributions were introduced even before the emergence of the Kumaraswamy distribution thus, providing us with several alternative distributions to try should any lack of fit occasion the use of either the beta distribution or the Kumaraswamy distribution. Some of these distributions includes the Topp-Leone distribution [4]. The unit-gamma distribution [5] is one of the few unit interval distributions we are aware of, as well as the following more recent ones: Log-Lindley distribution [6], the two-parameter unit logistic distribution [7], the two-parameter unit-Birnbaum–Saunders distribution [8], unit-Gompertz [9], unit-inverse Gaussian distribution [10], unit-Rayleigh distribution [11], the unit-Weibull distribution [12], unit modified Burr-III distribution [13] and log-weighted exponential(WE) distribution [14], transmuted unit Rayleigh quantile regression model [15], unit generalized half-normal quantile regression model [16], unit generalized log Burr XII distribution [17] and unit folded normal distribution [18].

As earlier mentioned, unit data or data on [0,1) can result from data scaling/transformation. In the same vein, unit distributions can as well be constructed by some random variable transformations. For instance, if the random variable X is defined on $(0,\infty)$, then the random variable $Y=\exp(-X)$ can result to a unit distribution and a typical example of this case is when Y is distributed either as the unit-Rayleigh distribution or the log-WE distribution provided that X follows either the Rayleigh distribution or the weighted exponential(WE) distribution by [19], respectively. In a different development, the random variable $Y=X/(1+X)$ is said to follow the unit-Lindley distribution [9] if the random variable X follows the Lindley distribution [20]; also for the case of the so called power transformation; let $\lambda>0$, the random variable $Y=X^{\lambda}$ is said to follow the 2-parameter Topp-Leone distribution [21] if the random variable X follows the Topp-Leone distribution [4].

In addition to data on $[0,1)$, in practice, we may encounter unit data that includes other boundary cases such as $(0,1]$ and $[0,1]$. That is, the data may involve in addition to observations in the interval of $(0,1)$, many occurrences of both 0's and 1's or either a preponderance of 0's or a preponderance 1's and in such situations, we can apply the so called hurdle models for analysis. Several hurdle models exist in the literature and some of them includes the inflated beta distributions [22], the inflated Kumaraswamy distributions [23] and the inflated unit-Birnbaum-Saunders distribution [24].

The remainder of this paper contains: the proposed distribution in Section 2; additional properties of the proposed distribution in Section 3, parameter estimation in Section 4, UUTW regression model in Section 5, proposed 0 and 1 inflated model in Section 6, application examples in Section 7, discussion of results in Section 8 and concluding remarks in Section 9.

2. The proposed distribution

Here, we introduce the unit upper truncated Weibull distribution; hereafter referred to as, the UUTW distribution following a random variable transformation based on the zero-truncated Poisson power function (ZTPPF) random variable [25]. We would like to point out that the ZTPPF distribution is actually the upper truncated Weibull distribution. A random variable Y is said to follow the ZTPPF distribution if its cdf and probability density function (pdf) are given by

$$F(y)=\frac{1-\exp\left(-\lambda\left[\frac{y}{\alpha}\right]^{\beta}\right)}{1-\exp[-\lambda]}$$

and

$$f(y)=\frac{\lambda\beta y^{\beta-1}\exp\left(-\lambda\left[\frac{y}{\alpha}\right]^{\beta}\right)}{\alpha^{\beta}\left(1-\exp[-\lambda]\right)},$$

respectively, where $0<y<\alpha$, $\alpha>0$ and $\beta>0$.

Now, suppose that the random variable X follows the ZTPPF distribution, we define a new random variable Y by the transformation $Y=1-X/\alpha$ and the cdf of Y is defined by

$$F(y)=1-\frac{1-\exp\left(-\lambda[1-y]^{\beta}\right)}{1-\exp(-\lambda)}, \tag{1}$$

for $0\le y<1$, and the corresponding pdf given by

$$f(y)=\begin{cases}\frac{\lambda\beta}{1-\exp(-\lambda)}(1-y)^{\beta-1}\exp\left(-\lambda[1-y]^{\beta}\right), & 0\le y<1,\\ 0, & \text{elsewhere},\end{cases} \tag{2}$$

where $\lambda>0$ and $\beta>0$ are the shape parameters.

To emphasize the importance of the above transformation, we admit that the ZTPPF distribution can easily assume a unit distribution when the scale parameter α is equal to 1 but, this approach is restrictive in practice. However, the recommended transformation that resulted in Equation (1) enhances more flexibility in the construction of the unit distribution in that, α is allowed to take any value on the entire positive domain, i.e., $\alpha \in (0, \infty)$. In the application section, we compare the UUTW distribution whose pdf appear in Equation (2) with that obtained by simply setting $\alpha = 1$ in the ZTPPF distribution through different real-data examples. Also, the lifetime of many manufactured equipment and devices are sometimes characterized by the J-shaped hazard rate function, which usually involves two phases, namely; the phase of long useful life and the phase of wear-out. We have emphasized that there are many continuous distributions on the unit support however, most of these distributions mainly exhibits "bathtub shape" or the "upside-down bathtub shape" hazard rate characteristics or both. For instance, the bathtub hazard rate is important for describing the failure rate of some lifetime phenomena e.g., human. The bathtub shape involves three phases: early failure phase (e.g., infant mortality), chance failure phase (e.g., sudden death due to accidents) and the wear-out failure phase (i.e.; death due to the accumulation of natural impacts) [26]. There are limited number of distributions with J-shape hazard rate hence, the main motivation for this paper.

3. Additional properties

The reliability function is the most widely applied function in lifetime analysis and reliability engineering and it gives the probability that a functional device or item will operate over a certain period of time without failure. The reliability function of the UUTW distribution is given by Equation (3)

$$R(y) = \frac{1 - \exp(-\lambda[1-y]^{\beta})}{1 - \exp(-\lambda)}. \tag{3}$$

The hazard rate function (hrf) quantifies the likelihood of failure for any operating device or item hence; larger value of the hazard function implies higher probability of imminent failure. In other words, the hazard function represents the probability of failure over a short interval of time t_0 to $t_0 + \Delta t_0$ given that the device or item has survived up to t_0 thus, it is sometimes referred to as the instantaneous failure rate. The hazard function of the UUTW distribution is given by Equation (4)

$$h(y) = \frac{\lambda\beta[1-y]^{\beta-1}}{\exp(\lambda[1-y]^{\beta}) - 1}. \tag{4}$$

The plots of the cdf, pdf, reliability function and the hrf of the UUTW distribution are shown in Fig. 2 and from there, we could see that the pdf can be unimodal, L-shaped or J-shaped while the hrf is either increasing or J-shaped.

Apart from the tractable cdf of the proposed distribution, the J-shape characteristic of its hrf is another strength and motivation of this paper because, several existing survival models often present increasing hrf but, lack the relatively constant phase of the hrf. However, this flat region of the J-shaped hrf otherwise known as the long useful period is important and perhaps one of the most significant phase for reliability prediction and evaluation activities, as it explains the normal life span of the component or system. Therefore, having a model that can adequately capture this flat region is very essential. Our model is suitable for modeling the long useful and the final (wear-out) phases of many lifetime phenomena. The wear-out periods are usually as a result of the natural accumulation of adverse impacts.

3.1. Asymptotic behavior

The limiting behaviors for the pdf and hrf of the UUTW distribution are presented as follows. For any $\lambda \in (0, \infty)$ and $\beta \in (1, \infty)$, we have that $\lim_{y\to 0^+} f(y) = \lim_{y\to 0^+} h(y) = \frac{\lambda\beta}{\exp\lambda - 1}$; $\lim_{y\to 1^-} f(y) = 0$ and $\lim_{y\to 1^-} h(y) = \infty$. Also, for any $\lambda \in (0, \infty)$ and $\beta \in (0, 1]$, we have that $\lim_{y\to 0^+} f(y) = \lim_{y\to 0^+} h(y) = \frac{\lambda\beta}{\exp\lambda - 1}$ and $\lim_{y\to 1^-} f(y) = \lim_{y\to 1^-} h(y) = \infty$. When $\beta = 1$, the asymptotic behavior of $f(y) = \frac{\lambda\exp(\lambda y)}{\exp(\lambda)-1}$ (i.e., $f(y)$ is increasing) and the asymptotic behavior of $h(y) = \frac{\lambda}{\exp(\lambda[1-y])-1}$ (i.e., $h(y)$ is increasing) with the corresponding limit behaviors as follows: $\lim_{y\to 0^+} f(y) = \lim_{y\to 0^+} h(y) = \frac{\lambda}{\exp(\lambda)-1}$ and $\lim_{y\to 1^-} f(y) = \frac{\lambda\exp(\lambda)}{\exp(\lambda)-1}$ and $\lim_{y\to 1^-} h(y) = \infty$.

Clearly, for any $\lambda > 0$ and $\beta > 0$, the following result from fundamental calculus: $\lim_{y\to 0^+} f(y) = f(0) = \frac{\lambda\beta}{\exp\lambda - 1}$ suggests that the UUTW distribution is continuous at the lower end point; thus, the distribution could have a bounded lower support.

3.2. Shape property

The first derivative of the natural logarithm of the pdf of the UUTW distribution in Equation (2) with respect to y is given by

$$\frac{d\log f(y)}{dy} = -\frac{\beta-1}{1-y} + \lambda\beta(1-y)^{\beta-1}. \tag{5}$$

The critical value of the pdf denoted by y_0 is given by

$$y_0 = 1 - \left(\frac{\beta-1}{\lambda\beta}\right)^{\frac{1}{\beta}}; \ \lambda, \beta > 1. \tag{6}$$

From Equation (5), for $\lambda \le 1$ the pdf of the UUTW distribution can be strictly monotone on $y \in [0, 1)$ depending on the value of β. Specifically, the pdf is decreasing when $\beta > 1$ and increasing when $\beta \le 1$. If the critical value y_0 in Equation (6) exists, the pdf of the UUTW distribution can increase on the interval $[0, y_0)$ and decrease on the interval $(y_0, 1)$ suggesting that it's maximum only occur at y_0. In probability theory and statistics, y_0 is called the mode of the distribution. The mode describes the observation that has the

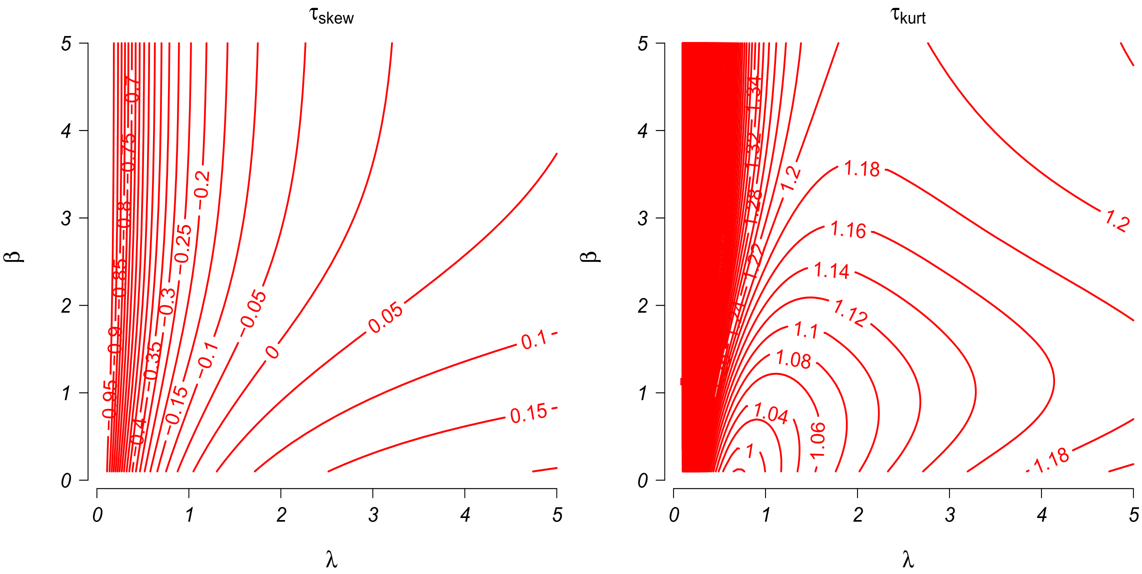

Fig. 1. Contour plots for the: skewness (τ_{skew}) and kurtosis (τ_{kurt}) of the UUTW for a wide range of λ and β values.

highest frequency (most occurring) in a set of data. In other words, the mode corresponds to the observation for which the pdf of the random variable Y has its maximum.

3.3. The quantile function

In probability theory, the quantile function of an independent and identically distributed random variable has many theoretical and applied uses. For instance, if $F(y)$ is the cdf, the quantile function can be used to generate random observations from $F(y)$. This idea serves as a basis for sampling or simulating from an arbitrary distribution. The quantile function of the UUTW distribution is defined by

$$Y_Q = 1 - \left[-\frac{1}{\lambda}\log\left(1-(1-Q)[1-\exp(-\lambda)]\right)\right]^{\frac{1}{\beta}}\ ;\ 0 < Q < 1. \tag{7}$$

We can simulate random sample of size n from the UUTW distribution by first assuming that Q follows the standard uniform distribution, i.e., $Q \sim Unif(0,1)$ and evaluating Y_Q in Equation (7) thus, $Y = Y_Q$ follows the UUTW distribution with parameter λ and β.
In statistics and probability theory, the median (second quartile) is an important measure of central tendency and it has a unique advantage over the mean when it comes to data description because, it is not affected by outliers thus, it gives a better representation of a typical value. In robust statistics, the median is the most resistant statistic. The median is of the UUTW distribution can be obtained by evaluating Equation (7) at $Q = \frac{1}{2}$ and it is given by Equation (8)

$$Y_{0.5} = 1 - \left[-\frac{1}{\lambda}\log\left(\frac{1}{2}[1+\exp(-\lambda)]\right)\right]^{\frac{1}{\beta}}. \tag{8}$$

Similarly, other quantiles like the first quartile and the third quartile can be calculated by setting Q to $\frac{1}{4}$ and $\frac{3}{4}$, respectively in Equation (7). The well-known Bowley's skewness (τ_{skew}) and Moor's kurtosis (τ_{kurt}) which are important measures of symmetry and tailedness, respectively; can be calculated by using the first, second, third and fourth quartiles and some octiles. Fig. 1 shows the contour plots of the Bowley's skewness and Moor's kurtosis for varying values of β and λ.

From Fig. 1 (τ_{skew}), we can see that the pdf of the UUTW distribution can either be skewed (right-skewed or left-skewed) or symmetric depending on the values of the parameters while Fig. 1 (τ_{kurt}) indicate that the tail of the pdf of the UUTW distribution is leptokurtic (fatter than that of the normal distribution).

3.4. The kth ordinary moment

The kth ordinary moment of any continuous random variable Y is defined by $\mathbb{E}(y^k) = \int_{y\in\mathrm{R}} y^k f(y)dy$ thus, if Y follows the UUTW distribution, its kth ordinary moment is given by

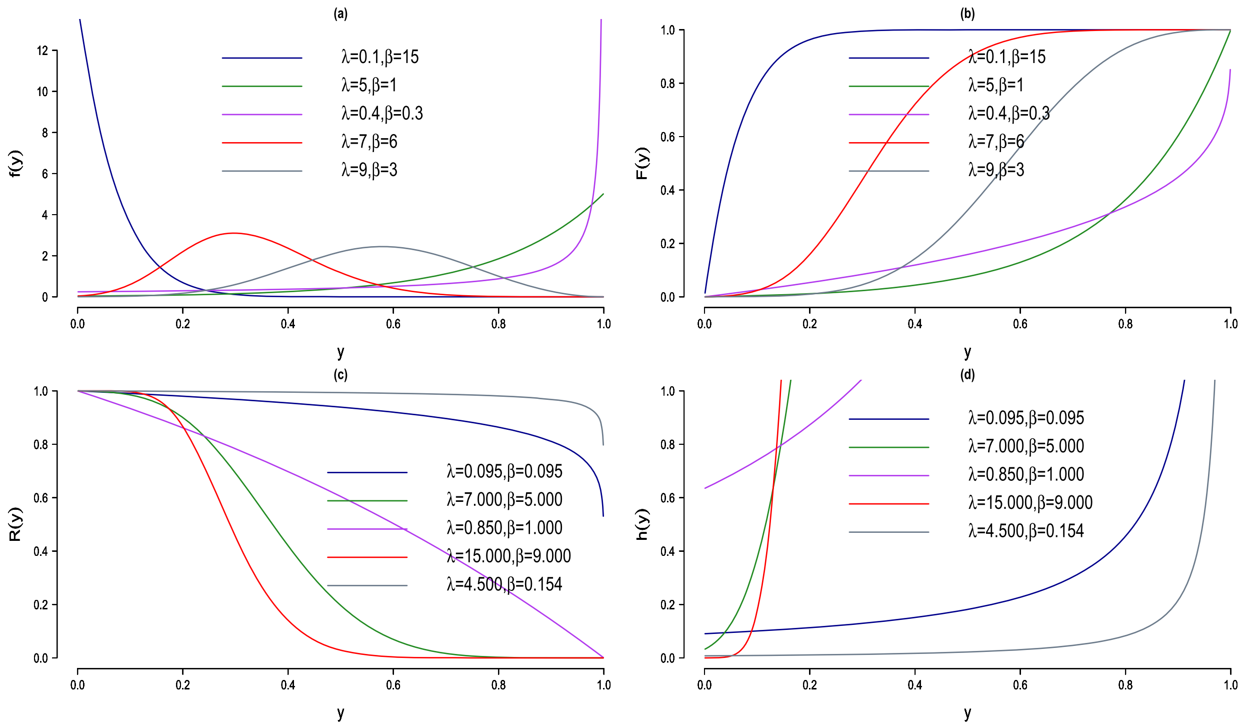

Fig. 2. Some plots of the pdf (a), cdf (b), Survival function (c) and hazard rate function (d) of the proposed distribution for selected values of λ and β.

$$\begin{aligned}\mathbb{E}(y^k) &= \int_0^1 y^k \frac{\lambda\beta}{1-\exp(-\lambda)}[1-y]^{\beta-1}\exp\left(-\lambda[1-y]^{\beta}\right)dy \\ &= \frac{\lambda\beta}{1-\exp(-\lambda)}\int_0^1 y^k[1-y]^{\beta-1}\sum_{i=0}^{\infty}\frac{(-\lambda)^i[1-y]^{\beta i}}{i!}dy \\ &= \frac{\lambda\beta\Gamma(k+1)}{1-\exp(-\lambda)}\,{}_1\Psi_1\left[\begin{array}{c|c} (\beta,\beta) & \\ & -\lambda \\ (\beta+k+1,\beta) & \end{array}\right]. \end{aligned} \tag{9}$$

Equation (9) was obtained after some trivial algebra. Where ${}_1\Psi_1(\cdot)$ is the special case of the multivariate ${}_m\Psi_n(\cdot)$ function defined by

$${}_m\Psi_n\left[\begin{array}{c|c} (\alpha_1,A_1),\cdots,(\alpha_m,A_m) & \\ & -z \\ (\gamma_1,B_1),\cdots,(\gamma_n,B_n) & \end{array}\right] = \sum_{i=0}^{\infty}\frac{\prod_{j=1}^{m}\Gamma(\alpha_j+A_j i)}{\prod_{k=1}^{n}\Gamma(\gamma_k+B_k i)}\frac{z^i}{i!}$$

for $-\infty < z < \infty$, where $-\infty < \alpha_j < \infty$, $-\infty < \gamma_k < \infty$, $A_j \neq 0$ and $B_k \neq 0$ for $j = 1, 2, \cdots, m$ and $k = 1, 2, \cdots, n$, see [27], [28], [29] and [30]-Equation (1.9).

Let denote $\mathbb{E}(y^k)$ by μ'_k, we list the first four order moments about zero of the UUTW distribution as follows:

$$\mu'_1 = \frac{\lambda\beta}{1-\exp(-\lambda)}\,{}_1\Psi_1\left[\begin{array}{c|c} (\beta,\beta) & \\ & -\lambda \\ (\beta+2,\beta) & \end{array}\right],$$

$$\mu'_2 = \frac{2\lambda\beta}{1-\exp(-\lambda)}\,{}_1\Psi_1\left[\begin{array}{c|c} (\beta,\beta) & \\ & -\lambda \\ (\beta+3,\beta) & \end{array}\right],$$

$$\mu'_3 = \frac{6\lambda\beta}{1-\exp(-\lambda)}\,{}_1\Psi_1\left[\begin{array}{c|c} (\beta,\beta) & \\ & -\lambda \\ (\beta+4,\beta) & \end{array}\right]$$

and

$$\mu_4' = \frac{24\lambda\beta}{1-\exp(-\lambda)}\, {}_1\Psi_1\left[\begin{array}{c|c} (\beta,\beta) & \\ & -\lambda \\ (\beta+5,\beta) & \end{array}\right].$$

Using these moments, one can easily calculate the variance, skewness, kurtosis and other important measures for the UUTW distribution.

3.5. Inequality measures

Popular measures of inequality in economics, insurance, health and social sciences are the Lorenz curve, the Bonferroni curve and the Gini index. These indices are used to quantify the distribution of income or resources in a population.

3.5.1. Lorenz curve

Using the quantile function (Equation (7)) and the first order ordinary moment μ_1', we obtain the Lorenz curve of the UUTW distribution by using Equation (10) as follows.

$$L(y) = \frac{1}{\mu_1'} \int_0^y Y_Q dQ \tag{10}$$

setting $x = Y_Q$, $Q = F(x)$ and $dQ = f(x)dx$ we have that

$$\begin{aligned} L(y) &= \frac{1}{\mu_1'} \int_0^y x \frac{\lambda\beta}{1-\exp(-\lambda)} [1-x]^{\beta-1} \exp\left(-\lambda[1-x]^{\beta}\right) dx \\ &= \frac{\lambda\beta}{\mu_1'[1-\exp(-\lambda)]} \int_0^y x[1-x]^{\beta-1} \exp(-\lambda[1-x]^{\beta}) dx \\ &= \frac{\lambda\beta y^2}{2\mu_1'[1-\exp(-\lambda)]} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} {}_2F_1(2, 1-\beta-\beta k; 3; y). \end{aligned} \tag{11}$$

Equation (11) was obtained after some algebra and using the equivalence relation $B(x,a,b) = \frac{x^a}{a} {}_2F_1(a, 1-b; a+1; x)$, where $B(x,a,b)$ denotes the incomplete beta function defined by $\int_0^x t^{a-1}(1-t)^{b-1} dt$ for $a > 0$, $b > 0$ and $0 < x < 1$ and ${}_2F_1(a, 1-b; a+1; x)$ denotes the Gauss' hypergeometric function defined as

$$\sum_{\ell=0}^{\infty} \frac{(a)_\ell (b)_\ell}{(c)_\ell \ell!} x^\ell = 1 + \frac{ab}{c}x + \frac{a(a+1)b(b+1)}{c(c+1)2!}x^2 + \cdots = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{\ell=0}^{\infty} \frac{\Gamma(a+\ell)\Gamma(b+\ell)}{\Gamma(c+\ell)\ell!} x^\ell.$$

The Gauss' hypergeometric function converges if c is nonnegative integer $\forall\, |x| < 1$ and on the unit circle if $\mathscr{R}[c-a-b] > 0$ where, $(a)_n$ denotes the Pochhammer symbol.

3.5.2. Bonferroni curve

Following Equation (11), we derive the Bonferroni curve as follows.

$$B(y) = \begin{cases} \frac{L(y)}{y}, & \text{if } 0 < y \le 1 \\ 0, & \text{if } y = 0, \end{cases}$$

thus,

$$B(y) = \frac{\lambda\beta y}{2\mu_1'[1-\exp(-\lambda)]} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} {}_2F_1(2, 1-\beta-\beta k; 3; y).$$

The plots of the Lorenz and Bonferroni curves represent the percentage of the population on the horizontal axis and the percentage of their total income or wealth on the vertical axis.

3.5.3. Gini index

If the random variable Y follows the UUTW distribution, then using Equation (11) we can calculate the Gini index $G \in (0,1)$ as follows.

$$G = 1 - 2\int_0^1 L(y) dy. \tag{12}$$

The integral in the Gini inequality measure of the UUTW distribution (Equation (12)) can be easily evaluated numerically.

4. Maximum likelihood estimation for UUTW distribution

In this section, we describe the method of maximum likelihood estimation aka the MLE for the estimation of the two parameters (λ and β) of the UUTW distribution.

Let Y be a random variable with the UUTW distribution and $\mathbf{\Theta} = (\lambda, \beta)'$ be the parameter vector. The log-likelihood function $\mathscr{L}(\mathbf{\Theta}|\mathbf{y})$ based on a random sample of size n with the corresponding observations $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ is given by

$$\mathscr{L}(\mathbf{\Theta}|\mathbf{y}) = n\log(\lambda\beta) - n\log(1-\exp[-\lambda]) + (\beta-1)\sum_{i=1}^{n}\log(1-y_i) - \lambda\sum_{i=1}^{n}(1-y_i)^{\beta}. \tag{13}$$

The partial derivative of $\mathscr{L}(\mathbf{\Theta}|\mathbf{y})$ given in Equation (13) with respect to λ is

$$\frac{\partial\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda} = \frac{n}{\lambda} - \frac{n}{\exp(\lambda)-1} - \sum_{i=1}^{n}(1-y_i)^{\beta}, \tag{14}$$

and the partial derivative of $\mathscr{L}(\mathbf{\Theta}|\mathbf{y})$ with respect to β is

$$\frac{\partial\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\log(1-y_i) - \lambda\sum_{i=1}^{n}(1-y_i)^{\beta}\log(1-y_i). \tag{15}$$

Notice that Equation (14) and Equation (15) are nonlinear in the parameters λ and β, respectively. Therefore there is no analytical solutions for the simultaneous equations arising from setting the two equations to zero; however, we can conveniently obtain the MLEs for λ and β, denoted by $\hat{\lambda}$ and $\hat{\beta}$ by some numerical approach, such as the Newton-type algorithm by minimizing $\mathscr{L}(\mathbf{\Theta}|\mathbf{y})$. Several packages in some statistical and mathematical software like, the nlm and the optim functions in the R software [31] are available for implementing the Newton-type algorithm. For the standard error and the asymptotic confidence interval for λ and β we need to calculate the expected observed Fisher information matrix (Hessian matrix). First, we suppose that the MLEs are asymptotically Gaussian distributed based on the conventional large sample approximation.

To find the Hessian matrix, we require the second order derivatives of the log-likelihood equation in Equation (13) with respect to the parameters. All the second derivatives exist and they are given by

$$\frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda\partial\beta} = \frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\beta\partial\lambda} = -\sum_{i=1}^{n}(1-y_i)^{\beta}\log(1-y_i), \tag{16}$$

$$\frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda^2} = -\frac{n}{\lambda^2} + \frac{n\exp(\lambda)}{[\exp(\lambda)-1]^2}, \tag{17}$$

and

$$\frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\beta^2} = -\frac{n}{\beta^2} - \lambda\sum_{i=1}^{n}(1-y_i)^{\beta}[\log(1-y_i)]^2. \tag{18}$$

In Equation (17), $\frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda^2} < 0$ and in Equation (18), $\frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\beta^2} < 0$ which indicates that the profile log-likelihood for λ and β are concave and that $\hat{\lambda}$ and $\hat{\beta}$ corresponds to the local maximum. However, since there is only one local maximum for each parameter, we conclude that in each case, the local maximum is exactly the global maximum and by implication the MLE.

The Hessian matrix for the UUTW distribution is given by

$$\hat{I}(\mathbf{\Theta}) = -\mathbb{E}\begin{pmatrix} \hat{I}_{\lambda\lambda} & \hat{I}_{\lambda\beta} \\ \hat{I}_{\lambda\beta} & \hat{I}_{\beta\beta} \end{pmatrix}$$

with the corresponding asymptotic variance-covariance matrix of the MLEs given by

$$\begin{pmatrix} \hat{I}^{\star}_{\lambda\lambda} & \hat{I}^{\star}_{\lambda\beta} \\ \hat{I}^{\star}_{\lambda\beta} & \hat{I}^{\star}_{\beta\beta} \end{pmatrix} = -\mathbb{E}\begin{pmatrix} \hat{I}_{\lambda\lambda} & \hat{I}_{\lambda\beta} \\ \hat{I}_{\lambda\beta} & \hat{I}_{\beta\beta} \end{pmatrix}^{-1}.$$

Therefore, the asymptotic distribution of the estimators is given by

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\beta} \end{pmatrix} \sim N_2\left(\begin{bmatrix} \bar{\hat{\lambda}} \\ \bar{\hat{\beta}} \end{bmatrix}, \begin{bmatrix} \hat{I}^{\star}_{\lambda\lambda} & \hat{I}^{\star}_{\lambda\beta} \\ \hat{I}^{\star}_{\lambda\beta} & \hat{I}^{\star}_{\beta\beta} \end{bmatrix}\right).$$

Note that the entries of the Hessian matrix are given by $\hat{I}_{\lambda\beta} = \frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda\partial\beta}$, $\hat{I}_{\lambda\lambda} = \frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\lambda^2}$ and $\hat{I}_{\beta\beta} = \frac{\partial^2\mathscr{L}(\mathbf{\Theta}|\mathbf{y})}{\partial\beta^2}$ in Equations (16)-(18), respectively.

Considering that $\hat{\mathbf{\Theta}} \sim N_2(\bar{\hat{\mathbf{\Theta}}}, \hat{I}(\mathbf{\Theta})^{-1})$, where $\hat{I}(\mathbf{\Theta})$ is the estimated information matrix, the approximate $100(1-\gamma)\%$ confidence interval for λ and β are given by $\hat{\lambda} \pm Z_{\frac{\gamma}{2}}\sqrt{\hat{I}^{\star}_{\lambda\lambda}}$ and $\hat{\beta} \pm Z_{\frac{\gamma}{2}}\sqrt{\hat{I}^{\star}_{\beta\beta}}$, respectively, where $Z_{\frac{\gamma}{2}}$ is the γth percentile of the standard normal distribution, $\sqrt{\cdot}$ denote the standard error (SE) of the parameter. The SEs corresponds to the square roots of the diagonal elements of the variance-covariance matrix.

4.1. Simulation experiment for the MLE of the parameters of the UUTW distribution

In the previous section, we employed the MLE method to estimate the parameters of the UUTW distribution, and here we investigate the performance of this method through a Monte-Carlo simulation. The simulation involves different sample sizes (n), specifically: 25, 50, 75, 100, 150, 200, 250 and 300, with different parameter combinations as shown in Table 1, and 5000 replications in R software. Table 1 shows the values of the mean estimates, standard errors (SEs), biases, and mean square errors (MSEs) for λ and β. We consider a wide range of initial guesses for the parameters and we did not notice any convergence issues, so any set of initial values will give a similar result in all scenarios.

The Monte-Carlo algorithm is detailed as follows.

(i) For specific parameter values of $\boldsymbol{\Theta}$, simulate a random sample of size n from the UUTW distribution through the inverse transformation method.
(ii) Estimate the parameters of the UUTW distribution by the MLE method.
(iii) Perform 5000 replications of steps (i)-(ii).
(iv) For each of the two parameters calculate the mean, SE, bias and MSE of the 5000 parameter estimates in (iii). The mean, SE, bias and MSE for the parameters are expressed as: $\bar{\hat{\boldsymbol{\Theta}}} = \frac{1}{5000}\sum_{i=1}^{5000}\hat{\theta}_i$; $\mathrm{SE}_{\bar{\hat{\boldsymbol{\Theta}}}} = \sqrt{\frac{1}{5000}\sum_{i=1}^{5000}(\hat{\theta}_i - \bar{\hat{\theta}})^2}$; $\mathrm{Bias}_{\bar{\hat{\boldsymbol{\Theta}}}} = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\theta}_i - \theta)$ and $\mathrm{MSE}_{\bar{\hat{\boldsymbol{\Theta}}}} = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\theta}_i - \theta)^2$; respectively, where $\hat{\theta}_{i's}$ represents the MLEs of λ or β for the ith iteration under specific sample size n, $\bar{\hat{\theta}}$ corresponds to the mean of the parameter estimates, i.e., $\hat{\lambda}_{i's}$ and $\hat{\beta}_{i's}$ and θ denotes the actual values of the parameters.

Essentially, from Table 1, we note that the ML estimators for λ and β are efficient because the estimated values are equivalent to the actual values. Also, the ML estimators are consistent since their SEs, biases, and MSEs decrease with an increase in the sample size (n).

5. The UUTW regression model

The beta regression model is often used in generalized linear model (GLM) to analyze bivariate or multivariate data whose response variable is defined on the unit interval. Here, we introduce an alternative to the beta regression model.

The median of the UUTW distribution (Equation (8)) has a nice closed form analytical expression; so, we can re-specify it as ξ and make it a function of the explanatory variables by means of a link function; considerably, the logit-link function. The logit-link function come in handy when it comes to connecting the explanatory variables $\boldsymbol{x}^T$ say, and the response variable $y \in [0,1)$. Since $\xi \in (0,1)$, we can now define the logit-link function as

$$\xi_i = \frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\theta})}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\theta})};\ i = 1,2,3,\cdots,n,$$

where $\boldsymbol{x}_i^T = (1, x_{i1}, x_{i2}, \cdots, x_{ip})$ is the vector of p explanatory variables, $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{p-1})^T$ is the vector of p unknown regression coefficients and n is the sample size. The explanatory variables and the regression coefficients can take any value on the entire real line.

After a slight re-parametrization of the median and substituting into the pdf in Equation (2), we obtain the pdf of the UUTW regression model as

$$f(y,\xi;\lambda) = \frac{\lambda\beta^{\star}(\lambda,\xi)}{1-\exp(-\lambda)}(1-y)^{\beta^{\star}(\lambda,\xi)-1}\exp\left(-\lambda[1-y]^{\beta^{\star}(\lambda,\xi)}\right), \tag{19}$$

where $\beta^{\star}(\lambda,\xi) = \log\left(-\frac{1}{\lambda}\log\left[\frac{1}{2}(1+\exp[-\lambda])\right]\right)[\log(1-\xi)]^{-1}$.

The primary assumptions of the UUTW regression model (Equation (19)) are: the dependent variable Y must be a UUTW random variable, the p independent variables must be uncorrelated (i.e., absence of multicollinearity), no influential data (outlier) and the residuals are independent and identically distributed.

We recommend the MLE method for estimating the parameters of the UUTW regression model (λ and $\boldsymbol{\theta}$). The MLE for UUTW regression model is straightforward and its mathematical detail is analogous to that in Section 3 therefore, we omit it to avoid unnecessary repetition.

Next, we discuss two prominent diagnostic tools for the GLMs.

5.1. Cox-Snell residual

In general, one way to validate the adequacy of any fitted regression model is to check whether the residuals of the model are well behaved. In reliability studies, the Cox-Snell residual by Cox and Snell [32] is a kind of standardized residuals that is widely used for assessing the goodness-of-fit of the fitted regression model and it is given by

$$\hat{e}_i = -\ln[1 - F(y_i|\hat{\boldsymbol{\theta}}, \hat{\lambda})];\ i = 1,2,\cdots,n,$$

where $F(\cdot)$ is the theoretical cdf and in this case, the cdf of the UUTW regression model whose pdf is given in Equation (19). If the fitted distribution is valid for the data, the residuals will approximately follow the exponential distribution, with $\kappa = 1$, i.e., $\hat{\epsilon} \sim \exp(1)$.

Table 1
Simulation result for the unit upper truncated Weibull distribution.

Actual			Estimates							
β	λ	n	$\bar{\hat{\beta}}$	$\bar{\hat{\lambda}}$	$\mathrm{SE}_{\bar{\hat{\beta}}}$	$\mathrm{SE}_{\bar{\hat{\lambda}}}$	$\mathrm{Bias}_{\bar{\hat{\beta}}}$	$\mathrm{Bias}_{\bar{\hat{\lambda}}}$	$\mathrm{MSE}_{\bar{\hat{\beta}}}$	$\mathrm{MSE}_{\bar{\hat{\lambda}}}$
0.80	1.70	25	0.8526	1.9926	0.2156	1.2775	0.0526	0.2926	0.0493	1.7176
		50	0.8255	1.8274	0.1491	0.8738	0.0255	0.1274	0.0229	0.7797
		75	0.8169	1.7903	0.1201	0.7072	0.0169	0.0903	0.0147	0.5082
		100	0.8125	1.7625	0.1041	0.6026	0.0125	0.0625	0.0110	0.3670
		150	0.8088	1.7476	0.0844	0.4887	0.0088	0.0476	0.0072	0.2411
		200	0.8048	1.7230	0.0733	0.4258	0.0048	0.0230	0.0054	0.1818
		250	0.8043	1.7244	0.0655	0.3805	0.0043	0.0244	0.0043	0.1454
		300	0.8038	1.7207	0.0594	0.3424	0.0038	0.0207	0.0035	0.1177
8.00	7.00	25	8.4300	8.4688	1.4472	4.0280	0.4300	1.4688	2.2791	18.3801
		50	8.2082	7.6065	0.9622	2.0116	0.2082	0.6065	0.9691	4.4141
		75	8.1424	7.4074	0.7931	1.5603	0.1424	0.4074	0.6493	2.6003
		100	8.1046	7.2895	0.6777	1.3037	0.1046	0.2895	0.4702	1.7834
		150	8.0680	7.1853	0.5401	1.0066	0.0680	0.1853	0.2963	1.0475
		200	8.0441	7.1322	0.4626	0.8598	0.0441	0.1322	0.2160	0.7567
		250	8.0376	7.1038	0.4120	0.7557	0.0376	0.1038	0.1712	0.5818
		300	8.0313	7.0806	0.3781	0.6902	0.0313	0.0806	0.1439	0.4828
3.00	1.00	25	3.3142	1.3567	0.9014	1.1186	0.3142	0.3567	0.9111	1.3783
		50	3.1426	1.1639	0.6174	0.8022	0.1426	0.1639	0.4015	0.6704
		75	3.0883	1.0948	0.5089	0.6709	0.0883	0.0948	0.2668	0.4590
		100	3.0668	1.0731	0.4468	0.5955	0.0668	0.0731	0.2041	0.3600
		150	3.0352	1.0316	0.3667	0.4914	0.0352	0.0316	0.1357	0.2424
		200	3.0277	1.0216	0.3183	0.4246	0.0277	0.0216	0.1021	0.1808
		250	3.0259	1.0271	0.2854	0.3842	0.0259	0.0271	0.0821	0.1483
		300	3.0161	1.0142	0.2612	0.3546	0.0161	0.0142	0.0685	0.1259
5.00	0.45	25	5.7765	0.9374	1.6137	0.9931	0.7765	0.4874	3.2067	1.2236
		50	5.4172	0.7161	1.0873	0.6892	0.4172	0.2661	1.3562	0.5457
		75	5.2764	0.6206	0.8898	0.5783	0.2764	0.1706	0.8680	0.3635
		100	5.2049	0.5845	0.7745	0.5173	0.2049	0.1345	0.6418	0.2856
		150	5.1409	0.5372	0.6395	0.4363	0.1409	0.0872	0.4288	0.1980
		200	5.0973	0.5097	0.5645	0.3919	0.0973	0.0597	0.3281	0.1571
		250	5.0756	0.4960	0.5172	0.3610	0.0756	0.0460	0.2731	0.1324
		300	5.0569	0.4874	0.4695	0.3320	0.0569	0.0374	0.2236	0.1116
1.00	0.60	25	1.1365	1.0395	0.3100	1.0016	0.1365	0.4395	0.1147	1.1963
		50	1.0689	0.8239	0.2156	0.7306	0.0689	0.2239	0.0512	0.5838
		75	1.0477	0.7548	0.1772	0.6218	0.0477	0.1548	0.0337	0.4106
		100	1.0334	0.7058	0.1572	0.5511	0.0334	0.1058	0.0258	0.3148
		150	1.0223	0.6751	0.1286	0.4653	0.0223	0.0751	0.0170	0.2221
		200	1.0140	0.6402	0.1146	0.4170	0.0140	0.0402	0.0133	0.1755
		250	1.0108	0.6279	0.1019	0.3746	0.0108	0.0279	0.0105	0.1411
		300	1.0090	0.6224	0.0947	0.3504	0.0090	0.0224	0.0091	0.1233

We can verify whether the Cox-Snell residuals follow approximately the exponential distribution with unit parameter by modeling $\hat{\epsilon}$ with the exponential distribution with pdf specified as $f(\hat{\epsilon}) = \kappa \exp(-\kappa\hat{\epsilon})$, where $\hat{\epsilon} > 0$ and $\kappa > 0$. Once κ is estimated, we can plot the cdf of the residuals based on the estimated exponential distribution, i.e., $\exp(\hat{\kappa})$ versus the cdf of the set of n randomly generated observations from the $\exp(1)$. We expect to see a straight line running through the origin if the fitted model is valid as shown in the Cox-Snell residual plots of the fitted regression models for the household food expenditure data in Fig. 7.

5.2. Randomized quantile residual

One can alternatively use the randomized quantile residuals [33] to access the adequacy of the fitted regression model. The randomized quantile residuals are given by

$$\hat{\eta}_i = \Phi^{-1}[\hat{F}(y_i|\hat{\theta}, \hat{\lambda})];\ i = 1, 2, \cdots, n,$$

where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. If the fitted distribution is a valid model for the data, the residuals will follow the standard Gaussian distribution, i.e., $\hat{\eta} \sim N(0, 1)$ as shown in the quantile residual plots of the fitted regression models for the household food expenditure data in Fig. 7 and by the p-values of the following normality tests: Jarque Bera test, Pearson chi-square test, Shapiro-Francia test, Anderson-Darling test, Cramer-von Mises test and the Lilliefors (Kolmogorov-Smirnov) test for the quantile residuals of the fitted regression models for the household food expenditure data in Tables 6 and 7.

6. 0 and 1 inflated unit upper truncated Weibull distribution

As we know, the UUTW distribution is best suited for modeling random variables that are defined on [0,1) but without many $0'$s however, in practice, we may encounter some random variables that generates data that include few 0's and 1's or even a preponderance of 0's and 1's and in such case, the support of the random variable is [0,1] and we need a suitable model to capture the inflation of $0'$s and $1'$s.

There are several instances in the literature where the support of the pdf of some notable continuous distributions excludes the endpoint(s) even when the distribution clearly have densit(y/ies) at such point(s), few examples includes: the exponential distribution with pdf defined as $f(y) = \lambda \exp(-\lambda y),\ y, \lambda > 0$ in both [34] and [35] omits 0 at the support of y even when $f(0) = \lambda$; the Marshall–Olkin generalized exponential distribution with pdf defined as $f(x) = p\theta \exp(-\theta y)[1-(1-p)\exp(-\theta y)]^2,\ y, \theta > 0$ and $p \in (0,1]$ in [36] omits 0 at the support of y even when $f(0) = \theta/p$; the unit Lindley distribution with pdf defined as $f(x) = \theta^2/(1+\theta)(1-y)^{-3}\exp(-\theta y/[1-y]),\ \theta > 0,\ y \in (0,1)$ in [37] omits 0 at the support of y even when it is clear that $f(0) = \theta^2/(1+\theta)$; the unit-improved second-degree Lindley distribution with pdf defined as $f(x) = \lambda^3(1-y)^{-2}/(\lambda^2+2\lambda+2)(1+y/[1-y])^2\exp(-y\lambda/(1-y)),\ \lambda > 0,\ y \in (0,1)$ in [38] omits 0 at the support of y even when it is obvious that $f(0) = \lambda^3/(\lambda^2+2\lambda+2)$; and the unit-Gompertz distribution with pdf defined as $f(y) = \alpha\beta y^{-(\beta+1)}\exp[-\alpha(y^{-\beta}-1)],\ \alpha, \beta > 0$, and $y \in (0,1)$ in [9] omits 1 at the support of y even when $f(1) = \alpha\beta$. Also, there are two main variants of the support of the beta distribution. Some authors include the endpoints (i.e., $y \in [0,1]$) in the support of the beta distribution while others choose to exclude them (i.e., $y \in (0,1)$). For instance, refer to https://www.sciencedirect.com/topics/mathematics/beta-distribution. The inflated beta distribution by [22] for modeling 0 and 1 inflations was developed around the beta distribution with support on $(0,1)$.

In Subsection 3.1, we showed that the behavior of the pdf of the UUTW distribution when it is evaluated at $y = 0$; i.e., $f(0)$ is the same as when y approaches 0; i.e., $\lim_{y\to 0^+} f(y)$ therefore, in this section, without compromising the validity of the pdf, we capitalize on the convenience that comes with specifying the support of the pdf of the UUTW distribution as $y \in (0,1)$ instead of $y \in [0,1)$. Hence, the random variable Y could be said to follow the UUTW distribution if its cdf is defined as

$$F(y) = 1 - \frac{1-\exp\left(-\lambda[1-y]^\beta\right)}{1-\exp(-\lambda)},\ 0 < y < 1, \tag{20}$$

and pdf defined as

$$f(y) = \frac{\lambda\beta}{1-\exp(-\lambda)}(1-y)^{\beta-1}\exp\left(-\lambda[1-y]^\beta\right),\ 0 < y < 1. \tag{21}$$

From this point through to the end of this section, Equation (20) and Equation (21) would be referred to as the cdf and pdf, respectively of the UUTW distribution.

Suppose the random variable Y generates many 0's and 1's alongside other values between 0 and 1, we cannot model this sort of data by the proposed UUTW distribution, instead we develop a special model for it by extending the UUTW distribution in Equation (21). We name the latest model the 0-1 inflated UUTW (ZOIUUTW for short) distribution. The ZOIUUTW distribution is a mixture model whose cdf is defined by Equation (22)

$$F_{\text{ZOIUUTW}}(y;\phi,\psi,\beta,\lambda) = \phi\text{Ber}(y;\psi) + (1-\phi)F(y;\lambda,\beta), \tag{22}$$

and pdf defined by

$$f_{\text{ZOIUUTW}}(y;\phi,\psi,\beta,\lambda) = \begin{cases} \phi(1-\psi); & \text{if } y = 0 \\ (1-\phi)f(y;\lambda,\beta); & \text{if } y \in (0,1) \\ \phi\psi; & \text{if } y = 1 \end{cases} \tag{23}$$

where λ and β are as defined before, $y \in [0,1]$, $\psi \in [0,1]$ is the proportion of degenerate values that are equal to 1, $\phi \in (0,1)$ is the mixing parameter (i.e., the proportion of y that are either 0 or 1), $\text{Ber}(y;\psi)$ denote the cdf of a Bernoulli random variable with parameter ψ, $F(\cdot)$ and $f(\cdot)$ denote the cdf and pdf of the UUTW distribution in Equation (20) and Equation (21), respectively. Fig. 3(a)-(d) shows the plots of the pdf of the ZOIUUTW distribution in Equation (23) for different parameter combinations.

If Y follows the ZOIUUTW distribution, its kth order ordinary moment is given by

$$\mathbb{E}(y^k) = \phi\psi + (1-\phi)\mu'_k$$

where μ'_k is the kth ordinary moment of the UUTW distribution in Equation (9).

Worthy of note is the fact that the pdf of the ZOIUUTW distribution in Equation (23) can alternatively be expressed as

$$\begin{aligned} f_{\text{ZOIUUTW}}(y;\phi,\psi,\beta,\lambda) &= \left[\phi\psi^y(1-\psi)^{1-y}\right]^{\mathbb{1}_{\{0,1\}}(y)} \times [(1-\phi)f(y;\lambda,\beta)]^{1-\mathbb{1}_{\{0,1\}}(y)} \\ &= \left[\phi^{\mathbb{1}_{\{0,1\}}(y)}(1-\phi)^{1-\mathbb{1}_{\{0,1\}}(y)}\right] \times \left[\psi^y(1-\psi)^{1-y}\right]^{\mathbb{1}_{\{0,1\}}(y)} \\ &\quad \times f(y;\lambda,\beta)^{1-\mathbb{1}_{\{0,1\}}(y)}, \end{aligned} \tag{24}$$

where $\mathbb{1}_{\{0,1\}}(y)$ is an indicator function that takes the value of 1 if $y \in \{0,1\}$ and the value of 0 if $y \notin \{0,1\}$.

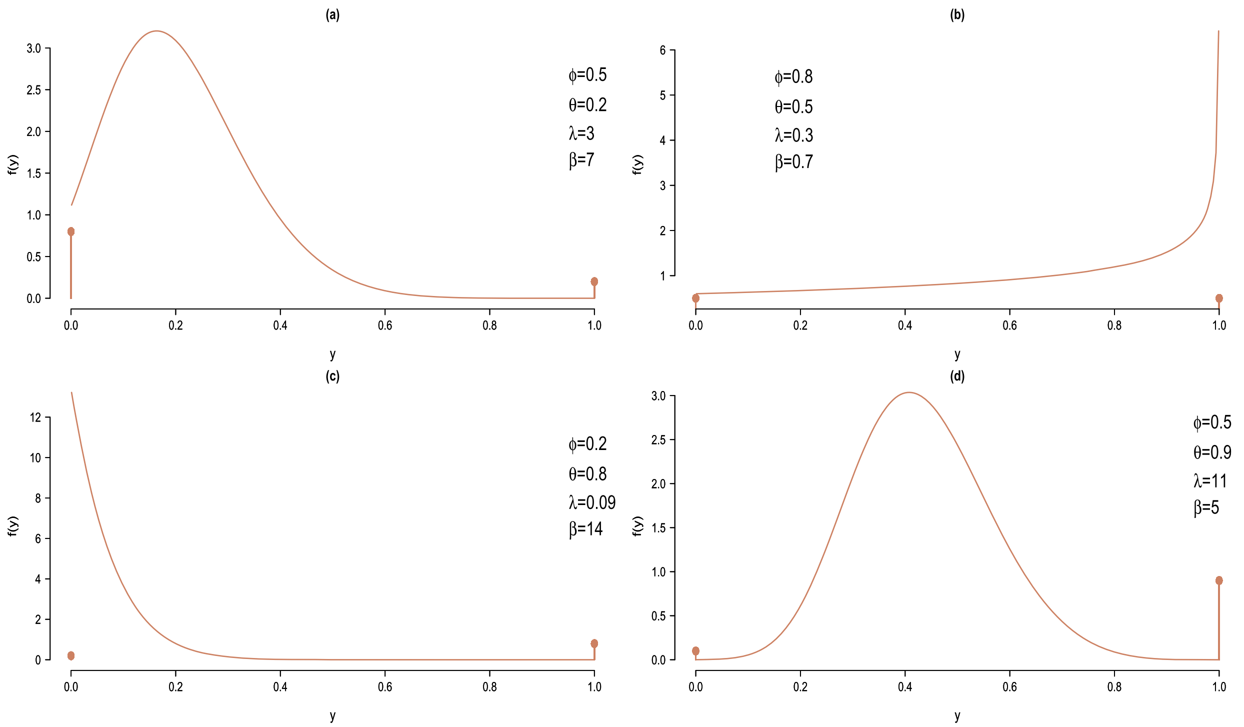

Fig. 3. Some plots of the pdf of the proposed ZOIUUTW distribution for selected values of ϕ, ψ, λ and β.

6.1. The MLE for ZOIUUTW distribution

Suppose we draw the random sample $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$ from the ZOIUUTW distribution, the likelihood function for $\boldsymbol{\Phi} = (\psi, \phi, \lambda, \beta)'$ based on Equation (24) is given by

$$L(\boldsymbol{\Phi}|\boldsymbol{y}) = \prod_{i=1}^{n} f_{\text{ZOIUUTW}}(y_i; \phi, \psi, \beta, \lambda) = L_1(\phi|\boldsymbol{y}) \times L_2(\psi|\boldsymbol{y}) \times L_3(\lambda, \beta|\boldsymbol{y}), \tag{25}$$

where

$$L_1(\phi|\boldsymbol{y}) = \prod_{i=1}^{n} \phi^{\mathbb{1}_{\{0,1\}}(y_i)}(1-\phi)^{1-\mathbb{1}_{\{0,1\}}(y_i)} = \phi^{\sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i)}(1-\phi)^{n-\sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i)},$$

$$\begin{aligned} L_2(\psi|\boldsymbol{y}) &= \prod_{i=1}^{n} \left[\psi^{y_i}(1-\psi)^{1-y_i}\right]^{\mathbb{1}_{\{0,1\}}(y_i)} = \psi^{\sum_{i=1}^{n} y_i \mathbb{1}_{\{0,1\}}(y_i)}(1-\psi)^{\sum_{i=1}^{n}(1-y_i)\mathbb{1}_{\{0,1\}}(y_i)} \\ &= \psi^{\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(y_i)}(1-\psi)^{[\sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i) - \sum_{i=1}^{n} \mathbb{1}_{\{1\}}(y_i)]} \text{ and} \end{aligned}$$

$$L_3(\lambda, \beta|\boldsymbol{y}) = \prod_{\substack{i=1 \\ y_i \in (0,1)}} f(y_i; \beta, \lambda) = \prod_{\substack{i=1 \\ y_i \in (0,1)}} \frac{\lambda\beta}{1-\exp(-\lambda)}(1-y_i)^{\beta-1}\exp(-\lambda[1-y_i]^{\beta}).$$

The log-likelihood function can be expressed as the natural logarithm of Equation (25) as in Equation (26):

$$\mathscr{L}(\boldsymbol{\Phi}|\boldsymbol{y}) = \ln \prod_{i=1}^{n} f_{\text{ZOIUUTW}}(y_i; \phi, \psi, \beta, \lambda) = \mathscr{L}_1(\phi|\boldsymbol{y}) + \mathscr{L}_2(\psi|\boldsymbol{y}) + \mathscr{L}_3(\lambda, \beta|\boldsymbol{y}), \tag{26}$$

where

$$\mathscr{L}_1(\phi|\boldsymbol{y}) = \ln(\phi)\sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i) + \ln(1-\phi)\left[n - \sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i)\right],$$

$$\mathscr{L}_2(\psi|\boldsymbol{y}) = \ln(\psi)\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(y_i) + \ln(1-\psi)\left[\sum_{i=1}^{n} \mathbb{1}_{\{0,1\}}(y_i) - \sum_{i=1}^{n} \mathbb{1}_{\{1\}}(y_i)\right] \text{ and}$$

$$\mathscr{L}_3(\lambda, \beta|\boldsymbol{y}) = \sum_{\substack{i=1 \\ y_i \in (0,1)}} \ln(\lambda\beta) - \sum_{\substack{i=1 \\ y_i \in (0,1)}} \ln[1-\exp(-\lambda)] + (\beta-1)\sum_{\substack{i=1 \\ y_i \in (0,1)}} \ln(1-y_i) - \lambda \sum_{\substack{i=1 \\ y_i \in (0,1)}} (1-y_i)^{\beta}.$$

We define the score function as $J(\Phi) = [J_\phi(\phi), J_\psi(\psi), J_\lambda(\lambda), J_\beta(\beta)]$, where

$$J_\phi(\phi) = \frac{\partial \mathscr{L}_1(\phi|\mathbf{y})}{\partial\phi} = \frac{1}{\phi}\sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i) - \frac{1}{1-\phi}\left[n - \sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)\right],$$

$$J_\psi(\psi) = \frac{\partial \mathscr{L}_1(\psi|\mathbf{y})}{\partial\psi} = \frac{1}{\psi}\sum_{i=1}^{n}\mathbb{1}_{\{1\}}(y_i) - \frac{1}{1-\psi}\left[\sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i) - \sum_{i=1}^{n}\mathbb{1}_{\{1\}}(y_i)\right],$$

$$J_\lambda(\lambda) = \frac{\partial \mathscr{L}_1(\lambda|\mathbf{y})}{\partial\lambda} = \frac{n - \sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)}{\lambda} - \frac{n - \sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)}{\exp(\lambda)-1} - \sum_{\substack{i=1\\ y_i\in(0,1)}}(1-y_i)^\beta \text{ and}$$

$$J_\beta(\beta) = \frac{\partial \mathscr{L}_1(\beta|\mathbf{y})}{\partial\beta} = \frac{n - \sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)}{\beta} + \sum_{\substack{i=1\\ y_i\in(0,1)}}\ln(1-y_i) - \lambda\sum_{\substack{i=1\\ y_i\in(0,1)}}(1-y_i)^\beta\ln(1-y_i).$$

Notice that $J_\phi(\phi)$ and $J_\psi(\psi)$ involve a single parameter each ϕ and ψ, respectively. Therefore, the maximum likelihood estimators of ϕ and ψ can be obtained by setting $J_\phi(\phi)$ to 0 and solving for ϕ and analogously setting $J_\psi(\psi)$ to 0 and solving for ψ hence, the two estimators are given by $\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)$ and $\hat{\psi} = \frac{\sum_{i=1}^{n}\mathbb{1}_{\{1\}}(y_i)}{\sum_{i=1}^{n}\mathbb{1}_{\{0,1\}}(y_i)}$, respectively. Each of the remaining two equations $J_\lambda(\lambda)$ and $J_\beta(\beta)$ involves two parameters; specifically, λ and β thus, to obtain the maximum likelihood estimators $\hat{\lambda}$ and $\hat{\beta}$, we set both equations to 0 and solve them simultaneously through the Newton-Raphson's method.

The Fisher information matrix for the parameters of the ZOIUUTW distribution can be specified as

$$\hat{K}(\Phi) = -\mathbb{E}\begin{pmatrix} \hat{K}_{\phi\phi} & 0 & 0 & 0\\ 0 & \hat{K}_{\psi\psi} & 0 & 0\\ 0 & 0 & \hat{K}_{\lambda\lambda} & \hat{K}_{\lambda\beta}\\ 0 & 0 & \hat{K}_{\lambda\beta} & \hat{K}_{\beta\beta}\end{pmatrix}$$

where $\hat{K}_{\phi\phi} = -\frac{n}{\phi(1-\phi)}$, $\hat{K}_{\psi\psi} = -\frac{n\phi}{\psi(1-\psi)}$, $\hat{K}_{\lambda\lambda} = -\frac{n(1-\phi)}{\lambda^2} + \frac{n(1-\phi)}{1-\exp(-\lambda)}$, $\hat{K}_{\beta\beta} = -\frac{n(1-\phi)}{\beta^2} - \lambda\sum_{\substack{i=1\\ y_i\in(0,1)}}(1-y_i)^\beta[\ln(1-y_i)]^2$ and $\hat{K}_{\lambda\beta} = -\sum_{\substack{i=1\\ y_i\in(0,1)}}(1-y_i)^\beta\ln(1-y_i)$.

Confidence intervals for the parameters $\hat{\mathbf{\Phi}}$ can be constructed as before. For large sample size, $\hat{\mathbf{\Phi}}$ is asymptotically multivariate normal distributed in particular, $\hat{\mathbf{\Phi}} \sim N_4(\bar{\hat{\mathbf{\Phi}}}, \hat{K}(\mathbf{\Phi})^{-1})$, where $\hat{K}(\mathbf{\Phi})$ is the Hessian matrix. The approximate $100(1-\gamma)\%$ confidence interval for ϕ, ψ, λ and β are given by $\hat{\phi} \pm Z_{\frac{\gamma}{2}}\sqrt{\mathrm{SE}_\phi}$, $\hat{\psi} \pm Z_{\frac{\gamma}{2}}\sqrt{\mathrm{SE}_\psi}$, $\hat{\lambda} \pm Z_{\frac{\gamma}{2}}\sqrt{\mathrm{SE}_\lambda}$ and $\hat{\beta} \pm Z_{\frac{\gamma}{2}}\sqrt{\mathrm{SE}_\beta}$ respectively, where $Z_{\frac{\gamma}{2}}$ is the γth percentile of the standard normal distribution.

We would like to conclude this section by pointing out that, with slight modification to Equation (24), that it can be amenable for cases of either only 0-inflation (i.e., $y \in [0,1)$) or only 1-inflation (i.e., $y \in (0,1]$) and the corresponding inferential treatments remain more or less the same as in this section. Therefore, we omit such analogous developments to avoid an unnecessarily lengthy article.

6.2. Simulation experiment for the MLE of the parameters of the ZOIUUTW distribution

In this section, we adopt analogous procedure to the simulation exercise in Section 4.1. The objective is to carryout a numerical experiment to assess the maximum likelihood estimators of the parameters (β, λ, ψ and ϕ) of the ZOIUUTW distribution. We use the same sample sizes as in Section 4.1 but, we fix different parameter values as shown in Table 9. The algorithm for generating random numbers from the ZOIUUTW distribution is detailed below.

Algorithm 1: An algorithm for generating data from ZOIUUTW distribution.

Initialize: n, ϕ, ψ, β, λ;
$n_1 \leftarrow n \times \phi$;
$n_2 \leftarrow n - n_1$;
$y \leftarrow \mathrm{rbern}(n_1, \psi)$; /* Sample n_1 from Bernoulli (ψ) */
$Q \leftarrow \mathrm{runif}(n_2)$; /* Sample n_2 from Uniform (0,1) */
$z \leftarrow 1 - [-1/\lambda \times \log(1-(1-Q)\times[1-\exp(-\lambda)])]^{(1/\beta)}$; /* UUTW draws */
Data$\leftarrow c(y, z)$; /* n realizations from ZOIUUTW */

We can see from Table 9, that the MLE estimators for β, λ, ψ and ϕ are efficient since the estimated values are equivalent to the actual values and the estimators are consistent because their corresponding SEs, biases, and MSEs decrease with an increase in the sample size.

7. Practical data examples

Here, we demonstrate the usefulness of both the UUTW and ZOIUUTW distributions. In Section 7.1, we give two real-life examples for the UUTW distribution and in Section 7.2, we give one real-life example for the ZOIUUTW distribution.

7.1. Examples based on the UUTW distribution

In this section, we give some practical examples of possible applications of the proposed model with two different data-sets and in each case, we compare the goodness-of-fit of the proposed model with those of the seven well-known models whose pdfs are given below. The first data-set represents the household food expenditure data for 38 households ([39], Table 15.4); the data has three variables, the first one is the response variable representing the proportion of household income that was spent only on food (i.e., food divided by income) Y and the remaining two are the explanatory variables corresponding to the household income (previously mentioned) X_1 and the number of individuals living in the household (household size) X_2. The descriptive plots of the household food expenditure data are shown in Fig. 4 and from there, it could be seen from the density plots that the pdf of the proportion of household income spent on food is not only unimodal but right skewed, while the values of the Pearson's correlation coefficients and the scatter plots, indicate statistically significant negative and positive correlations between the proportion of household income spent on food and the household income and household size, respectively; whereas, no statistically significant correlation exists between household income and household size. The data Y is right-skewed with skewness statistic of 0.9427343 and leptokurtic with kurtosis statistic of 3.859457; also, the $\min(Y) = 0.1075258$ and $\max(Y) = 0.5612430$ suggesting that Y can be modeled by the UUTW distribution. The second data is on the maximum flood levels of a certain river in Pennsylvania in millions of cubic feet per second (mlcf/s) [40]. Let Y denote the maximum flood levels data, Y is right-skewed with skewness statistic of 1.067324 and leptokurtic with kurtosis statistic of 3.598898; also the $\min(Y) = 0.265$ and $\max(Y) = 0.740$ thus, implying that Y can be modeled by the UUTW distribution.

To illustrate the flexibility of the UUTW distribution, we compare its fitting capability for modeling the household food expenditure data and the maximum flood level data with the fitting capabilities of the following competing unit distributions.

1. 1-parameter Topp-Leone distribution [4]

$$f(y) = 2\beta(1-y)[y(2-y)]^{\beta-1};\ \beta > 0.$$

2. unit Rayleigh distribution [11]

$$f(y) = -\frac{2\beta}{y}\log(y)\exp(-\beta[\log(y)]^2);\ \beta > 0.$$

3. 2-parameter Topp-Leone distribution [21]

$$f(y) = \beta[y^\lambda(2-y^\lambda)]^{\beta-1}[\lambda y^{\lambda-1}(2-y^\lambda) - \lambda y^{2\lambda-1}];\ \lambda > 0,\ \beta > 0.$$

4. log-WE distribution [14]

$$f(y) = \frac{\beta+1}{\beta}\lambda y^{\lambda-1}(1-y^{\beta\lambda});\ \lambda > 0,\ \beta > 0.$$

5. Kumaraswamy distribution [2]

$$f(y) = \lambda\beta y^{\lambda-1}(1-y^\lambda)^{\beta-1};\ \lambda > 0,\ \beta > 0.$$

6. beta distribution

$$f(y) = \frac{\Gamma(\lambda+\beta)}{\Gamma(\lambda)\Gamma(\beta)}y^{\lambda-1}(1-y)^{\beta-1};\ \lambda > 0,\ \beta > 0.$$

7. unit ZTPPF distribution [25]

$$f(y) = \frac{\lambda\beta y^{\beta-1}}{\exp(\lambda)-1};\ \lambda > 0,\ \beta > 0.$$

All the seven competing distributions take values between 0 and 1. To check whether or not each of the distributions fit any of the data well, we use the Kolmogorov–Smirnov test with test statistic defined by

$$\mathrm{KS} = \max_{y\in\mathrm{Data}}\left|F_n(y) - \widehat{F}(y)\right|.$$

Noting that large KS p-value (>0.05) implies that the fitted distribution provides reasonably good fit for the data. To discriminate among the fitted distributions, we use the following information criteria statistics.

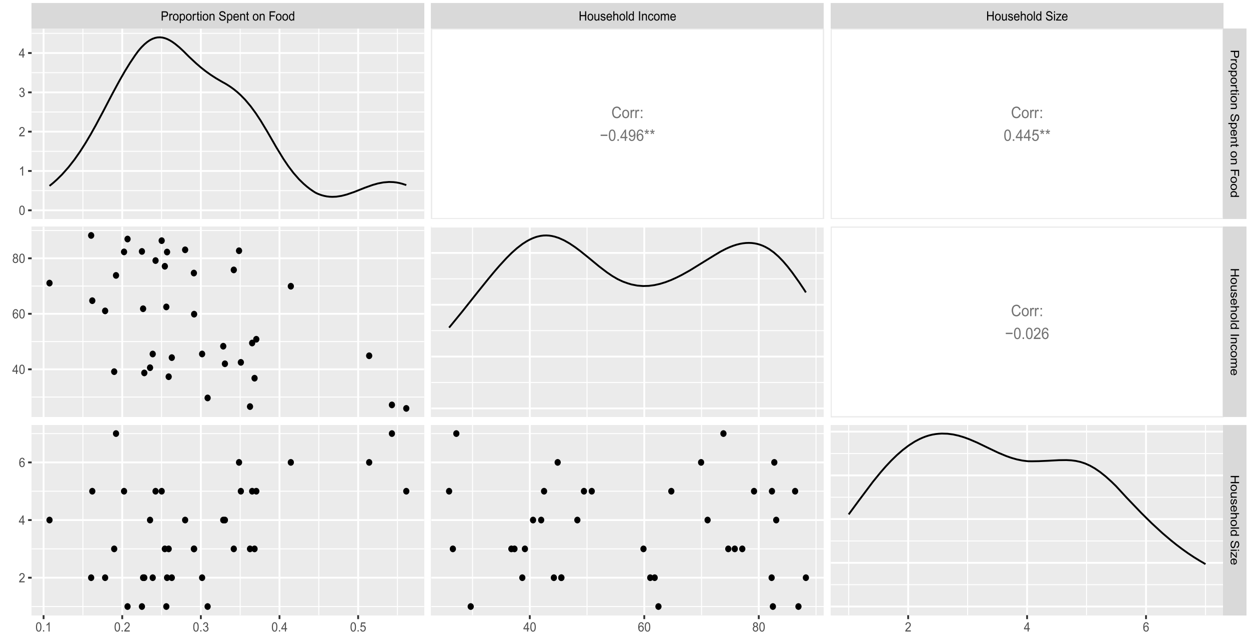

Fig. 4. Descriptive plots for the household food expenditure data.

Table 2
Fit results for different distributions for the data on the proportion of household income spent on food.

Distribution(s)	Parameter Estimate [SE] $\hat{\beta}$	$\hat{\lambda}$	K-S test Statistic	p-value
UUTW distribution	9.0435 [1.1675]	13.3416 [4.3577]	0.0964	0.8389
beta	6.0716 [1.3586]	14.8221 [3.3988]	0.1101	0.7056
Kumaraswamy	26.9552 [10.8244]	2.9541 [0.3693]	0.6945	2.2×10^{-16}
log-WE	0.1885 [0.5556]	1.4149 [0.3572]	0.3464	0.0001
1-para Topp-Leone	1.0000 [0.1622]	–	0.2910	0.0024
unit Rayleigh	0.5567 [0.0903]	–	0.3171	0.0007
2-para Topp-Leone	0.4163 [0.5607]	2.2363 [2.5575]	0.3774	2.2×10^{-5}
unit ZTPPF	3.0129 [0.3537]	29.8293 [11.2662]	0.1184	0.6193

1. Akaike information criterion (AIC) due to [41] defined by

$$\mathrm{AIC} = -2\widehat{\mathscr{L}} + 2k,$$

2. Bayes information criterion (BIC) due to [42] defined by

$$\mathrm{BIC} = -2\widehat{\mathscr{L}} + k\log(n),$$

3. Akaike information criterion with a correction (AICc) due to [43] defined by

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1},$$

where $\widehat{\mathscr{L}}$ and k denote the maximized log likelihood value and the number of parameters, respectively. Any distribution with the smallest AIC, BIC and AICc values is considered to be the better model.

We obtained the Hessian matrices for the fitted UUTW distribution for the two data-sets as

Household income	Maximum flood levels
$\left.\begin{pmatrix} 2.9734047 & -0.6913883 \\ -0.6913883 & 0.2134250 \end{pmatrix}\right\rvert_{\text{UUTW distribution}}$	$\left.\begin{pmatrix} 4.23034890 & -0.43842038 \\ -0.43842038 & 0.05497373 \end{pmatrix}\right\rvert_{\text{UUTW distribution}}$

Table 3
Model fit discrimination for the data on the proportion of household income spent on food.

		Information Criteria		
Distribution(s)	$-\hat{\mathscr{L}}$	AIC	BIC	AICc
UUTW distribution	-36.1400	-68.2801	-65.0049	-67.9372
beta	-35.3464	-66.6929	-63.4177	-66.3500
Kumaraswamy	-33.4891	-62.9782	-59.7030	-62.6353
log-WE	-14.5450	-25.0899	-21.8147	-24.7470
1-para Topp-Leone	-12.9134	-23.8268	-22.1892	-23.7156
unit Rayleigh	-23.7528	-45.5056	-43.8680	-45.3945
2-para Topp-Leone	-9.5086	-15.0172	-11.7421	-14.6744
unit-ZTPPF	-33.7835	-63.5670	-60.2918	-63.2242

Table 4
Fit results for different distributions for the maximum flood levels data.

	Parameter Estimate [SE]		K-S test	
Distribution(s)	$\hat{\beta}$	$\hat{\lambda}$	Statistic	p-value
UUTW distribution	6.2224 [1.1673]	19.0738 [10.2398]	0.1636	0.6582
beta	6.7587 [2.0951]	9.1142 [2.8526]	0.1988	0.4083
Kumaraswamy	11.7919 [5.3617]	3.3634 [0.6034]	0.4841	0.0002
log-WE	0.0063 [0.5770]	2.1997 [0.7142]	0.3108	0.0420
1-para Topp-Leone	1.0000 [0.2236]	–	0.4598	0.0004
unit Rayleigh	1.0000 [0.2236]	–	0.2419	0.1925
2-para Topp-Leone	1434.1760 [2219.3910]	2.8×10^{-2} [0.0223]	0.2810	0.0850
unit ZTPPF	3.5260 [0.5648]	14.4518 [5.9650]	0.1987	0.4084

Table 5
Model fit discrimination for the maximum flood levels data.

		Information Criteria		
Distribution(s)	$-\hat{\mathscr{L}}$	AIC	BIC	AICc
UUTW distribution	-15.0725	-26.1450	-21.9233	-25.9381
beta	-14.0622	-24.1245	-19.9027	-23.9176
Kumaraswamy	-12.8662	-21.7324	-17.5106	-21.5255
log-WE exp	-6.6692	-9.3383	-5.1166	-9.1314
1-para Topp-Leone	-2.2864	-2.5728	-0.4619	-2.5050
unit Rayleigh	-10.9249	-19.8497	-18.8540	-19.6275
2-para Topp-Leone	-10.9113	-17.8226	-13.6008	-17.6157
unit-ZTPPF	-13.2640	-22.5280	-20.5366	-21.8221

while the Hessian matrix for the UUTW regression model with the intercept for the income data was obtained as

$$\left.\begin{pmatrix} 2.261381 \times 10^{2} & 1.339075 \times 10^{4} & 7.795482 \times 10^{2} & 1.276036 \times 10^{-1} \\ 1.339075 \times 10^{4} & 8.819196 \times 10^{5} & 4.588221 \times 10^{4} & 8.78320609 \\ 7.795482 \times 10^{2} & 4.588221 \times 10^{4} & 3.317699 \times 10^{3} & 8.656662 \times 10^{-1} \\ 1.276036 \times 10^{-1} & 8.78320609 & 8.656662 \times 10^{-1} & 0.02133125 \end{pmatrix}\right|_{\text{Regression with intercept}},$$

and the Hessian matrix for the UUTW regression model without the intercept for the income data was obtained as

$$\left.\begin{pmatrix} 8.033747 \times 10^{5} & 4.196223 \times 10^{4} & 7.81888694 \\ 4.196223 \times 10^{4} & 3.044179 \times 10^{3} & 0.71319795 \\ 7.81888694 & 0.71319795 & 0.03334324 \end{pmatrix}\right|_{\text{Regression without intercept}}.$$

The inverse of the Hessian matrices results to the corresponding variance-covariance matrices. The square root of the diagonal elements of each of the variance-covariance matrix gives the standard errors of the parameters in Tables 2, 4, 6 and 7.

Table 6
Fitted regression model with the intercept and the residual analysis for the household expenditure on food data.

	$-\mathscr{L}$	θ_0	θ_1	θ_2	λ
Estimates	-44.6518	-0.6749	-0.0110	0.1039	20.1741
SE	–	0.2552	0.0034	0.0401	6.9090
$\lvert z\rvert$[a]	–	2.6446	3.2353	2.5910	2.9200
Pr(> $\lvert z\rvert$)	–	0.0082	0.0012	0.0096	0.0035
Normality test for the estimated regression residuals $\hat{\eta}_i$					
Test	statistic	df	p-value		
Jarque Bera	5.0373	2	0.5214		
Pearson chi-square	4.1579	–	0.6553		
Shapiro-Francia	0.9415	–	0.0467		
Anderson-Darling	0.7524	–	0.0459		
Cramer-von Mises	0.1341	–	0.0367		
Lilliefors (Kolmogorov-Smirnov)	0.1455	–	0.0412		

[a] z is equal to the model coefficient divided by the corresponding standard error (SE); it is called the z-score and it follows the standard normal distribution.

Table 7
Fitted regression model without the intercept and the residual analysis for the household expenditure on food data.

	$-\mathscr{L}$	θ_1	θ_2	λ
Estimates	-41.3455	-0.0182	0.0466	16.8988
SE	–	0.0021	0.0343	5.4917
$\lvert z\rvert$[a]	–	8.6667	1.3586	3.0772
Pr(> $\lvert z\rvert$)	–	0.0000	0.1743	0.0021
Normality test for the estimated regression residuals $\hat{\eta}_i$				
Test	statistic	df	p-value	
Jarque Bera	1.2645	2	0.5314	
Pearson chi-square	1.7895	–	0.9380	
Shapiro-Francia	0.9830	–	0.7304	
Anderson-Darling	0.2156	–	0.8349	
Cramer-von Mises	0.0302	–	0.8388	
Lilliefors (Kolmogorov-Smirnov)	0.0638	–	0.9602	

[a] z is equal to the model coefficient divided by the corresponding standard error (SE); it is called the z-score and it follows the standard normal distribution.

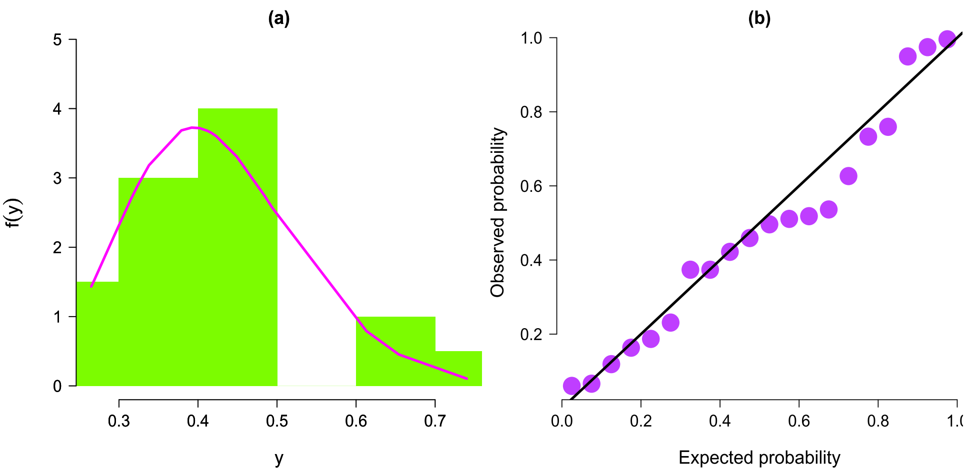

Fig. 5. (a) Plot of the estimated pdf of the UUTW distribution superimposed on the empirical density of the data on the maximum flood level. (b) P-P plot of the fitted UUTW distribution for the data on the maximum flood level.

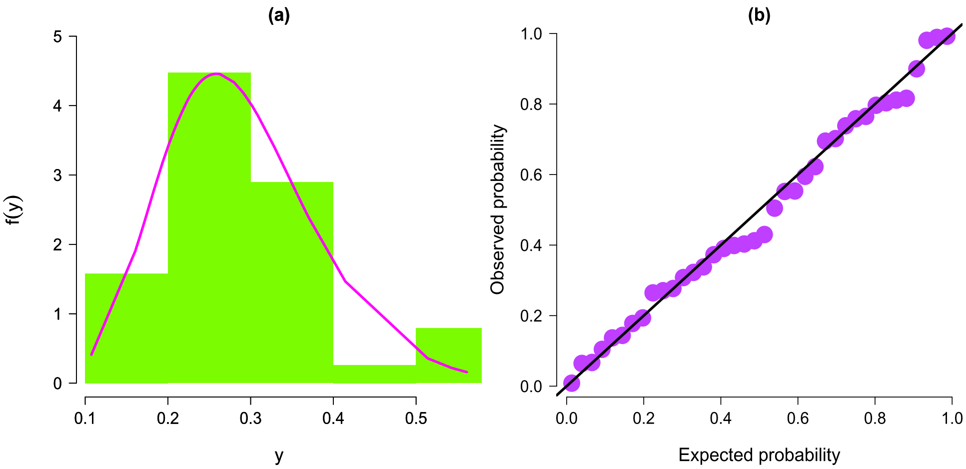

Fig. 6. (a) Plot of the estimated pdf of the UUTW distribution superimposed on the empirical density of the data on the proportion of income spent on food. (b) P-P plot of the fitted UUTW distribution for the data on the proportion of income spent on food.

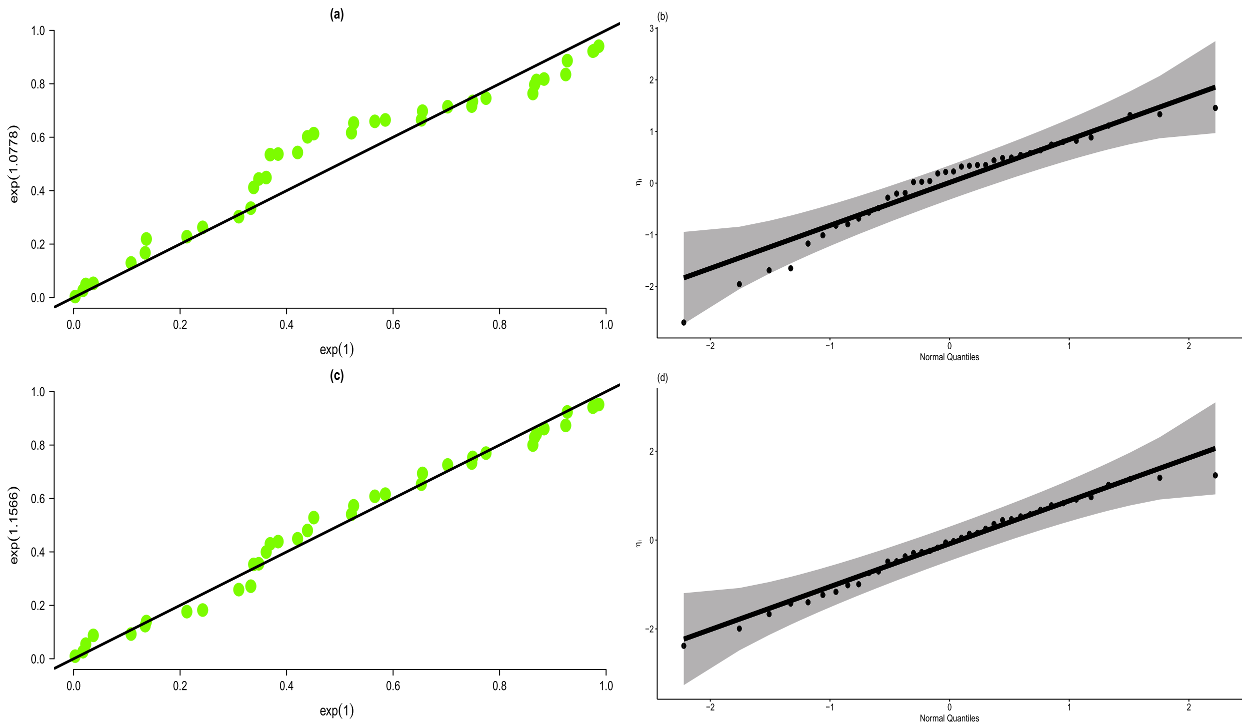

Fig. 7. Cox-Snell residual plots: (a) for the regression model with the intercept and (c) for the regression model without the intercept. The randomized quantile residual plots: (b) for the regression model with the intercept and (d) for the regression model without the intercept.

7.2. Example based on the ZOIUUTW distribution

In this section, we use the CD34+ data which involve 239 patients at the Edmonton Hematopoietic Stem Cell Lab in Cross Cancer Institute - Alberta Health Services from 2003 to 2008. The data can be obtained from the R software package "simplexreg" [44]. The data consists of five variables in columns but, we only use the variable in the fourth column corresponding to the adjusted age "ageadj" of the patients who underwent chemotherapy in the cancer institute. In order to obtain a set of data on $[0,1]$, we divided the adjusted age data by the maximum adjusted age value (i.e., 31) and henceforth, we shall simply refer to this scaled data as "the age data". Out of the 239 observations in the age data, 48 of them are either 0 or 1. The frequencies of 0's and 1's in the age data are 46 and 2, respectively. Therefore, the data is considered to be 0 and 1 inflated hence, can be suitably modeled by the ZOIUUTW distribution. The empirical density plot of the data is depicted in Fig. 5(a) and the data is left-skewed with skewness statistic of -0.1600658 and leptokurtic with kurtosis statistic of 1.715061. We compare the fit of the ZOIUUTW distribution to the fits of the zero and one inflated beta (ZOIbeta) distribution [22] and the zero and one inflated Kumaraswamy (ZOIKum) distribution [23].

Table 8
Fit results for different 0 & 1 inflated distributions for the 0 and 1 inflated age data.

	Parameter Estimate [SE]				Log-likelihood & Information Criteria			
Distribution(s)	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\psi}$	$\hat{\phi}$	$-\mathscr{L}$	AIC	BIC	AICc
ZOIUUTW	1.7372 [0.1340]	2.6693 [0.4425]	0.0417 [2.8×10^{-2}]	0.2008 [2.5×10^{-2}]	111.4232	230.8465	244.7523	231.0174
ZOIbeta	1.7824 [1.7×10^{-1}]	1.5348 [1.4×10^{-1}]	0.0417 [2.8×10^{-2}]	0.2008 [2.5×10^{-2}]	113.2475	234.4951	248.4009	234.6660
ZOIKum	1.7280 [0.1484]	1.5908 [0.1669]	0.0416 [2.8×10^{-2}]	0.2009 [2.5×10^{-2}]	112.8925	233.7850	247.6908	233.9559

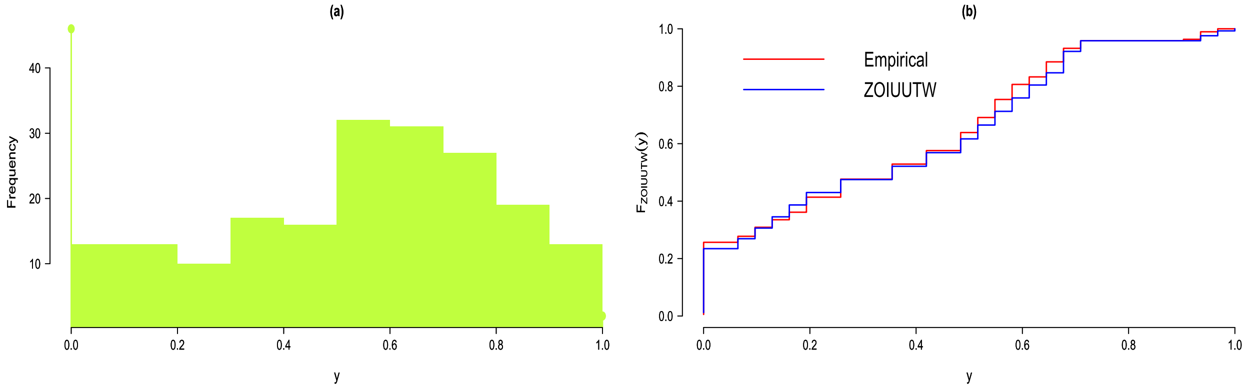

Fig. 8. (a) Empirical density plot of the age data. (b) Estimated cdf plot of the ZOIUUTW distribution superimposed on the empirical cdf plot of the age data.

The pdf of the ZOIbeta distribution is given by Equation (27)

$$f_{\text{ZOIbeta}}(y;\phi,\psi,\beta,\lambda)=\begin{cases}\phi(1-\psi); & \text{if } y=0\\ (1-\phi)f(y;\lambda,\beta); & \text{if } y\in(0,1)\\ \phi\psi; & \text{if } y=1\end{cases} \tag{27}$$

where $f(y;\lambda,\beta)$ denote the pdf of the beta distribution which is defined as $\dfrac{y^{\lambda}(1-y)^{\beta-1}}{B(\lambda,\beta)}$; $\lambda,\beta>0$ and $y\in(0,1)$ where $B(\cdot,\cdot)$ denote the beta function defined by $\dfrac{\Gamma(\lambda)\Gamma(\beta)}{\Gamma(\lambda+\beta)}$ and $\Gamma(\cdot)$ denote the gamma function defined by $\Gamma(v)=\int_0^{\infty} x^{v-1}\exp(-x)dx$.

The pdf of the ZOIKum distribution can be specified as in Equation (28)

$$f_{\text{ZOIKum}}(y;\phi,\psi,\beta,\lambda)=\begin{cases}\phi(1-\psi); & \text{if } y=0\\ (1-\phi)f(y;\lambda,\beta); & \text{if } y\in(0,1)\\ \phi\psi; & \text{if } y=1\end{cases} \tag{28}$$

where $f(y;\lambda,\beta)$ denote the pdf of the Kumaraswamy distribution which is defined as $\lambda\beta y^{\lambda-1}(1-y^{\lambda})^{\beta-1}$; $\lambda,\beta>0$ and $y\in(0,1)$.

We chose to compare the fitting capability of the ZOIUUTW distribution with only the fitting capabilities of the ZOIbeta distribution and ZOIKum distribution because of the long history, popularity and extensive theoretical and application developments of their respective baseline distributions, i.e., the beta distribution and the Kumaraswamy distribution.

The Hessian matrix for the fitted ZOIUUTW distribution is given below and computing the inverse of the matrix would result to the variance-covariance matrix; where, the square root of the diagonal elements of the corresponding variance-covariance matrix yields the standard errors of the parameters in Table 8.

$$\left.\begin{pmatrix} 1.256420\times10^{2} & -2.839572\times10^{1} & 3.552714\times10^{-9} & 0.000000\times10^{0}\\ -2.839572\times10^{1} & 11.52549 & 0.000000\times10^{0} & 0.000000\times10^{0}\\ 3.552714\times10^{-9} & 0.00000\times10^{0} & 1.201843\times10^{3} & -3.552714\times10^{-9}\\ 0.000000\times10^{0} & 0.00000\times10^{0} & -3.552714\times10^{-9} & 1.489207\times10^{3}\end{pmatrix}\right|_{\text{ZOIUUTW distribution}}.$$

Table 9
Simulation result for the 0 & 1 inflated unit upper truncated Weibull distribution.

n	$\bar{\bar{\beta}}$	$\bar{\bar{\lambda}}$	$\bar{\bar{\psi}}$	$\bar{\bar{\phi}}$	$SE_{\bar{\bar{\beta}}}$	$SE_{\bar{\bar{\lambda}}}$	$SE_{\bar{\bar{\psi}}}$	$SE_{\bar{\bar{\phi}}}$	$Bias_{\bar{\bar{\beta}}}$	$Bias_{\bar{\bar{\lambda}}}$	$Bias_{\bar{\bar{\psi}}}$	$Bias_{\bar{\bar{\phi}}}$	$MSE_{\bar{\bar{\beta}}}$	$MSE_{\bar{\bar{\lambda}}}$	$MSE_{\bar{\bar{\psi}}}$	$MSE_{\bar{\bar{\phi}}}$
	Actual values: $\beta = 3.0000$, $\lambda = 1.0000$, $\psi = 0.5000$ and $\phi = 0.6000$															
75	3.2609	1.2995	0.5001	0.6000	0.8085	1.0305	0.0737	0.0012	0.2609	0.2995	1×10^{-4}	0.0000	0.7217	1.1516	0.0054	0.0000
100	3.1776	1.1964	0.5002	0.6000	0.6847	0.8835	0.0646	0.0011	0.1776	0.1964	2×10^{-4}	0.0000	0.5003	0.8191	0.0042	0.0000
150	3.1165	1.1199	0.4997	0.6000	0.5674	0.7380	0.0531	0.0007	0.1165	0.1199	-3×10^{-4}	0.0000	0.3355	0.5590	0.0028	0.0000
200	3.0886	1.0951	0.4999	0.6000	0.4942	0.6511	0.0455	0.0006	0.0886	0.0951	-1×10^{-4}	0.0000	0.2521	0.4329	0.0021	0.0000
250	3.0642	1.0672	0.4996	0.6000	0.4497	0.5896	0.0412	0.0005	0.0642	0.0672	-4×10^{-4}	0.0000	0.2064	0.3521	0.0017	0.0000
300	3.0488	1.0500	0.5001	0.6000	0.4070	0.5403	0.0371	0.0004	0.0488	0.0500	1×10^{-4}	0.0000	0.1680	0.2944	0.0014	0.0000
	Actual values: $\beta = 5.0000$, $\lambda = 4.0000$, $\psi = 0.7000$ and $\phi = 0.4000$															
75	5.1370	4.2856	0.6998	0.4000	0.7257	1.2263	0.0837	1×10^{-4}	0.1370	0.2856	-2×10^{-4}	0.0000	0.5453	1.5852	0.0070	0.0000
100	5.0932	4.1963	0.7007	0.4000	0.6272	1.0213	0.0723	0.0000	0.0932	0.1963	7×10^{-4}	0.0000	0.4021	1.0814	0.0052	0.0000
150	5.0609	4.1250	0.6999	0.4000	0.5027	0.8010	0.0591	0.0000	0.0609	0.1250	-1×10^{-4}	0.0000	0.2564	0.6572	0.0035	0.0000
200	5.0450	4.0855	0.7000	0.4000	0.4332	0.6806	0.0514	0.0000	0.0450	0.0855	0.0000	0.0000	0.1896	0.4705	0.0026	0.0000
250	5.0367	4.0793	0.7003	0.4000	0.3886	0.6088	0.0460	1×10^{-4}	0.0367	0.0793	3×10^{-4}	0.0000	0.1523	0.3769	0.0021	0.0000
300	5.0368	4.0726	0.7003	0.4000	0.3509	0.5541	0.0416	1×10^{-4}	0.0368	0.0726	3×10^{-4}	0.0000	0.1244	0.3123	0.0017	0.0000
	Actual values: $\beta = 0.5000$, $\lambda = 0.1000$, $\psi = 0.9000$ and $\phi = 0.2000$															
75	0.5458	0.4035	0.9012	0.1997	0.0981	0.5295	0.0772	0.0029	0.0458	0.3035	0.0012	-3×10^{-4}	0.0117	0.3725	0.0060	0.0000
100	0.5407	0.3714	0.8998	0.1998	0.0846	0.4683	0.0680	0.0021	0.0407	0.2714	−0.0002	-2×10^{-4}	0.0088	0.2930	0.0046	0.0000
150	0.5320	0.3156	0.9000	0.2000	0.0690	0.3905	0.0549	0.0012	0.0320	0.2156	0.0000	0.0000	0.0058	0.1989	0.0030	0.0000
200	0.5272	0.2904	0.9005	0.2000	0.0597	0.3456	0.0476	0.0009	0.0272	0.1904	0.0005	0.0000	0.0043	0.1556	0.0023	0.0000
250	0.5226	0.2575	0.9003	0.2000	0.0535	0.3124	0.0419	0.0007	0.0226	0.1575	0.0003	0.0000	0.0034	0.1224	0.0018	0.0000
300	0.5196	0.2404	0.9007	0.2000	0.0484	0.2850	0.0384	0.0007	0.0196	0.1404	0.0007	0.0000	0.0027	0.1009	0.0015	0.0000
	Actual values: $\beta = 1.0000$, $\lambda = 0.5000$, $\psi = 0.5000$ and $\phi = 0.5000$															
75	1.0985	0.8189	0.4982	0.5067	0.2525	0.8190	0.0812	0.0016	0.0985	0.3189	−0.0018	0.0067	0.0735	0.7725	0.0066	0.0000
100	1.0797	0.7481	0.4988	0.5000	0.2192	0.7145	0.0704	0.0014	0.0797	0.2481	−0.0012	0.0000	0.0544	0.5720	0.0050	0.0000
150	1.0522	0.6692	0.5004	0.5000	0.1777	0.5990	0.0573	0.0012	0.0522	0.1692	0.0004	0.0000	0.0343	0.3874	0.0033	0.0000
200	1.0384	0.6239	0.5003	0.5000	0.1541	0.5350	0.0504	0.0011	0.0384	0.1239	0.0003	0.0000	0.0252	0.3016	0.0025	0.0000
250	1.0306	0.6006	0.4999	0.5000	0.1398	0.4814	0.0452	0.0008	0.0306	0.1006	−0.0001	0.0000	0.0205	0.2419	0.0020	0.0000
300	1.0242	0.5695	0.5002	0.5000	0.1286	0.4467	0.0411	0.0008	0.0242	0.0695	0.0002	0.0000	0.0171	0.2043	0.0017	0.0000

8. Discussion of results

From Table 2, it could be seen that only the UUTW distribution, followed by the beta distribution and the unit ZTPPF distribution provides good fit for the data on the proportion of household income spent on food. From Table 4 we have that the UUTW distribution, beta distribution, unit ZTPPF distribution and the unit Rayleigh distribution provide decent fits for the maximum flood level data; while in Tables 2 and 4, we observe that some of the competing models (based on their p-values) do not appear to provide adequate fit for the data-sets, but we examine them for the purpose of comparison and to demonstrate the worth of the suggested model in contrast to them. Moreover, the results in Table 3 and Table 5 indicate that the UUTW distribution with the smallest AIC, BIC and AICc values fits the two data-sets better than the rest of the distributions and the density plots and P-P plots in Fig. 5 and Fig. 6 also indicate that the UUTW distribution captured the two data-sets really well.

After demonstrating the suitability of the UUTW distribution and its better fitting performance viz-a-viz the beta distribution and the rest of the other competing distributions with respect to modeling of the response variable - the proportion of household income spent on food, we carried-out a regression analysis to investigate the impact of the household income and household size on the proportion of household income spent on food. First, we fitted two regression models, one with an intercept in Table 6 and another one without the intercept in Table 7. It could be seen from Table 6 that all the regression coefficients are statistically significant at 0.05 level whereas, after dropping the intercept in Table 7 we found that the regression coefficient for the number of individuals in the household became non-statistically significant at 0.05 level but, this appears to be a clear departure from the obvious. The residual analysis for the two regression models in Table 6 and Table 7 as well as the Cox-Snell residual plots and the randomized quantile residual plots in Fig. 7(a)-(d) did not indicate any model fit inadequacy for the fitted regression models; however, to decide which one between the model in Table 6 and Table 7 describes the data better enough, we use the likelihood ratio test. The likelihood ratio test allows us to segregate among two hierarchically nested models. The model in Table 6 reduces to the one in Table 7 when $\theta_0 = 0$ thus, using the likelihood ratio test, we tested $H_0 : \theta_0 = 0$ against $H_1 : \theta_0 \neq 0$. For this test, we got a p-value of $Pr(\chi_1^2 > 2[44.6518 - 41.3455]) = 0.010$, where χ_1^2 denotes a chi-square random variable with 1 degree of freedom. Therefore, we have enough evidence at 0.05 level of significance to reject H_0 and we must reject that $\theta_0 = 0$. Hence, the regression model without the intercept in Table 7 is not significantly better than the one with the intercept in Table 6.

Since we have selected the full model in Table 6, based on Section 5, the results therein can be expressed as the following regression equation,

$$\log\left(\frac{\hat{\epsilon}_i}{1-\hat{\epsilon}_i}\right) = -0.6749 - 0.0110\,\text{Household income}_i + 0.1039\,\text{Household size}_i. \tag{29}$$

Thus, the interpretation of Equation (29) is that, the higher the household income the lower the proportion of income spent on food and this result is consistent with the famous Engel's Law which states that *"the proportion of household income spent on food declines as the household income increases"*; however, the more the number of individuals in the household increases, the more the proportion of income spent on food increases.

The ZOIUUTW distribution was fitted to the age data alongside with the ZOIbeta distribution and the ZOIKum distribution as competitors. The fit results of the models are presented in Table 8 and from there, we could observe that the ZOIUUTW distribution gave the smallest AIC, BIC and AICc values compare to the ones produced by the competing distributions and the plot of the estimated cdf of the ZOIUUTW distribution superimposed on the empirical cdf plot in Fig. 8(b) indicates by their mimicking patterns that the ZOIUUTW distribution provides a good fit for the age data. Therefore, for this data, the ZOIUUTW distribution provides a better fit than both the ZOIbeta distribution and the ZOIKum distribution.

9. Concluding remarks

In this paper, we introduced and studied a unit upper truncated Weibull (UUTW) distribution with the inflated variant, where the inflation occurs at both 0 and 1. Aside from the pdf and cdf, other important mathematical properties of the proposed distribution were derived, such as the reliability function, hazard rate function (J-shaped), moments, quantile function, inequality measures. Statistical inferences on the parameters of the proposed unit distribution have been dealt with via the method of maximum likelihood estimation. Three real-data applications are used to demonstrate the effectiveness of the proposed model. The first two real-life examples are related to the field of economics and hydrology which involve data on the proportion of household income spent on food and the maximum flood level data, respectively, the UUTW model provides better results than the existing one/two-parameter Topp-Leone, unit Rayleigh, log-weighted exponential, Kumaraswamy, beta and the unit zero-truncated Poisson power function distributions. Then, the flexibility of the inflated version of the proposed model is illustrated using the CD34+ data, which involve 239 patients at the Edmonton Hematopoietic Stem Cell Lab in Cross Cancer Institute-Alberta Health Services from 2003 to 2008. Furthermore, we developed a regression model based on the UUTW distribution. As an extended development of the aforementioned application of the UUTW distribution to the univariate case of the proportion of household income spent on food, we demonstrate that given the household income and the household size, the UUTW regression model can adequately predict the proportion of household income spent on food.

Ethics declarations

Informed consent was not required for this study because it involves secondary data which are freely available from many sources.

CRediT authorship contribution statement

Idika E. Okorie: Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emmanuel Afuecheta:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation. **Hassan S. Bakouch:** Validation, Supervision, Methodology.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the Editor, Associate Scientific Editor and the three referees for careful reading and comments which greatly improved the paper. The authors would also like to acknowledge the support provided by Deanship of Scientific Research (DSR) at King Fahd University of Petroleum & Minerals for taking care of the APC.

References

[1] A.K. Gupta, S. Nadarajah, Handbook of Beta Distribution and Its Applications, CRC Press, Boca Raton, FL, ISBN 0-8247-5396-8, 2004.
[2] P. Kumaraswamy, A generalized probability density function for double-bounded random processes, J. Hydrol. 46 (1–2) (1980) 79–88, https://doi.org/10.1016/0022-1694(80)90036-0.
[3] M.C. Jones, Kumaraswamy's distribution: a beta-type distribution with some tractability advantages, Stat. Methodol. 6 (1) (2009) 70–81, https://doi.org/10.1016/j.stamet.2008.04.001.
[4] C.W. Topp, F.C. Leone, A family of J-shaped frequency functions, J. Am. Stat. Assoc. 50 (269) (1955) 209–219, https://doi.org/10.1080/01621459.1955.10501259.
[5] A. Grassia, On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma distribution and derived compound distributions, Aust. J. Stat. 19 (1977) 108–114, https://doi.org/10.1111/j.1467-842X.1977.tb01277.x.
[6] E. Gómez-Déniz, M.A. Sordo, E. Calderín-Ojeda, The log-Lindley distribution as an alternative to the beta regression model with applications in insurance, Insur. Math. Econ. 54 (2013) 49–57, https://doi.org/10.1016/j.insmatheco.2013.10.017.
[7] A.F.B. Menezes, J. Mazucheli, S. Dey, The unit-logistic distribution: different methods of estimation, Braz. Oper. Res. Soc. 38 (3) (2018) 555–578, https://doi.org/10.1590/0101-7438.2018.038.03.0555.
[8] J. Mazucheli, A.F.B. Menezes, S. Dey, The unit-Birnbaum-Saunders distribution with applications, Chil. J. Stat. (ISSN 0718-7920) 9 (2018) 47–57.
[9] J. Mazucheli, A.F.B. Menezes, S. Dey, Unit-Gompertz distribution with applications, Statistica 79 (2019) 25–43, https://doi.org/10.6092/issn.1973-2201/8497.
[10] M.E. Ghitany, J. Mazucheli, A.F.B. Menezes, F. Alqallaf, The unit-inverse Gaussian distribution: a new alternative to two-parameter distributions on the unit interval, Commun. Stat., Theory Methods 48 (2019) 3423–3438, https://doi.org/10.1080/03610926.2018.1476717.
[11] R.A. Bantan, C. Chesneau, F. Jamal, M. Elgarhy, M.H. Tahir, A. Ali , et al., Some new facts about the unit-Rayleigh distribution with applications, Mathematics 8 (11) (2020) 1954, https://doi.org/10.3390/math8111954.
[12] J. Mazucheli, F.B.A. Menezes, L.B. Fernandes, R.P. de Oliveira, M.E. Ghitany, The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, J. Appl. Stat. 47 (2020) 954–974, https://doi.org/10.1080/02664763.2019.1657813.
[13] M.A. Haq, S. Hashmi, K. Aidi, P.L. Ramos, F. Louzada, Unit modified Burr-III distribution: estimation, characterizations and validation test, Ann. Data Sci. (2020), https://doi.org/10.1007/s40745-020-00298-6.
[14] E. Altun, The log-weighted exponential regression model: alternative to the beta regression model, Commun. Stat., Theory Methods 50 (10) (2021) 2306–2321, https://doi.org/10.1080/03610926.2019.1664586.
[15] M.Ç. Korkmaz, C. Chesneau, Z.S. Korkmaz, Transmuted unit Rayleigh quantile regression model: alternative to beta and Kumaraswamy quantile regression models, UPB Sci. Bull., Ser. A, Appl. Math. Phys. (ISSN 1223-7027) 83 (2021) 149–158.
[16] J. Mazucheli, M.Ç. Korkmaz, A.F. Menezes, V. Leiva, The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics, and numerical applications, Soft Comput. 27 (1) (2023) 279–295, https://doi.org/10.1007/s00500-022-07278-3.
[17] F.A. Bhatti, A. Ali, G. Hamedani, M.Ç. Korkmaz, M. Ahmad, The unit generalized log Burr XII distribution: properties and application, AIMS Math. (2021), https://doi.org/10.3934/math.2021592.
[18] M.C. Korkmaz, C. Chesneau, Z.S. Korkmaz, The unit folded normal distribution: a new unit probability distribution with the estimation procedures, quantile regression modeling and educational attainment applications, J. Reliab. Stat. Stud. 261 (298) (2022), https://doi.org/10.13052/jrss0974-8024.15111.
[19] R.D. Gupta, D. Kundu, A new class of weighted exponential distributions, Statistics 43 (6) (2009) 621–634, https://doi.org/10.1080/02331880802605346.
[20] D.V. Lindley, Fiducial distributions and Bayes' theorem, J. R. Stat. Soc. Ser. B (Methodol.) (1958) 102–107, http://www.jstor.org/stable/2983909.
[21] K. Shekhawat, V.K. Sharma, An extension of J-shaped distribution with application to tissue damage proportions in blood, Sankhya B 83 (2) (2021) 548–574, https://doi.org/10.1007/s13571-019-00218-6.
[22] R. Ospina, S.L. Ferrari, Inflated beta distributions, Stat. Pap. 51 (2010) 111–126, https://doi.org/10.1007/s00362-008-0125-4.
[23] F. Cribari-Neto, J. Santos, Inflated Kumaraswamy distributions, An. Acad. Bras. Ciênc. 91 (2019), https://doi.org/10.1590/0001-3765201920180955.
[24] G. Martínez-Flórez, R. Tovar-Falón, C. Barrera-Causil, Inflated unit-Birnbaum-Saunders distribution, Mathematics 10 (4) (2022) 667, https://doi.org/10.3390/math10040667.
[25] I.E. Okorie, A.C. Akpanta, J. Ohakwe, D.C. Chikezie, C.U. Onyemachi, M.K. Rastogi, Zero-truncated Poisson-power function distribution, Ann. Data Sci. 8 (1) (2021) 107–129, https://doi.org/10.1007/s40745-019-00201-y.
[26] R.E. Glaser, Bathtub and related failure rate characterizations, J. Am. Stat. Assoc. 75 (371) (1980) 667–672, https://doi.org/10.1080/01621459.1980.10477530.
[27] E.M. Wright, The asymptotic expansion of the generalized hypergeometric function, J. Lond. Math. Soc. 1 (4) (1935) 286–293, https://doi.org/10.1112/jlms/s1-10.40.286.

[28] A.M. Mathai, R.K. Saxena, The H-Function with Applications in Statistics and Other Disciplines, John Wiley & Sons, ISBN 0470263806, 1978.
[29] H.M. Srivastava, K.C. Gupta, S.P. Goyal, The H-functions of one and two variables, with applications, South Asian Publishers, https://cir.nii.ac.jp/crid/1130000794835392768, 1982.
[30] A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, Theory and Applications of Fractional Differential Equations (vol. 204), Elsevier, 2006, ISBN-13:978-0-444-51832-3, ISBN-10:0-444-51832-0.
[31] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021, https://www.R-project.org/.
[32] D.R. Cox, E.J. Snell, A general definition of residuals, J. R. Stat. Soc., Ser. B, Methodol. 30 (2) (1968) 248–265, https://doi.org/10.1111/j.2517-6161.1968.tb00724.x.
[33] P.K. Dunn, G.K. Smyth, Randomized quantile residuals, J. Comput. Graph. Stat. 5 (3) (1996) 236–244, https://doi.org/10.1080/10618600.1996.10474708.
[34] N. Balakrishna, K. Shiji, On a class of bivariate exponential distributions, Stat. Probab. Lett. 85 (2014) 153–160, https://doi.org/10.1016/j.spl.2013.11.009.
[35] K. Okhli, M.J. Nooghabi, On the contaminated exponential distribution: a theoretical Bayesian approach for modeling positive-valued insurance claim data with outliers, Appl. Math. Comput. 392 (2021) 125712, https://doi.org/10.1016/j.amc.2020.125712.
[36] H. Torabi, F.L. Bagheri, E. Mahmoudi, Estimation of parameters for the Marshall–Olkin generalized exponential distribution based on complete data, Math. Comput. Simul. 146 (2018) 177–185, https://doi.org/10.1016/j.matcom.2017.11.005.
[37] S. Chakraborty, S. Bhattacharjee, Modeling proportion of success in high school leaving certificate examination-a comparative study of inflated unit Lindley and inflated beta distribution, J. Math. Comput. Sci. 11 (6) (2021) 7170–7198, https://doi.org/10.28919/jmcs/6248.
[38] E. Altun, G.M. Cordeiro, The unit-improved second-degree Lindley distribution: inference and regression modeling, Comput. Stat. 35 (2020) 259–279, https://doi.org/10.1007/s00180-019-00921-y.
[39] W.E. Griffiths, R.C. Hill, G.G. Judge, Learning and Practicing Econometrics, John Willey and Sons, Inc., New York, USA, ISBN 0471513644, 1993, ISBN: 0471599514.
[40] R. Dumonceaux, C.E. Antle, Discrimination between the log-normal and the Weibull distributions, Technometrics 15 (4) (1973) 923–926, https://doi.org/10.1080/00401706.1973.10489124.
[41] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.
[42] G.E. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464, https://www.jstor.org/stable/2958889.
[43] C.M. Hurvich, C.L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (2) (1989) 297–307, https://doi.org/10.1093/biomet/76.2.297.
[44] P. Zhang, Z. Qiu, C. Shi, simplexreg: an R package for regression analysis of proportional data using the simplex distribution, J. Stat. Softw. 71 (11) (2016), https://doi.org/10.18637/jss.v071.i11.