P-ISSN: 2790-6795	Annals of Human and Social Sciences	Oct-Dec 2023,Vol. 4, No. 4
O-ISSN:2790-6809	http://doi.org/10.35484/ahss.2023(4-IV)05	[51-66]

Annals of Human and Social Sciences
www.ahss.org.pk

RESEARCH PAPER

Unit Interval Exponentiated Exponential Distribution and Quantile Regression Model: Applications for the COVID-19 Data and Bounded Responses Data

[1]Bushra Masood, [2]Shakila Bashir* and [3]Nimra Masood

1. MPhil Scholar, Faculty of Computer and Mathematics Sciences, Forman Christian College (A Chartered University) Lahore, Punjab, Pakistan
2. Associate Professor, Faculty of Computer and Mathematics Sciences, Forman Christian College (A Chartered University) Lahore, Punjab, Pakistan
3. MPhil Scholar, Department of Statistics, Government College University Lahore, Punjab, Pakistan

Corresponding Author shakilabashir@fccollege.edu.pk

ABSTRACT

The aim of this research is to propose the bounded exponentiated exponential (BEE) distribution for modeling the datasets on the unit interval e.g., between [0,1]. The proposed unit interval distribution is used to develop a quantile regression model. Continuous probability distributions are very useful to model lifetime data sets. Every single probability distribution is not suitable for all kinds of data sets. Therefore, proposing a new density can always be useful if showing the versatility and flexibility in it. Bounded exponentiated exponential distribution is developed by transforming the variable. The proposed density function exhibits different shapes which show its flexibility over different kinds of data sets. Many statistical and reliability properties of the BEE distribution have been developed. Few estimations methods have been discussed to estimate the parameters of the BEE distribution and a monte Carlo simulation study has been done. Subsequently, the applications of the BEE distribution are illustrated using COVID-19 data. Finally, several properties of the quantile regression model are derived, and the model is also applied on a unit interval response variable data set. For the purpose of modeling dependence between measures in a dataset, a bivariate extension of the proposed distribution is also developed. Furthermore, the bivariate model can be extended toward the development of its properties and applications.

Keywords: BEE Distribution, Bivariate Extension, Bounded, Exponentiated Exponential, Quantile Regression

Introduction

Modeling and prediction for diseases are the crucial responsibilities of epidemiologists and researchers as well who are interested in the estimation and chances of the occurrence of diseases. Probability distributions play a vital role in implementing these responsibilities, by modeling the adaptability in the occurrence of diseases. Many researchers have proposed several new probability distributions discrete and continuous as well for modeling the number of infections, mortality rate, and recovery rate during the novel coronavirus disease (COVID-19) and its impact on humankind. Midst the proposed probability distributions for modeling diseases, those defined on unit interval play an important role to their appropriateness in the areas such as health, psychology, and epidemiology. For example, the researcher might be interested in modeling the mortality rate or recovery rate and in such situations, the variables are usually proportions, fractions, or rates which are defined in the unit interval. In such situations the modeled distribution should also be bound with the unit interval domain. Although beta distribution is the oldest distribution based on unit interval but due to the complexity of its cumulative distributions function (CDF) and quantile function, new probability distributions defined over unit interval are developed whose CDFs and quantile function are manageable. Recently proposed unit interval probability distributions are unit gamma/Gompertz distribution, Bantan et al. (2021); bounded odd inverse Pareto exponential distribution Nasiru et al. (2021); bounded shifted Gompertz distribution, Jodra (2020); unit modified Burr-III distribution Haq et al. (2020); unit generalized half normal distribution Korkmaz (2020); unit Lindley distribution Mazucheli et al. (2019); unit Gompertz distribution, Mazucheli et al. (2019); logit slash distribution, Korkmaz (2019); unit Weibull distribution, Mazucheli et al, (2018); unit inverse Gaussian distribution Ghitany et al. (2018).

In spite of the presence of many unit interval probability distributions in the literature, no single distribution is proficient in modeling all types of data afterward during the data generating process the produced data may have different features such as symmetric, skewed, varied degrees of kurtosis and monotonic and non-monotonic hazard rates. Here in this research, we proposed a new unit distribution named as bounded exponentiated exponential (BEE) distribution. The impetus for the newly proposed probability distribution is to provide a model for modeling intricate data on the unit interval that shows platykurtic, leptokurtic, reversed J, left/right skewed, bathtub and J shapes; to develop a bivariate distribution for modeling independence between random data on the unit interval; to develop a quantile regression model for understanding the relationship between a response variable and given covariates.

Methedology

Development of Bounded Exponentiated Exponential (BEE) Distribution

A random variable X follows the Exponentiated Exponential Distribution if the Cumulative density function (CDF) and the Probability density function (PDF) are as defined below

$$F(y) = \left(1 - e^{-\lambda y}\right)^{\alpha} \tag{1}$$

and

$$f(y) = \alpha\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right)^{\alpha-1} \tag{2}$$

Now, we propose a new distribution named bound exponentiated exponential distribution (BEED) by the transformation of Y = e^{-X} → X = -log(Y). The CD of the (BEED) is obtained as follows

$$F_Y(y;\ \alpha, \lambda) = P(e^{-X} \leq y)$$

$$F_Y(y;\ \alpha, \lambda) = 1 - F_X(-\log(Y);\ \alpha, \lambda)$$

Finally, we get, the CDF of the (BEED) is given below

$$F(y) = 1 - \left(1 - y^{\lambda}\right)^{\alpha},\ 0 < y < 1,\ \alpha, \lambda > 0 \tag{3}$$

Where α is the shape and λ is the scale parameter.

The probability density function (PDF) of (BEED) is given below

$$f(y) = \alpha\, \lambda y^{\lambda-1}\left(1 - y^{\lambda}\right)^{\alpha-1},\ 0 < y < 1,\ \alpha, \lambda > 0 \tag{4}$$

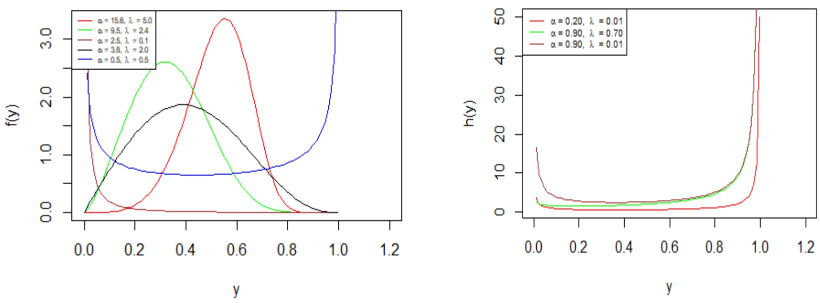

Figure 1. PDF (left) HRF (right) of the BEE distribution

From figure 1, from the PDF plot it can be observed that the BEE distribution exhibits a variety of shapes as symmetric, left and right skewed, U-shaped. The graph of hazard rate function (HRF) of the BEE distribution shows the bathtub and increasing trend for different values of parameters.

Results and findings: The finding and results of the article has been devided into some section explained as reliability measures, some statistical properties, bivariate extemsion, estimation of parameters, simulations, applications and quantile regression model.

Reliability measures

In this section a few reliability measures as survival (reliability) function, hazard rate function, cumulative hazard rate function, reversed hazard rate function, odd function, mills ratio, elasticity, and

The survival function of BEED is given as

$$S(y) = \left(1 - y^{\lambda}\right)^{\alpha} \tag{5}$$

The Hazard rate is the death rate of a subject of given age y. The hazard function of (BEED) drive below

$$h(y) = \frac{\alpha\,\lambda y^{\lambda-1}\left(1-y^{\lambda}\right)^{\alpha-1}}{\left(1-y^{\lambda}\right)^{\alpha}} \tag{6}$$

The cumulative hazard rate function of the BEED is

$$H(y) = -\alpha log\left[1 - y^{\lambda}\right]$$

The Reverse Hazard Function is the ratio of the life probability density to its distribution function. The reverse hazard function of BEED drive below

$$r_h(y) = \frac{\alpha\,\lambda y^{\lambda-1}\left(1-y^{\lambda}\right)^{\alpha-1}}{1-\left(1-y^{\lambda}\right)^{\alpha}} \tag{7}$$

The odd function of the BEED is

$$O(y) = \frac{1-\left[1-y^{\lambda}\right]^{\alpha}}{\left[1-y^{\lambda}\right]^{\alpha}} \tag{8}$$

The mills ratio for the BEED is

$$M(y) = \frac{(1-y^{\lambda})^{\alpha}}{\alpha\,\lambda y^{\lambda-1}(1-y^{\lambda})^{\alpha-1}} \quad (9)$$

The elasticity for the BEED is

$$\varepsilon(y) = \frac{\alpha\,\lambda y^{\lambda}(1-y^{\lambda})^{\alpha-1}}{1-(1-y^{\lambda})^{\alpha}} \quad (11)$$

Some Statistical Properties of BEED

In this section several important properties including quantile function, median, rth moments, mgf, mean, variance, incomplete moments, Lorenz and Bonferroni curves of BEED have been investigated. Numerical values for the first four moments, variance, standard deviation, measures of skewness and kurtosis, and coefficient of variation have been presented in table 1.

The Quantile function of the BEED is

$$Q(y;\alpha,\lambda) = y_q = \left[1 - (1-q)^{1/\alpha}\right]^{1/\lambda} \quad (12)$$

The median and Inter Quartile Range for the BEED can be calculated as Median = $y_{0.5}$ and IQR = $y_{0.75} - y_{0.25}$.

The rth moments of BEED are given by

$$\mu_r' = \alpha B\left(\frac{r}{\lambda}+1,\alpha\right) \quad (13)$$

The mean of the BEED is

$$\mu_1' = \frac{\alpha\Gamma(1/\lambda)\Gamma(\alpha)}{(1+\alpha\lambda)\Gamma\left(\frac{1}{\lambda}+\alpha\right)} \quad (14)$$

The variance of the BEED is

$$\sigma^2 = \frac{2\alpha\Gamma(2/\lambda)\Gamma(\alpha)}{(2+\alpha\lambda)\Gamma\left(\frac{2}{\lambda}+\alpha\right)} - \left(\frac{\alpha\Gamma(1/\lambda)\Gamma(\alpha)}{(1+\alpha\lambda)\Gamma\left(\frac{1}{\lambda}+\alpha\right)}\right)^2 \quad (15)$$

The SD (Standard deviation), CV (coefficient of variation), CS (Coefficient of Skewness), and CK (Coefficient of Kurtosis) from the Central moments from the following formulas given below

Standard deviation = $\sqrt{\mu_2}$

Coefficient of Skewness /β_1 = $\frac{\mu_3^2}{\mu_2^3}$

Coefficient of Kurtosis /β_2 = $\frac{\mu_4}{\mu_2^2}$

Table 1
Values of moment measures, including the SD, CV, CS and CK.

μ_r'	α = 0.4, λ = 2.5	α = 0.4, λ = 2.5	α = 20, λ = 1.5
μ_1'	0.845234	0.526464	0.119228
μ_2'	0.750028	0.305664	0.020337
μ_3'	0.683809	0.190082	0.004329
μ_4'	0.634227	0.124461	0.001079
SD	0.188701	0.168819	0.078241
CV	1.431727	1.450113	1.559052
CS	0.382065	0.011103	-0.02134
CK	0.008592	0.007327	0.010044

Table 1 indicates that CS is positive and negative as well for different parametric values, the BEED can be positively/negatively skewed. The CK shows values less than 3, it reveals that the BEED is showing the platykurtic trend.

Theorem 1. The r^{th} incomplete Central moment of EED is given below

$$\varphi_r = \alpha B\left(y^{\lambda}; \frac{r}{\lambda}+1, \alpha\right) \tag{16}$$

Where $B(z;\alpha,\beta) = \int_0^z y^{\alpha-1}(1-y)^{\beta-1}$dy, it is known as the Beta function.

Proof.

$$\varphi_r = E(y^r) = \int_0^y y^r f(y)\, dy$$

$$\varphi_r = \int_0^y y^r \alpha\, \lambda y^{\lambda-1}\left(1-y^{\lambda}\right)^{\alpha-1} dy$$

So, the above expression becomes the Beta Function given in eq. (16).

Theorem 2. The moment generating function of BEED is given below

$$M_y(t) = \alpha \sum_{n=0}^{\infty} \frac{t^n}{n!} B\left(\frac{r}{\lambda}+1, \alpha\right) \tag{17}$$

Proof.

$$M_y(t) = E(e^{ty}) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mu_r'$$

We know that $\mu_r' = \alpha B\left(\frac{r}{\lambda}+1, \alpha\right)$. So, the above expression gives the result in eq. (17).

Theorem 3. The Lorenz curve $L_F(y)$ for incomplete moments is defined as

$$\mathrm{L_F(y)} = \frac{\alpha}{\mu} B\left(y; \frac{1}{\lambda}+1, \alpha\right) \tag{18}$$

Where $B(z;\alpha,\beta) = \int_0^z y^{\alpha-1}(1-y)^{\beta-1}$dy, it is known as the Beta function.

Proof.

$$L_F(y) = \frac{1}{\mu}\int_0^y y\, f(y)\, dy$$

$$L_F(y) = \frac{1}{\mu}\int_0^y y\alpha\, \lambda y^{\lambda-1}\left(1-y^{\lambda}\right)^{\alpha-1}\, dy$$

So, the above expression becomes the function given in eq. (18).

Corollary 4. Bonferroni curve $B_F(y)$ is defined as

$$B_F(y) = \frac{L_F(y)}{F(y)}$$

$$B_F(y) = \frac{\frac{\alpha}{\mu}B\left(y;\frac{1}{\lambda}+1,\alpha\right)}{1-\left(1-y^{\lambda}\right)^{\alpha}} \tag{19}$$

Bivariate Extension

Modeling the relationship between two quantitative variables may be of interest to researchers. For example, one could be interested in modelling the link between an individual's age and BMI or might be interested in investigating independence/dependence between variables. Consequently, bivariate distributions can be utilized to get these estimations. The bivariate distributions can be used in reliability analysis, queuing theory, finance and indemnity risk analysis. This section proposed a bivariate extension of BEED abbreviated as BE-BEED. The CDF and PDF of BE-BEE distribution is shown below for a bivariate continuous random vector (X, Y).

$$F_{X,Y}(x,y;\boldsymbol{\eta}) = \frac{\left[1-\left(1-x^{\lambda}\right)^{\alpha}\right]\left[1-\left(1-y^{\lambda}\right)^{\alpha}\right]}{\left\{1-(\delta_1+\delta_3)\left(1-x^{\lambda}\right)^{\alpha}+(\delta_2+\delta_3)\left(1-y^{\lambda}\right)^{\alpha}\right\}^{-1}} \tag{20}$$

$$f_{X,Y}(x,y;\boldsymbol{\eta}) = \frac{(\alpha\lambda)^2(xy)^{\lambda-1}\left[\left(1-x^{\lambda}\right)\left(1-y^{\lambda}\right)\right]^{\alpha}\begin{bmatrix}(2\delta_2+2\delta_3)\left(1-y^{\lambda}\right)^{\alpha}-(2\delta_1+2\delta_3)\left(1-x^{\lambda}\right)^{\alpha}\\ -\delta_2+\delta_1+1\end{bmatrix}}{(1-x^{\lambda})(1-y^{\lambda})}$$

where $\alpha,\lambda > 0$, -1< $\delta_1+\delta_3$< 1, -1< $\delta_2+\delta_3$< 1, 0 < x < 1, 0 < y < 1 and

(21)

$\boldsymbol{\eta} = (\alpha,\lambda,\delta_1,\delta_2,\delta_3)^{\boldsymbol{T}}$. The parameters δ_1, δ_2 and δ_3 quantify the dependency between a BE-BEED random vectors of two variables.

Figure 2 shows the CDF graphs for the given parameter values.

i- α = 3.5, λ = 8.2, δ_1= 0.3, δ_2= 0.1, δ_3= 0.3;

ii- α = 2.5, λ = 0.8, δ_1= 0.5, δ_2= 0.4, δ_3= 0.2 and

iii- α = 0.5, λ = 4.8, δ_1= -0.3, δ_2= -0.7, δ_3= -0.1

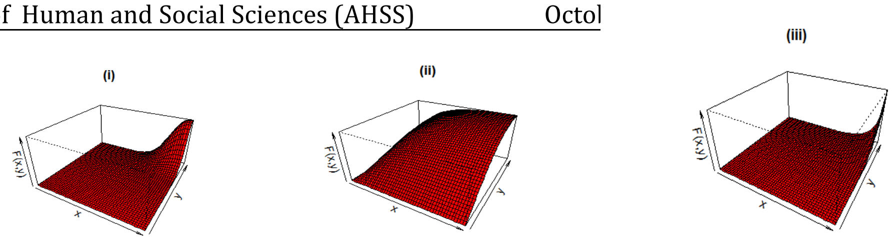

Figure 2. CDF plots of the BE-BEE distribution

Figure 2 shows the PDF graphs for the given parameter values.

i- α = 3.5, λ = 8.2, δ_1= 0.3, δ_2= 0.1, δ_3= 0.3.

ii- α = 2.5, λ = 0.8, δ_1= 0.5, δ_2= 0.4, δ_3= 0.2 and

iii- α = 0.5, λ = 4.8, δ_1= -0.3, δ_2= -0.7, δ_3= -0.1

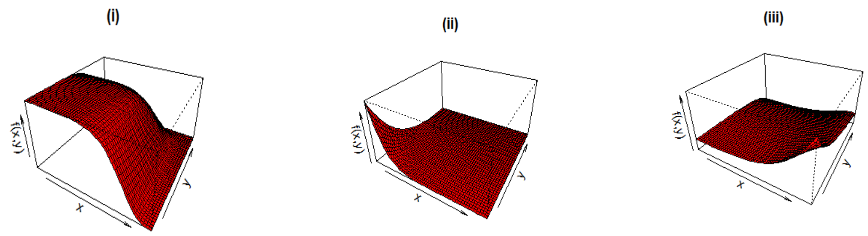

Figure 3. PDF plots of the BE-BEE distribution

Estimation of Parameters

This section describes six methods of estimation for estimating the parameters of the BEED distribution. Maximum likelihood estimation (MLE), Ordinary least squares (OLS), Weighted least squares (WLS), Cramér-von Mises (CVM), Percentile (PC) estimation, and Anderson-Darling (AD) approaches.

Maximum Likelihood Estimation

Suppose that Y follows the BEE distribution, then we have

$$l = log(\alpha\,\lambda) + (\lambda - 1)log(y) + (\alpha - 1)log\left(1 - y^{\lambda}\right) \qquad (22)$$

To find the value of the "$\boldsymbol{\alpha}$", taking derivative of Equation (22) with respect to $\boldsymbol{\alpha}$ and we obtain

$$\frac{d\ell}{d\alpha} = \frac{1}{\alpha} + log\left(1 - y^{\lambda}\right) \qquad (23)$$

And to find the value of the "**λ**", taking derivative of Equation (22) with respect to $\boldsymbol{\lambda}$ and we obtain

$$\frac{d\ell}{d\lambda} = \frac{1}{\lambda} + log(y) - \frac{\lambda y^{\lambda-1}(\alpha-1)}{(1-y^{\lambda})} \quad (24)$$

However, as a result the equations have not a closed form and can be solved numerically to find the parameter estimates.

4.2. Ordinary and Weighted Least Squares Estimation

Suppose that $Y_i, i = 1,2,\ldots,n$ denotes the order statistics from a sample of size n, and we have

$$E[F(Y_{(i)})] = \frac{i}{(n+1)}$$

The least square estimator parameters "α" and "λ" are estimated by minimizing

$$Q(\alpha,\lambda) = \sum_{i=1}^{n}\left[F(Y_{(i:n)}|\alpha,\lambda) - \frac{i}{(n+1)}\right]^2 \quad (25)$$

In the case of Bound Exponentiated Exponential distribution, eq. (25) becomes

$$Q(\alpha,\lambda) = \sum_{i=1}^{n}\left[1 - \left(1-y_{(i)}^{\lambda}\right)^{\alpha} - \frac{i}{(n+1)}\right]^2 \quad (26)$$

To find the estimates of the "$\boldsymbol{\alpha}$" and "$\boldsymbol{\lambda}$", take partial derivative of Equation (26) with respect to the parameters. The following equations are

$$\sum_{i=1}^{n}\left[1 - \left(1-y_{(i)}^{\lambda}\right)^{\alpha} - \frac{i}{(n+1)}\right]\Delta_s\left(y_{(i)}|\alpha,\lambda\right) = 0, s = 1,2. \quad (27)$$

where

$$\Delta_1\left(y_{(i)}|\alpha,\lambda\right) = ln\left(1-y_{(i)}^{\lambda}\right)\left(1-y_{(i)}^{\lambda}\right)^{\alpha}$$

and

$$\Delta_2\left(y_{(i)}|\alpha,\lambda\right) = \alpha y_{(i)}^{\lambda}\ln\left(y_{(i)}\right)\left(1-y_{(i)}^{\lambda}\right)^{\alpha-1}$$

The WLS estimates $\boldsymbol{\hat{\alpha}_{WLS}}$ and $\boldsymbol{\hat{\lambda}_{WLS}}$, can obtain by minimizing

$$WLS(\alpha,\lambda) = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F(Y_{(i:n)}|\alpha,\lambda) - \frac{i}{(n+1)}\right]^2 \quad (28)$$

$$WLS(\alpha,\lambda) = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1 - \left(1-y_{(i)}^{\lambda}\right)^{\alpha} - \frac{i}{(n+1)}\right]^2 \quad (29)$$

To find the estimates of the "$\boldsymbol{\hat{\alpha}}$" and "$\boldsymbol{\hat{\lambda}}$", take the partial derivative of Equation (29) with respect to the parameters. The following equations are

$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1 - \left(1-y_{(i)}^{\lambda}\right)^{\alpha} - \frac{i}{(n+1)}\right]\Delta_s\left(y_{(i)}|\alpha,\lambda\right) = 0, s = 1,2. \quad (30)$$

Where $\Delta_s\left(y_{(i)}|\alpha,\lambda\right)$ = 0, s = 1, 2 is define above

4.3. Cramér–Von Mises Estimation

Let $Y_i, i = 1,2,\dots,n$ be ordered observations taken from BEE random variables. The Cramér–Von Mises estimates of the parameters $\hat{\boldsymbol{\alpha}}_{\boldsymbol{CVM}}$ and $\hat{\boldsymbol{\lambda}}_{\boldsymbol{CVM}}$ are determined by minimizing the function that are given below

$$CVM(\alpha,\lambda) = \frac{1}{12n} + \sum_{i=1}^{n}\left[F\left(Y_{(i:n)}|\alpha,\lambda\right) - \frac{2i-1}{2n}\right]^2 \quad (31)$$

$$CVM(\alpha,\lambda 0) = \frac{1}{12n} + \sum_{i=1}^{n}\left[1 - \left(1 - y_{(i)}^{\lambda}\right)^{\alpha} - \frac{2i-1}{2n}\right]^2 \quad (32)$$

Differentiate the eq. (32) with respect to "$\boldsymbol{\alpha}$" and "$\boldsymbol{\lambda}$", the estimates of the parameters can be determined numerically by the following equations.

$$\sum_{i=1}^{n}\left[1 - \left(1 - y_{(i)}^{\lambda}\right)^{\alpha} - \frac{2i-1}{2n}\right]\Delta_s\left(y_{(i)}|\alpha,\lambda\right) = 0, s = 1,2 \quad (33)$$

Where $\Delta_s\left(y_{(i)}|\alpha,\lambda\right)$ are define in the section 5.2.

4.4. Anderson–Darling Estimation

Let $Y_i, i = 1,2,\dots,n$ be ordered observations from sample from n BEE random variables. The Anderson–Darling estimates of the parameters $\hat{\boldsymbol{\alpha}}_{\boldsymbol{AD}}$ and $\hat{\boldsymbol{\lambda}}_{\boldsymbol{AD}}$ are determined by minimizing the function that are given below

$$A(\alpha,\lambda) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{logF(x_{1:n}|\alpha,\lambda) + log\bar{F}(x_{n+1-i:n}|\alpha,\lambda)\right\} \quad (34)$$

These estimators can be derived by solving the non-linear equations given below

$$A(\alpha,\lambda) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{log\left(1 - \left(1 - y_{(i)}^{\lambda}\right)\right) + log\left(1 - y_{(i)}^{\lambda}\right)\right\} \quad (35)$$

with respect to parameters α and λ.

4.5. Percentile Estimation

Let $Y_i, i = 1,2,\dots,n$ be ordered observations from sample from n BEE random variables and

$u_i = \frac{i}{n+1}$ is an unbiased estimate of $F_Y\left(y_{(i)};\alpha,\lambda\right)$. The PC estimates of the BTCPE distribution parameters are derived by minimizing the following function:

$$PC(\alpha,\lambda) = \sum_{i=1}^{n}[y_i - \mathrm{Q}(y;\alpha,\lambda)]^2 \quad (36)$$

$$PC(\alpha,\lambda) = \sum_{i=1}^{n}\left[y_i - \left[1 - (1-q)^{1/\alpha}\right]^{1/\lambda}\right]^2 \quad (37)$$

with the parameters α and λ.

Simulation Study

In this section a Monte Carlo simulation study is conducted by taking different samples sizes as 20, 40, 100, 200 and 400 with10,000 iterations. Biases, averages biases, mean squared errors (MSE) and mean relative errors (MRE) are calculated for parameters of the proposed distribution with different estimation methods discussed above. From table 2, it can be observed that as sample size increases the MSE decreases for both parameters for MLE, AD, CVM, OLS and WLS. Overall MLE is showing best results, then AD and WLS are second better estimates, OLS is the third one and CVM is last one, in the context of MSE's.

Table 2
Average bias, bias, MSE and MRE for different parametric values

Models		$\alpha = 1$					$\beta = 2$				
		20	40	100	200	400	20	40	100	200	400
MLE	Average Bias	1.7862	1.6346	1.5615	1.5233	1.5112	1.1360	1.0584	1.0300	1.0074	1.0035
	Bias	0.5004	0.3091	0.1817	0.1246	0.0879	0.2539	0.1625	0.1015	0.0679	0.0477
	MSE	0.5742	0.1938	0.0569	0.0257	0.0123	0.1268	0.0451	0.0167	0.0075	0.0036
	MRE	0.3336	0.2061	0.1212	0.0831	0.0586	0.2539	0.1625	0.1015	0.0679	0.0477
AD	Average Bias	1.6769	1.5713	1.5230	1.5172	1.4990	1.0546	1.0158	1.0066	1.0020	0.9995
	Bias	0.4737	0.3119	0.1917	0.1278	0.0900	0.2308	0.1652	0.1047	0.0710	0.0502
	MSE	0.6358	0.1809	0.0583	0.0268	0.0130	0.0935	0.0445	0.0168	0.0083	0.0040
	MRE	0.3158	0.2080	0.1278	0.0852	0.0600	0.2308	0.1652	0.1047	0.0710	0.0502
CVM	Average Bias	1.9203	1.6610	1.5540	1.5186	1.5129	1.1553	1.0621	1.0257	1.0077	1.0059
	Bias	0.6506	0.3873	0.2102	0.1422	0.1020	0.2986	0.1947	0.1205	0.0751	0.0569
	MSE	1.2950	0.3113	0.0792	0.0334	0.0163	0.1831	0.0665	0.0227	0.0090	0.0051
	MRE	0.4337	0.2582	0.1401	0.0948	0.0680	0.2986	0.1947	0.1205	0.0751	0.0569
LS	Average Bias	1.6227	1.5248	1.5240	1.5092	1.5130	1.0164	0.9904	1.0024	1.0028	1.0037
	Bias	0.5126	0.3375	0.2114	0.1415	0.1050	0.2472	0.1708	0.1179	0.0793	0.0560
	MSE	0.9253	0.2103	0.0734	0.0330	0.0173	0.1054	0.0477	0.0225	0.0098	0.0050
	MRE	0.3417	0.2250	0.1409	0.0944	0.0700	0.2472	0.1708	0.1179	0.0793	0.0560
WLS	Average Bias	1.6722	1.5664	1.5352	1.5206	1.5076	1.0466	1.0164	1.0142	1.0067	1.0013
	Bias	0.5176	0.3023	0.1854	0.1340	0.0899	0.2521	0.1669	0.1016	0.0725	0.0495
	MSE	0.7019	0.1659	0.0608	0.0290	0.0131	0.1157	0.0449	0.0176	0.0083	0.0039
	MRE	0.3451	0.2015	0.1236	0.0894	0.0599	0.2521	0.1669	0.1016	0.0725	0.0495

Applications

In this section, the applications of the BEED distribution are demonstrated, and its performance is compared to that of other competing distributions defined in the unit interval like Unit Burr-III (UBIII), Modi et al. (2019), Bounded M-O Extended Exponential. Gosh et al. (2019), Unit Gompertz, Mazucheli et al. (2019), Unit Lindley, Mazucheli et al. (2018) and Unit Weibull. The model selection approaches used in arriving at the optimal model are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The best model for these selection procedures is the one with the lowest test statistic. The datasets represent COVID-19 patient death rates in Canada and the United Kingdom (UK), as well as COVID-19 patient recovery rates in Spain. The first two datasets of UK and Canada were recently reported by Nasiru et al. (2022), and the third dataset of Spain is available in Afify et al. (2022).

Table 3 shows descriptive information for COVID-19 mortality in the United Kingdom and Canada, as well as the recovery rate in Spain. Because of the kurtosis values, the datasets are platykurtic for each country. The mortality rate in the United Kingdom is skewed to the right, whereas it is skewed to the left in Canada. Spain's recovery rate is likewise skewed to the left. The boxplot of the datasets in Figure 4 also supports this.

Table 3
Descriptive statistics

Countries	UK	Canada	Spain
Minimum	0.0807	0.1159	0.4286
Maximum	0.5331	0.3347	0.8628
Mean	0.2888	0.2305	0.7240
Skewness	0.0489	-0.0873	-0.7049
Kurtosis	1.9616	2.6537	2.6021

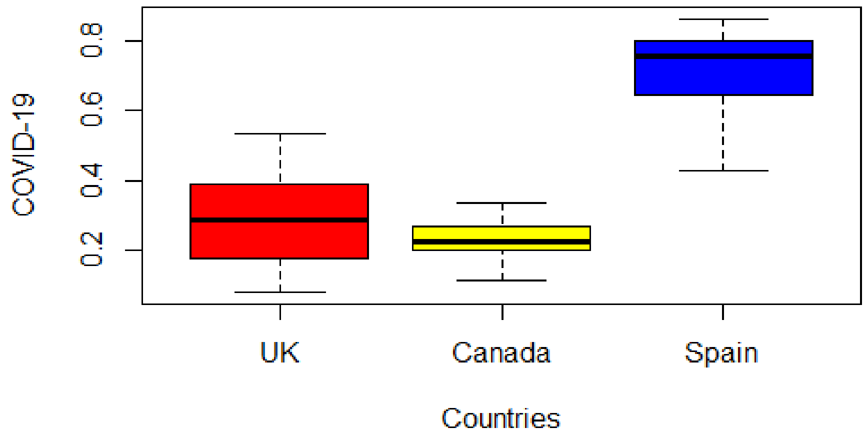

Figure 4. Box Plot of Covid-19 datasets

6.1. COVID-19 Mortality rate in UK

Table 4 shows the log-likelihood (ℓ), AIC and BIC for the fitted distributions, as well as the ML estimates of the parameters and their standard errors in brackets. Due to the lowest values of AIC, BIC, AD, CAID and the highest value of log-likelihood, the BEE distribution provides the best fit to the UK mortality dataset.

Table 4
Model Selection Criteria and Parameter estimates for UK

Model	Parameters	ℓ	AIC	BIC	AD	CAID
BEED	$\alpha = 2.6804$ (0.3008) $\lambda = 19.5817$ (6.4979)	45.8644	-87.7288	-83.5401	0.5956	0.5956
UBIII	$\alpha = 0.0758$ (0.0382) $\beta = 13.3691$ (6.5426)	38.9028	-73.8056	-69.6170	0.6784	0.6784
BMOEE	$\alpha = 3.5711$ (0.4035) $\beta = 101.7405$ (57.0214)	40.7201	-77.4402	-73.2515	1.3215	1.3215
UW	$\alpha = 3.1229$ (0.3047) $\beta = 0.2834$ (0.0602)	42.5622	-81.1244	-76.9357	1.1678	1.1678
UG	$\alpha = 1.8208$ (0.2198) $\beta = 0.0630$ (0.0274)	36.4368	-68.8736	-64.6849	1.8961	1.8961

6.2. COVID-19 Mortality rate in Canada.

Table 5 shows the ML estimates of the parameters, as well as the standard errors and model selection criteria for the fitted distributions. The BEE distribution, once again, gives the greatest fit to the Canada mortality dataset, as it has the highest log-likelihood and the lowest AIC, BIC, AD and CAID values.

Table 5
Model Selection Criteria and Parameter estimates for Canada.

Model	Parameters	ℓ	AIC	BIC	AD	CAID
BEED	$\alpha = 3.4100$ (0.2678) $\lambda = 112.7124$ (40.3173)	-80.649	-157.3	-153.25	0.3264	0.0569
UBIII	$\alpha = 0.0721$ (0.1606) $\beta = 11.2663$ (25.0812)	-30.886	-57.772	-53.721	0.3376	0.0589
BMOEE	$\alpha = 3.8284$ (0.2514) $\beta = 219.6406$ (73.8469)	-69.658	-135.32	-131.26	0.6014	0.0916
UW	$\alpha = 6.1130$ (0.5832) $\beta = 0.0567$ (0.0197)	-79.951	-155.9	-151.85	1.4315	0.2312
UG	$\alpha = 3.4261$ (0.1839) $\beta = 0.0040$ (0.0013)	-74.314	-144.63	-140.58	2.2076	0.3677

Figure 6 indicates that the BEE distribution provides a better fit to mortality for Canada than the other models because it better fits the PDF and CDF of the dataset than the other models.

6.3. COVID-19 Recovery rate in Spain.

Table 6 displays the ML estimates of the parameters, as well as their respective standard errors and model selection criteria for the fitted distributions. The BEE distribution provides the best fit to the Spain recovery rate dataset since it has the lowest AIC, BIC, AD and CAID values and the highest log-likelihood.

Table 6
Model Selection Criteria and Parameter estimates for Spain

Model	Parameters	ℓ	AIC	BIC	AD	CAID
BEED	$\alpha = 8.0783$ (0.9470) $\lambda = 7.7385$ (2.0187)	58.8343	-113.6686	-109.2893	0.8388	0.1366
UBIII	$\alpha = 5.4397$ (0.7948) $\beta = 2.0613$ (0.1723)	53.7963	-103.5927	-99.2134	1.5197	0.2575
BMOEE	$\alpha = 9.9962$ (1.2361) $\beta = 22.0547$ (9.8519)	51.4637	-98.9275	-94.5482	1.7569	0.301
UW	$\alpha = 2.2317$ (0.2036) $\beta = 8.6413$ (1.6964)	53.9658	-103.9316	-99.5523	1.5036	0.2538
UG	$\alpha = 3.8481$ (0.6024) $\beta = 0.2793$ (0.1059)	46.0284	-88.0569	-83.6776	2.2896	0.3857

Quantile Regression

In this section the quantile regression model is developed using the BEED. When the response variable specified in the unit interval is skewed or polluted with outliers, the beta regression model, which represents the response variable's conditional mean, becomes unreliable. To model the impact of variables on the response variable, a strong regression model is required. In this section, a quantile regression model is developed for modelling the response variable's conditional quantile. Given the BEED distribution's quantile function, the PDF may be re-parameterized in terms of its quantile function.

Suppose $\omega = Q(y;\, \alpha, \lambda)$ then $\lambda = \frac{log\left[1-(1-q)^{1/\alpha}\right]}{\log(\omega)}$. So, the PDF and CDF of re-parameterized distribution named as bounded exponentiated exponential quantile regression model (BEEDQRM) is given below

$$f(y;\alpha,\lambda) = \alpha\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}\right] y^{\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}.-1\right]}\left[1-y^{\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}.}\right]^{\alpha-1} \tag{38}$$

and

$$F(y;\alpha,\lambda) = 1-\left[1-y^{\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}}\right]^{\alpha} \tag{39}$$

Here "$\boldsymbol{\omega}$" is the parameter of quantile. The BEED quantile is express as

$$g(\omega_i) = z_i'\theta$$

where z_i' = $\left(1, z_{i1}, z_{i2}, \dots, z_{ip}\right)$ are the i^th^ covariate vectors, θ = $\left(\theta_o, \theta_1, \dots, \theta_p\right)'$ is the vectors of unknown parameters.

The covariates are linked to the conditional median of the dependent variable Y using the link function. The logit link function is used to link the covariates to the conditional quantile

$y \in$ [0,1]. So, we have

$$g(\omega_i) = logit(\omega_i) = log\left(\frac{\omega_i}{1-\omega_i}\right)$$

we can write further as

$$\omega_i = \frac{exp(z_i'\theta)}{1+exp\left(z_i'\theta\right)}$$

Substitute the ω_i in the Eq. (38) and we get

$$f(y;\alpha,\lambda) = \alpha\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}\right] y^{\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}.-1\right]}[1-z_i]^{\alpha-1} \tag{40}$$

where $z_i = y_i^{\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}\right]}$.

The log likelihood for determining the BEEDQRM parameters is provided by

$$\ell = \sum_{i=1}^{n} log\left[(\alpha)\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega_i)}\right] + \sum_{i=1}^{n}\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega_i)} - 1\right] log(y_i) + (\alpha-1)log(1-z_i) \tag{41}$$

where z_i is define above.

The parameters of the regression equation are estimated by directly maximizing the log likelihood function. The parameters will be denoted by $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\theta}}$ of α and θ respectively.

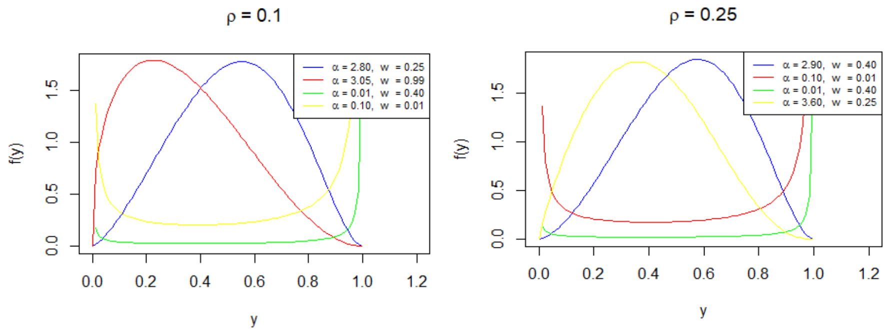

Figure 8. PDF plot of BEEQRM for some selected parameter and quantile values

From figure 8, it can be observed that BEEDQRM is showing a variety of shapes for different values of parameters. BEEDQRM shows right/left skewed, symmetric and U shaped.

The survival function and the hazard function of BEEQRM are given as

$$S(y) = \left[1 - y^{\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}}\right]^{\alpha} \tag{42}$$

$$h(y) = \frac{\alpha\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}\right] y^{\left[\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}-1\right]}\left[1-y^{\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}}\right]^{\alpha-1}}{\left[1-y^{\frac{log\left\{1-(1-q)^{1/\alpha}\right\}}{\log(\omega)}}\right]^{\alpha}} \tag{43}$$

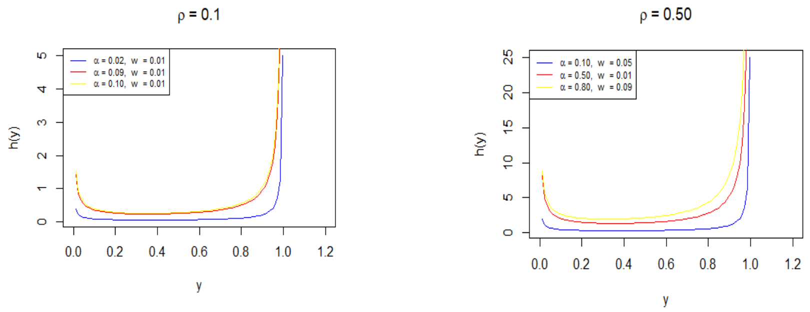

Figure 9. HRF plot of BEEQRM for some selected parameter and quantile values

Applications of Quantile Regression for BEED

In this section the applications of quantile regression are presented using the bounded exponentiated exponential quantile regression model (BEEDQRM). The data is taken from Petterle et al (2020). The response variable is fat percentage (as said earlier that it should be with in unit interval), the response variable is recorded in five regions including android, arms, gynoids, legs, and trunk, and independent variables are age (in years), body mass index (in kg/m^2), gender (male/female), and IPAQ (sedentary (S), insufficiently active (I), or active (A)). Table 7 presents ML estimates, standard errors (SE), and p-value. A level of significance of 5% is used

Table 7
ML estimates, standard errors, p-value lower limit (LL), upper limit (UL), and t score for quantile regression

Quantiles	Estimates		SE	t score	LL (95%)	UL (95%)	P value
$q = 0.1$	$\hat{\theta}_o$ (Intercept)	-0.3983	0.1504	-2.6479	-0.7339	-0.1417	0.0085
	$\hat{\theta}_1$ (BMI)	1.5509	0.3055	5.0767	0.9807	2.1831	0.0000
	$\hat{\theta}_2$ (IPAQ)	-0.5503	0.2366	-2.3260	-1.0769	-0.1456	0.0207
	$\hat{\theta}_3$ (ARMS)	0.1981	0.0445	4.4493	0.1174	0.2926	0.0000
$q = 0.25$	$\hat{\theta}_o$ (Intercept)	-0.6946	0.1024	-6.7831	-0.8453	-0.4422	0.0000
	$\hat{\theta}_1$ (BMI)	2.5972	0.2633	9.8631	2.0383	3.0747	0.0000
	$\hat{\theta}_2$ (IPAQ)	-1.9125	0.2261	-8.4590	-2.3388	-1.4489	0.0000
	$\hat{\theta}_3$ (ARMS)	0.3846	0.0707	5.4364	0.1957	0.4742	0.0000
$q = 0.5$	$\hat{\theta}_o$ (Intercept)	-0.6946	0.1024	-6.7831	-0.8453	-0.4422	0.0000
	$\hat{\theta}_1$ (BMI)	2.5972	0.2633	9.8631	2.0383	3.0747	0.0000
	$\hat{\theta}_2$ (IPAQ)	-1.9125	0.2261	-8.4590	-2.3388	-1.4489	0.0000
	$\hat{\theta}_3$ (ARMS)	0.3846	0.0707	5.4364	0.1957	0.4742	0.0000
$q = 0.75$	$\hat{\theta}_o$ (Intercept)	-0.6946	0.1024	-6.7831	-0.8453	-0.4422	0.0000
	$\hat{\theta}_1$ (BMI)	2.5972	0.2633	9.8631	2.0383	3.0747	0.0000
	$\hat{\theta}_2$ (IPAQ)	-1.9125	0.2261	-8.4590	-2.3388	-1.4489	0.0000
	$\hat{\theta}_3$ (ARMS)	0.3846	0.0707	5.4364	0.1957	0.4742	0.0000
$q = 0.9$	$\hat{\theta}_o$ (Intercept)	-0.6946	0.1024	-6.7831	-0.8453	-0.4422	0.0000
	$\hat{\theta}_1$ (BMI)	2.5972	0.2633	9.8631	2.0383	3.0747	0.0000
	$\hat{\theta}_2$ (IPAQ)	-1.9125	0.2261	-8.4590	-2.3388	-1.4489	0.0000
	$\hat{\theta}_3$ (ARMS)	0.3846	0.0707	5.4364	0.1957	0.4742	0.0000

Conclusion

In this research the bounded exponentiated exponential distribution is proposed for unit interval data sets. The shape of BEE distribution is symmetric, left skewed, right skewed, reversed J shape. The HRF shows the bathtub and increasing shapes. These variety of shapes makes the BEE distribution a suitable model for modeling data sets that exhibit such traits. Few properties of the BEE distribution, including cdf, quantile function, median, moments, inequality measures, reliability measures and order statistics have been studied. Six estimations methods to estimate the parameters of the BEE distribution have been discussed. To check the performance of the estimators a Monte Carlo simulation had been done. A bivariate extension of the BEE distribution has also been discussed, and only pdf and cdf for it has been shown with their graphs as well. The applications of the BEE distribution have been shown on the three datasets on the mortality rates and recovery rates of COVID-19, in UK, Canada and Spain. The three data sets revealed that the proposed distribution (BEED) performs better that the other competing distributions. Finally, a quantile regression model for studying the relationship between the conditional quantiles of a bounded response variable and a set of covariates is proposed. The survival and hazard functions of the BEEDQRM have been derived. The graphical shapes for the pdf, cdf and HRF for the BEEDQRM have been shown. An application for quantile regression is presented to show the applicability of BEED quantile regression.

Reference

Abouammoh, A. M., Alshingiti, & Arwa M. (2009). Reliability estimation of generalized inverted exponential distribution. *Journal of Statistical Computation and Simulation*, *79(11),* 1301–1315.

Afify, A.Z., Nassar, M., Kumar, D., & Cordeiro, G.M. (2022). A new unit distribution: Properties and applications. *Electron. J. Appl. Stat. Anal., 15,* 460–484.

Bantan, R. Jamal, F., Chesneau, C., & Elgarhy, M. (2021). Theory and Applications of the Unit Gamma/Gompertz Distribution. *Mathematics, 9*, 1850.

Ghitany, M.E., Mazucheli, J., Menezes, A.F.B., & Alqallaf, F. (2018). The unit-inverse Gaussian distribution: A new alternative to two parameter distributions on the unit interval. *Commun. Stat.-Theory Methods,* 48, 3423–3438.

Ghosh, I., Dey, S., & Kumar, D. (2019). Bounded M-O Extended Exponential Distribution with Applications. *Stochastics and Quality Control., 34,* 35–51.

Haq, M.A.U., Hashmi, S., Aidi, K., Ramos, P.L., & Louzada, F. (2020). Unit Modified Burr-III Distribution: Estimation, Characterizations and Validation Test. *Ann. Data Sci., 99,* 1– 26.

Jodrá, P. A. (2020). Bounded distribution derived from the shifted Gompertz law. *J. King Saud Univ.-Sci., 32,* 523–536.

Korkmaz, M.C. (2019). A new heavy-tailed distribution defined on the bounded interval. *J. Appl. Stat., 47,* 2097–2119.

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

Mazucheli, J., Menezes, A. F. B., & Chakraborty, S. (2018), On the one parameter unit-Lindley distribution and its associated regression model for proportion data*. Journal of Applied Statistics., 46,* 700–714.

Mazucheli, J., Menezes, A. F. B., & Dey, S. (2019). Unit-Gompertz Distribution with Applications. *STATISTICA., 79(1),* 25-43.

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

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

Modi, K., & Gill, V. (2019). Unit Burr-III distribution with application. *Journal of Statistics and Management Systems., 23*, 579–592.

Nasiru, S., Abubakari, A.G., & Angbing, I.D. (2021). Bounded Odd Inverse Pareto Exponential Distribution: Properties, Estimation, and Regression. Int. *J. Math. Math. Sci.,* 9955657.

Nasiru, S., Abubakari, A.G., & Chesneau, C. (2021). New Lifetime Distribution for Modeling Data on the Unit Interval: Properties, Applications and Quantile Regression. *Math. Comput. Appl., 27*, 105.

Petterle, R.R., Bonat, W.H., Scarpin, C.T., Jonasson, T., and Borba, V.Z.C. (2020). Multivariate quasi-beta regression models for continuous bounded data. *Int. J. Biostat., 17,* 39–53.