Research article

A novel quantile regression for fractiles based on unit logistic exponential distribution

Hanan Haj Ahmad[1,2,*] and Kariema A. Elnagar[3]

[1] Department of Basic Science, The General Administration of Preparatory Year, King Faisal University, Hofuf 31982, Al-Ahsa, Saudi Arabia

[2] Department of Mathematics and Statistics, College of Science, King Faisal University, Hofuf 31982, Al-Ahsa, Saudi Arabia

[3] Basic Sciences Department, Misr Higher Institute of Commerce and Computers, Mansoura, Egypt

* **Correspondence:** Email: hhajahmed@kfu.edu.sa.

Abstract: Continuous developments in unit interval distributions have shown effectiveness in modeling proportional data. However, challenges persist in diverse dispersion characteristics in real-world scenarios. This study introduces the unit logistic-exponential (ULE) distribution, a flexible probability model built upon the logistic-exponential distribution and designed for data confined to the unit interval. The statistical properties of the ULE distribution were studied, and parameter estimation through maximum likelihood estimation, Bayesian methods, maximum product spacings, and least squares estimates were conducted. A thorough simulation analysis using numerical techniques such as the quasi-Newton method and Markov chain Monte Carlo highlights the performance of the estimation methods, emphasizing their accuracy and reliability. The study reveals that the ULE distribution, paired with tools like randomized quantile and Cox-Snell residuals, provides robust assessments of goodness of fit, making it well-suited for real-world applications. Key findings demonstrate that the unit logistic-exponential distribution captures diverse data patterns effectively and improves reliability assessment in practical contexts. When applied to two real-world datasets—one from the medical field and the other from the economic sector—the ULE distribution consistently outperforms existing unit interval models, showcasing lower error rates and enhanced flexibility in tail behavior. These results underline the distribution's potential impact in areas requiring precise proportions modeling, ultimately supporting better decision-making and predictive analyses.

Keywords: logistic exponential distribution; quantile regression model; maximum likelihood estimation; Bayes estimation; Markov Chain Monte Carlo; randomized quantile residual; Cox-Snell residual
Mathematics Subject Classification: 60E05, 62E10, 62E15

1. Introduction

Unit distributions can be obtained using various transformations of well-known continuous distributions, including the negative exponential function transformation. Unit distributions give more flexibility to the original distribution along the unit interval without adding new parameters. Several unit distributions have been used to model data for percentages in many areas such as biology, mortality, recovery rates, economics, health, and reliability analysis, among others. New distributions have been proposed on the unit interval, such as the unit generalized half normal by Korkmaz [1], unit inverse Gaussian by Ghitany et al. [2], unit Gamma distribution by Consul and Jain [3], unit Weibull distribution by Mazucheli et al. [4], unit Gompertz distribution by Mazucheli et al. [5], unit Omega distribution by Abd El-Monsef et al. [6], unit Burr-XII distribution by Korkmaz and Chesneau [7], the unit-half normal distribution by Bakouch et al. [8], and unit exponential Pareto distribution by Haj Ahmad et al. [9]. The unit-exponentiated half-logistic distribution, and the bounded power Lomax were introduced by Hassan et al. ([10] and [11], respectively). Fayomi et al. [12] and [13] studied the unit-exponentiated Lomax distribution and the unit–power Burr X distribution.

The exponential distribution is widely utilized for modeling real-life data, primarily due to its memoryless feature and analytical simplicity. However, its utility is somewhat constrained because of its constant hazard rate and decreasing density function. To address these limitations and enhance its flexibility, numerous researchers have developed various modifications of the exponential distribution; for example, Hassan et al. [14] introduced a new four-parameter extended exponential distribution. Recent extensions include the exponentiated exponential [15], beta exponential [16], transmuted generalized exponential [17], Kumaraswamy transmuted-G family of distributions by [18]. A new three-parameter extension of the exponential distribution was introduced by [19]; also, [20] studied a new method for generating distributions with an application to the exponential distribution. Alpha power exponential distribution was discussed by [21], and extended exponential distribution and its applications were discussed by [22]. The Marshall-Olkin alpha power family of distributions was studied by [23], the new three-parameter exponential distribution by [24], and finally, Topp-leone exponential distribution was studied by [25].

Lan and Leemis created a new generalized exponential distribution [26]. It was called the logistic-exponential (LE) distribution with explicit-form density and distribution functions. LE is a distribution with two parameters that possess constant, increasing, decreasing, bathtub, and upside-down bathtub failure rate shapes. The only disadvantage of this distribution is that it does not yield explicit-form expressions for the moments, so it must be computed numerically using software like Mathematica, R, or Matlab.

For the LE distribution, the cumulative distribution function (cdf) is given by

$$F(x) = 1 - (1 + (e^{\beta x} - 1)^{\theta})^{-1}, x > 0, \tag{1.1}$$

with scale parameter $\beta > 0$ and shape parameter $\theta > 0$. The probability density function (pdf) is given by

$$f(x) = \frac{\theta \beta e^{\beta x}(e^{\beta x} - 1)^{\theta-1}}{(1 + (e^{\beta x} - 1)^{\theta})^2}, x > 0. \tag{1.2}$$

This paper focuses on a unique special case of the LE distribution, known as the unit-logistic-exponential distribution, confined to the unit interval. The ULE distribution is particularly valuable

for its capability to model long-tailed characteristics observed in many real-world datasets. Beyond its utility in capturing long-tailed behavior, the ULE distribution also plays a crucial role in uncovering patterns and trends within the data. By examining the distribution's shape, one can derive insights into the underlying factors influencing the data, which can guide resource allocation decisions or improve the accuracy of future predictions.

Traditional regression methods typically assess differences in outcome variables between populations by focusing on the mean (such as in ordinary least squares regression) or by evaluating a population's average effect (as seen in logistic regression models), after accounting for other explanatory variables. Quantile regression, however, offers the flexibility to examine how the slopes of the regression line vary across different quantiles of the data distribution. For instance, while the median line might remain stable, the 90th quantile prediction line may show a significant upward trend, whereas the 10th quantile prediction line could indicate a notable downward trend.

The primary objective of the research is to explore the new ULE distribution, examine its statistical characteristics, and demonstrate its effectiveness by implementing it in two real-world data examples. The motivations for studying the ULE distribution are as follows: (i) it is a unique model because it is defined on the unit interval [0, 1] instead of the positive real numbers; (ii) it exhibits significant flexibility in tail behavior, making it useful in risk assessment with relatively better outcomes; and (iii) the strength of this new distribution lies in its ability to model and fit various real-world data with lower error rates compared to other competing models.

This new model demonstrates its effectiveness in fitting the recovery rate for the CD34+ cells, which is a crucial indicator of the sufficiency of peripheral blood stem cells. A more recent purging technique involves the positive selection of CD34+ cells. By focusing on positive CD34+ cell selection, the variability in antigen expression on tumor cells becomes irrelevant, allowing this method to be applied for purging across almost all tumor types, assuming that the tumor cells do not express the CD34+ antigen. The most advanced method currently available for the positive selection of CD34+ cells, enabling their clinical isolation with extremely high purity, is magnetic-activated cell sorting (MACS).

Beyond its applications in modeling the recovery rates, it has a high impact on modeling failure times of components in economic and engineering contexts. Cost-effectiveness analysis (CEA) offers a systematic framework for regulators to compare the quantified benefits of legislative or regulatory decisions against their associated costs. When applied correctly, CEA compels policymakers and regulators to rigorously quantify the health or environmental benefits of potential government actions aimed at reducing risks. It also provides clear metrics for decision-makers, facilitating comparisons among different strategies for addressing the same issue, such as mitigating risks to human health, public safety, or the environment.

The ULE model, while promising and flexible, has some limitations that should be acknowledged as follows: (a) The estimation of parameters in the ULE model, particularly using methods like Bayesian approaches and maximum product spacings, can be computationally intensive. It requires unlimited computational resources and expertise in advanced numerical methods. (b) Although the ULE model was derived from the LE distribution, finding certain statistical properties, such as moments, often requires numerical solutions, such as the quasi-Newton and Markov chain Monte Carlo techniques, that can increase the computational burden and may introduce numerical stability issues in certain scenarios. (c) The ULE model is specifically designed for data within the unit interval (0, 1), limiting its applicability to datasets that can be naturally transformed or confined to this range.

The remainder of this work is arranged as follows: In Section 2, the ULE distribution is presented and its properties are discussed. In Section 3, the maximum likelihood, Bayes, maximum product spacings, least squares estimate of θ and β, and the Fisher information matrix are formulated. In Section 4, simulation studies are introduced. In Section 5, a quantile regression model based on ULE distribution is obtained. In Section 6, the two datasets are studied. Finally, the concluding remarks are provided in Section 7.

2. Unit logistic exponential distribution

Suppose X to be a continuous random variable with LE distribution having two parameters θ and β, then the cdf of the random variable $Y = e^{-X}$ can be represented by

$$F_{ULE}(y) = (1 + (y^{-\beta} - 1)^{\theta})^{-1}, 0 < y < 1, \beta, \theta > 0. \tag{2.1}$$

The pdf for the ULE distribution is defined by

$$f_{ULE}(y) = \frac{\theta \beta y^{-(\beta+1)}(y^{-\beta} - 1)^{\theta-1}}{(1 + (y^{-\beta} - 1)^{\theta})^2}, 0 < y < 1. \tag{2.2}$$

The hazard rate function (HRF) of ULE distribution is as follows

$$h_{ULE}(y) = \frac{\theta \beta y^{-(\beta+1)}}{(y^{-\beta} - 1)(1 + (y^{-\beta} - 1)^{\theta})}. \tag{2.3}$$

More statistical properties and related functions of the ULE distribution are presented in the following subsections.

### 2.1. Shape of density function and hazard rate function

The first derivative of the pdf of ULE distribution is

$$\frac{\partial f_{ULE}(x)}{\partial x} = \frac{\beta \theta \left(x^{-\beta} - 1\right)^{\theta} \Psi_1(x)}{x^2 \left(x^{\beta} - 1\right)^2 \left(\left(x^{-\beta} - 1\right)^{\theta} + 1\right)^3},$$

where $\Psi_1(x) = -\beta\theta + (\beta\theta - 1)\left(x^{-\beta} - 1\right)^{\theta} + (\beta + 1)x^{\beta}\left(\left(x^{-\beta} - 1\right)^{\theta} + 1\right) - 1$.

The above equations require a numerical evaluation to obtain the local maximum, minimum, and inflection points.

Figure 1 illustrates various possible pdf shapes for specific values of the parameters θ and β. The pdf can exhibit different forms, such as decreasing (D), increasing (I), unimodal, bathtub-shaped, or increasing-decreasing-increasing, depending on the chosen parameter values. This shape versatility makes the ULE distribution well-suited for analyzing data within the unit interval. Additionally, the left panel of Figure 2 provides a visual representation of the cdf of the ULE distribution, as defined in Eq (2.1).
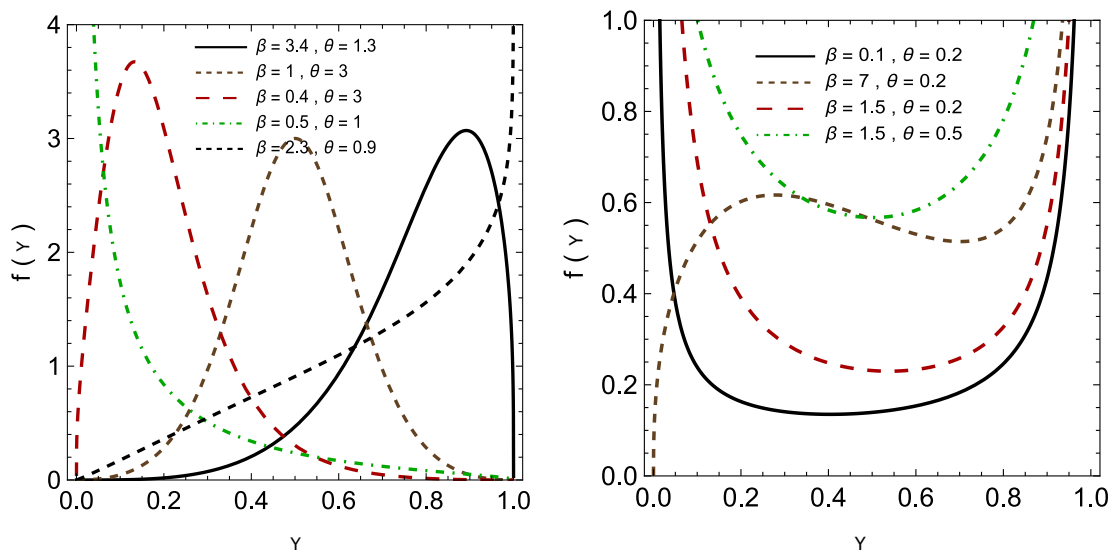
Figure 1. Plots of the ULE distribution for pdf for various parameters' values.
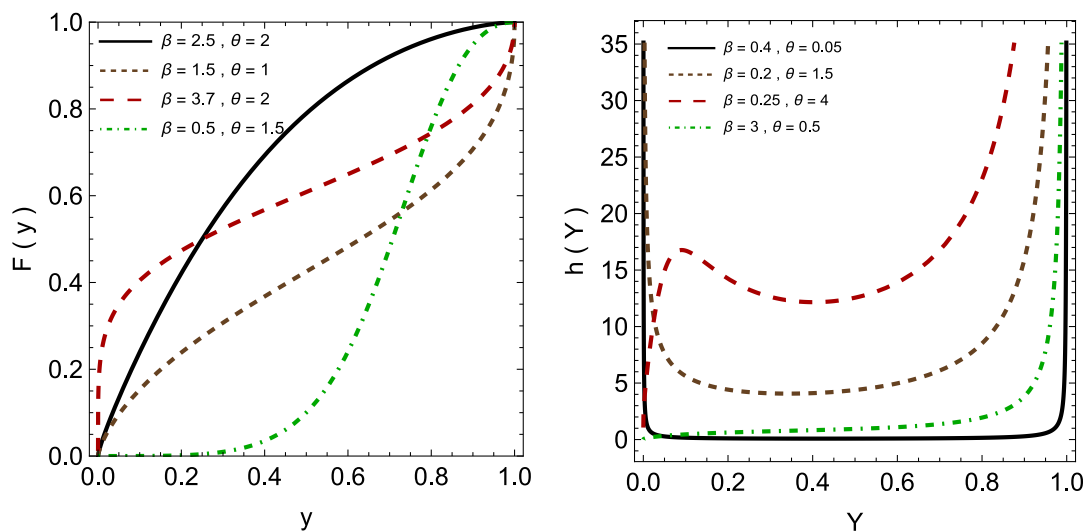
Figure 2. Plots of ULE distribution for cdf and HRF.

For the hazard rate function of ULE distribution, we have

$$\lim_{y \to 1^-} h_{ULE}(y) = \infty$$

and

$$\lim_{y \to 0^+} h_{ULE}(y) = 0 \quad \text{for} \quad \beta > 1; \beta\theta > 1.$$

The first derivative for $h_{ULE}(y)$ is given by

$$\frac{\partial\, h_{ULE}(y)}{\partial y} = \frac{\beta\theta\Psi_2(y)}{y^2 \left(y^\beta - 1\right)^2 \left(\left(y^{-\beta} - 1\right)^\theta + 1\right)^2},$$

where $\Psi_2(y) = (\beta\theta - 1)\left(y^{-\beta} - 1\right)^{\theta} + (\beta + 1)y^{\beta}\left(\left(y^{-\beta} - 1\right)^{\theta} + 1\right) - 1$. Clearly, the sign of $\frac{\partial\, h_{ULE}(y)}{\partial y}$ depends on the sign of $\Psi(y)$. The HRF (2.3) is increasing for $\beta > 1$ and $\beta\theta > 1$. Additionally, the right panel of Figure 2 shows that for different values of the parameters β and θ, the HRF for ULE can accommodate bathtub and IDI shapes.

2.2. Quantile function and moments

The τ_{th} quantile of the ULE distribution has a simple closed form and will be utilized later to define a new quantile regression model based on ULE distribution. The quantile regression model for ULE distribution is given by

$$Q(\tau) = \left[1 + (\frac{1}{\tau} - 1)^{\frac{1}{\theta}}\right]^{\frac{-1}{\beta}}. \tag{2.4}$$

A random variate can be generated via the inverse transformation method by

$$Y = \left[1 + (\frac{1}{U} - 1)^{\frac{1}{\theta}}\right]^{\frac{-1}{\beta}}, \quad \text{where} \quad U \sim Uniform(0, 1).$$

To analyze the effect of the parameters β and θ on the skewness and kurtosis of the ULE distribution, we check the trend of the Galton skewness (sk) and the Moors kurtosis (ku), as described in the following expressions:

$$sk = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{2}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)},$$

and

$$ku = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)},$$

where $Q(.)$ is the quantile function (2.4). Plots for both measures as functions of β and θ are given in Figure 3. It is observed that the skewness decreases with an increase of β. For $\beta = 1$, the ULE distribution tends to be symmetric, and for $0 < \beta < 1$, it is positively skewed and decreases toward 0 as θ increases. Conversely, for $\beta > 1$, the ULE distribution has negative skewness, which increases toward 0 as θ increases. Additionally, the kurtosis increases with θ, while it initially decreases and then increases as β increases.
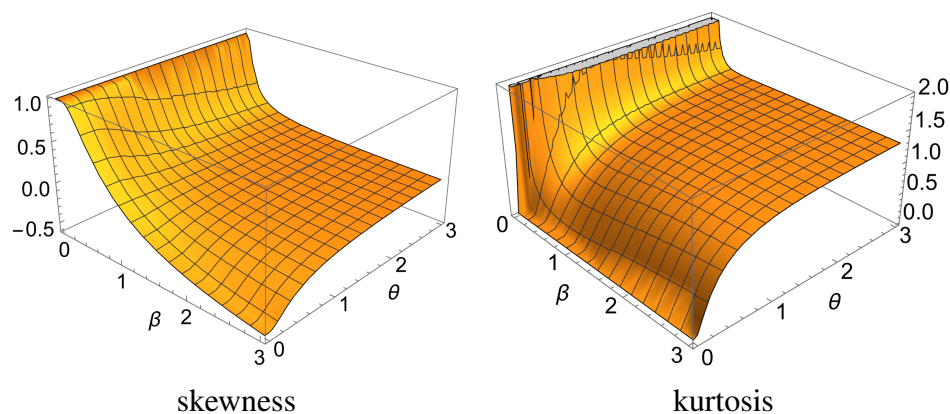
Figure 3. Plots for skewness and kurtosis of the ULE distribution.

As stated by [26], the moment for the logistic exponential distribution exists but cannot be expressed in closed form. Also, the ULE distribution does not have a closed-form expression for its moment and must be computed numerically via

$$\mu'_r = E(y^r) = \int_0^1 y^r f_{ULE}(y)dy,$$

where μ'_r denotes the r_{th} moment for ULE distribution.

2.3. Order statistics

When simulating different lifetime systems with some component structures, order statistics, also called ordered random variables, must be considered. Here, along with their typical distributional properties, the order statistics of the ULE distribution are presented. Let $y_1, y_2, ..., y_n$ denote a random sample of size n from the ULE having cdf $F(y)$ and pdf $f(y)$. Then, the order statistics are denoted by, $Y_{(1)} \leq Y_{(2)} \leq ... \leq Y_{(n)}$, $Y_{(1)} = min(Y_1, Y_2, ..., Y_n)$, and $Y_{(n)} = max(Y_1, Y_2, ..., Y_n)$. The kth order statistic's probability density function is given by

$$f_{Y_{(r)}}(y) = \frac{1}{\beta(r, n-r+1)} f(y)[F(y)]^{r-1}[1-F(y)]^{n-r}$$

$$= \frac{1}{\beta(r, n-r+1)} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} f(y)[F(y)]^{j+r-1}$$

$$= \frac{\theta\beta}{\beta(r, n-r+1)} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} y^{-(\beta+1)}(y^{-\beta}-1)^{\theta-1}(1+(y^{-\beta}-1)^\theta)^{-j-r-1},$$

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

In particular, the pdf of Y_1, can be expressed as

$$f_{Y_{(1)}}(y) = \frac{\theta\beta}{\beta(1, n)} \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} y^{-(\beta+1)}(y^{-\beta}-1)^{\theta-1}(1+(y^{-\beta}-1)^\theta)^{-j-2}.$$

Also, the pdf of Y_n, can be expressed as

$$f_{Y_{(n)}}(y) = \frac{\theta\beta}{\beta(n, 1)} y^{-(\beta+1)}(y^{-\beta}-1)^{\theta-1}(1+(y^{-\beta}-1)^\theta)^{-n-1}.$$

2.4. Stochastic ordering

Let Y_1 and Y_2 be two random variables with pdfs $f_1(y)$ and $f_2(y)$, respectively. Let us assume that Y_1 is stochastically less than Y_2 in terms of the likelihood ratio order ($Y_1 \leq_{lr} Y_2$) if the ratio $f_2(y)/f_1(y)$ is non-decreasing concerning y.

The ULE distribution has a stochastic ordering as illustrated in the next preposition.

Preposition 2.1. *Let Y_1 and Y_2 be two random variables, such that $Y_1 \sim ULE(\theta_1, \beta)$ and $Y_2 \sim ULE(\theta_2, \beta)$. If $\theta_1 \geq \theta_2$, then $Y_1 \leq_{lr} Y_2$.*

Proof. The proof is straightforward by taking the first derivative of $f_2(u; \theta_2, \beta)/f_1(y; \theta_1, \beta)$. □

Likelihood ratio order implies a hazard rate order ($Y_1 \leq_{hr} Y_2$), which in turn implies the usual stochastic order ($Y_1 \leq_{st} Y_2$); for further details on stochastic orders, see [27].

3. Methods of estimation

In this research, four estimation methods—namely maximum likelihood estimation, Bayesian estimation, maximum product spacings, and the least squares estimate—are employed to ensure robust and comprehensive parameter estimation. Each method offers unique statistical properties and benefits, contributing to a whole evaluation of the model's performance. The maximum likelihood estimation is a widely used method known for its asymptotic efficiency and consistency, producing parameter estimates that maximize the likelihood of observing the given data. This makes it particularly useful when seeking estimates with desirable large-sample properties. Bayesian estimation, on the other hand, incorporates prior information along with the observed data, providing a flexible framework for parameter inference and enabling the calculation of credible intervals that offer a probabilistic interpretation of parameter uncertainty. The maximum product spacings method is used as an alternative to the maximum likelihood method, particularly effective when dealing with datasets that include extreme or non-uniformly distributed observations. It often provides more stable estimates in cases where traditional likelihood methods may face convergence issues. The least squares estimation method is selected for its simplicity and its ability to minimize the sum of squared differences between observed and predicted values, making it valuable for reducing overall estimation error.

3.1. Maximum likelihood estimation

The maximum likelihood (ML) estimation method is the most well-known classical inference in statistics. The ML estimates are obtained based on maximizing the log-likelihood function of the ULE distribution. Variance-covariance matrix of the unknown population parameters is utilized to construct the asymptotic confidence intervals for θ and β. Numerical methods are used to determine the required estimators, specifically adopting the well-known quasi-Newton method.

Suppose $y_1, y_2, ..., y_n$ represent a random sample of size n with ULE distribution. The likelihood function for ULE distribution is given as

$$L(\theta, \beta) = \prod_{i=}^{n} \frac{\theta \beta y_i^{-(\beta+1)} (y_i^{-\beta} - 1)^{\theta-1}}{(1 + (y^{-\beta} - 1)^\theta)^2} = \theta^n \beta^n \prod_{i=1}^{n} y_i^{-(\beta+1)} (y_i^{-\beta} - 1)^{\theta-1} \left[1 + (y_i^{-\beta} - 1)^\theta\right]^{-2}, \quad (3.1)$$

and log-likelihood function of the random sample as

$$lnL(\theta, \beta) = nln(\theta) + nln(\beta) - (\beta + 1) \sum_{i=1}^{n} ln(y_i) + (\theta - 1) \sum_{i=1}^{n} ln(y_i^{-\beta} - 1) - 2 \sum_{i=1}^{n} ln\left[1 + (y_i^{-\beta} - 1)^\theta\right]. \quad (3.2)$$

Taking the partial derivatives of Eq (3.2) with respect to β and θ, the ML estimates ($\hat{\beta}$ and $\hat{\theta}$) are obtained by solving the following nonlinear systems of equations, as follows:

$$\frac{\partial}{\partial \beta} \ln L(\theta, \beta) = \frac{n}{\beta} - \sum_{i=1}^{n} ln y_i - (\theta - 1) \sum_{i=1}^{n} \frac{y_i^{-\beta} ln(y_i)}{(y_i^{-\beta} - 1)} + 2 \sum_{i=1}^{n} \frac{\theta(y_i^{-\beta} - 1)^{\theta-1} y_i^{-\beta} ln(y_i)}{\left[1 + (y_i^{-\beta} - 1)^\theta\right]} \quad (3.3)$$

and

$$\frac{\partial}{\partial \theta} \ln L(\theta, \beta) = \frac{n}{\theta} + \sum_{i=1}^{n} ln(y_i^{-\beta} - 1) - 2 \sum_{i=1}^{n} \frac{(y_i^{-\beta} - 1)^\theta ln(y_i^{-\beta} - 1)}{\left[1 + (y_i^{-\beta} - 1)^\theta\right]}. \quad (3.4)$$

The log-likelihood equations defined in Eqs (3.3) and (3.4) do not provide a closed-form solution. Since obtaining a solution is challenging, it is often more practical to employ nonlinear optimization techniques, such as the quasi-Newton algorithm, to numerically maximize the log-likelihood function. Additionally, explicit confidence intervals for these parameters cannot be directly constructed, necessitating the use of approximate methods to determine the confidence intervals for θ and β. To achieve this, we first calculate the second-order partial derivatives, which are essential for deriving the Fisher information matrix.

So, the Fisher information matrix is given by

$$I^{-1} = \begin{bmatrix} \frac{-\partial^2}{\partial\theta^2}lnL(\theta,\beta) & \frac{-\partial^2}{\partial\theta\partial\beta}lnL(\theta,\beta) \\ \frac{-\partial^2}{\partial\beta\partial\theta}lnL(\theta,\beta) & \frac{-\partial^2}{\partial\beta^2}lnL(\theta,\beta) \end{bmatrix}^{-1} = \begin{bmatrix} var(\hat{\theta}) & cov(\hat{\theta},\hat{\beta}) \\ cov(\hat{\beta},\hat{\theta}) & var(\hat{\beta}) \end{bmatrix}^{-1}, \tag{3.5}$$

where

$$\frac{\partial^2}{\partial\theta^2}ln(L(\theta,\beta)) = \frac{-n}{\alpha^2} - 2\sum_{i=1}^{n} \frac{(ln(y_i^{-\beta} - 1))^2(y_i^{-\beta} - 1)}{\left[1 + (y_i^{-\beta} - 1)^\theta\right]^2},$$

$$\frac{\partial^2}{\partial\theta\partial\beta}ln(L(\theta,\beta)) = \frac{\partial^2}{\partial\lambda\partial\theta}ln(L)$$

$$= -\sum_{i=1}^{n} \frac{y_i^{-\beta}ln(y_i)}{(y_i^{-\beta} - 1)} - 2\sum_{i=1}^{n}(y_i^{-\beta})^2 ln(y_i)\frac{(y_i^{-\beta} - 1)^\theta\left[1 + (y_i^{-\beta} - 1) + \theta ln(y_i^{-\beta} - 1))^\theta\right]}{(y_i^{-\beta} - 1)(1 + (y_i^{-\beta} - 1)^\theta)^2},$$

and

$$\frac{\partial^2}{\partial\beta^2}ln(L(\theta,\beta)) = \frac{-n}{\beta^2} + \beta(\theta - 1)\sum_{i=1}^{n} ln(y_i)\left[\frac{y_i^{-\beta-1}}{y_i^{-\beta} - 1} + \frac{y_i^{-2\beta-1}}{(y_i^{-\beta} - 1)^2}\right]$$

$$+ 2\sum_{i=1}^{n}\theta(ln(y_i))^2\frac{(y_i^{-\beta} - 1)^\theta\left[y_i^{-\beta}(1 + (y_i^{-\beta} - 1)^\theta) - \theta\right]}{(y_i^{-\beta} - 1)(1 + (y_i^{-\beta} - 1)^\theta)^2}.$$

The $(1 - \zeta)100\%$ confidence interval for the parameters β and θ can be presented as

$$(\hat{\theta}_L, \hat{\theta}_U) = \hat{\theta} \pm z_{1-\frac{\zeta}{2}}\sqrt{var(\hat{\theta})} \quad and \quad (\hat{\beta}_L, \hat{\beta}_U) = \hat{\beta} \pm z_{1-\frac{\zeta}{2}}\sqrt{var(\hat{\beta})},$$

where $\hat{\theta}$ and $\hat{\beta}$ are the maximum likelihood estimators of θ and β, $z_{1-\frac{\zeta}{2}}$ is the percent of the standard normal distribution, and $var(\hat{\theta}), var(\hat{\beta})$ are the asymptotic variances of ML computed utilizing the inverse for the information matrix.

3.2. Bayes estimation

This section addresses the challenge of deriving Bayes estimators (BE) for the shape and scale parameters of the ULE distribution. Depending on the information known about the parameter, one can choose between using informative or non- informative priors. However, when sample information about the parameters is available, it is preferable to utilize an informative prior. The Gamma prior is

preferred in many Bayesian analyses because it is mathematically convenient, conceptually attractive, and computationally friendly. Its flexibility and conjugacy properties make it a powerful tool for representing prior knowledge and updating beliefs in light of new data; see Zellner et al. [28]. The following Gamma prior distributions are applied when θ and β are independent parameters.

$$\pi_1(\theta) \propto \theta^{a_1-1}exp(-b_1\theta), \theta > 0$$

and

$$\pi_2(\beta) \propto \beta^{a_2-1}exp(-b_2\beta), \beta > 0,$$

where a_i and b_i, $i = 1, 2$ are supposed to be known hyper-parameters and selected to reflect the prior distribution of the unknown parameters. Thus, we propose to use piecewise independent Gamma priors for both the shape and scale parameters for the ULE distribution because the Gamma distribution is very flexible; see Dey et al. [29] and Kundu and Howlader [30]. The joint prior distribution for θ and β is

$$\pi(\theta, \beta) = \theta^{a_1-1}\beta^{a_2-1}exp(-b_1\theta - b_2\beta). \tag{3.6}$$

The posterior distribution is derived from the likelihood function Eq (3.1) and the prior distribution Eq (3.6). The joint posterior distribution of θ and β is denoted as

$$\pi^*(\theta, \beta \mid \underline{y}) = \theta^{a_1-1}\beta^{a_2-1}exp(-b_1\theta - b_2\beta) \prod_{i=0}^{n} y_i^{-(\beta+1)}(y_i^{-\beta})^{\theta-1} \left[1 + (y_i^{-\beta} - 1)^\theta\right]^{-2}.$$

The conditional posterior densities of θ and β is presented as

$$\pi_1^*(\theta \mid \beta, \underline{y}) = \theta^{a_1-1}exp(-b_1\theta) \prod_{i=0}^{n} (y_i^{-\beta})^{\theta-1} \left[1 + (y_i^{-\beta} - 1)^\theta\right]^{-2} \tag{3.7}$$

and

$$\pi_2^*(\beta \mid \theta, \underline{y}) = \beta^{a_2-1}exp(-b_2\beta) \prod_{i=0}^{n} y_i^{-(\beta+1)}(y_i^{-\beta})^{\theta-1} \left[1 + (y_i^{-\beta} - 1)^\theta\right]^{-2}. \tag{3.8}$$

The Metropolis-Hastings sampler is necessary for implementing the MCMC technique because the conditional posteriors of θ and β in the previous equations do not conform to any standard distribution. Tierney [31] introduced the Metropolis-Hastings (M-H) algorithm within Gibbs sampling to generate posterior samples, and the process is outlined as follows:

1) Suggest initial values to be $(\theta^{(0)}, \beta^{(0)})$.
2) Let $j = 1$.
3) Using the M-H algorithm, create $\theta^{(j)}$ and $\beta^{(j)}$ by using M-H algorithm below using Eqs (3.7) and (3.8) with the normal distribution $N(\theta^{j-1}, var(\theta))$ and $N(\beta^{j-1}, var(\beta))$ and the inverse Fisher information matrix is used to calculate $var(\theta)$ and $var(\beta)$.

   (a) We set a $N(\theta^{j-1}, var(\theta))$ and $N(\beta^{j-1}, var(\beta))$ as a proposal distribution for θ^* and β^* respectively.

(b) The acceptance rule is defined by the probabilities ρ_θ and ρ_β, which compare the proposed samples against the current samples. We use the acceptance probability:

$$\rho_\theta = min\left[1, \frac{\pi_1^*(\theta^* \mid \beta^{j-1}, x)}{\pi_1^*(\alpha^{j-1} \mid \beta^{j-1}, x)}\right],$$

$$\rho_\beta = min\left[1, \frac{\pi_2^*(\beta^* \mid \alpha^j, , x)}{\pi_2^*(\beta^{j-1} \mid \alpha^j, x)}\right].$$

(c) Set u_1 and u_2 from a uniform $(0, 1)$ distribution.

(d) If $u_1 < \rho_\theta$, the request is approved, so make a decision $\theta^j = \theta^*$ or else specify $\theta^j = \theta^{j-1}$.

(e) If $u_2 < \rho_\beta$, the request is approved, so make a decision $\beta^j = \beta^*$ or else specify $\beta^j = \beta^{j-1}$.

4) Calculating $\theta^{(j)}$ and $\beta^{(j)}$.

5) Set $j = j + 1$.

6) Steps from 3 to 5 must be repeated N times.

7) To find the credible intervals (CRIs) for α and β for ψ_k^j, $j = 1, 2, .., N$, $k = 1, .., 4$ and $(\psi_1, \psi_2) = (\theta, \beta)$ as $\psi_k^1 < \psi_k^2 < ... < \psi_k^N$ then the $(1 - \gamma)100\%$ CRIs of ψ_k is

$$(\psi_k(\frac{\gamma}{2}(N - M)), \psi_k((1 - \frac{\gamma}{2})(N - M))).$$

For practical application, we employ a burn-in period to allow the Markov chain to stabilize. Additionally, thinning reduces autocorrelation in the sampled chains, ensuring that the final posterior estimates represent independent samples. Specifically, after an initial burn-in of M samples, we retain every $k - th$ sample to construct the posterior distribution, as indicated by $\psi_k^{(j)}$, $j = M + 1, ..., N$, for sufficiently large N.

The variances of the proposal distributions, $var(\theta)$ and $var(\beta)$, are adjusted based on pilot runs to reach an acceptance rate that optimizes the performance of the M-H algorithm. Typically, an acceptance rate between 20% and 30% is targeted to balance convergence speed and sampling efficiency [32, 33].

In Bayesian analysis, there are four basic elements: the data, the model, the prior, and the loss function. A good Bayesian estimator must minimize the loss function. So, in practical studies, using loss functions in Bayesian inference is important for several reasons, namely: (a) it helps in making decisions that are optimal under uncertainty, and (b) it quantifies the consequences of making incorrect inferences or decisions. By trying to minimize the error itself and the expected cost of errors, which is often more relevant in practice, (c) it can lead to more robust statistical methods. Several authors discussed many loss functions in Bayesian inference, while the squared error loss function was widely used.

The squared error loss (SEL) function is used to get the approximate Bayes estimates of $\hat{\theta}$ and $\hat{\beta}$, being defined as $L\left(\hat{\phi} - \phi\right) = \left(\hat{\phi} - \phi\right)^2$. The SEL function is symmetric, meaning that it gives equal weights to both over and under-estimation. In real life, we encounter many situations where over-estimation may be more serious than under-estimation, or vice versa. The Bayesian estimation under the SEL is given as:

$$\hat{\theta}_{BS} = \frac{1}{N - M} \sum_{j=M+1}^{N} \theta^{(j)},$$

and

$$\hat{\beta}_{BS} = \frac{1}{N - M} \sum_{j=M+1}^{N} \beta^{(j)}.$$

3.3. Maximum product spacings

Cheng and Amin [34] and Ranneby [35], introduced the maximum product of spacings (MPS) estimation approach as an alternative ML estimation. They examined the properties of MPS estimators and demonstrated that in certain situations where ML estimators fail to provide consistent and asymptotically efficient estimates, MPS succeeds. These situations include instances where the likelihood function is unbounded from above, heavy-tailed distributions with unspecified scale and location parameters [36] and mixture distributions. Consequently, the MPS method overcomes the limitations of the ML method while preserving nearly all of its desirable properties in large sample sizes [37]. Assume the parameter vector $\upsilon = (\theta, \beta)$, then the MPS function is presented as

$$MPS(\upsilon) = \frac{1}{n+1} \sum_{i=1}^{n+1} log[F(y_{(i)}, \theta, \beta) - F(y_{(i-1)}, \theta, \beta)].$$

By maximizing $MPS(\upsilon)$ concerning υ, one can obtain MPS estimates. The following non-linear equations' simultaneous solutions are also provided for them:

$$\frac{\partial MPS(\upsilon)}{\partial \theta} = \frac{1}{n+1} \sum_{i=1}^{n+1} \left[\frac{F(y_i, \theta, \beta)'^{;\theta} - F(y_{i-1}, \theta, \beta)'^{;\theta}}{F(y_{(i)}\theta, \beta) - F(y_{(i-1)}\theta, \beta)} \right] = 0$$

and

$$\frac{\partial MPS(\upsilon)}{\partial \beta} = \frac{1}{n+1} \sum_{i=1}^{n+1} \left[\frac{F(y_i, \theta, \beta)'^{;\beta} - F(y_{i-1}, \theta, \beta)'^{;\beta}}{F(y_{(i)}\theta, \beta) - F(y_{(i-1)}\theta, \beta)} \right] = 0,$$

where

$$F(y_i, \theta, \beta)'^{;\theta} = -(y_i^{-\beta} - 1)^{\theta} ln(y_i^{-\beta} - 1)(1 + (y_i^{-\beta} - 1)^{\theta})^{-2} \tag{3.9}$$

and

$$F(y_i, \theta, \beta)'^{;\beta} = y_i^{-\beta} ln(y_i)(y_i^{-\beta} - 1)^{\theta-1}(1 + (y_i^{-\beta} - 1)^{\theta})^{-2}. \tag{3.10}$$

The asymptotic confidence intervals for the parameters β and θ, derived from the MPS estimation, rely on the established asymptotic equivalence between the ML and the MPS method, as noted by [34, 38, 39]. Thus, the $100(1 - \zeta)\%$ confidence interval for the parameters using MPS is formulated as follows:

$$\hat{\theta}_{MPS} \pm z_{1-\frac{\zeta}{2}} \sqrt{var(\hat{\theta}_{MPS})} \ and \qquad \hat{\beta}_{MPS} \pm z_{1-\frac{\zeta}{2}} \sqrt{var(\hat{\beta}_{MPS})},$$

where $\hat{\theta}_{MPS}$ and $\hat{\beta}_{MPS}$ are the MPS estimates of θ and β, $z_{1-\frac{\zeta}{2}}$ is the percent of the standard normal distribution, and $var(\hat{\theta}_{MPS}), var(\hat{\beta}_{MPS})$ are the asymptotic variances computed utilizing the inverse of the observed information matrix (3.5).

3.4. Least squares estimation

Least-squares methods produce the estimated parameters with the highest probability (maximum likelihood) of being true if several critical conditions are guaranteed. Least-squares parameter estimation is a basic procedure for evaluating the confidence intervals for the unknown model's parameters. It discusses the practical ways of applying least-squares techniques to experimental data.

The least square estimates (LSE) for $\hat{\theta}_{LSE}$ and $\hat{\beta}_{LSE}$ of θ and β, respectively, are observed by minimizing the function:

$$LSE(\upsilon) = \sum_{i=1}^{n}(F(y_i, \theta, \beta) - E[F(y_i, \theta, \beta)])^2,$$

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

$$\frac{\partial LSE(\upsilon)}{\partial \theta} = 2\sum_{i=1}^{n}F(y_i, \theta, \beta)^{';\theta}\left(F(y_i, \theta, \beta) - \frac{i}{(n+1)}\right) = 0$$

and

$$\frac{\partial LSE(\upsilon)}{\partial \beta} = 2\sum_{i=1}^{n}F(y_i, \theta, \beta)^{';\beta}\left(F(y_i, \theta, \beta) - \frac{i}{(n+1)}\right) = 0,$$

where $F(y_i, \theta, \beta)^{';\theta}$ and $F(y_i, \theta, \beta)^{';\beta}$ as mentioned in Eqs (3.9) and (3.10).

The bootstrap percentile method (PCI) is commonly used to construct confidence intervals. The following procedure describes the steps to construct PCI:

i) Calculate (LSE) $\hat{\beta}LSE$ and $\hat{\theta}LSE$ for the ULE distribution.
ii) Generate a bootstrap sample by using the estimates $(\hat{\beta}LSE, \hat{\theta}LSE)$, and then obtain the bootstrap estimate, denoted as $(\hat{\beta}b, \hat{\theta}b)$, based on the bootstrap sample.
iii) Repeat step (ii) B times to obtain the estimates $(\hat{\beta}b_1, \hat{\theta}b_1)$, $(\hat{\beta}b_2, \hat{\theta}b_2)$, ... $(\hat{\beta}b_B, \hat{\theta}b_B)$.
iv) Construct the $100(1-\zeta)\%$ bootstrap confidence intervals for the parameters β and θ by using the $\frac{\zeta}{2}$ and $1-\frac{\zeta}{2}$ quantiles of the empirical distribution of the bootstrap estimates for β and θ, respectively.

4. Simulation

The simulation study provides a comprehensive assessment of the estimation methods for estimating the parameters θ and β. By generating synthetic datasets with known parameter values, one can evaluate how well these estimation methods converge to the true values as the dataset size increases. Measures such as mean squared errors (MSEs) and average biases (ABs) are used to assess the convergence and accuracy of the point estimators. The performance of 95% confidence intervals is measured based on average widths (AW) and coverage probabilities (CP).

To generate observations, we used the ULE distribution and applied the inverse transformation method with various values of the parameters $(\theta, \beta) = (1, 3), (0.45, 0.5), (3, 0.4), (0.25, 1)$, and $(1.9, 0.8)$. These cases represent various trends of the pdf (2.2) for the ULE distribution. The simulation is redone 1000 times, with varying sample sizes of $n = 50, 100$, and 150. For each simulation iteration, we generate 1000 bootstrap samples and accumulate 10000 observations for the Markov

chain Monte Carlo (MCMC) algorithm. We used the first 1000 observations as burn-in to minimize initial distribution effects. Additionally, in the thinning process, every fifth observation is selected to reduce dependence and enhance the robustness of the analysis. The simulation results are presented in Tables 1–5. Mathematica scripts were used for all simulation results.

The findings from these tables regarding accuracy and biased-ness are summarized as follows:

- The MSE for all estimation methods decreases as the sample size increases from 50 to 150, indicating improved accuracy with larger sample sizes.
- The ABs generally reduce with larger sample sizes, suggesting that estimators become less biased.

For comparison of estimators from Tables 1–5, the following points are indicated:

- ML estimates show relatively low MSEs and consistent CPs, making it a reliable choice across sample sizes.
- MPSE has competitive MSEs and ABs, but its CPs for θ can be slightly lower, especially at smaller n.
- Bayesian estimation provides good MSEs and high CPs, often outperforming other methods in maintaining high coverage.
- The average width of the confidence intervals decreases as the sample size increases for all estimation techniques indicating more precise estimates with larger datasets. Also, it is observed that the approximate confidence interval (ACI) for MPS has smaller interval widths than the other estimation approaches.
- The coverage probabilities are close to the nominal level, showing that the confidence intervals produced by these methods effectively contain the true parameter values. It is observed that LSE is showing coverage probabilities higher than the nominal level.

Overall, Bayesian estimation offers strong performance with high CPs and reasonable AWs. ML estimation consistently performs well, with solid MSEs, ABs, and CPs, making it a dependable choice. While small-sample scenarios are not the primary focus of this study, we acknowledge their practical importance. Bootstrap or Bayesian approaches will be more appropriate for small-sample analyses rather than the asymptotic confidence intervals.

Table 1. Simulation results for $\beta = 1, \theta = 3$.

n	Est.	ML		MPSE		Bayesian		LSE	
		β	θ	β	θ	β	θ	β	θ
	MSEs	0.0037	0.148	0.0037	0.142	0.00349327	0.076	0.0037	0.127
50	ABs	0.0019	0.073	0.0024	-0.130	0.006	0.056	-0.002	-0.105
	AWs	0.228	1.441	0.239	1.352	0.229	1.28257	0.248	1.517
	CPs	0.936	0.942	0.945	0.88	0.947	0.973	0.96	0.93
	MSEs	0.0017	0.067	0.0016	0.071	0.0017	0.054	0.0019	0.063
100	ABs	0.00027	-0.0023	-0.0014	-0.103	0.0019	0.0077	0.007	-0.027
	AWs	0.1629	0.995	0.166	0.964	0.163	0.943	0.170	1.115
	CPs	0.95	0.942	0.955	0.893	0.953	0.953	0.925	0.965
	MSEs	0.0011	0.043	0.001	0.049	0.0012	0.035	0.001	0.048
150	ABs	0.0047	0.016	0.004	-0.090	0.008	-0.011	0.0001	0.003
	AWs	0.133	0.817	0.136	0.790	0.133	0.775481	0.135	0.925
	CPs	0.958	0.952	0.968	0.905	0.973	0.967	0.935	0.965

Table 2. Simulation output for $\beta = 0.45, \theta = 0.5$.

n	Est.	ML		MPSE		Bayesian		LSE	
		β	θ	β	θ	β	θ	β	θ
	MSEs	0.0108	0.005	0.011	0.005	0.027	0.0124	0.026	0.007
50	ABs	0.0217	0.009	0.020	-0.021	0.069	0.017	0.020	0.0006
	AWs	0.425	0.306	0.404	0.268	0.654	0.419	0.589	0.330
	CPs	0.976	0.956	0.973	0.918	0.973	0.973	0.91	0.93
	MSEs	0.0055	0.003	0.005	0.003	0.0134	0.007	0.011	0.004
100	ABs	0.009	-0.0004	0.004	-0.018	0.032	0.011	0.011	-0.003
	AWs	0.286	0.206	0.278	0.194	0.424	0.289	0.391	0.235
	CPs	0.942	0.946	0.94	0.896	0.953	0.947	0.93	0.95
	MSEs	0.0034	0.0017	0.003	0.002	0.005	0.0019	0.0057	0.0022
150	ABs	0.0121	-0.0003	0.002	-0.013	0.018	0.0005	0.013	-0.005
	AWs	0.231	0.166	0.226	0.160	0.236	0.168	0.317	0.191
	CPs	0.968	0.958	0.963	0.93	0.887	0.933	0.96	0.93

Table 3. Simulation output for $\beta = 3, \theta = 0.4$.

n	Est.	ML		MPSE		Bayesian		LSE	
		β	θ	β	θ	β	θ	β	θ
	MSEs	0.521	0.004	0.523	0.003	0.103	0.003	2.276	0.005
50	ABs	0.163	0.007	0.161	-0.018	0.035	0.012	0.242	-0.0009
	AWs	3.009	0.254	2.831	0.220	1.960	0.218	4.488	0.286
	CPs	0.976	0.954	0.973	0.905	1.	0.953	0.92	0.92
	MSEs	0.2443	0.002	0.264	0.002	0.138	0.0017	0.669	0.003
100	ABs	0.067	-0.0009	0.032	-0.014	0.023	0.007	0.097	-0.002
	AWs	2.006	0.170	1.934	0.159	1.588	0.158	2.981	0.202
	CPs	0.968	0.948	0.935	0.905	0.973	0.967	0.93	0.95
	MSEs	0.178	0.001	0.154	0.001	0.094	0.001	0.309	0.002
150	ABs	0.082	0.00017	0.017	-0.010	0.019	0.003	-0.005	0.005
	AWs	1.616	0.137	1.577	0.132	1.379	0.131	2.297	0.166
	CPs	0.95	0.946	0.965	0.925	0.967	0.94	0.95	0.95

Table 4. Simulation output for $\beta = 0.25, \theta = 1$.

n	Est.	ML		MPSE		Bayesian		LSE	
		β	θ	β	θ	β	θ	β	θ
	MSEs	0.0017	0.020	0.002	0.018	0.001	0.0133	0.002	0.021
50	ABs	0.006	0.024	0.007	-0.045	0.006	0.011	0.003	-0.009
	AWs	0.158	0.527	0.161	0.484	0.162	0.518	0.178	0.551
	CPs	0.956	0.944	0.958	0.903	0.973	0.96	0.91	0.93
	MSEs	0.0008	0.009	0.0007	0.010	0.0010	0.008	0.001	0.011
100	ABs	0.002	0.0004	-0.00004	-0.036	0.008	0.0001	0.002	-0.012
	AWs	0.110	0.362	0.110	0.347	0.113	0.360	0.123	0.395
	CPs	0.958	0.936	0.96	0.898	0.933	0.953	0.93	0.95
	MSEs	0.0006	0.006	0.0005	0.006	0.0005	0.007	0.0007	0.007
150	ABs	0.0045	0.003	0.0002	-0.026	0.0043	0.0011	0.0004	-0.013
	AWs	0.0899	0.295	0.090	0.287	0.090	0.295	0.098	0.319
	CPs	0.95	0.956	0.955	0.925	0.953	0.92	0.92	0.92

Table 5. Simulation results for $\beta = 1.9, \theta = 0.8$.

n	Est.	ML		MPSE		Bayesian		LSE	
		β	θ	β	θ	β	θ	β	θ
	MSEs	0.128	0.012	0.130	0.011	0.063	0.011	0.210	0.015
50	ABs	0.054	0.018	0.061	-0.036	0.020	0.021	0.037	-0.005
	AWs	1.397	0.438	1.411	0.400	1.226	0.420	1.672	0.461
	CPs	0.948	0.952	0.973	0.898	0.973	0.973	0.91	0.93
	MSEs	0.0609	0.006	0.059	0.007	0.051	0.006	0.098	0.008
100	ABs	0.022	-0.0007	0.005	-0.028	0.036	0.004	0.022	-0.008
	AWs	0.972	0.301	0.968	0.288	0.924	0.297	1.144	0.329
	CPs	0.952	0.938	0.948	0.91	0.947	0.953	0.93	0.96
	MSEs	0.040	0.004	0.039	0.004	0.036	0.004	0.049	0.005
150	ABs	0.0408	0.0007	0.003	-0.0201	0.050	0.004	0.033	-0.007
	AWs	0.795	0.245	0.785	0.238	0.771	0.242	0.929	0.271
	CPs	0.964	0.958	0.958	0.928	0.96	0.953	0.96	0.93

5. Quantile regression model for ULE distribution

This section proposes a new quantile regression model based on ULE distribution for component response variables, offering a preference for commonly used beta, Kumaraswamy, and unit-Weibull regression models. When incorporating covariate information into regression analysis of a probability distribution, it is common to connect the mean response to the covariates. In contrast to modeling the mean of the response, quantile regression models were pioneered by [40] to model the conditional quantiles of the response variable as a function of the covariates without any distributional assumptions on the error term, and these models are also considered robust regressions.

Since the mean of the ULE distribution does not have a tractable closed-form expression, quantile regression models are a solid substitute. The quantile function of the ULE distribution has a simple treatable form, allowing its cdf and pdf to be expressed using a re-parametrization established on the quantile function in Eq (2.4). Let $\mu = Q(\tau, \theta, \beta)$ and the parameter $\beta = -\dfrac{\log\left[\left(\frac{1}{\tau}-1\right)^{1/\theta}+1\right]}{\log(\mu)}$. Then, the cdf and pdf of the re-parametrized ULE distribution are given by

$$G(y, \mu, \theta) = \left[1 + \left(y^{\frac{\log\left[\left(\frac{1}{\tau}-1\right)^{1/\theta}+1\right]}{\log(\mu)}} - 1\right)^{\theta}\right]^{-1} \tag{5.1}$$

and

$$g(y,\mu,\theta) = \theta[\log(1/\mu)]^{-1}\log\left((1/\tau-1)^{1/\theta}+1\right)y^{\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu)}-1}$$
$$\times\left[y^{\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu)}}-1\right]^{\theta-1}\left[\left(y^{\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu)}}-1\right)^{\theta}+1\right]^{-2},\tag{5.2}$$

where $0 < y < 1, 0 < \mu < 1$ and $0 < \tau < 1$.

Given independent random variables $X_i; i = 1,\ldots,n$ from the re-parameterized ULE distribution with pdf in Eq (5.2), it is feasible to establish the ULE quantile regression. This regression assumes the functional relation of the median of Y_i

$$g(\mu_i) = \boldsymbol{\delta}^T\boldsymbol{x_i},$$

where $\boldsymbol{x_i} = (1, x_{1i}, x_{2i},\ldots, x_{ki})$ represents the covariates vector; also, $\boldsymbol{\delta} = (\delta_0, \delta_1,\ldots, \delta_k)^T$ is the coefficients vector in the regression model. Various selections for the link function $g(.)$ can be treated; however, in this context, we only consider the logit link function:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right).$$

Thus we have

$$\mu_i = \frac{e^{\boldsymbol{\delta}^T\boldsymbol{x_i}}}{1+e^{\boldsymbol{\delta}^T\boldsymbol{x_i}}},\ i = 1,\ldots,n.\tag{5.3}$$

When $\tau = 0.5$, it becomes obvious that the conditional median response is obtained.

5.1. Maximum likelihood estimation

Referring to Eqs (5.2) and (5.3), the log-likelihood function of the ULE quantile regression model is given by

$$\ell(\boldsymbol{\Omega}) = n\log(\theta) - \sum_{i=1}^{n}\log(\log(1/\mu_i))$$
$$+ n\log\left(\log\left((1/\tau-1)^{1/\theta}+1\right)\right) + \sum_{i=1}^{n}\log(y_i)\left[\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu_i)}-1\right]$$
$$+ (\theta-1)\sum_{i=1}^{n}\log\left[y^{\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu)}}-1\right] - 2\sum_{i=1}^{n}\log\left[\left(y^{\frac{\log\left((\frac{1}{\tau}-1)^{1/\theta}+1\right)}{\log(\mu)}}-1\right)^{\theta}+1\right],$$

where $\boldsymbol{\Omega} = (\theta, \boldsymbol{\delta}^T)^T$ is the vector of the unknown parameters. The ML $\hat{\boldsymbol{\Omega}} = (\hat{\theta}, \hat{\boldsymbol{\delta}}^T)^T$ can be obtained by directly maximizing the log-likelihood function $\ell(\boldsymbol{\Omega})$ using mathematical packages such as Mathematica or R.

In the classical approach, the confidence intervals for the parameters were constructed based on asymptotic normality, i.e., $\hat{\boldsymbol{\Omega}}$ can be approximated by a $(k + 1)-$variate normal distribution with zero means and covariance matrix $I^{-1}(\hat{\boldsymbol{\Omega}})$, where $I(\boldsymbol{\Omega})$ is the observed information matrix defined by

$$I(\boldsymbol{\Omega}) = -\left[\frac{\partial^2\ell(\boldsymbol{\Omega})}{\partial\boldsymbol{\Omega}\partial\boldsymbol{\Omega}^T}\right].$$

The asymptotic $(1 - \alpha)100\%$ confidence interval for the vector parameter Ω_j; $j = 1, \ldots, k + 1$ is $\hat{\Omega}_j \pm z_{\frac{\alpha}{2}} se(\hat{\Omega}_j)$, where $z_{\frac{\alpha}{2}}$ represents the upper $\frac{\alpha}{2}$ percentile with a standard normal distribution and $se(\hat{\Omega}_j)$ is the square root of the jth diagonal entry of the Fisher matrix $I^{-1}(\hat{\Omega})$.

5.2. Model sufficiency

Residual analysis plays a vital role in testing the goodness of fit for a regression model. For this purpose, two types of residuals are employed: randomized quantile residual (RQR) [41] and Cox-Snell residual (CSR) [42]. The RQRs are computed as follows:

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

where $G(y_i, \mu_i, \theta)$ denotes the cdf of the ULE quantile regression model in Eq (5.1), and $\Phi^{-1}[.]$ represents the inverse cdf of the standard normal distribution. The adequacy of the fitted model is indicated by the RQRs ($\hat{r}_i$) following a standard normal distribution.

On the other side, the Cox-Snell residuals are written as:

$$\hat{e}_i = -\log[\bar{G}(y_i, \mu_i, \theta)],$$

where $\bar{G}(y_i, \mu_i, \theta) = 1 - G(x_i, \mu_i, \theta)$ represents the survival function of the ULE quantile model of regression. The Cox-Snell residuals ($\hat{e}_i$) are expected to follow an exponential distribution with a scale parameter of 1.

In summary, both the randomized quantile residual and Cox-Snell residual are valuable tools to evaluate the goodness of fit for the ULE quantile regression model and can provide insights into the appropriateness of the model for a given dataset.

5.3. Simulation

This subsection outlines a simulation study to assess the effectiveness of maximum likelihood estimators (MLEs) and their corresponding asymptotic confidence intervals (CIs) for the parameters in the ULE quantile regression model. The evaluation metrics include mean squared error (MSE) and average bias to assess the point estimates of the parameters, while coverage probability (CPs) along with average width (AWs) to assess the efficiency of the constructed 95% CIs. Two distinct scenarios are considered within the simulation design:

(a) *The case of one covariate:* the simulation study is performed as

$$logit(\mu_i) = \delta_0 + \delta_1 y_{i1}, \ i = 1, \ldots, n.$$

In this case, we consider $(\delta_0, \delta_1) = (-2, 2)$.

(b) *The case of two covariates:* the simulation study is performed as

$$logit(\mu_i) = \delta_0 + \delta_1 y_{i1} + \delta_2 y_{i2}, \ i = 1, \ldots, n.$$

In this case, we consider $(\delta_0, \delta_1, \delta_2) = (2, -2, 2)$.

In both cases, the true value for the parameter θ is 0.5 and the quantile values are τ = 0.1, 0.25, 0.5, 0.75, and 0.9. The covariate x_{i1} is generated from Bernoulli distribution with parameter 0.5 while the covariate x_{i2} is generated from the standard normal distribution for n =50,100 and 150. For each scenario, the simulation is repeated 1000 times.

The simulation results are given in Tables 6 and 7. The results in these tables confirm desirable properties for the maximum likelihood estimates for the parameters of the ULE quantile regression. From Tables 6 and 7, we observe that the MSE, ABs, and AWs decrease with increasing sample size. These metrics are consistently attaining their lowest values for the parameter θ. The MSEs and ABs attain their highest values when $\tau = 0.9$, particularly for small sample sizes. The coverage probabilities remain close to the nominal 95% confidence level.

Table 6. Results of the simulation for $\theta = 0.5, \delta_0 = 2, \delta_1 = -2, \delta_2 = 2$.

τ	n	θ				δ_0				δ_1				δ_2			
		MSEs	ABs	AWs	CPs	MSEs	ABs	AWs	CPs	MSEs	ABs	AWs	CPs	MSEs	ABs	AWs	CPs
0.1	50	0.007	0.016	0.303	0.943	0.097	0.033	1.202	0.943	0.251	0.080	1.960	0.951	0.051	0.023	0.871	0.96
	100	0.003	0.006	0.207	0.966	0.053	0.012	0.835	0.946	0.102	0.012	1.252	0.937	0.038	0.010	0.725	0.946
	150	0.002	-0.002	0.167	0.946	0.035	0.024	0.705	0.931	0.072	0.004	1.009	0.926	0.024	0.011	0.593	0.963
0.25	50	0.008	0.018	0.304	0.908	0.092	0.063	1.202	0.944	0.281	-0.033	1.986	0.952	0.0553	0.016	0.882	0.944
	100	0.003	0.004	0.206	0.944	0.053	0.023	0.833	0.924	0.0998	0.016	1.243	0.964	0.038	0.010	0.710	0.944
	150	0.002	-0.002	0.167	0.932	0.033	0.006	0.703	0.948	0.061	0.003	1.013	0.952	0.024	0.008	0.604	0.944
0.5	50	0.008	0.015	0.300	0.912	0.121	0.047	1.381	0.948	0.293	-0.031	2.006	0.956	0.058	0.023	0.897	0.936
	100	0.003	0.004	0.205	0.94	0.065	0.023	0.942	0.936	0.099	0.016	1.246	0.968	0.039	0.011	0.716	0.936
	150	0.002	-0.002	0.166	0.928	0.042	0.016	0.789	0.944	0.062	0.005	1.015	0.952	0.024	0.007	0.609	0.948
0.75	50	0.006	0.016	0.273	0.928	0.437	-0.026	2.367	0.932	0.269	-0.017	2.0686	0.968	0.081	0.077	0.991	0.932
	100	0.002	0.007	0.192	0.96	0.163	0.005	1.614	0.976	0.120	-0.037	1.285	0.952	0.047	0.050	0.794	0.952
	150	0.002	-0.001	0.153	0.932	0.126	0.041	1.310	0.924	0.060	-0.018	1.033	0.984	0.032	0.026	0.648	0.94
0.9	50	0.006	0.048	0.256	0.952	0.810	-0.420	3.339	0.968	0.260	0.014	2.213	0.964	0.097	0.060	1.181	0.952
	100	0.004	0.041	0.187	0.852	0.586	-0.403	2.327	0.876	0.156	0.002	1.396	0.94	0.082	0.040	0.958	0.94
	150	0.003	0.036	0.148	0.876	0.329	-0.313	1.836	0.892	0.079	-0.017	1.104	0.96	0.047	-0.024	0.759	0.932

Table 7. Results of the simulation for $\theta = 0.5, \delta_0 = 2, \delta_1 = -2$.

τ	n	θ				δ_0				δ_1			
		MSEs	ABs	AWs	CPs	MSEs	ABs	AWs	CPs	MSEs	ABs	AWs	CPs
0.1	50	0.007	0.011	0.296	0.916	0.338	-0.025	2.504	0.968	0.480	0.078	2.831	0.968
	100	0.003	0.005	0.206	0.956	0.158	0.013	1.592	0.948	0.199	0.020	1.895	0.972
	150	0.002	-0.002	0.167	0.948	0.103	0.0003	1.336	0.956	0.150	-0.002	1.569	0.948
0.25	50	0.007	0.012	0.297	0.908	0.349	-0.046	2.520	0.976	0.518	0.099	2.846	0.968
	100	0.003	0.005	0.206	0.952	0.188	-0.007	1.602	0.912	0.230	0.045	1.905	0.956
	150	0.002	-0.001	0.167	0.920	0.118	-0.020	1.344	0.972	0.163	0.030	1.574	0.948
0.5	50	0.007	0.016	0.299	0.929	0.425	-0.030	2.806	0.954	0.440	0.039	2.889	0.974
	100	0.003	0.004	0.206	0.954	0.217	-0.016	1.837	0.940	0.240	0.017	1.954	0.960
	150	0.002	-0.002	0.167	0.951	0.153	0.010	1.519	0.943	0.170	0.027	1.596	0.949
0.75	50	0.007	0.012	0.298	0.936	2.089	-0.254	5.308	0.944	0.974	0.242	3.753	0.940
	100	0.003	0.005	0.206	0.952	0.650	-0.064	3.424	0.956	0.346	0.061	2.360	0.936
	150	0.002	-0.003	0.166	0.936	0.561	0.012	2.785	0.928	0.237	0.031	1.899	0.936
0.9	50	0.007	0.021	0.309	0.928	4.413	0.488	3.186	0.898	1.563	-0.764	1.728	0.902
	100	0.003	0.008	0.214	0.980	1.707	0.817	2.539	0.902	0.986	-0.893	1.068	0.880
	150	0.002	0.007	0.173	0.960	1.518	0.873	1.606	0.926	0.989	-0.925	0.630	0.928

6. Applications

This section explores the advantages of the practical use of ULE distribution and its regression model compared to other models commonly used for modeling data within unit intervals. The unit models used in this comparison are the Beta, Kumaraswamy, Topp-Leon, unit-Burr XII, and unit half-logistic geometric (UHLG) distributions.

To assess and compare the fitted distributions, we employ various goodness-of-fit measures. These include the Kolmogorov-Smirnov distance (KS) with corresponding p-value, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic ($A^\star$), and Cramer-von Mises statistic ($W^\star$).

Additionally, two graphical methods are used to assess the goodness of fit for the regression models:

i) The quantile-quantile (QQ) plot, which compares randomized quantile residuals against normal quantiles with a simulated enclosure, as proposed by Atkinson [43].

ii) The probability-probability (PP) plot compares the empirical probabilities of Cox-Snell residuals with those from the standard exponential distribution.

The model is considered the best fit for the data if its values are closest to the diagonal bar in both plots. Mathematica scripts were used for all numerical calculations and to generate the figures.

6.1. Recovery rate applications

The recovery rate of CD34+ cells is a crucial factor in assessing the sufficiency of peripheral blood stem cell (PBSC) collection for bone marrow transplantation. This procedure facilitates rapid hematologic recovery following myeloablative therapy for various malignant hematological diseases.

This study analyzed data from 239 patients who underwent autologous PBSC transplantation, focusing on CD34+ cell recovery rates as documented in [44]. Additionally, we incorporated relevant covariates that could potentially influence CD34+ cell viability recovery:

- Gender (y_1): 0 denotes female and 1 denotes male;
- Chemotherapy (y_2): 0 for a single-day chemotherapy regimen and 1 for a three-day regimen;
- Age (y_3): Adjusted patient's age, calculated as the current age minus 40.

A univariate modeling approach was employed to assess the suitability of the ULE distribution for modeling CD34+ cell recovery rates. The ML estimates of the ULE distribution's parameters and various goodness-of-fit statistics were calculated and are presented in Table 8. The results from this table strongly suggest that the ULE distribution provides the best fit for the given dataset and it generally outperforms other models across multiple criteria (AIC, BIC, -Loglik, $W^\star$, $A^\star$), indicating that it best fits the CD34+ cell recovery data. The Unit-Omega and Unit-BurrXII models perform well but do not surpass the ULE model according to AIC, BIC, or -Loglik. To further visualize this relationship, Figure 4 was generated. This figure includes histogram plots of the data, overlaid with the estimated densities of the ULE distribution, as well as the Kaplan-Meier curve and the estimated survival curves.

Table 8. MLEs (std. errors) and discriminant criteria for recovery rate of CD34+ cells data.

| Model	Estimates (Standard Errors)		K-S (p-value)	AIC	BIC	-Loglik	$W^\star$	$A^\star$
	β	θ						
ULE	3.2454	1.77959	0.0500323	-383.997	-377.044	-193.998	0.0675539	0.48204
	(0.142222)	(0.100898)	(0.587827)					
Beta	2.28593	8.66714	0.0650045	-379.734	-372.782	-191.867	0.14045	0.87598
	(0.458761)	(0.332228)	(0.264737)					
Kumaraswamy	2.43553	6.69423	0.0722763	-377.528	-370.575	-190.764	0.191987	1.14739
	(0.458761)	(0.332228)	(0.16457)					
Unit-Omega	3.6080	7.7337	0.05331	-383.955	-377.002	-193.977	0.0820	0.5276
	(0.3322)	(0.4588)	(0.5055)					
Unit- BurrXII	1.73211	10.076	0.0522282	-383.005	-376.052	-193.503	0.0888576	0.582411
	(0.458761)	(0.332228)	(0.532105)					
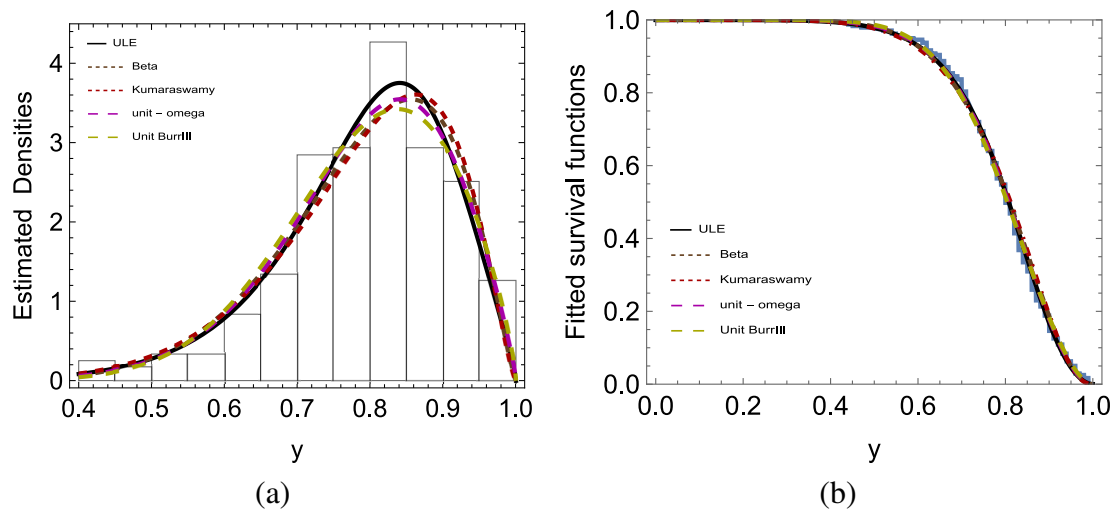
Figure 4. (a) Histogram and estimated densities. (b) Estimating the probability of CD34+ cell recovery over time using Kaplan-Meier survival analysis.

To derive Bayesian estimates and credible intervals, we run the MCMC algorithm 55,000 times, discarding the first 5,000 iterations as burn-in and selecting every 5th value for thinning. The burn-in period is determined based on pilot runs of the Metropolis-Hastings algorithm. Specifically, we monitor the convergence behavior of the chains for the parameters θ and β, examining trace plots to identify when the Markov chain stabilizes around the posterior distribution Figure 5 shows a smooth histogram of the marginal posterior density, a trace plot where each chain remains within a similar region without distinct trends, and results from the first 100 lags of the autocorrelation function (ACF). These observations reflect the efficient convergence of the MCMC chain.

Percentile bootstrap confidence intervals were computed from 1,000 bootstrap samples of the parameters. Table 9 displays the maximum likelihood estimates (MLE), bootstrap standard errors (BSE), and the associated 95% confidence intervals. It is evident that the MLE, BSE, Mean Posterior Standard Errors (MPSE), and LSE for the parameter θ are closely aligned, with confidence intervals that are nearly of the same length.

Table 9. Point estimates and confidence intervals for the ULE parameters related to the recovery rate of CD34+ cells data.

Parameter	MLE		MPSE		BSE		LSE	
	MLE	ACI_{ML}	MPSE	ACI_{MPS}	BSE	CrI	LSE	BCI
θ	1.77959	(1.5818,1.9774)	1.8044	(1.6114,2.0062)	1.7469	(1.5527,1.9412)	1.8551	(1.6090,2.0701)
β	3.2454	(2.9667,3.5241)	3.2254	(2.9628,3.5129)	3.2453	(2.9631,3.5275)	3.1989	(2.9563,3.4995)
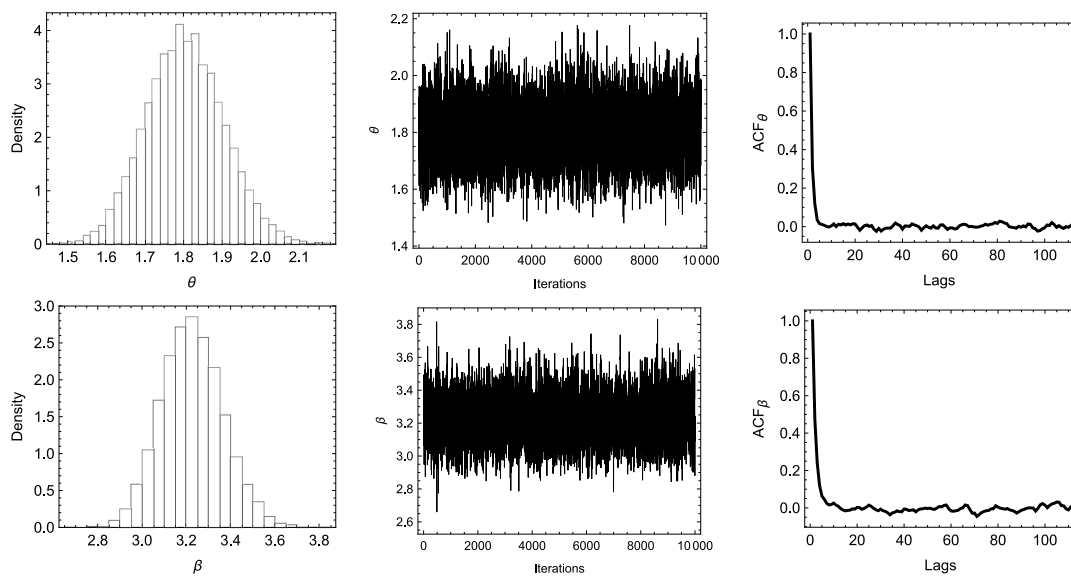
Figure 5. Histogram for the marginal posterior density, trace plot, and the first 100 lags of autocorrelation values of MCMC results for recovery rate data.

A multivariate regression model is now employed to examine the effects of gender, chemotherapy, and age on the recovery rate. The proposed regression model is:

$$logit(\mu_i) = \delta_0 + \delta_1 y_{i1} + \delta_2 y_{i2} + \delta_3 y_{i3}; i = 1, \ldots, 239.$$

Table 10 presents the fitting results for beta, Kumaraswamy, and ULE regression models. Table 10 shows MLE for α, and $\delta_i, i = 0, 1, 2, 3$, along with their respective standard errors (SE) and p-values. Additionally, each regression model is accompanied by AIC and BIC statistics.

According to Table 10, there is a statistically significant positive effect of the type of chemotherapy and age on the recovery rate of CD34+ cells, whereas the gender parameter does not significantly affect the recovery rate in any of the regression models.

Additionally, among the beta and Kumaraswamy regression models, the ULE regression model offers the best fit to the data, as indicated by its lower AIC and BIC values. This finding is further explored in Figure 6, which displays the P-P plot of Cox-Snell residuals and Q-Q plots of the randomized quantile residuals with a simulated envelope for all the fitted regression models.

Table 10. The outcomes of the fitted regression models using discriminant criteria for the recovery rate of CD34+ cells data.

Parameters	Beta			Kumaraswamy			ULE		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
δ_0	0.9990	0.1291	< 0.0000	1.1997	0.1397	< 0.0000	1.04811	0.135555	< 0.0000
δ_1	0.0659	0.0939	0.483	0.0418	0.0955	0.6619	0.100013	0.07255	0.169365
δ_2	0.2116	0.1038	0.0425	0.1833	0.1150	0.1123	0.21845	0.104438	.037553
δ_3	0.0142	0.0054	0.0088	0.0107	0.0059	0.0692	0.01709	0.00547946	0.002043
α	11.3447	1.0181	< 0.0000	6.7274	0.4543	< 0.0000	1.82216	0.10317	< 0.0000
AIC	-381.79			-375.66			-388.659		
BIC	-364.41			-358.28			-371.276		
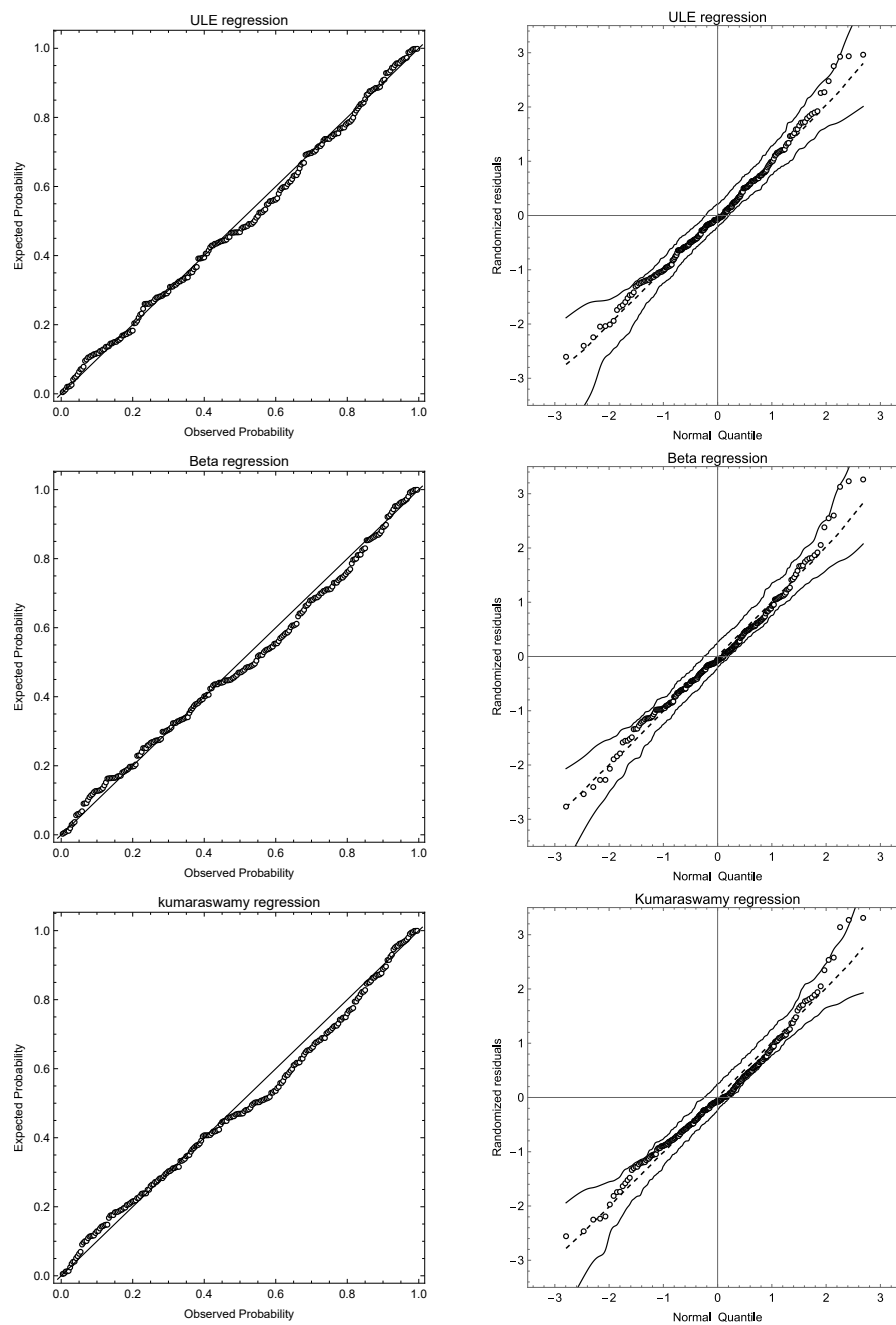
Figure 6. P-P plots of Cox-Snell residuals (shown in the left side) and Q-Q plots of randomized quantile residuals (shown in the right side) for the recovery rate data.

6.2. Data on the cost-effectiveness of risk management

The dataset in question is presented by Schmit and Roth [45], which includes 73 responses from a survey distributed to 374 risk managers of major North American corporations. Schmit and Roth [45] sought to assess the cost-effectiveness of different management strategies in reducing a company's risk of property losses and accidents, taking into account specific company characteristics like size and industry. The response variable, y (Firmcost), represents the firm-specific ratio of premiums plus
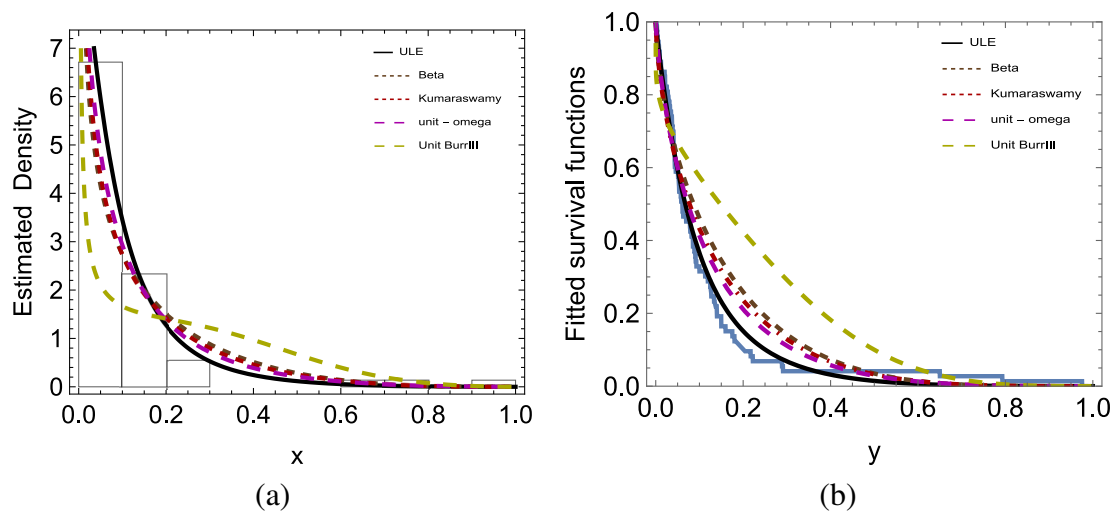
uninsured losses relative to total assets. The covariates associated with this response variable are as follows:

- Suppose (y_1): the firm-specific ratio of the total per-occurrence retention levels, as assessed by the corporate risk manager.
- Cap (y_2): 1 if the firm utilizes a captive insurance, and zero if it does not.
- Sizelog (y_3): the logarithm of the firm's total asset value.
- Indcost (y_4): the industry average of premiums and uninsured losses relative to total assets, as reported in the 1985 cost of risk survey (a risk measurement).
- Central (y_5): the role of the local manager in determining local retention levels, as evaluated by the corporate risk manager.
- Sizelog (y_6): the significance of analytical tools in risk management decision-making, as assessed by the corporate risk manager.

Initially, univariate modeling is applied to the response variable, risk management cost-effectiveness data, to evaluate the performance of the ULE distribution. Table 11 presents the ML estimates for the parameters along with the goodness-of-fit statistics. The findings in Table 11 indicate that the ULE model provides the best overall fit to the risk management cost-effectiveness data, as evidenced by its favorable AIC, BIC, -Loglik, and goodness-of-fit test results (K-S, Cramér-von Mises, and Anderson-Darling). The unit-Omega model performs reasonably well but does not surpass the ULE model across the proposed criteria. This conclusion is further illustrated in Figure 7, which displays histogram plots of the data with the estimated densities, as well as the Kaplan-Meier curve with the estimated survival curves.

Table 11. MLEs (std. errors) and discriminant criteria for risk management cost-effectiveness data.

| Model	Estimates (Standard Errors)		K-S (p-value)	AIC	BIC	-Loglik	Cramér-von Mises	Anderson-Darling
	β	θ						
ULE	0.257	2.582	0.068	-170.611	-166.03	-87.306	0.096	0.874
	(0.142)	(0.101)	(0.893)					
Beta	3.798	0.613	0.181	-148.235	-143.654	-76.118	0.697	3.961
	(0.459)	(0.332)	(0.017)					
Kumaraswamy	3.441	0.665	0.154	-153.308	-148.727	-78.654	0.502	3.097
	(0.459)	(0.332)	(0.064)					
Unit Omega	5.377	0.787	0.128	-163.383	-158.802	-83.691	0.336	2.153
	(0.928)	(0.073)	(0.180)					
Unit-BurrXII	0.348	2.841	0.338	-89.013	-84.432	-46.507	2.662	12.874
	(0.0625)	(0.421)	(0.0)					

Figure 7. (a) Histogram and estimated densities. (b) Kaplan-Meier and estimated survival functions for risk management cell data.

To compute the Bootstrap errors (BE) and confidence intervals (CI), we ran the MCMC algorithm 55,000 times, discarding the initial 5,000 iterations as burn-in and selecting every 5th value for thinning. As mentioned in the previous section, the pilot runs of the Metropolis-Hastings algorithm determine the burn-in period through monitoring convergence trace, as shown in Figure 8. This figure illustrates a smooth and well-behaved histogram of the marginal posterior density, a trace plot showing that each chain consistently explores the same region without evident trends, and the results from the first 100 lags of the autocorrelation function (ACF). These findings illustrate fast convergence performance for the MCMC chain. Percentile bootstrap confidence intervals were obtained from 1,000 bootstrap samples of the parameters. Table 12 presents the ML estimates, bootstrap standard errors (BSE), and the corresponding 95% confidence intervals. It is noted that the MLE, BSE, mean posterior standard errors (MPSE), and LSE of parameter θ are similar, with their confidence intervals nearly the same length. The regression model for μ_i is presented as

$$logit(\mu_i) = \delta_0 + \delta_1 y_{i1} + \delta_2 y_{i2} + \delta_3 y_{i3} + \delta_4 y_{i4} + \delta_5 y_{i5} + \delta_6 y_{i6}; \ i = 1, \ldots, 73.$$

Table 12. Point estimates and interval estimates of ULE parameters for risk management cost-effectiveness data.

Parameter	MLE		MPSE		BSE		LSE	
	MLE	ACI_{ML}	MPSE	ACI_{MPS}	BSE	CrI	LSE	BCI
θ	2.5823	(2.3845,2.7800)	2.4194	(1.9312,2.9077)	2.5500	(2.0647,3.0615)	3.0267	(2.4038,3.6891)
β	0.2572	(-0.0215,0.5359	0.2579	(0.2290,0.2868)	0.2581	(0.2314,0.2873)	0.2536	(0.2303,0.279)
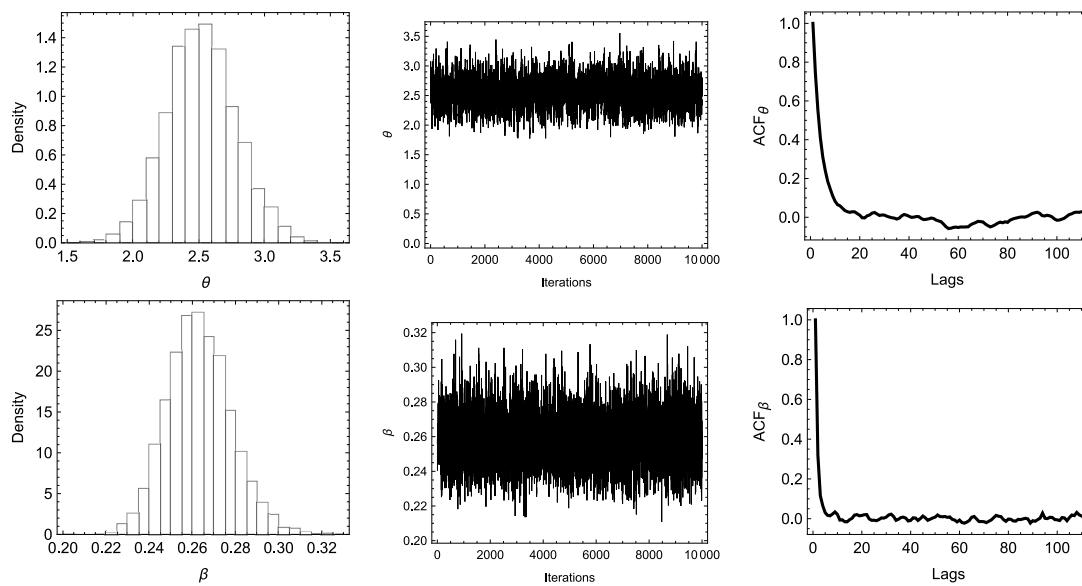
Figure 8. Histogram for the marginal posterior density, trace plot, and the first 100 lags of autocorrelation values of MCMC results for risk management data.

Table 13 presents the fitting results for beta, Kumaraswamy, and ULE regression models. It shows ML estimates for α, and $\delta_i, i = 0, 1, 2, 3, 4, 5, 6$, along with their respective standard errors (SE) and p-values. Additionally, each regression model is accompanied by AIC and BIC statistics.

Based on Table 13, we can notice that there is a statistically significant positive impact of the type of chemotherapy and age on the risk management cost-effectiveness data, while the parameter gender has no statistically significant impact on the response variable recovery rate at the usual level for all regression models.

Table 13. Maximum likelihood estimates for the parameters of the fitted regression models, along with AIC and BIC values, for the risk management cost-effectiveness data.

Parameters	Beta			Kumaraswamy			ULE		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
δ_0	1.888	1.172	0.112	2.539	1.550	0.106	3.823	1.463	0.011
δ_1	-0.001	0.014	0.932	-0.036	0.018	0 .043	-0.014	0.012	0.232
δ_2	0.178	0.232	0.445	0.596	0.390	0.131	0.142	0.250	0.573
δ_3	-0.512	0.123	< 0.000	-0.798	0.161	< 0.000	-0.872	0.149	< 0.000
δ_4	1.236	0.459	0.009	5.257	1.436	0.0005	1.799	0.444	0.0001
δ_5	-0.012	0.088	0.890	-0.028	0.120	0.817	-0.070	0.101	0.493
δ_6	-0.004	0.021	0.855	-0.027	0.032	0.396	0.006	0.024	0.808
α	6.331	1.123	< 0.000	0.978	0.106	< 0.000	3.503	0.357	< 0.000
AIC	-159.446			-181.653			-200.186		
BIC	-141.122			-163.33			-181.863		

Furthermore, compared to the beta and Kumaraswamy regression models, the ULE regression model provides the most optimal fit to the data as it has the least AIC and BIC. Figure 9 provides a deeper analysis of this result, displaying the P-P plot of Cox-Snell residuals and the Q-Q plots of randomized quantile residuals, complete with a simulated envelope, for all the regression models fitted.
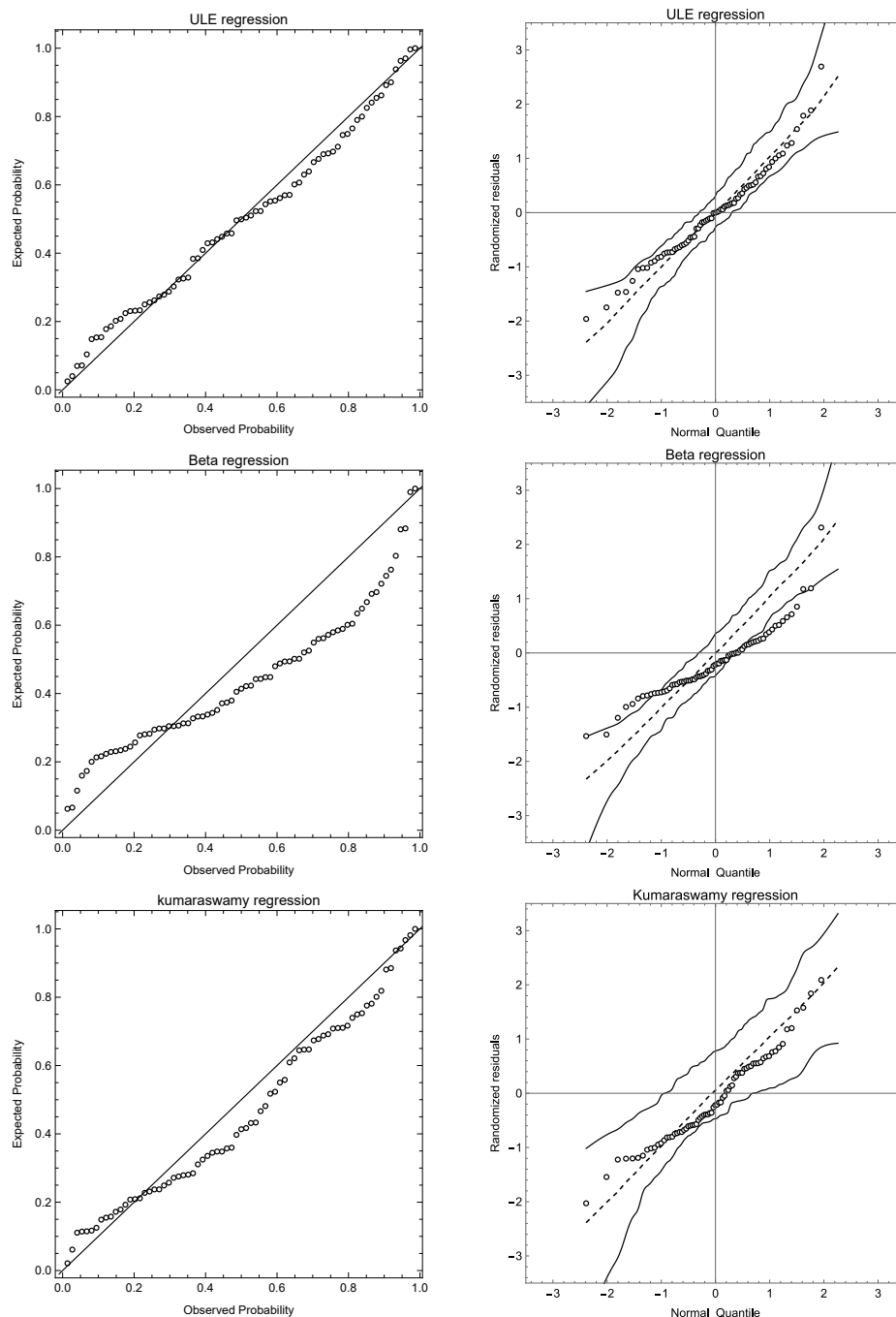
Figure 9. P-P plots of Cox-Snell residuals (shown on the left side) and Q-Q plots of randomized quantile residuals (shown on the right side) for the risk management cost-effectiveness data.

7. Conclusions

This paper introduced a versatile two-parameter unit distribution called the ULE distribution, designed for modeling datasets confined to the interval $(0, 1)$. We derived and analyzed the survival function and hazard rate function of the ULE distribution, presenting their behavior through graphical representations. We investigated various methods for parameter estimation of the ULE distribution, including maximum likelihood estimation, maximum product spacing, least squares, and Bayesian estimation. An extensive simulation study reveals that Bayesian estimation demonstrates robust performance, characterized by high CPs and acceptable AWs. ML estimate consistently shows strong results with reliable MSEs, ABs, and CPs, making it a trustworthy option. LSE and MPSE may exhibit greater variability with smaller sample sizes, but their performance improves and remains satisfactory as n increases. Additionally, we proposed a novel quantile regression model based on the ULE distribution for unit response variables, providing an alternative to the commonly used beta, Kumaraswamy, and Unit-Weibull regression models. It was found that the randomized quantile residual and Cox-Snell residual are useful tools for assessing the goodness of fit of the ULE quantile regression model, offering valuable insights into the model's suitability for a specific dataset. Our empirical analysis focused on datasets related to the recovery rate of CD34+ cells and the cost-effectiveness of risk management. The numerical analysis with the help of goodness of fit tests emphasizes the suitability of the ULE distribution to model these data. One of the limitations of this model is the time and cost constraints, as we consider a complete data set, units may take a long time to fail, or it can be expensive to perform a complete sample experiment, especially in medical and biological fields. Censoring schemes can be employed to adjust this limitation. Hence, our plan for future work is to implement censored samples instead of complete ones. Another limitation is that even though the ULE distribution is proposed as effective for fitting real-world data, its performance might vary depending on the nature and characteristics of the datasets. It may not always outperform existing models in all scenarios, especially if the data do not exhibit the specific long-tailed behavior for which the ULE distribution is suited. The ULE distribution is defined on the unit interval $(0, 1)$. This limitation implies that it may not be suitable for datasets that extend beyond this interval or where transformations to fit this range introduce bias or distortions.

Author contributions

Hanan Haj Ahmad: Methodology, validation, investigation, resources, writing–review & editing, funding acquisition; Kariema A. Elnagar: Methodology, conceptualization, investigation, resources, data curation, writing–original draft, Writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Acknowledgments

Conflicts of interest

The authors declare no conflicts of interest.

References

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

2. 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, *Comm. Statist. Theory Methods*, **48** (2019), 3423–3438. https://doi.org/10.1080/03610926.2018.1476717

3. P. C. Consul, G. C. Jain, On the log-gamma distribution and its properties, *Statistische Hefte*, **12** (1971), 100–106. https://doi.org/10.1007/BF02922944

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

5. J. Mazucheli, A. F. Menezes, S. Dey, Unit-Gompertz distribution with applications, *Statistica*, **79** (2019), 25–43. https://doi.org/10.6092/issn.1973-2201/8497

6. M. M. E. Abd El-Monsef, M. M. El-Awady, M. M. Seyam, A new quantile regression model for modeling child mortality, *Int. J. Biomath.*, **15** (2022), 2250031. https://doi.org/10.1142/S1793524522500310

7. M. Ç. Korkmaz, C. Chesneau, On the unit Burr-XII distribution with the quantile regression modeling and applications, *Comp. Appl. Math.*, **40** (2021), 29. https://doi.org/10.1007/s40314-021-01418-5

8. H. S. Bakouch, A. S. Nik, A. Asgharzadeh, H. S. Salinas, A flexible probability model for proportion data: Unit-half-normal distribution, *Comm. Statist. Case Stud. Data Anal. Appl.*, **7** (2021), 271–288. https://doi.org/10.1080/23737484.2021.1882355

9. H. Haj Ahmad, E. M. Almetwally, M. Elgarhy, D. A. Ramadan, On unit exponential Pareto distribution for modeling the recovery rate of COVID-19, *Processes*, **11** (2023), 232. https://doi.org/10.3390/pr11010232

10. A. S. Hassan, A. Fayomi, A. Algarni, E. M. Almetwally, Bayesian and non-Bayesian inference for unit-exponentiated half-logistic distribution with data analysis, *Appl. Sci.*, **12** (2022), 11253. https://doi.org/10.3390/app122111253

11. A. S. Hassan, A. M. Khalil, H. F. Nagy, Data analysis and classical estimation methods of the bounded power Lomax distribution, *Reliab. Theory Appl.*, **19** (2024), 770–789.

12. A. Fayomi, A. S. Hassan, E. M. Almetwally, Inference and quantile regression for the unit-exponentiated Lomax distribution, *Plos one*, **18** (2023), e0288635. https://doi.org/10.1371/journal.pone.0288635

13. A. Fayomi, A. S. Hassan, H. Baaqeel, E. M. Almetwally, Bayesian inference and data analysis of the unit-power Burr X distribution, *Axioms*, **12** (2023), 297. https://doi.org/10.3390/axioms12030297

14. A. S. Hassan, R. E. Mohamed, O. Kharazmi, H. F. Nagy, A new four-parameter extended exponential distribution with statistical properties and applications, *Pak. J. Stat. Oper. Res.*, **18** (2022), 179–193. https://doi.org/10.18187/pjsor.v18i1.3872

15. R. D. Gupta, D. Kundu, Generalized exponential distribution: Different method of estimations, *J. Stat. Comput. Simul.*, **69** (2001), 315–337. https://doi.org/10.1080/00949650108812098

16. M. C. Jones, Families of distributions arising from distributions of order statistics, *Test*, **13** (2004), 1–43. https://doi.org/10.1007/BF02602999

17. M. S. Khan, R. King, I. L. Hudson, Transmuted generalized exponential distribution: A generalization of the exponential distribution with applications to survival data, *Commun. Statist.*, **46** (2017), 4377–4398. https://doi.org/10.1080/03610918.2015.1118503

18. A. Z. Afify, G. M. Cordeiro, H. M. Yousof, Z. M. Nofal, A. Alzaatreh, The Kumaraswamy transmuted-G family of distributions: Properties and applications, *J. Data Sci.*, **14** (2016), 245–270. https://doi.org/10.6339/JDS.201604_14(2).0004

19. A. J. Lemonte, G. M. Cordeiro, G. Moreno-Arenas, A new useful three-parameter extension of the exponential distribution, *Statistics*, **50** (2016), 312–337. https://doi.org/10.1080/02331888.2015.1095190

20. A. Mahdavi, D. Kundu, A new method for generating distributions with an application to exponential distribution, *Comm. Statist. Theory Methods*, **46** (2017), 6543–6557. https://doi.org/10.1080/03610926.2015.1130839

21. M. Nassar, A. Z. Afify, M. K. Shakhatreh, Estimation methods of alpha power exponential distribution with applications to engineering and medical data, *Pak. J. Stat. Oper. Res.*, **16** (2020), 149–166. http://dx.doi.org/10.18187/pjsor.v16i1.3129

22. A. Z. Afify, O. A. Mohamed, A new three-parameter exponential distribution with variable shapes for the hazard rate: Estimation and applications, *Mathematics*, **8** (2020), 135. https://doi.org/10.3390/math8010135

23. M. Nassar, D. Kumar, S. Dey, G. M. Cordeiro, A. Z. Afify, The Marshall-Olkin alpha power family of distributions with applications, *J. Comput. Appl. Math.*, **351** (2019), 41–53. https://doi.org/10.1016/j.cam.2018.10.052

24. M. A. Aldahlan, A. Z. Afify, A new three-parameter exponential distribution with applications in reliability and engineering, *J. Nonlinear Sci. Appl.*, **13** (2020), 258–269.

25. S. Abbas, A. Jahngeer, S. H. Shahbaz, A. Z. Afify, M. Q. Shahbaz, Topp-Leone moment exponential distribution: Properties and applications, *J. Natl. Sci. Found. Sri Lanka*, **48** (2020), 265–274.

26. Y. Lan, L. M. Leemis, The logistic-exponential survival distribution, *Naval Res. Logist.*, **55** (2008), 252–264. https://doi.org/10.1002/nav.20279

27. M. Shaked, J. G. Shanthikumar, *Stochastic orders*, New York: Springer, 2007. https://doi.org/10.1007/978-0-387-34675-5

28. A. Zellner, Bayesian estimation and prediction using asymmetric loss functions, *J. Amer. Statist. Assoc.*, **81** (1986), 446–451.

29. S. Dey, S. Ali, C. Park, Weighted exponential distribution: Properties and different methods of estimation, *J. Stat. Comput. Simul.*, **85** (2015), 3641–3661. https://doi.org/10.1080/00949655.2014.992346

30. D. Kundu, H. Howlader, Bayesian inference and prediction of the inverse Weibull distribution for Type-II censored data, *Comput. Statist. Data Anal.*, **54** (2010), 1547–1558. https://doi.org/10.1016/j.csda.2010.01.003

31. L. Tierney, Markov chains for exploring posterior distributions, *Ann. Statist.*, **22** (1994), 1701–1728.

32. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian data analysis*, New York: Chapman and Hall/CRC, 1995. https://doi.org/10.1201/9780429258411

33. G. O. Roberts, J. S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60** (1998), 255–268. https://doi.org/10.1111/1467-9868.00123

34. R. C. H. Cheng, N. A. K. Amin, Estimating parameters in continuous univariate distributions with a shifted origin, *J. R. Stat. Soc. Ser. B Stat. Methodol*, **45** (1983), 394–403. https://doi.org/10.1111/j.2517-6161.1983.tb01268.x

35. B. Ranneby, The maximum spacing method. An estimation method related to the maximum likelihood method, *Scand. J. Statist.*, **11** (1984), 93–112.

36. E. J. G. Pitman, *Some basic theory for statistical inference: Monographs on applied probability and statistics*, New York: Chapman & Hall, 1979. https://doi.org/10.1201/9781351076777

37. R. C. H. Cheng, L. Traylor, Non-regular maximum likelihood problems, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57** (1995), 3–24. https://doi.org/10.1111/j.2517-6161.1995.tb02013.x

38. K. Ghosh, S. R. Jammalamadaka, A general estimation method using spacings, *J. Statist. Plann. Inference*, **93** (2001), 71–82. https://doi.org/10.1016/S0378-3758(00)00160-9

39. S. Anatolyev, G. Kosenok, An alternative to maximum likelihood based on spacings, *Econometric Theory*, **21** (2005), 472–476. https://doi.org/10.1017/S0266466605050255

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

41. P. K. Dunn, G. K. Smyth, Randomized quantile residuals, *J. Comput. Graph. Statist.*, **5** (1996), 236–244.

42. D. R. Cox, E. J. Snell, A general definition of residuals, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **30** (1968), 248–265. https://doi.org/10.1111/j.2517-6161.1968.tb00724.x

43. A. C. Atkinson, Two graphical displays for outlying and influential observations in regression, *Biometrika*, **68** (1981), 13–20. https://doi.org/10.1093/biomet/68.1.13

44. P. Zhang, Z. Qiu, C. Shi, simplexreg: An R package for regression analysis of proportional data using the simplex distribution, *J. Statist. Softw.*, **71** (2016), 1–21. https://doi.org/10.18637/jss.v071.i11

45. J. T. Schmit, K. Roth, Cost effectiveness of risk management practices, *J. Risk Insurance*, **57** (1990), 455–470.