

Article

Unit Exponential Probability Distribution: Characterization and Applications in Environmental and Engineering Data Modeling

Hassan S. Bakouch ^{1,2} , Tassaddaq Hussain ³ , Marina Tošić ⁴ , Vladica S. Stojanović ^{5,*} 
and Najla Qarmalah ^{6,*} 

- ¹ Department of Mathematics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia; h.bakouch@qu.edu.sa
- ² Department of Mathematics, Faculty of Science, Tanta University, Tanta 31111, Egypt; hassan.bakouch@science.tanta.edu.eg
- ³ Department of Statistics, Mirpur University of Science and Technology, Mirpur 10250, Pakistan; tassaddaq.stat@must.edu.pk
- ⁴ Department of Mathematics, Faculty of Sciences & Mathematics, University of Priština in Kosovska Mitrovica, 38220 Kosovska Mitrovica, Serbia; marina.tosic@pr.ac.rs
- ⁵ Department of Informatics & Computer Sciences, University of Criminal Investigation and Police Studies, 11060 Belgrade, Serbia
- ⁶ Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
- * Correspondence: vladica.stojanovic@kpu.edu.rs (V.S.S.); nmubinqarmalah@pnu.edu.sa (N.Q.); Tel.: +381-60-390-0552 (V.S.S.); +966-11-823-6238 (N.Q.)

Abstract: Distributions with bounded support show considerable sparsity over those with unbounded support, despite the fact that there are a number of real-world contexts where observations take values from a bounded range (proportions, percentages, and fractions are typical examples). For proportion modeling, a flexible family of two-parameter distribution functions associated with the exponential distribution is proposed here. The mathematical and statistical properties of the novel distribution are examined, including the quantiles, mode, moments, hazard rate function, and its characterization. The parameter estimation procedure using the maximum likelihood method is carried out, and applications to environmental and engineering data are also considered. To this end, various statistical tests are used, along with some other information criterion indicators to determine how well the model fits the data. The proposed model is found to be the most efficient plan in most cases for the datasets considered.

Keywords: unit distribution; statistical model; hazard function; characterizations; estimation; simulation; application

MSC: 60E05; 62E15; 62F10



Citation: Bakouch, H.S.; Hussain, T.; Tošić, M.; Stojanović, V.S.; Qarmalah, N. Unit Exponential Probability Distribution: Characterization and Applications in Environmental and Engineering Data Modeling. *Mathematics* **2023**, *11*, 4207. <https://doi.org/10.3390/math11194207>

Academic Editors: Francesca Condino and Domma Filippo

Received: 8 August 2023

Revised: 30 September 2023

Accepted: 7 October 2023

Published: 9 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proportional variables are often encountered in data science, where they are used as stochastic models that describe, for instance, the number of successes divided by the number of attempts, party votes, the proportion of money spent on a cause, or the attendance rate of public events. Therefore, proportion analysis is necessary in various fields such as healthcare, economics, and engineering. Usually, to model the behavior of such random variables (RVs), distributions defined on a unit interval are used, which are highly valuable in modeling proportions and percentages. It is conceivable to model and forecast such variables, but one must look outside the traditional model because the data are limited to the range $(0, 1)$. For further study, readers are referred to [1–3].

In this context, the beta model was proposed by Bayes [4], which in many fields of statistics is a convenient and helpful model widely used for modeling percentages and proportions. However, there are a number of scenarios where it seems to not be a suitable one. Therefore, alternatively, several distributions have been developed for modeling bounded

variables like proportions, indices, and rates, such as the unit distribution studied in [5], the unit Johnson distribution proposed in [6], the four-parameter distribution introduced in [7], the distribution proposed in [8], the Topp–Leone distribution studied in [9], and the unit gamma distribution introduced in [10]. More recently, many other unit interval distribution functions have been introduced, such as the cumulative distribution function (CDF) quantile distribution [11], new unit interval distribution [12], the unit-inverse Gaussian distribution [13], the log-xgamma distribution [14], unit Gompertz, unit Lindley, and unit Weibull distributions [15–17], the log-weighted exponential distribution [18], the unit Johnson SU distribution [19], the unit log–log distribution [20], the new unit distribution [21], the unit–power Burr X distribution [22], and the unit Teissier distribution [23], while in [24], the unit interval distribution via the conditional distribution approach was studied. Notice that all of these distributions are potential candidates for describing proportions. It is worth noting that the approaches mentioned above are mainly based on conventional strategies, namely the following:

- (i) Log transformation approaches;
- (ii) The CDF and quantile methodology;
- (iii) Reciprocal transformation;
- (iv) Exponential transformation;
- (v) The conditional distribution methodology;
- (vi) The T-X family approach.

However, all of the earlier models and others seem to be casual ways of generating unit interval distributions. In the current study, our motivational strategies begin with recalling the epsilon function examined in [25], which is defined as

$$\varepsilon_{\lambda,a}(x) = \begin{cases} \left(\frac{a+x}{a-x}\right)^{\frac{\lambda a}{2}}, & -a < x < a \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $\lambda \in \mathbb{R} \setminus \{0\}$ and $a > 0$. For comprehensive details about the above and its bounded version, readers are referred to [25]. The function $y = \varepsilon_{\lambda,a}(x)$ is the solution of an epsilon differential equation of the first order:

$$y' = \frac{\lambda a^2 y}{a^2 - x^2},$$

In addition, it satisfies the following property of the exponential limit:

$$\lim_{a \rightarrow +\infty} \varepsilon_{\lambda,a}(x) = e^{\lambda x}, \quad \forall x \in (-a, +a).$$

Furthermore, it is also related to the CDF class proposed in [7], which is based on the exponential function. However, the unit interval variants thus proposed differ from the design of our CDF. As will be seen, the distribution proposed here is much more flexible and exhibits both positive and negative skewness. Moreover, as will be seen below, the hazard rate function (HRF) of the proposed model purely yields an increasing failure rate (IFR) behavior, or all values of $\lambda > 0$ thus belong to the decreasing mean residual life (DMRL) class.

The rest of the manuscript is organized as follows. In the next section, the basic stochastic properties of the proposed distribution are presented. The mode, quantiles, HRF, and characterization of the new distribution, among other properties, are examined. Section 3 shows the procedure for estimating the parameters of the proposed distribution using the maximum likelihood (ML) method, along with a Monte Carlo simulation study. Applications to a number of real-world datasets are given in Section 4, while the last section provides some concluding remarks.

2. The Proposed Unit Exponential Distribution

Let X be a bounded RV, and without loss of generality, it is convenient that values of X belong to the unit interval $[0, 1]$. Also, suppose that the CDF of the RV X is defined by the following equality:

$$F(x) = \begin{cases} 1 - \exp\left[\alpha\left(1 - \left(\frac{1+x}{1-x}\right)^\beta\right)\right], & 0 \leq x < 1; \\ 1, & x = 1; \end{cases} \quad (2)$$

where $\alpha, \beta > 0$. The CDF given by Equation (2) is called the unit exponential distribution (UED) (with the parameters α and β) and referred to as the UED (α, β) . Note that the UED is related to the epsilon function defined in Equation (1). Indeed, when taking $a = 1$ and $\beta = \lambda/2$, Equation (2) becomes

$$F(x) = 1 - \exp[\alpha(1 - \varepsilon_{2\beta,1}(x))],$$

when $0 \leq x < 1$. Note that in this form, the function $F(x)$ represents the composition of the CDF of the so-called one-shifted exponential distribution [26] and the epsilon function mentioned above. At the same time, it is obvious that $F(x)$ approaches 0 and 1 when $x \rightarrow 0$ and $x \rightarrow 1$, respectively, and thus represents a valid unit CDF. Graphical representations of the CDFs of the UED for different parameters α and β are shown in Figure 1. It portrays that for $\alpha \rightarrow 0$ and $\beta \geq 3$, the CDF curve is concave (bent inward), while for $\alpha \rightarrow 1$, the CDF curve is convex (bent outward).

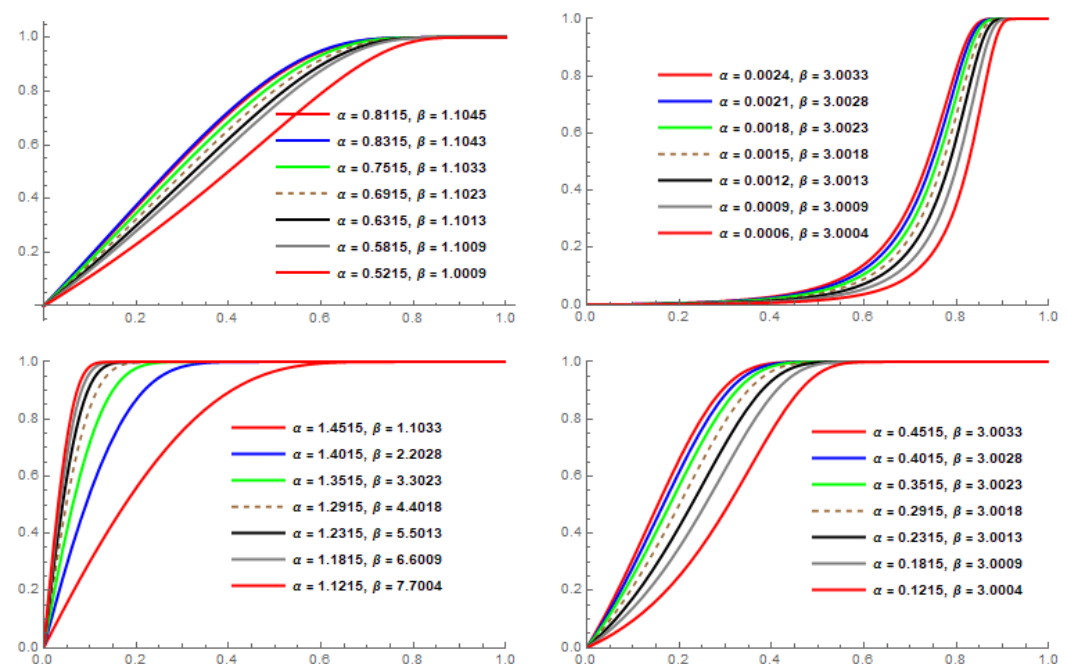


Figure 1. Plots of the CDFs of the UED for varying parameters.

By differentiating the CDF given by Equation (2), the probability density function (PDF) of the UED when $0 \leq x < 1$ can be easily obtained as follows:

$$f(x) = \frac{2\alpha\beta}{1-x^2} \left(\frac{1+x}{1-x}\right)^\beta \bar{F}(x). \quad (3)$$

Here, $\bar{F}(x) = 1 - F(x)$ is the tail of the CDF $F(x)$. Notice that the UED has two parameters $\alpha, \beta > 0$, where one is like a dispersion and the other is like a shape parameter. Also, this PDF structure is similar to one of the simpler forms of the so-called proper dispersion models introduced in [7], but it does not belong to that class.

2.1. Properties of the Model

In practice, it is required that the proposed UED, whose PDF is defined by Equation (3), presents flexibility to describe the data adequately. In this regard, it exhibits negative and positive skewness for all values of $\alpha > 0$ and $\beta > 0$. The flexibility property of the UED can be visualized as in Figure 2, where the various cases of the appropriate PDF are shown, depending on the parameter values α and $\beta > 0$. These plots show the different skewness possibilities and the existence of modes of the UED that can be used to fit some real-world datasets.

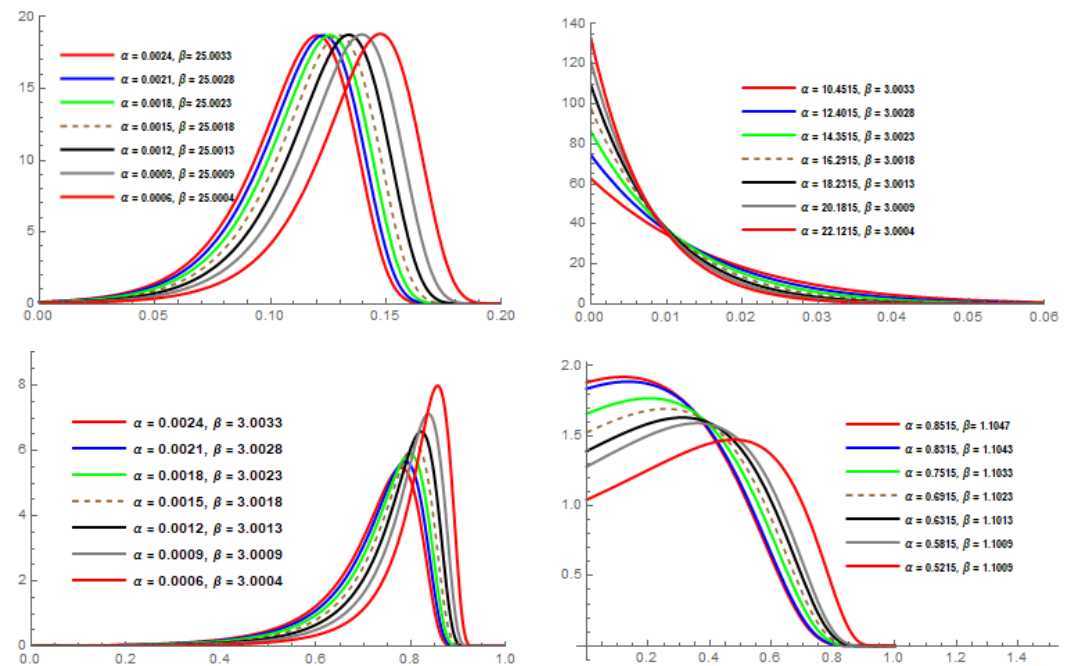


Figure 2. Plots of the PDFs of the UED for varying parameters.

2.1.1. Quantile

As a first property, the quantile function of the UED is quite manageable. By inverting the CDF $F(x)$, given by Equation (2), the quantile function is determined as follows:

$$Q(y) = F^{-1}(y) = \frac{(1 - \ln(1 - y)/\alpha)^{1/\beta} - 1}{(1 - \ln(1 - y)/\alpha)^{1/\beta} + 1}, \quad y \in (0, 1).$$

Thanks to this function, the median of the UED is given by

$$M_e = Q(1/2) = \frac{(1 + \ln 2/\alpha)^{1/\beta} - 1}{(1 + \ln 2/\alpha)^{1/\beta} + 1}.$$

Using $Q(y)$, we are able to define various measures of skewness and kurtosis, as well as important actuarial measures (see, for example, [2,27]).

2.1.2. Mode

Note that Figure 2 shows that the PDF of the proposed model can have (at most) one mode. To identify this property, we should prove the following result, which collects these findings and their implications:

Proposition 1. The PDF $f(x)$, given by Equation (3), has a unique mode if and only if $0 < \alpha < 1$. Otherwise, the UED does not have any modes.

Proof. The mode of the PDF $f(x)$ is a solution of the equation $f'(x) = 0$, which after certain calculations and simplification becomes

$$x + \beta - \alpha\beta\left(\frac{1+x}{1-x}\right)^\beta = 0. \quad (4)$$

If we denote by $\psi(x)$ the left-hand side of Equation (4), then the following is easily obtained:

$$\lim_{x \rightarrow 1^-} \psi(x) < 0 \quad \text{and} \quad \lim_{x \rightarrow 0^+} \psi(x) = \beta(1 - \alpha).$$

Obviously, the inequalities $0 < \alpha < 1$ and $\beta > 0$ give $\beta(1 - \alpha) > 0$. Then, Equation (4) has real solutions, which guarantee that $f(x)$ has at least one mode. Next, the function $\psi(x)$ defined above has the derivative

$$\psi'(x) = 1 - \frac{2\alpha\beta^2}{1-x^2} \left(\frac{1+x}{1-x}\right)^\beta.$$

Note that $\psi'(x)$ is strictly decreasing because

$$\psi''(x) = -\frac{4\alpha\beta^2(x+\beta)}{(1-x^2)^2} \left(\frac{1+x}{1-x}\right)^\beta < 0.$$

This fact then implies that the previously detected mode is unique. $\square$

2.1.3. Behavior of the PDF at $x \rightarrow 0^+$ and $x \rightarrow 1^-$

The behavior of the PDF $f(x)$ at the ends of the unit interval (i.e., when $x \rightarrow 0^+$ and $x \rightarrow 1^-$) indicates how $f(x)$ converges or not in these limits. In terms of data modeling, these facts would reflect the empirical limits on the extremes that the data show. At the limit $x \rightarrow 0^+$, according to Equations (2) and (3), the following is easily obtained:

$$\lim_{x \rightarrow 0^+} f(x) = 2\alpha\beta.$$

On the other hand, to analyze the limit of $f(x)$ at $x \rightarrow 1^-$, we observe the function $\ln f(x)$, which can be written as

$$\begin{aligned} \ln f(x) &= \ln(2\alpha\beta) + (\beta - 1) \ln(1+x) - (\beta + 1) \ln(1-x) + \alpha \left(1 - \left(\frac{1+x}{1-x}\right)^\beta\right) \\ &= \frac{1}{(1-x)^\beta} \left((1-x)^\beta (\ln(2\alpha\beta) + (\beta - 1) \ln(1+x) - (\beta + 1) \ln(1-x) + \alpha) \right. \\ &\quad \left. - \alpha(1+x)^\beta \right). \end{aligned}$$

Hence, we obtain

$$\lim_{x \rightarrow 1^-} (1-x)^\beta \ln f(x) = -\alpha 2^\beta,$$

which implies that in a data representation, the data would decay at exponential rates when $x \rightarrow 1^-$.

2.1.4. Moments

Let X be an RV with the CDF given by Equation (2). Then, the r th moment of X , using partial integration, can be expressed as follows:

$$\begin{aligned} E(X^r) &= \int_0^1 x^r dF(x) = \int_1^0 x^r d(1 - F(x)) = r \int_0^1 x^{r-1} (1 - F(x)) dx \\ &= r \exp(\alpha) \int_0^1 x^{r-1} \exp \left[-\alpha \left(\frac{1+x}{1-x} \right)^\beta \right] dx. \end{aligned}$$

This integral can be determined numerically with the use of many pieces of software, such as R, MATHEMATICA, and MATLAB. The following result proposes a series expansion of $E(X^r)$ that can be used for numerical approximation:

Proposition 2. The r th moment of X can be expanded as follows:

$$E(X^r) = \frac{2r\alpha^{1/\beta} \exp(\alpha)}{\beta} \sum_{k=0}^{r-1} \sum_{\ell=0}^{+\infty} \binom{r-1}{k} \binom{-(r+1)}{\ell} (-1)^k \alpha^{(k+\ell+1)/\beta} \Gamma \left(-\frac{k+\ell+1}{\beta}, \alpha \right),$$

where $\Gamma(a, x)$ denotes the upper incomplete gamma function (i.e., $\Gamma(a, x) = \int_x^{+\infty} t^{a-1} \exp(-t) dt$).

Proof. By applying the change in the variable $y = (1+x)/(1-x)$, we have

$$E(X^r) = 2r \exp(\alpha) \int_1^{+\infty} \frac{(y-1)^{r-1}}{(y+1)^{r+1}} \exp(-\alpha y^\beta) dy. \quad (5)$$

Then, using the “generalized version” of the binomial formula two times in a row, since $y > 1$, we find

$$\begin{aligned} \frac{(y-1)^{r-1}}{(y+1)^{r+1}} &= y^{-2} \frac{(1-1/y)^{r-1}}{(1+1/y)^{r+1}} \\ &= y^{-2} \left[\sum_{k=0}^{r-1} \binom{r-1}{k} (-1)^k y^{-k} \right] \left[\sum_{\ell=0}^{+\infty} \binom{-(r+1)}{\ell} y^{-\ell} \right] \\ &= \sum_{k=0}^{r-1} \sum_{\ell=0}^{+\infty} \binom{r-1}{k} \binom{-(r+1)}{\ell} (-1)^k y^{-(k+\ell+2)}. \end{aligned} \quad (6)$$

Also, with the change in the variable $z = \alpha y^\beta$, the following is obtained:

$$\begin{aligned} \int_1^{+\infty} y^{-(k+\ell+2)} \exp(-\alpha y^\beta) dy &= \frac{\alpha^{(k+\ell+1)/\beta}}{\beta} \int_\alpha^{+\infty} z^{-(k+\ell+1)/\beta-1} \exp(-z) dz \\ &= \frac{\alpha^{(k+\ell+1)/\beta}}{\beta} \Gamma \left(-\frac{k+\ell+1}{\beta}, \alpha \right). \end{aligned} \quad (7)$$

Therefore, by substituting Equations (6) and (7) into Equation (5), as well as by inverting the sign of the integral and the sum, the desired result is obtained. $\square$

2.1.5. Failure (Hazard) Rate Function

The HRF of the UED is given by

$$h(x) = \frac{f(x)}{\bar{F}(x)} = \frac{2\alpha\beta}{1-x^2} \left(\frac{1+x}{1-x} \right)^\beta. \quad (8)$$

When $x \rightarrow 0^+$, the limit of $h(x)$ is $2\alpha\beta > 0$, and when $x \rightarrow 1^-$, the limit is $+\infty$. Thus, this function is strictly increasing, as can be seen in Figure 3, meaning that when x increases, the frequency at which an engineered system or component fails also increases.

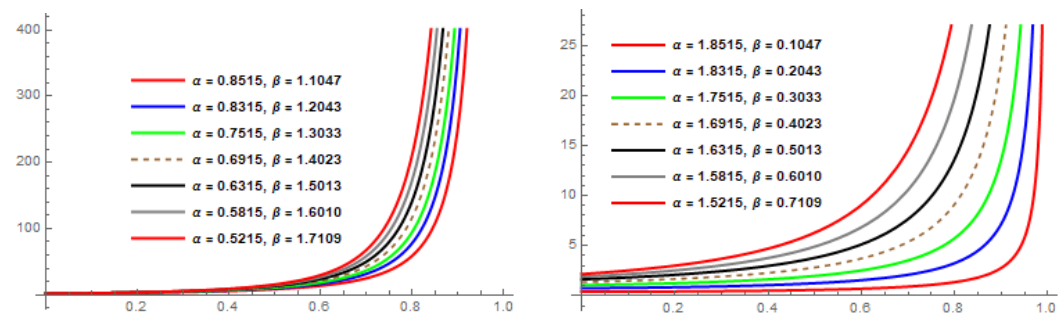


Figure 3. Plots of the HRFs of the UED for varying parameters.

2.2. Characterizations

To interpret the HRF realistically, we shall try to characterize Equation (3) with hazard and mean residual life functions. Characterization in general terms implies that under certain conditions, a family of distributions is the only one possessing a designated property. Researchers can identify the actual probability distribution with the help of characterization. For detailed study, readers are referred to the works of Ahsanullah et al. [28,29] and Hamedani [30]. In this regard, we characterize the proposed model with the HRF and truncated moments, and the characterizing conditions are defined as follows:

Proposition 3. The RV $X : \Omega \rightarrow (0, +\infty)$ has a continuous PDF $f(x)$ if and only if the HRF $h(x)$ satisfies the following equation:

$$\frac{f'(x)}{f(x)} = \frac{h'(x)}{h(x)} - h(x). \quad (9)$$

Proof. According to the definition of the HRF, given by the first equality in Equation (8), it follows that

$$\frac{h'(x)}{h(x)} = \frac{f'(x)\bar{F}(x) + f^2(x)}{\bar{F}^2(x)} \cdot \frac{\bar{F}(x)}{f(x)} = \frac{f'(x)}{f(x)} + h(x).$$

Thus, the statement of proposition immediately follows. $\square$

Proposition 4. The RV $X : \Omega \rightarrow (0, +\infty)$ has a UED (α, β) if and only if the HRF $h(x)$, defined by Equation (8), satisfies the following equation:

$$\frac{h'(x)}{(h(x))^2} = \frac{x + \beta}{\alpha\beta} \left(\frac{1-x}{1+x} \right)^\beta. \quad (10)$$

Proof. *Necessity:* Assume that $X \sim \text{UED}(\alpha, \beta)$, with the PDF $f(x)$ defined by Equation (3). Then, the logarithm of this PDF, in the same way as in Section 2.1.3, can be expressed as:

$$\ln(f(x)) = \ln(2\alpha\beta) + (\beta - 1)\ln(1+x) - (\beta + 1)\ln(1-x) + \alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right).$$

By differentiating both sides of this equality with respect to x , we obtain

$$\frac{f'(x)}{f(x)} = \frac{\beta - 1}{1+x} + \frac{\beta + 1}{1-x} - \frac{2\alpha\beta}{(1-x)^2} \left(\frac{1+x}{1-x} \right)^{\beta-1} = \frac{2}{1-x^2} \left(x + \beta - \alpha\beta \left(\frac{1+x}{1-x} \right)^\beta \right). \quad (11)$$

Thus, according to Equations (8) and (9), it follows that

$$\frac{h'(x)}{h(x)} = \frac{f'(x)}{f(x)} + h(x) = \frac{2(x + \beta)}{1-x^2},$$

which after certain simplification yields Equation (10).

Sufficiency: Suppose that Equation (10) holds. After integration, it can be rewritten as follows:

$$\int \frac{h'(x)}{(h(x))^2} dx = \int \frac{x + \beta}{\alpha\beta} \left(\frac{1-x}{1+x} \right)^\beta dx,$$

That is, we have

$$-\frac{1}{h(x)} = \frac{x^2 - 1}{2\alpha\beta} \left(\frac{1-x}{1+x} \right)^\beta.$$

From the above equation, we obtain the HRF $h(x)$ as shown in Equation (8). Furthermore, by replacing this function in Equation (9), and after integration, we obtain

$$\begin{aligned} \int \frac{f'(x)}{f(x)} dx &= 2 \int \left[\frac{x + \beta}{1-x^2} - \frac{\alpha\beta}{1-x^2} \left(\frac{1+x}{1-x} \right)^\beta \right] dx + C_1 \\ &= (\beta - 1) \ln(1+x) - (\beta + 1) \ln(1-x) - \alpha \left(\frac{1+x}{1-x} \right)^\beta + C_1, \end{aligned}$$

that is, we have

$$f(x) = \frac{\exp \left[C_1 - \alpha \left(\frac{1+x}{1-x} \right)^\beta \right]}{1-x^2} \left(\frac{1+x}{1-x} \right)^\beta.$$

Another integration implies that

$$F(x) = \int f(x) dx + C_2 = -\frac{\exp \left[C_1 - \alpha \left(\frac{1+x}{1-x} \right)^\beta \right]}{2\alpha\beta} + C_2,$$

whereby from the conditions $F(0) = 0$ and $F(1) = 1$, the constants $C_1 = \alpha + \ln(2\alpha\beta)$ and $C_2 = 1$ are obtained. Thus, the function $F(x)$ is indeed the CDF from $UED(\alpha, \beta)$, which completes the proof. $\square$

The following theorem was used in [31] as well as [28,29] in order to characterize different univariate continuous distributions:

Theorem 1. Let $(\Omega; \mathcal{F}; P)$ be a given probability space, and let $\mathbb{H} = [a, b]$ be an interval for some $a < b$, where $a = -\infty$ and $b = +\infty$ might as well be allowed. Also, let $X : \Omega \rightarrow \mathbb{H}$ be a continuous RV with the CDF $F(x)$ and $g(x)$ and $t(x)$ be two real functions defined on $\mathbb{H}$ and such that

$$\mathbb{E}[g(X)|X \geq x] = \xi(x) \mathbb{E}[t(X)|X \geq x], \quad x \in \mathbb{H}$$

is defined with some real function $\xi(x)$. Assume that $g(x), t(x) \in C^1(\mathbb{H})$, $\xi(x) \in C^2(\mathbb{H})$, and $F(x)$ is a twice continuously differentiable and strictly monotone function on the set $\mathbb{H}$. Finally, assume that the equation $t(x)\xi(x) = g(x)$ has no real solution in the interior of $\mathbb{H}$. Then, $F(x)$ is uniquely determined by the functions $g(x), t(x)$, and $\xi(x)$ as follows:

$$F(x) = C \int_0^x \left| \frac{\xi'(u)}{\xi(u)t(u) - g(u)} \right| e^{-s(u)} du, \quad (12)$$

where the function $s(x)$ is a solution of the differential equation

$$s'(x) = \frac{\xi'(x)t(x)}{\xi(x)t(x) - g(x)},$$

and C is a constant such that $\int_{\mathbb{H}} dF(x) = 1$.

Now, we discuss the characterization of the UED based on Theorem 1 and some simple relationship between two functions and the RV $X \sim UED(\alpha, \beta)$.

Proposition 5. Let $X : \Omega \rightarrow [0, 1)$ be a continuous RV and

$$\begin{aligned} t(x) &= 3 \exp \left[2\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], & x \in [0, 1) \\ g(x) &= 2 \exp \left[\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], & x \in [0, 1). \end{aligned}$$

The RV X has a PDF defined by Equation (3) if and only if there exists a function $\xi(x)$, defined as in Theorem 1, that satisfies the differential equation

$$\frac{\xi'(x)}{\xi(x)t(x) - g(x)} = \frac{2\alpha\beta}{1-x^2} \left(\frac{1+x}{1-x} \right)^\beta \exp \left[-2\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], \quad 0 \leq x < 1. \quad (13)$$

Proof. *Necessity:* For the RV $X \sim \text{UED}(\alpha, \beta)$, with the CDF and PDF given by Equations (2) and (3), respectively, after a certain computation, we obtain

$$\begin{aligned} (1 - F(x))\mathbb{E}[t(X)|X \geq x] &= 3e^{ar(x;\beta)} \int_x^1 \frac{2\alpha\beta}{1-u^2} \left(\frac{1+u}{1-u} \right)^\beta e^{3ar(u;\beta)} du \\ &= \exp \left[4\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], \\ (1 - F(x))\mathbb{E}[g(X)|X \geq x] &= 2e^{ar(x;\beta)} \int_x^1 \frac{2\alpha\beta}{1-u^2} \left(\frac{1+u}{1-u} \right)^\beta e^{2ar(u;\beta)} du \\ &= \exp \left[3\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], \end{aligned}$$

where $0 < x < 1$ and $r(x) := 1 - \left(\frac{1+x}{1-x} \right)^\beta$. This implies that

$$\xi(x) := \frac{\mathbb{E}(g(x)|X \geq x)}{\mathbb{E}(t(x)|X \geq x)} = \exp \left[-\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right], \quad 0 < x < 1, \quad (14)$$

that is, we have

$$\xi(x)t(x) - g(x) = 3e^{ar(x;\beta)} - 2e^{ar(x;\beta)} = \exp \left[\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right) \right] > 0, \quad 0 < x < 1.$$

Hence, Equation (13) clearly holds.

Sufficiency: If the function $\xi(x)$ satisfies the differential Equation (13), then it follows that

$$s'(x) = \frac{\xi'(x)t(x)}{\xi(x)t(x) - g(x)} = \frac{6\alpha\beta}{1-x^2} \left(\frac{1+x}{1-x} \right)^\beta, \quad 0 < x < 1,$$

Therefore, one can take

$$s(x) = -3\alpha \left(1 - \left(\frac{1+x}{1-x} \right)^\beta \right).$$

Using Equation (12), it is easy to obtain that the RV X has a PDF given by Equation (3). $\square$

According to the previous proposition, one immediately obtains the following:

Corollary 1. Let $X : \Omega \rightarrow [0, +\infty)$ be a continuous RV and functions $t(x)$ and $g(x)$ be given as in Proposition 5. Then, $X \sim \text{UED}(\alpha, \beta)$, with the PDF as shown in Equation (3) if and only if the function $\xi(x)$ has the form in Equation (14).

3. Estimation and Simulation Procedures

Let us assume that $x_1, \dots, x_n$ are observed values of the sample of size n taken from the $UED(\alpha, \beta)$. We propose the maximum likelihood method for estimating the pair of parameters (α, β) . This means that the estimates of those parameters are the ones that maximize the likelihood function

$$L(\alpha, \beta | x_1, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

As is known, this solution also corresponds to the one that maximizes the log-likelihood function; in other words, it maximizes

$$l = l(\alpha, \beta | x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i).$$

By differentiating the function l with respect to each parameter, the estimators of α and β can be obtained by solving the coupled equations

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \left(1 - \left(\frac{1+x_i}{1-x_i} \right)^\beta \right) = 0 \\ \frac{\partial l}{\partial \beta} &= \frac{n}{\beta} + \sum_{i=1}^n \ln \left(\frac{1+x_i}{1-x_i} \right) - \alpha \sum_{i=1}^n \left(\frac{1+x_i}{1-x_i} \right)^\beta \ln \left(\frac{1+x_i}{1-x_i} \right) = 0. \end{aligned}$$

From the first equation, we obtain

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

and by replacing this output in the second coupled equation, we obtain

$$\frac{n}{\beta} + \sum_{i=1}^n \ln \left(\frac{1+x_i}{1-x_i} \right) + \frac{\sum_{i=1}^n \left(\frac{1+x_i}{1-x_i} \right)^\beta \ln \left(\frac{1+x_i}{1-x_i} \right)}{1 - \frac{1}{n} \sum_{i=1}^n \left(\frac{1+x_i}{1-x_i} \right)^\beta} = 0.$$

Obviously, the last equation has only β as an unknown parameter. Now, by denoting $z_i = (1+x_i)/(1-x_i) > 1, i = 1, \dots, n$, and

$$\mathcal{L}(\beta) = \frac{n}{\beta} + \sum_{i=1}^n \ln z_i + \frac{\sum_{i=1}^n z_i^\beta \ln z_i}{1 - \frac{1}{n} \sum_{i=1}^n z_i^\beta},$$

then by applying the L'Hopital's rule, one obtains

$$\begin{aligned} \lim_{\beta \rightarrow 0^+} \mathcal{L}(\beta) &= \sum_{i=1}^n \ln z_i + n \lim_{\beta \rightarrow 0^+} \frac{\sum_{i=1}^n (1 - z_i^\beta + \beta z_i^\beta \ln z_i)}{\beta \sum_{i=1}^n (1 - z_i^\beta)} \\ &= \sum_{i=1}^n \ln z_i + n \lim_{\beta \rightarrow 0^+} \frac{\sum_{i=1}^n (-z_i^\beta \ln z_i + \ln z_i)}{\sum_{i=1}^n (1 - z_i^\beta - \beta z_i^\beta \ln z_i)} \\ &= \sum_{i=1}^n \ln z_i + n \lim_{\beta \rightarrow 0^+} \frac{\sum_{i=1}^n (-z_i^\beta \ln^2 z_i)}{\sum_{i=1}^n (-z_i^\beta \ln z_i - \ln z_i)} \\ &= \sum_{i=1}^n \ln z_i + \frac{n}{2} \cdot \frac{\sum_{i=1}^n \ln^2 z_i}{\sum_{i=1}^n \ln z_i} > 0. \end{aligned}$$

On the other hand, assuming that $z_1 > \max\{z_2, \dots, z_n\}$, it follows that

$$\begin{aligned}\lim_{\beta \rightarrow +\infty} \mathcal{L}(\beta) &= \sum_{i=1}^n \ln z_i + \lim_{\beta \rightarrow +\infty} \frac{\ln z_1 + \sum_{i=2}^n \left(\frac{z_i}{z_1}\right)^\beta \ln z_i}{z_1^{-\beta} - \frac{1}{n} - \frac{1}{n} \sum_{i=2}^n \left(\frac{z_i}{z_1}\right)^\beta} \\ &= \sum_{i=1}^n \ln z_i - n \ln z_1 < 0.\end{aligned}$$

Hence, equation $\mathcal{L}(\beta) = 0$ has at least one solution, and it can be solved numerically, for instance, by using the Newton–Raphson algorithm. This task may be performed using the function “uniroot” available in the statistical programming software “R” (version 4.3.1). Once β is estimated, this output can be used for estimating α .

For computing the interval estimators for $\theta = (\alpha, \beta)'$ and testing hypotheses with these parameters, we find the observed matrix information:

$$I(\theta) = - \begin{pmatrix} \frac{\partial^2 l(\theta)}{\partial \alpha^2} & \frac{\partial^2 l(\theta)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l(\theta)}{\partial \beta \partial \alpha} & \frac{\partial^2 l(\theta)}{\partial \beta^2} \end{pmatrix},$$

where

$$\begin{aligned}\frac{\partial^2 l(\theta)}{\partial \alpha^2} &= -\frac{n}{\alpha^2} \\ \frac{\partial^2 l(\theta)}{\partial \alpha \partial \beta} &= \frac{\partial^2 l(\theta)}{\partial \beta \partial \alpha} = -\sum_{i=1}^n \left(\frac{1+x_i}{1-x_i}\right)^\beta \ln\left(\frac{1+x_i}{1-x_i}\right) \\ \frac{\partial^2 l(\theta)}{\partial \beta^2} &= -\frac{n}{\beta^2} - \alpha \sum_{i=1}^n \left(\frac{1+x_i}{1-x_i}\right)^\beta \ln^2\left(\frac{1+x_i}{1-x_i}\right).\end{aligned}$$

Note that $I(\hat{\theta})$ is a consistent estimator of the expected Fisher information matrix $E[I(\theta)]$ (see, for example, [32]). Under some suitable conditions, the approximation to a normal distribution $\hat{\theta} \approx \mathcal{N}(\theta, I(\hat{\theta})^{-1})$ holds, and more generally, we have

$$a' \hat{\theta} \approx \mathcal{N}(a' \theta, a' I(\hat{\theta})^{-1} a),$$

for any vector $a = (a_1, a_2)'$. By choosing $a = (1, 1)'$, we find the $100 \times (1 - \delta) \%$ confidence interval:

$$\theta_i \pm z_{\delta/2} \sqrt{(I(\hat{\theta})^{-1})_{ii}},$$

where $0 < \delta < 1$ and $z_{\delta/2}$ is the $1 - \delta/2$ quantile of the standard normal distribution.

Simulation Study

In this part, we shall discuss the effectiveness of the proposed MLE procedure, which will be used in application for better predictions of a phenomenon. In this regard, we considered four sets of parameters and conducted a Monte Carlo simulation with 20,000 replications in order to generate samples of various sizes (i.e., $n = 25, 50, 150, 350, 500$) from the UED (α, β) . The parameter combinations are listed below:

Set-I: $\alpha = 0.9856, \beta = 0.2178$;
 Set-II: $\alpha = 1.8986, \beta = 0.3218$;
 Set-III: $\alpha = 2.4390, \beta = 2.5145$;
 Set-IV: $\alpha = 0.4390, \beta = 1.5145$.

For all of them, the MLE estimates were obtained by using MATHEMATICA 13.0 software. The simulation results are portrayed in Tables 1–4, where they are compiled according to the following definitions:

$$\text{Bias} := \mathbb{E}(\hat{\Theta}) - \Theta;$$

$$\text{Mean square error (MSE)} := \mathbb{E}((\hat{\Theta} - \Theta)^2);$$

$$\text{Lower Confidence Limit} := \text{LCL} = \hat{\Theta} - z_{\frac{\delta}{2}} \frac{\sqrt{\text{Var}(\hat{\Theta})}}{n};$$

$$\text{Upper Confidence Limit} := \text{UCL} = \hat{\Theta} + z_{\frac{\delta}{2}} \frac{\sqrt{\text{Var}(\hat{\Theta})}}{n},$$

where $\Theta = (\alpha, \beta)$. From these tables, there is evidence that both the bias and MSE of the MLE estimates tended toward zero as the sample sizes increased, whereas the 95% confidence limits became compact as the sample size increased.

Table 1. Mean, bias, MSE, LCL, and UCL for Set-I.

Sample Size	Parameter	Estimate	Bias	MSE	LCL	UCL
$n = 25$	α	0.9417	−0.0439	0.0111	0.9335	0.9499
	β	0.2180	0.0013	0.00008	0.2172	0.2189
$n = 50$	α	0.9511	−0.0344	0.0068	0.9479	0.9544
	β	0.2189	0.0012	0.00007	0.2186	0.2193
$n = 150$	α	0.9655	−0.0200	0.0032	0.9648	0.9663
	β	0.2192	0.0014	0.00004	0.2192	0.2192
$n = 350$	α	0.9685	−0.0171	0.0022	0.9683	0.9688
	β	0.2194	0.0016	0.00003	0.2194	0.2194
$n = 500$	α	0.9729	−0.0126	0.0015	0.9728	0.9732
	β	0.2194	0.0016	0.00002	0.2194	0.2194

Moreover, Table 1 shows a downward bias for $\hat{\alpha}$ and an upward one for $\hat{\beta}$. Similarly, the MSE approached zero as the sample size increased.

Table 2. Mean, bias, MSE, LCL, and UCL for Set-II.

Sample Size	Parameter	Estimate	Bias	MSE	LCL	UCL
$n = 25$	α	1.8547	−0.0438	0.0472	1.8376	1.8717
	β	0.3140	−0.0077	0.0004	0.3125	0.3155
$n = 50$	α	1.8902	−0.0084	0.0230	1.8843	1.8962
	β	0.3149	−0.0068	0.0003	0.3143	0.3156
$n = 150$	α	1.9204	0.0217	0.0117	1.9161	1.9246
	β	0.3167	−0.0050	0.0001	0.3163	0.3172
$n = 350$	α	1.9346	0.0359	0.0071	1.9341	1.9352
	β	0.3171	−0.0046	0.0001	0.3172	0.3172
$n = 500$	α	1.9337	0.0351	0.0063	1.9334	1.9342
	β	0.3171	−0.0047	0.00008	0.3170	0.3171

Also, Table 2 portrays a downward bias for $\hat{\alpha}$ and $\hat{\beta}$ when the sample sizes were less than or equal to 50, while there was an upward bias for $\hat{\alpha}$ when the sample size increased. However, the MSE approached zero as the sample size increased.

Table 3. Mean, bias, MSE, LCL, and UCL for Set-III.

Sample Size	Parameter	Estimate	Bias	MSE	LCL	UCL
$n = 25$	α	2.3256	−0.1133	0.0578	2.307	2.3445
	β	2.4995	−0.0149	0.0126	2.4908	2.5084
$n = 50$	α	2.3539	−0.0850	0.0316	2.3472	2.3620
	β	2.4871	−0.0274	0.0136	2.4826	2.4916
$n = 150$	α	2.3915	−0.0475	0.0123	2.3900	2.3929
	β	2.5121	−0.0023	0.0048	2.5113	2.5132
$n = 350$	α	2.4044	−0.0345	0.0065	2.4040	2.40491
	β	2.5155	0.0010	0.0028	2.5152	2.5158
$n = 500$	α	2.4051	−0.0338	0.0052	2.4048	2.4055
	β	2.5180	0.0035	0.0023	2.5179	2.5183

In the case of Set-III, shown in Table 3, the bias was downward for $\hat{\alpha}$ for all sample sizes. On the contrary, it was upward for $\hat{\beta}$ when the sample sizes were higher, usually for those greater than 150. Notice that all biases were negligibly small and approached zero as the sample size increased.

Table 4. Mean, bias, MSE, LCL, and UCL for Set-IV.

Sample Size	Parameter	Estimate	Bias	MSE	LCL	UCL
$n = 25$	α	0.4173	−0.0217	0.0024	0.4134	0.4212
	β	1.5168	0.0023	0.0037	1.5120	1.5215
$n = 50$	α	0.4251	−0.0138	0.0013	0.4237	0.4264
	β	1.5166	0.0021	0.0027	1.5146	1.5186
$n = 150$	α	0.4285	−0.0105	0.0007	0.4281	0.4288
	β	1.5205	0.0060	0.0015	1.5200	1.5210
$n = 350$	α	0.4314	−0.0076	0.0004	0.4313	0.4315
	β	1.5236	0.0091	0.0009	1.5234	1.5238
$n = 500$	α	0.4332	−0.0058	0.0003	0.4332	0.4333
	β	1.5203	0.0058	0.0009	1.5202	1.5203

Similarly, Table 4 shows a downward bias for $\hat{\alpha}$ and an upward bias for $\hat{\beta}$, but all biases were negligibly small and approached zero as the sample size increased, and the same is true for the MSE. In summary, the above results show that the MLE is a suitable estimation method for realistic forecasting.

4. Model Compatibility and Its Application to Real-World Data

Here, the possibility of applying the UED model in terms of modeling empirical distributions of some real-world processes is discussed in more detail. To that end, by using several typical statistical indicators, the quality of fitting with the UED was also checked. The obtained results were also compared with the results of fitting using some of the previously known unit interval probability distributions, which additionally checked the possibility of applying the UED.

4.1. Measures of Goodness-of-Fit

In order to test the null hypothesis $H_0 : F_n(x) = F_0(x)$, where $F_n(x)$ is the empirical CDF and $F_0(x)$ is the CDF of some specified (theoretical) distribution, usually some well-known statistical tests are used. In order to test the hypothesis that some real-world data are taken from the UED (i.e., from some other stochastic distribution), the following statistical tests are used here:

- The Kolmogorov–Smirnov (KS) test, whose test-statistics are defined by

$$KS = \max_{1 \leq i \leq k} \left\{ \frac{i}{k} - z_i, z_i - \frac{i-1}{k} \right\},$$

where k denotes the number of classes and z_i represents the values of the theoretical CDF.

- The Anderson–Darling (AD_0^*) test, which usually attaches more mass to the distributions tails and whose test-statistics are

$$A_0^* = \left(\frac{2.25}{k^2} + \frac{0.75}{k} + 1 \right) \left\{ -k - \frac{1}{k} \sum_{i=1}^k (2i-1) \ln(z_i(1-z_{k-i+1})) \right\}.$$

- The Cramér–von Mises (CVM_0^*)-test is a derived version of the KS test, with test-statistics defined by

$$W_0^* = \sum_{i=1}^K \left(z_i - \frac{2i-1}{2k} \right)^2 + \frac{1}{12k}.$$

Additionally, in order to check the quality of fitting certain real-world data using the UED (i.e., some other distribution), the following indicators were used:

- The Akaike information criterion (AIC), defined as

$$AIC = 2m - 2\ell(\hat{\Theta}),$$

where m denotes the number of parameters.

- The corrected Akaike information criterion (AICc), expressed as

$$AICc = AIC + \frac{2m(m+1)}{n-m-1}.$$

- The Bayesian information criterion (BIC), which is defined as

$$BIC = m \ln(n) - 2\ell(\hat{\Theta}).$$

- The Hannan–Quinn information criterion (HQIC), expressed as

$$HQIC = -2\ell(\hat{\Theta}) + 2m \ln(\ln(m)).$$

- The consistent Akaike information criterion (CAIC), given as

$$CAIC = -2\ell(\hat{\Theta}) + m(\ln(n) + 1).$$

- The Vuong test was also used for model selection purposes.

For comprehensive details about these measures, readers may refer to Akaike [33], Hussain et al. [34], Murthy et al. [35], and Vuong [36], respectively.

4.2. Comparative Models

We also compared the proposed UED model with well-known unit interval models defined by the following PDFs:

- The beta distribution (BD) [4]:

$$f_{\alpha}^{BD}(x) = \frac{1}{B(\alpha, \beta)} x^{\beta-1} (1-x)^{\alpha-1}, \quad \alpha, \beta > 0, \quad 0 < x < 1,$$

- The Johnson S_B distribution (JSBD) [6]:

$$f_{\alpha,\beta}^{\text{JSBD}}(x) = \frac{\beta \exp\left[-\frac{1}{2}(\alpha + \beta \ln(\frac{x}{1-x}))^2 - \beta x\right]}{\sqrt{2\pi x(1-x)}}, \quad \alpha, \beta > 0, \quad 0 < x < 1,$$

- The Kumaraswamy distribution (KwD) [8]:

$$f_{\alpha,\beta}^{\text{KwD}}(x) = \alpha \beta x^{\alpha-1} (1-x^\alpha)^{\beta-1}, \quad \alpha, \beta > 0, \quad 0 < x < 1,$$

- The unit Gompertz distribution (UGoMD) [15]:

$$f_{\alpha,\beta}^{\text{UGoMD}}(x) = \alpha \beta x^{-\alpha-1} e^{-\beta(x^{-\alpha}-1)}, \quad \alpha, \beta > 0, \quad 0 < x < 1.$$

In order to compare the fitting results, we considered four different real-world datasets classified into two sections: (1) environmental and (2) engineering. The results obtained from the statistical analysis of these datasets are discussed below.

4.3. Environmental Datasets

Datasets I and II. The first two datasets were reported by Maiti [37], and they represent the following measured values:

- Soil moisture (Dataset I): 0.0179, 0.0798, 0.0959, 0.0444, 0.0938, 0.0443, 0.0917, 0.0882, 0.0439, 0.049, 0.0774, 0.0171, 0.0305, 0.0757, and 0.0468;
- Permanent wilting points (PWP) (Dataset II): 0.0821, 0.0561, 0.0202, 0.051, 0.0041, 0.0226, 0.0556, 0.0829, 0.0062, 0.0695, 0.0557, 0.0243, 0.0083, 0.0532, and 0.0118.

In this regard, we compiled both the descriptive and theoretical (UED) statistics, which are listed in Tables 5 and 6, respectively. Note that the descriptive statistics of all datasets include the sample size (SS), mean, median, standard deviation (SD), skewness (SK), and kurtosis (KU).

Table 5. Descriptive statistics for Datasets I and II.

Dataset	SS	Mean	Median	SD	SK	KU
I	15	0.0598	0.0490	0.0277	−0.1083	1.6247
II	15	0.0402	0.0510	0.0277	0.1083	1.6247

Table 6. Theoretical statistics from the UED.

Dataset	SS	Mean	Median	SD	SK	KU
I	15	0.0606	0.0621	0.0254	−0.2107	2.3825
II	15	0.0406	0.0384	0.0247	0.2942	2.3050

In addition, the total test time (TTT) plot, introduced in [38], is portrayed in Figure 4 for both datasets. Notice that, in particular, the TTT plot indicates the empirical HRF, portraying an IFR. Tables 5 and 6 also reveal that the theoretical UED statistics as well as the observed descriptive statistics showed remarkable closeness to each other, and it appears that both sets of data can be simulated by the proposed model. Furthermore, it is evident from Figure 5 that neither dataset contained any outliers.

Table 7 portrays that the model proposed by the UED is the best strategy for analyzing the observed dataset (Dataset I) in relation to all other distributions of unit intervals. Namely, although the p value of the KS statistics for the KwD was the highest, the other nonparametric tests, CVM_0^* and AD_0^* , indicate that for the UED, the minimum tested values were obtained. Also, based on the estimated values of the Vuong statistics, given in Table 8, the KwD and UED had an indecisive status. Thus, the UED is the best strategy, which is also confirmed by Figure 6. Similarly, Table 9 portrays that the proposed UED model is also one of the best strategies for the analysis of Dataset II in all aspects. Namely, the test

statistics, including the KS test, CVM_0^* , and AD_0^* , had the lowest values compared with all the selected, previously known interval models. In addition, the Vuong statistic, which compares models based on the likelihood ratio phenomenon, openly supported the UED. Finally, Figure 6 also confirms our claim that the UED is the best strategy. Moreover, Tables 10 and 11 yield the lowest information criterion values for the UED compared with the competing models.

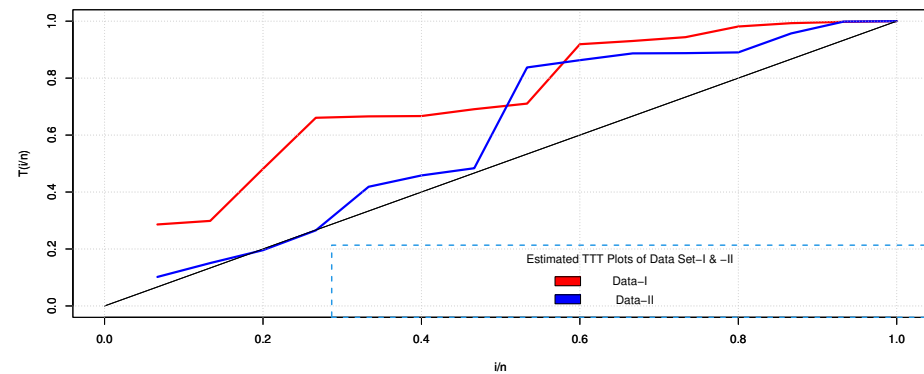


Figure 4. TTT plots of Datasets I and II.

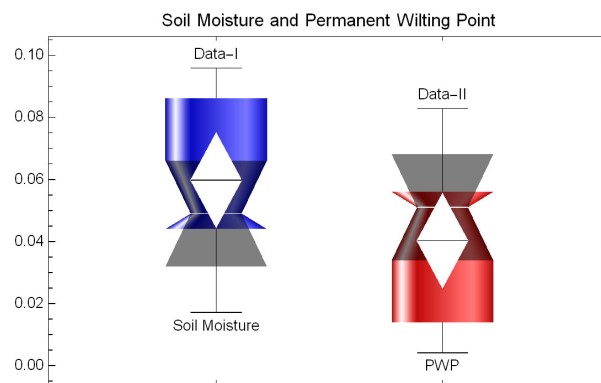


Figure 5. Box plots for datasets I and II.

Table 7. ML estimates and goodness-of-fit statistics for Dataset I.

Distribution	$\hat{\beta}$	$\hat{\alpha}$	CVM_0^*	AD_0^*	KS	p -Value
UED	18.4218	0.0773	0.6239	0.1026	0.2079	0.5361
BD	3.8233	60.2492	0.6858	0.1041	0.2099	0.5232
KwD	719.3842	2.4408	0.6887	0.1109	0.2003	0.5844
JSBD	4.9859	1.7279	0.7751	0.1117	0.2128	0.5056
UGoMD	1.6525	0.0048	1.0587	0.1613	0.2353	0.3769

Table 8. Vuong test statistics for Datasets I and II.

Models	Dataset I	Suitability	Dataset II	Suitability
UED-BD	1.4601	UED	2.5935	UED
UED-KwD	0.9738	Indecisive	3.4585	UED
UED-JSBD	1.5427	UED	1.6793	UED
UED-UGoMD	2.2142	UED	1.5955	UED

Table 9. MLE and goodness-of-fit statistics for Dataset II.

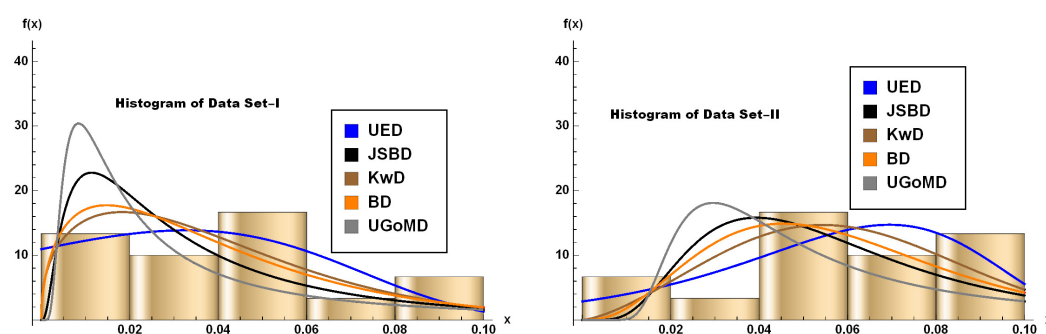
Distribution	$\hat{\beta}$	$\hat{\alpha}$	CVM_0^*	AD_0^*	KS	p -Value
UED	11.8676	0.4607	0.6239	0.1096	0.1960	0.6118
BD	1.5370	36.8071	0.6869	0.1199	0.2481	0.3142
KwD	78.9162	1.4011	0.7074	0.1224	0.2409	0.3487
JSBD	3.5837	1.0177	0.8112	0.1364	0.2619	0.2549
UGoMD	0.9497	0.0219	0.9011	0.1499	0.2386	0.3603

Table 10. Estimates of the maximum log-likelihood and information criteria for Dataset I.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
UED	33.8617	−63.7233	−62.7233	−62.3072	−63.7384	−60.3072
BD	32.8026	−61.6052	−60.6052	−60.1891	−61.6203	−58.1891
KwD	33.3796	−62.7592	−61.7592	−61.3431	−62.7743	−59.3431
JSBD	32.0631	−60.1262	−59.1262	−58.7101	−60.1413	−56.7101
UGoMD	29.6463	−55.2925	−54.2925	−53.8764	−55.3076	−51.8764

Table 11. Estimates of the maximum log-likelihood and information criteria for Dataset II.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
UED	35.2604	−66.5208	−65.5208	−65.1047	−66.5359	−63.1047
BD	34.1097	−64.2194	−63.2194	−62.8033	−64.2345	−60.8033
KwD	34.3392	−64.6784	−63.6784	−63.2623	−64.6935	−61.2623
JSBD	33.0448	−62.0896	−61.0896	−60.6735	−62.1047	−58.6735
UGoMD	31.1648	−58.3296	−57.3296	−56.9135	−58.3447	−54.9135

**Figure 6.** Datasets I and II (given by histograms) fitted via unit interval distributions (given by lines).

4.4. Engineering Datasets

Datasets III and IV. The third and fourth datasets were first introduced and studied in [39] for Burr measurements on iron sheets. For the third dataset of 50 observations of the Burr measurements (in the unit of millimeters), the hole diameter was 12 mm, and the sheet thickness was 3.15 mm. For the fourth dataset of 50 observations, the hole diameter and sheet thickness were 9 mm and 2 mm, respectively. Hole diameter readings were taken for jobs with respect to one hole and then selected and fixed as per a predetermined orientation. These two datasets refer to two different machines being compared, and one can see [39] for the technical details of measuring the datasets. Note that both datasets were also analyzed in [19,40–42]. The descriptive statistics of these datasets, as well as the corresponding theoretical statistics for the UED, are presented in Tables 12 and 13,

respectively. The TTT plot and box plots of the observed data are given in Figures 7 and 8, respectively. It can be observed that Datasets III and IV were positively skewed and platykurtic in nature, which is confirmed by Tables 12 and 13. In addition, from Figure 8, it is evident that the empirical and theoretical aspects of these datasets, in terms of the absence of outliers, are in close agreement and indicate that the proposed model can be used effectively. Such findings are also consolidated within Tables 14 and 15, which show that the UED exhibited minimal values in the almost all cases for the goodness-of-fit statistic, which ensures that the UED is one of the best strategies.

Table 12. Descriptive statistics for Datasets III and IV.

Dataset	SS	Mean	Median	SD	SK	KU
III	50	0.1632	0.1600	0.0810	0.0723	2.2166
IV	50	0.1520	0.1600	0.0785	0.0061	2.3012

Table 13. Theoretical statistics from the UED.

Dataset	SS	Mean	Median	SD	SK	KU
III	50	0.1633	0.1641	0.0809	0.0259	2.2511
IV	50	0.1519	0.1521	0.0777	0.0262	2.2521

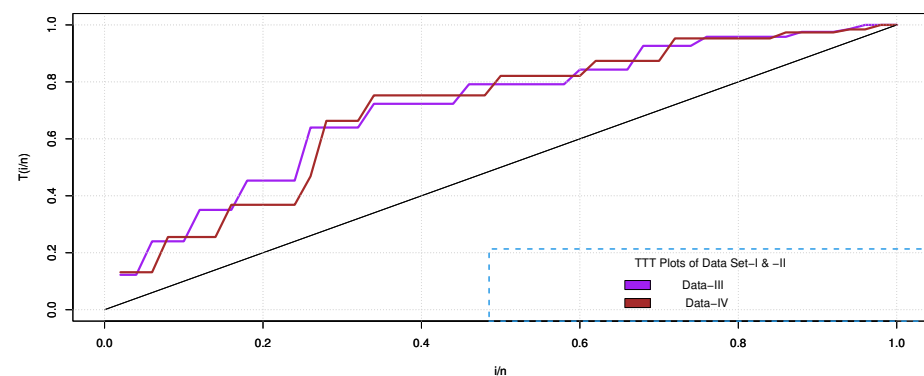


Figure 7. TTT plots of Datasets III and IV.

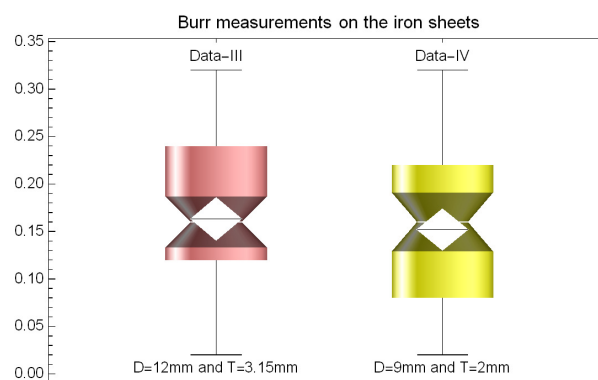


Figure 8. Box plots for Datasets III and IV.

Table 14. MLEs and goodness-of-fit statistics for Dataset III.

Distribution	$\hat{\beta}$	$\hat{\alpha}$	CVM_0^*	AD_0^*	KS	p -Value
UED	4.7879	0.1756	0.3274	0.0419	0.1242	0.9881
BD	2.6824	13.8640	0.1538	0.9120	0.1414	0.5555
KwD	1.0746	0.0925	12.2879	2.3943	0.7222	0.0000
JSBD	2.3767	1.3175	0.2495	1.4647	0.1740	0.0968
UGoMD	0.0924	1.0747	0.5213	3.0810	0.2046	0.0304

Table 15. MLEs and goodness-of-fit statistics for Dataset IV.

Distribution	$\hat{\beta}$	$\hat{\alpha}$	CVM_0^*	AD_0^*	KS	p -Value
UED	4.8518	0.1996	0.3224	0.0339	0.1239	0.9928
BD	2.4003	13.5218	0.2871	1.5649	0.1981	0.7340
KwD	1.9606	31.3769	0.2093	1.2683	0.1691	0.8825
JSBD	2.3682	1.2374	0.4145	2.2458	0.2285	0.5579
UGoMD	0.0916	1.0250	0.6091	3.4278	0.2312	0.5426

However, the likelihood aspects and information criterion values also favor the proposed UED model, which can be seen in Tables 16 and 17, respectively. Furthermore, the shape of our proposed model, as shown in Figure 9, matched the data in a better way compared with the other competing models. Finally, the Vuong statistic, as depicted in Table 18, also shows the capability of the proposed model.

Table 16. Estimates of the maximum log-likelihood and information criteria for Dataset III.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
UED	−57.0712	−110.142	−109.887	−106.318	−108.686	−104.318
BD	−54.6066	−105.213	−104.958	−101.389	−103.757	−99.3892
KwD	−56.0686	−108.137	−107.882	−104.313	−106.681	−102.313
JSBD	−51.3231	−98.6462	−98.3909	−94.8222	−97.19	−92.8222
UGoMD	−40.672	−77.344	−77.0887	−73.52	−75.8878	−71.52

Table 17. Estimates of the maximum log-likelihood and information criteria for Dataset IV.

Distribution	$-l$	AIC	AICC	BIC	HQIC	CAIC
UED	−59.3536	−114.707	−114.452	−110.883	−113.251	−108.883
BD	−55.9312	−107.862	−107.607	−104.038	−106.406	−102.038
KwD	−57.5214	−111.043	−110.788	−107.219	−109.587	−105.219
JSBD	−52.305	−100.61	−100.355	−96.786	−99.1538	−94.786
UGoMD	−42.6099	−81.2198	−80.9645	−77.3957	−79.7636	−75.3957

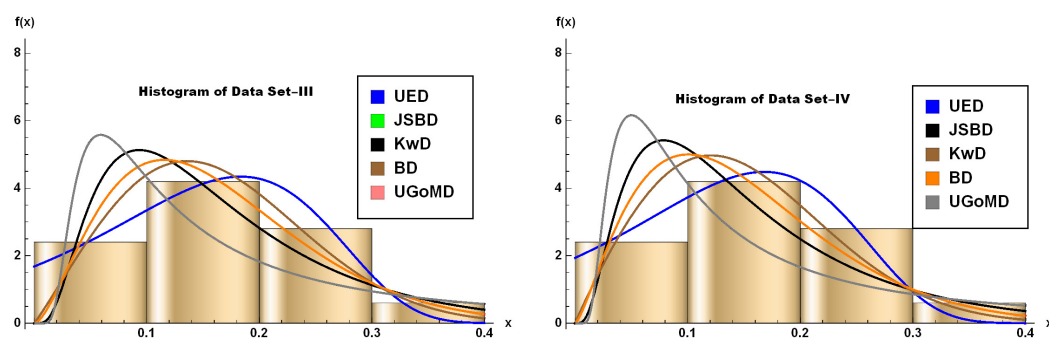


Figure 9. Datasets III and IV (given by histograms) fitted via unit interval distributions (given by lines).

Table 18. Vuong test statistic for Datasets III and IV.

Models	Dataset III	Suitability	Dataset IV	Suitability
UED-BD	0.4137	Indecisive	3.5339	UED
UED-KwD	−2.3203	KwD	3.9633	UED
UED-JSBD	2.1336	UED	3.4202	UED
UED-UGoMD	4.9679	UED	4.0306	UED

5. Concluding Remarks

We introduced a two-parameter bounded model called the unit exponential distribution (UED), which is appropriate for modeling skewed and IFR data. Some of its mathematical properties were studied, including the moments, quantiles, and other distributional behavior. A characterization of the UED via the HRF was made, which provided the identification requirements of the distribution and thus provided a reliable prediction compared with the well-known unit domain models. The model parameters were estimated with the MLE method. We also provided a guide line to choose the best model by using various goodness-of-fit statistics. Applications of the newly defined distribution showed that the proposed models have better modeling abilities than competitive models. For this purpose, we used four datasets in two different disciplines, namely environmental and engineering, and it was found that the proposed strategy was the best one in the unit interval domain. Moreover, in a further study, the proposed model could also be generalized over the interval $[0, s]$ by introducing the function

$$F(x) = 1 - \exp \left[\alpha \left(1 - \left(\frac{s+x}{s-x} \right)^{\beta s} \right) \right].$$

where, obviously, $F(x) = 0$ when $x = 0$ and $F(x) = 1$ when $x = s$.

Author Contributions: Conceptualization, H.S.B. and T.H.; methodology, H.S.B., M.T. and N.Q.; software, H.S.B. and T.H.; validation, H.S.B., M.T. and V.S.S.; formal analysis, H.S.B., T.H. and V.S.S.; data curation, H.S.B. and T.H.; writing—original draft preparation, H.S.B., T.H. and V.S.S.; writing—review and editing, M.T., V.S.S. and N.Q.; visualization, T.H. and M.T.; supervision, H.S.B. and V.S.S.; project administration, M.T. and N.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R376), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia for the financial support for this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fleiss, J.L.; Levin, B.; Paik, M.C. *Statistical Methods for Rates and Proportions*, 3rd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1993.
2. Gilchrist, W. *Statistical Modelling with Quantile Functions*; CRC Press: Abingdon, UK, 2000.
3. Seber, G.A.F. *Statistical Models for Proportions and Probabilities*; Springer: Berlin/Heidelberg, Germany, 2013.
4. Bayes, T. An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F.R.S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philos. Trans. R. Soc.* **1763**, *53*, 370–418. [\[CrossRef\]](#)
5. Leipnik, R.B. Distribution of the Serial Correlation Coefficient in a Circularly Correlated Universe. *Ann. Math. Stat.* **1947**, *18*, 80–87. [\[CrossRef\]](#)
6. Johnson, N. Systems of Frequency Curves Derived From the First Law of Laplace. *Trabajos Estadística* **1955**, *5*, 283–291. [\[CrossRef\]](#)
7. Jørgensen, B. Proper Dispersion Models. *Braz. J. Probab. Stat.* **1997**, *11*, 89–128.
8. Kumaraswamy, P. A Generalized Probability Density Function for Double-Bounded Random Processes. *J. Hydrol.* **1980**, *46*, 79–88. [\[CrossRef\]](#)
9. Topp, C.W.; Leone, F.C. A Family of J-Shaped Frequency Functions. *J. Am. Stat. Assoc.* **1955**, *50*, 209–219. [\[CrossRef\]](#)
10. Consul, P.C.; Jain, G.C. On the Log-Gamma Distribution and Its Properties. *Stat. Hefte* **1971**, *12*, 100–106. [\[CrossRef\]](#)
11. Smithson, M.; Shou, Y. CDF-Quantile. Distributions for Modelling RVs on the Unit Interval. *Br. J. Math. Stat. Psychol.* **2017**, *70*, 412–438. [\[CrossRef\]](#)
12. Nakamura, L.R.; Cerqueira, P.H.R.; Ramires, T.G.; Pescim, R.R.; Rigby, R.A.; Stasinopoulos, D.M. A New Continuous Distribution on the Unit Interval Applied to Modelling the Points Ratio of Football Teams. *J. Appl. Stat.* **2019**, *46*, 416–431. [\[CrossRef\]](#)
13. Ghitany, M.E.; Mazucheli, J.; Menezes, A.F.B.; Alqallaf, F. The Unit-Inverse Gaussian Distribution: A New Alternative to Two-Parameter Distributions on the Unit Interval. *Commun. Stat. Theory Methods* **2019**, *48*, 3423–3438. [\[CrossRef\]](#)
14. Altun, E.; Hamedani, G. The Log-Xgamma Distribution with Inference and Application. *J. Soc. Fr. Stat.* **2018**, *159*, 40–55.
15. Mazucheli, J.; Menezes, A.F.; Dey, S. Unit-Gompertz Distribution with Applications. *Statistica* **2019**, *79*, 25–43. [\[CrossRef\]](#)
16. Mazucheli, J.; Menezes, A.F.B.; Chakraborty, S. On the One Parameter Unit-Lindley Distribution and Its Associated Regression Model for Proportion Data. *J. Appl. Stat.* **2019**, *46*, 700–714. [\[CrossRef\]](#)
17. Mazucheli, J.; Menezes, A.F.B.; Fernandes, L.B.; de Oliveira, R.P.; Ghitany, M.E. The Unit-Weibull Distribution as an Alternative to the Kumaraswamy Distribution for the Modeling of Quantiles Conditional on Covariates. *J. Appl. Stat.* **2019**, *47*, 954–974. [\[CrossRef\]](#)
18. Altun, E. The Log-Weighted Exponential Regression Model: Alternative to the Beta Regression Model. *Commun. Stat. Theory Methods* **2020**, *50*, 2306–2321. [\[CrossRef\]](#)
19. Gündüz, S.; Mustafa, Ç.; Korkmaz, M.C. A New Unit Distribution Based on the Unbounded Johnson Distribution Rule: The Unit Johnson SU Distribution. *Pak. J. Stat. Oper. Res.* **2020**, *16*, 471–490. [\[CrossRef\]](#)
20. Korkmaz, M.C.; Korkmaz, Z.S. The Unit Log-log Distribution: A New Unit Distribution with Alternative Quantile Regression Modeling and Educational Measurements Applications. *J. Appl. Stat.* **2023**, *50*, 889–908. [\[CrossRef\]](#)
21. Afify, A.Z.; Nassar, M.; Kumar, D.; Cordeiro, G.M. A New Unit Distribution: Properties and Applications. *Electron. J. Appl. Stat.* **2022**, *15*, 460–484.
22. Fayomi, A.; Hassan, A.S.; Baaqeel, H.; Almetwally, E.M. Bayesian Inference and Data Analysis of the Unit–Power Burr X Distribution. *Axioms* **2023**, *12*, 297. [\[CrossRef\]](#)
23. Krishna, A.; Maya, R.; Chesneau, C.; Irshad, M.R. The Unit Teissier Distribution and Its Applications. *Math. Comput. Appl.* **2022**, *27*, 12. [\[CrossRef\]](#)
24. Biswas, A.; Chakraborty, S. A new method for constructing continuous distributions on the unit interval. *arXiv* **2021**, arXiv:2101.04661.
25. Dombi, J.; Jónás, T.; Tóth, Z.E. The Epsilon Probability Distribution and its Application in Reliability Theory. *Acta Polytech. Hung.* **2018**, *15*, 197–216.
26. Aslam, M.; Noor, F.; Ali, S. Shifted Exponential Distribution: Bayesian Estimation, Prediction and Expected Test Time Under Progressive Censoring. *J. Test. Eval.* **2020**, *48*, 1576–1593. [\[CrossRef\]](#)
27. Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D. Coherent Measures of Risk. *Math. Financ.* **1999**, *9*, 203–228. [\[CrossRef\]](#)
28. Ahsanullah, M.; Shakil, M.; Kibria, B.M.G. Characterizations of Continuous Distributions by Truncated Moment. *J. Mod. Appl. Stat. Methods* **2016**, *15*, 316–331. [\[CrossRef\]](#)
29. Ahsanullah, M.; Ghitany, M.E.; Al-Mutairi, D.K. Characterization of Lindley Distribution by Truncated Moments. *Commun. Stat. Theory Methods* **2017**, *46*, 6222–6227. [\[CrossRef\]](#)
30. Hamedani, G.G. Characterizations of Univariate Continuous Distributions Based on Truncated Moments of Functions of Order Statistics. *Stud. Sci. Math. Hung.* **2010**, *47*, 462–468. [\[CrossRef\]](#)
31. Glänzel, W. A Characterization Theorem Based on Truncated Moments and Its Application to Some Distribution Families. In *Mathematical Statistics and Probability Vol. B*; Bauer, P., Konecny, F., Wertz, W., Eds.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1987; pp. 75–84.
32. Lindsay, B.G.; Li, B. On second-order optimality of the observed Fisher information. *Ann. Stat.* **1997**, *25*, 2172–2199. [\[CrossRef\]](#)
33. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Autom. Control* **1974**, *9*, 716–723. [\[CrossRef\]](#)
34. Hussain, T.; Bakouch, H.S.; Chesneau, C. A New Probability Model with Application to Heavy-Tailed Hydrological Data. *Environ. Ecol. Stat.* **2019**, *26*, 127–151. [\[CrossRef\]](#)
35. Murthy, D.N.P.; Xie, M.; Jiang, R. *Weibull Models*; John Wiley and Sons: Hoboken, NJ, USA, 2004.
36. Vuong, Q.H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **1989**, *57*, 307–333. [\[CrossRef\]](#)

37. Maity, R. *Statistical Methods in Hydrology and Hydroclimatology*; Springer Nature Singapore Pte Ltd.: Singapore, 2018.
38. Aarset, M.V. How to Identify a Bathtub Hazard Rate. *IEEE Trans. Reliab.* **1987**, *36*, 106–108. [[CrossRef](#)]
39. Dasgupta, R. On the Distribution of Burr with Applications. *Sankhya B* **2011**, *73*, 1–19. [[CrossRef](#)]
40. Dey, S.; Mazucheli, J.; Anis, M. Estimation of Reliability of Multicomponent Stress–strength for a Kumaraswamy Distribution. *Commun. Stat. Theory Methods* **2017**, *46*, 1560–1572. [[CrossRef](#)]
41. Dey, S.; Mazucheli, J.; Nadarajah, S. Kumaraswamy Distribution: Different Methods of Estimation. *Comput. Appl. Math.* **2018**, *37*, 2094–2111. [[CrossRef](#)]
42. ZeinEldin, R.A.; Chesneau, C.; Jamal, F.; Elgarhy, M. Different Estimation Methods for Type I Half-Logistic Topp–Leone Distribution. *Mathematics* **2019**, *7*, 985. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.