©2026 Ada Academica https://adac.ee

Eur. J. Stat. 6 (2026) 19
doi: 10.28924/ada/stat.6.19

A Novel Quantile Regression Approach Based on the Zubair-Unit Rayleigh Model

Mahdi Rasekhi[1,2,*]

[1] *Department of Statistics, Faculty of Mathematical Sciences and Statistics, Malayer University, Malayer, Iran*

[2] *Department of Artificial Intelligence and Data Science, Faculty of Interdisciplinary Science and Technology, Malayer University, Malayer, Iran*

[*] *Correspondence: m.rasekhi@malayeru.ac.ir, rasekhimahdi@gmail.com*

Abstract. In this paper, we introduce a novel two-parameter distribution on the unit interval, termed the Zubair-Unit Rayleigh distribution. We derive its fundamental statistical properties and evaluate various parameter estimation techniques through a comprehensive simulation study. Building upon this foundation, we develop a new quantile regression model based on the proposed distribution. Applications to real-world datasets demonstrate that both the new distribution and its corresponding regression model provide a superior fit and greater flexibility compared to existing alternatives.

1. Introduction

Unit interval distributions are widely employed to model percentages and proportions across diverse fields, including biology, economics, public health, risk analysis, and measurement sciences. The Beta distribution remains the most prominent choice for modeling and statistical inference in this domain. However, this classical model can prove inadequate for certain complex datasets, thereby limiting analytical accuracy. Consequently, recent literature has seen growing interest in developing novel unit interval models. These new models are typically derived either by applying mathematical transformations to well-established continuous distributions or by mixing general distribution families with existing unit distributions. A primary advantage of these approaches is the enhanced flexibility they provide over the unit interval. For example, recent literature has introduced the Log-Lindley [1], Log-Weighted Exponential [2], Modified Kies Topp-Leone [3] and Unit Log-Logistic Exponential [4] distributions, among others.

Bantan et al. (2020) [5] introduced a simple one-parameter model on the unit interval, known as the unit-Rayleigh distribution. This model is highly flexible and unimodal, capable of exhibiting both increasing and bathtub-shaped hazard rate functions. Separately, Ahmad (2020) [6]

Received: 4 Aug 2026.

Key words and phrases. stochastic ordering; stress-strength model; entropy; regression model; unit distribution; air conditioning failures dataset; educational attainment.

developed a general class of distributions termed the Zubair-G family, which incorporates an additional shape parameter. In this paper, we propose a novel two-parameter unit distribution by applying the Zubair-G family framework to the unit-Rayleigh distribution, which we term the Zubair-Unit Rayleigh (ZUR) distribution. We subsequently derive and investigate its fundamental mathematical and statistical properties.

Regression typically models the relationship between response and explanatory variables by focusing on the conditional mean. For data bounded on the [0,1] interval, beta regression is standard, though alternatives like the unit Weibull, unit Gompertz, and unit inverse Gaussian address its limitations. However, mean-based models remain highly sensitive to skewness and outliers, which easily distort the mean. To overcome this, quantile regression (QR) offers a robust alternative by modeling conditional quantiles. QR avoids strict distributional assumptions about the error term and provides superior flexibility for diverse datasets. Consequently, recent literature has introduced several novel unit-interval QR models to better capture bounded responses. Notable examples include the unit log-log [7], the new bound unit weibull [8] and Owen-based QR [9] models. Consequently, this paper proposes a novel ZUR-based quantile regression model and examines its diagnostic procedures.

The remainder of this paper is organized as follows. Section 2 introduces the proposed unit distribution and derives its fundamental statistical properties, including first-order stochastic dominance, Tsallis entropy and the stress-strength parameter. Section 3 details the parameter estimation procedures, specifically maximum likelihood, least squares, and weighted least squares estimation. Section 4 presents a simulation study designed to evaluate the finite-sample performance of these estimators. Section 5 develops the new quantile regression model based on the proposed distribution, addressing both parameter estimation and residual analysis. Section 6 demonstrates the practical utility of the proposed frameworks through two real-world applications: univariate modeling of an air conditioning failures dataset and quantile regression modeling of an educational attainment dataset. Finally, Section 7 concludes the paper with a summary of key findings.

2. The Proposed Unit Distribution and Its Statistical Properties

This section presents a new variant of the Unit Rayleigh distribution, named the Zubair-Unit Rayleigh (ZUR) distribution, constructed via the Zubair-G family of distributions introduced by Ahmad (2020) [6]. The respective cumulative distribution function (CDF) and probability distribution function (PDF) of the Zubair-G family are expressed as

$$F(x;\alpha,\xi)=\frac{e^{\alpha G(x;\xi)^2}-1}{e^{\alpha}-1},\quad \alpha,\xi>0,\ x\in\mathbb{R}$$

and

$$f(x;\alpha,\xi)=\frac{2\alpha g(x;\xi)G(x;\xi)e^{\alpha G(x;\xi)^2}}{e^{\alpha}-1},\quad x\in\mathbb{R},$$

where $g(x;\xi)$ and $G(x;\xi)$ are the PDF and CDF of the baseline distribution with parameters ξ, and α is the family's shape parameter. Also, the Unit Rayleigh (UR) distribution arises from transforming the well-known one-parameter Rayleigh distribution. Following Bantan et al. (2020) [5], the CDF for a random variable X following the UR distribution is defined as:

$$F(x,\beta) = \exp\left\{-\beta[\log(x)]^2\right\}, \quad x \in (0,1),$$

where $\beta > 0$ denotes the shape parameter. Its PDF is expressed as:

$$f(x,\beta) = -\frac{2\beta}{x}\log(x)\exp\left\{-\beta[\log(x)]^2\right\}, \quad x \in (0,1).$$

Therefore, the PDF and CDF of the proposed distribution based on these two distributions are expressed as:

$$f(x;\alpha,\beta) = \frac{2\alpha\beta|\log(x)|}{x(e^{\alpha}-1)}\exp\left\{-\beta[\log(x)]^2 + \alpha\exp\left\{-\beta[\log(x)]^2\right\}\right\}, \quad x \in (0,1), \tag{1}$$

and

$$F(x;\alpha,\beta) = \frac{\exp\left\{\alpha\exp\left\{-\beta[\log(x)]^2\right\}\right\} - 1}{e^{\alpha}-1}, \quad x \in (0,1), \quad \alpha,\beta > 0. \tag{2}$$

where α and β respectively denote the shape and scale parameters. Figure 1(a) illustrates the PDF of the proposed ZUR distribution, which exhibits diverse shapes including decreasing, unimodal with varying degrees of skewness, and approximately symmetric forms. This flexibility demonstrates that the proposed distribution is well-suited for modeling data exhibiting these characteristics. Quantile functions play a crucial role in simulating random variates and deriving

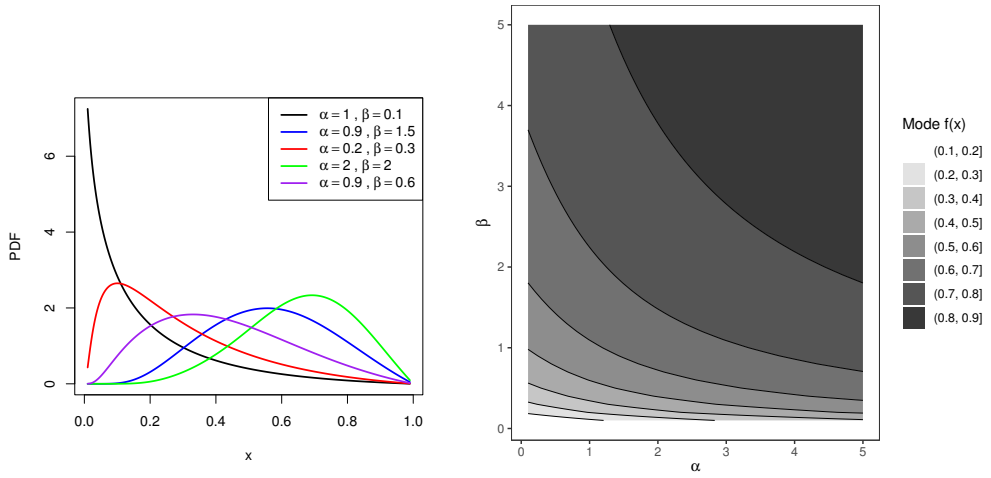


FIGURE 1. (a) PDF plots of the Zubair-Unit Rayleigh. (b) The location of mode in the Zubair-Unit Rayleigh.

measures of shape and dispersion. By inverting the CDF of the proposed ZUR distribution, its

quantile function (QF) is given by:

$$Q(u;\alpha,\beta)=\exp\left(-\sqrt{-\frac{1}{\beta}\ln\left(\frac{\ln[u(e^{\alpha}-1)+1]}{\alpha}\right)}\right),\quad u\in(0,1).$$

The mode of the ZUR distribution cannot be expressed in closed form but is defined implicitly by

$$x_{\text{mode}}=\exp\left(-\sqrt{\frac{u^*}{\beta}}\right),$$

where u^* satisfies

$$2u^*(1+\alpha e^{-u^*})=1.$$

Therefore, numerical techniques are required to compute the mode for given α and β. The variation of the mode with respect to the parameters is depicted in Figure 1(b). It is evident that the mode is sensitive to changes in the parameters, moving right as the parameter increases. The effects of the shape and scale parameters on skewness and kurtosis can be evaluated using quantile-based measures. We derive these measures using the QF of the ZUR distribution. Bowley's measure of skewness for X is given by

$$\text{Skewness}=\frac{Q(3/4)+Q(1/4)-2Q(1/2)}{Q(3/4)-Q(1/4)},$$

while Moors' measure of kurtosis is defined as

$$\text{Kurtosis}=\frac{Q(7/8)-Q(5/8)+Q(3/8)-Q(1/8)}{Q(6/8)-Q(2/8)}.$$

These quantile-based measures are robust to outliers and remain well-defined even for distributions that lack finite moments. Figure 2 illustrates the skewness and kurtosis of the ZUR distribution across various parameter values. The r-th raw moment of the ZUR distribution is

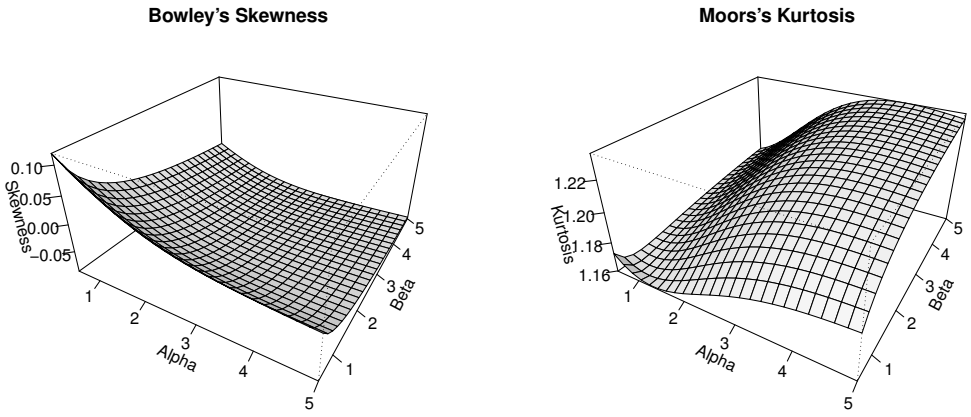


FIGURE 2. (a) Skewness plots of the Zubair-Unit Rayleigh. (b) Kurtosis plots of the Zubair-Unit Rayleigh.

given by:

$$\mu'_r = \frac{1}{e^{\alpha}-1}\sum_{i=0}^{\infty}\frac{\alpha^{i+1}}{i!}\left[\frac{1}{i+1} - \frac{r\sqrt{\pi}}{2\sqrt{2}(i+1)^{3/2}\sqrt{\beta}}\exp\left(\frac{r^2}{8(i+1)\beta}\right)\operatorname{erfc}\left(\frac{r}{2\sqrt{2(i+1)\beta}}\right)\right]$$

and the moment generating function (MGF) of the ZUR distribution is expressed as:

$$M_X(t) = \frac{1}{e^{\alpha}-1}\sum_{i=0}^{\infty}\sum_{r=0}^{\infty}\frac{t^r\alpha^{i+1}}{r!\,i!} \times \left[\frac{1}{i+1} - \frac{r\sqrt{\pi}}{2\sqrt{2}(i+1)^{3/2}\sqrt{\beta}}\exp\left(\frac{r^2}{8(i+1)\beta}\right)\operatorname{erfc}\left(\frac{r}{2\sqrt{2(i+1)\beta}}\right)\right],$$

where $\operatorname{erfc}(\cdot)$ denotes the complementary error function defined as $\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^{\infty} e^{-t^2}\,dt$.

2.1. **First-order stochastic dominance.** First-order stochastic dominance is a straightforward way to compare probability distributions using their CDFs. Distribution A dominates distribution B if its CDF is always less than or equal to B's, meaning $F_A(x) \leq F_B(x)$ for all values of x. This simple rule is widely used in biometrics, reliability, econometrics, and actuarial science to evaluate risks and outcomes. For more details on this concept, see Shaked and Shanthikumar (2007) [10].

Proposition 1. *Let $X \sim \mathrm{ZUR}(\alpha_1, \beta)$ and $Y \sim \mathrm{ZUR}(\alpha_2, \beta)$, where ZUR denotes the Zubair-Unit-Rayleigh distribution with CDF:*

$$F(x;\alpha,\beta) = \frac{\exp\left\{\alpha\exp\left\{-\beta[\log(x)]^2\right\}\right\} - 1}{e^{\alpha}-1}, \quad x \in (0,1), \alpha, \beta > 0.$$

If $\alpha_1 \geq \alpha_2$, then X first-order stochastically dominates Y. That is,

$$F(x;\alpha_1,\beta) \leq F(x;\alpha_2,\beta) \quad \forall x \in (0,1).$$

Proof. The CDF is given by

$$F(x;\alpha,\beta) = \frac{\exp\{\alpha\exp\{-\beta[\log(x)]^2\}\} - 1}{e^{\alpha}-1}.$$

For fixed $x \in (0,1)$ and $\beta > 0$, define $u = \exp\{-\beta[\log(x)]^2\} \in (0,1)$. Consider $g(\alpha) = F(x;\alpha,\beta)$ as a function of α. Differentiating yields $g'(\alpha) < 0$ for all $\alpha > 0$, since the numerator of the derivative is negative when $u \in (0,1)$. Thus, $F(x;\alpha,\beta)$ is strictly decreasing in α. Consequently, $\alpha_1 \geq \alpha_2$ implies $F(x;\alpha_1,\beta) \leq F(x;\alpha_2,\beta)$ for all $x \in (0,1)$, confirming first-order stochastic dominance. □

2.2. **Tsallis entropy.** While the entropy of the ZUR distribution can be computed through several approaches, this section specifically adopts the Tsallis entropy. General background on Tsallis entropy is provided by Amigo et al. (2018) [11]. The following proposition presents a series expansion of this entropy measure within the framework of the ZUR distribution.

Proposition 2. *Consider the Zubair Unit-Rayleigh distribution. For any entropic index $\tau < 1$, the Tsallis entropy is well-defined and admits the following integral representation:*

$$T(\tau,\alpha,\beta) = \frac{1}{\tau-1}\left[1-\left(\frac{2\alpha\beta}{e^{\alpha}-1}\right)^{\tau}\int_0^{+\infty} y^{\tau}e^{(\tau-1)y}\exp\left\{-\tau\beta y^2+\tau\alpha e^{-\beta y^2}\right\}dy\right].$$

Furthermore, when τ is restricted to the interval $(0,1)$, the entropy satisfies the following analytical upper bound:

$$T(\tau,\alpha,\beta) \leq \frac{1}{\tau-1}\left[1-\left(\frac{2\alpha\beta}{e^{\alpha}-1}\right)^{\tau}\frac{e^{\tau\alpha}}{2(\tau\beta)^{\frac{\tau+1}{2}}}\Gamma\left(\frac{\tau+1}{2}\right)\right],$$

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

Proof. By definition, the Tsallis entropy for the ZUR distribution is given by

$$T(\tau,\alpha,\beta) = \frac{1}{\tau-1}\left[1-\int_0^1 f(x;\alpha,\beta)^{\tau}dx\right].$$

Substituting the PDF of the ZUR distribution and applying the transformation $y = -\log(x)$ (where $dx = -e^{-y}dy$), the integral term transforms into

$$\int_0^1 f(x;\alpha,\beta)^{\tau}dx = \left(\frac{2\alpha\beta}{e^{\alpha}-1}\right)^{\tau}\int_0^{+\infty} y^{\tau}e^{(\tau-1)y}\exp\left\{-\tau\beta y^2+\tau\alpha e^{-\beta y^2}\right\}dy,$$

which establishes the integral representation stated in the proposition.

To derive the upper bound for $\tau \in (0,1)$, we observe that $(\tau-1) < 0$ implies $e^{(\tau-1)y} \leq 1$ for all $y \geq 0$. Additionally, since $e^{-\beta y^2} \leq 1$, it follows that $\exp\{\tau\alpha e^{-\beta y^2}\} \leq e^{\tau\alpha}$. Applying these inequalities to the integrand yields:

$$\int_0^{+\infty} y^{\tau}e^{(\tau-1)y}\exp\left\{-\tau\beta y^2+\tau\alpha e^{-\beta y^2}\right\}dy \leq e^{\tau\alpha}\int_0^{+\infty} y^{\tau}e^{-\tau\beta y^2}dy.$$

The remaining integral is evaluated using the substitution $t = \tau\beta y^2$, resulting in $\frac{1}{2}(\tau\beta)^{-\frac{\tau+1}{2}}\Gamma\left(\frac{\tau+1}{2}\right)$. Substituting this bound back into the entropy definition, and noting that $\frac{1}{\tau-1} < 0$ reverses the inequality sign, we obtain the stated upper bound:

$$T(\tau,\alpha,\beta) \leq \frac{1}{\tau-1}\left[1-\left(\frac{2\alpha\beta}{e^{\alpha}-1}\right)^{\tau}\frac{e^{\tau\alpha}}{2(\tau\beta)^{\frac{\tau+1}{2}}}\Gamma\left(\frac{\tau+1}{2}\right)\right].$$

□

Since the integral in the Tsallis entropy lacks an analytical expression, both the integral and the resulting entropy must be evaluated numerically.

2.3. **Stress-strength parameter.** In reliability analysis, the stress-strength parameter is a fundamental metric quantifying system performance. It represents the probability that a component's strength exceeds the applied stress, ensuring satisfactory operation. For comprehensive theoretical background, see Surles and Padgett (2001) [12]. The subsequent proposition derives a closed-form expression for this reliability measure specifically under the ZUR distribution, facilitating practical estimation and inference in lifetime data modeling contexts.

Proposition 3. *Suppose X_1 and X_2 are mutually independent random variables following the Zubair Unit-Rayleigh distribution with parameters (α_1, β) and (α_2, β), respectively, where the shape parameter $\beta > 0$ is shared. The stress-strength reliability, defined as $R(\alpha_1, \alpha_2, \beta) = P(X_2 \leq X_1)$, admits the following closed-form expression:*

$$R(\alpha_1, \alpha_2, \beta) = \frac{\alpha_1}{(e^{\alpha_1} - 1)(e^{\alpha_2} - 1)} \left[\frac{e^{\alpha_1+\alpha_2} - 1}{\alpha_1 + \alpha_2} - \frac{e^{\alpha_1} - 1}{\alpha_1} \right].$$

Furthermore, in the symmetric case where $\alpha_1 = \alpha_2 = \alpha$, the reliability parameter simplifies exactly to $R(\alpha, \alpha, \beta) = \frac{1}{2}$.

Proof. The stress-strength reliability is defined as $R = P(X_2 \leq X_1) = \int_0^1 F_{X_2}(x) f_{X_1}(x)\, dx$. Substituting the CDF and PDF of the ZUR distribution into this integral and applying the transformation $y = -\log(x)$, the domain shifts to $(0, +\infty)$.

Next, we apply the substitution $u = \beta y^2$, which simplifies the exponential terms, followed by $t = e^{-u}$. This sequence of transformations reduces the integral to a manageable form over the interval $(0, 1)$:

$$R = \frac{\alpha_1}{(e^{\alpha_1} - 1)(e^{\alpha_2} - 1)} \int_0^1 \left(e^{(\alpha_1+\alpha_2)t} - e^{\alpha_1 t} \right) dt.$$

Evaluating this integral yields the closed-form expression stated in the proposition. For the symmetric case where $\alpha_1 = \alpha_2 = \alpha$, substituting these parameters into the derived formula and utilizing the identity $e^{2\alpha} - 1 = (e^{\alpha} - 1)(e^{\alpha} + 1)$ leads to the simplification $R(\alpha, \alpha, \beta) = 1/2$. □

3. Estimation of Parameters

This section outlines three distinct estimation methods for the parameters of the ZUR distribution, which are detailed in the following subsections

3.1. Maximum likelihood estimation.

In this subsection, we derive parameter estimates for α and β using maximum likelihood (ML) estimation. Let $X_1, X_2, \ldots, X_n$ be a random sample from the ZUR(α, β) distribution with observed values $x_1, x_2, \ldots, x_n$, and let $\boldsymbol{\Upsilon} = (\alpha, \beta)^T$ denote the parameter vector. The log-likelihood function is:

$$\begin{aligned} \ell(\boldsymbol{\Upsilon}) = n \log(2\alpha\beta) + \sum_{i=1}^{n} \log |\log(x_i)| - \sum_{i=1}^{n} \log(x_i) \\ - n \log(e^{\alpha} - 1) - \beta \sum_{i=1}^{n} [\log(x_i)]^2 + \alpha \sum_{i=1}^{n} \exp\{-\beta [\log(x_i)]^2\}. \end{aligned}$$

The ML estimates $\hat{\alpha}$ and $\hat{\beta}$ are obtained by solving the score equations:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} - \frac{n e^{\alpha}}{e^{\alpha} - 1} + \sum_{i=1}^{n} \exp\{-\beta [\log(x_i)]^2\} = 0$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^{n} [\log(x_i)]^2 + \alpha \sum_{i=1}^{n} [\log(x_i)]^2 \exp\{-\beta [\log(x_i)]^2\} = 0.$$

From the first equation, $\alpha = n\left[\frac{ne^{\alpha}}{e^{\alpha}-1} - \sum_{i=1}^{n}\exp\{-\beta[\log(x_i)]^2\}\right]^{-1}$. Substituting this into the log-likelihood yields the profile log-likelihood $\ell(\beta)$ for β. Since closed-form solutions are unavailable, numerical optimization methods are required to obtain $\hat{\beta}$, and subsequently $\hat{\alpha}$.

Under standard regularity conditions, the MLEs are asymptotically bivariate normal with mean (α, β) and covariance matrix I^{-1}, where I is the Fisher information matrix. Approximate $100(1-\vartheta)\%$ confidence intervals are:

$$\hat{\alpha} \pm z_{\vartheta/2}\hat{s}_{\alpha} \quad \text{and} \quad \hat{\beta} \pm z_{\vartheta/2}\hat{s}_{\beta}$$

where $\hat{s}_{\alpha} = \sqrt{J^{-1}_{\alpha\alpha}}$, $\hat{s}_{\beta} = \sqrt{J^{-1}_{\beta\beta}}$, and J is the observed information matrix.

3.2. **Least Squares Estimation.** The least squares estimates (LSEs), denoted as $\hat{\alpha}_{LSE}$ and $\hat{\beta}_{LSE}$, are obtained by minimizing the sum of squared differences between the theoretical CDF and its expected value at the order statistics. The objective function is defined as:

$$LSE(\boldsymbol{\Upsilon}) = \sum_{i=1}^{n}\left[F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1}\right]^2$$

where $x_{(i)}$ represents the i-th order statistic, and $E[F(X_{(i)})] \approx i/(n+1)$ is the standard plotting position approximation.

To find the estimators, we differentiate $LSE(\boldsymbol{\Upsilon})$ with respect to α and β and set the results to zero, yielding the following normal equations:

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

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

For the ZUR distribution, the required partial derivatives of the CDF are:

$$\frac{\partial F}{\partial \alpha} = \frac{e^{-\beta[\log x]^2}\exp\left\{\alpha e^{-\beta[\log x]^2}\right\}(e^{\alpha}-1) - e^{\alpha}\left(\exp\left\{\alpha e^{-\beta[\log x]^2}\right\} - 1\right)}{(e^{\alpha}-1)^2}$$

$$\frac{\partial F}{\partial \beta} = -\frac{\alpha[\log x]^2\exp\left\{\alpha e^{-\beta[\log x]^2} - \beta[\log x]^2\right\}}{e^{\alpha}-1}$$

Since these equations are non-linear and cannot be solved analytically, numerical optimization algorithms must be employed to obtain $\hat{\alpha}_{LSE}$ and $\hat{\beta}_{LSE}$.

3.3. **Weighted Least Squares Estimation.** The weighted least squares estimates (WLSEs), denoted by $\hat{\alpha}_{WLSE}$ and $\hat{\beta}_{WLSE}$, are obtained by minimizing the weighted sum of squared deviations:

$$WLSE(\boldsymbol{\Upsilon}) = \sum_{i=1}^{n}\frac{1}{V[F(X_{(i)}, \alpha, \beta)]}\left[F(x_{(i)}, \alpha, \beta) - \frac{i}{n+1}\right]^2$$

where the variance term is given by $V[F(X_{(i)}, \alpha, \beta)] = \frac{i(n-i+1)}{(n+2)(n+1)^2}$ for $i = 1, 2, \ldots, n$.

Differentiating $WLSE(\boldsymbol{\Upsilon})$ with respect to α and β and setting the derivatives to zero yields the following estimating equations:

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

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

where the partial derivatives $\frac{\partial F}{\partial \alpha}$ and $\frac{\partial F}{\partial \beta}$ are given in Section 3.2. These non-linear equations do not admit closed-form solutions and must be solved numerically to obtain $\hat{\alpha}_{WLSE}$ and $\hat{\beta}_{WLSE}$.

4. Simulation Study

This section presents a Monte Carlo simulation study to assess the performance of the proposed ZUR estimators for various sample sizes n. We generate $N = 1000$ random samples of sizes $n \in \{50, 100, \ldots, 800\}$ using the true parameters $\alpha = 0.5$ and $\beta = 0.3$, with random variates obtained via the model's quantile function. All estimates are computed using the L-BFGS-B algorithm within the `optim` function in R. To compare the estimation methods, we evaluate the empirical bias, root mean square error (RMSE), and mean relative estimate (MRE). The bias, RMSE and MRE for α and β are calculated across all N samples as:

$$\text{Bias} = \sum_{j=1}^{N} \frac{\hat{\Psi}_{i,j} - \Psi_i}{N},$$

$$\text{RMSE} = \sqrt{\sum_{j=1}^{N} \frac{(\hat{\Psi}_{i,j} - \Psi_i)^2}{N}},$$

$$\text{MRE} = \sum_{j=1}^{N} \frac{\hat{\Psi}_{i,j}/\Psi_i}{N}, \quad i = 1, 2$$

where $\Psi \in \{\alpha, \beta\}$. Ideally, the empirical means should approximate the true parameter values, indicated by biases and MSEs approaching zero and MREs converging to one. The simulation results are summarized in Fig. 3.

Figure 3 confirms the consistency of all estimators: as sample size increases, RMSE and bias tend toward zero while MRE approaches one, consistent with theoretical expectations. All estimators are also asymptotically unbiased. The LS method initially shows larger biases for both parameters compared to other methods, whereas MLE and WLSE perform similarly and outperform LS across all criteria. As sample size grows, however, the performance differences among methods diminish. These results indicate that WLSE is particularly well-suited for inference with small samples, while all methods are viable alternatives for large-sample estimation of the proposed model. Notably, MLE retains key advantages for inference due to its desirable asymptotic properties (consistency, efficiency, and normality). Similar conclusions hold under alternative parameter configurations.



FIGURE 3. The results of simulation study.

5. New Quantile Regression Model

Given the tractable analytic form of the quantile function (qf) for the ZUR distribution, it is straightforward to re-parameterize its pdf and cdf in terms of specific quantiles. By defining the quantile parameter $\mu = Q(u;\alpha,\beta)$ and substituting the expression for β, we obtain:

$$\beta(\alpha,\mu) = \frac{\ln\left(\frac{\alpha}{\ln[u(e^{\alpha}-1)+1]}\right)}{[\ln(\mu)]^2}.$$

Consequently, the pdf and cdf of the re-parameterized ZUR distribution are given by

$$f(x;\alpha,\mu) = \frac{2\alpha\beta(\alpha,\mu)|\ln(x)|}{x(e^{\alpha}-1)} \exp\left\{-\beta(\alpha,\mu)[\ln(x)]^2 + \alpha\exp\left\{-\beta(\alpha,\mu)[\ln(x)]^2\right\}\right\}, \qquad (3)$$

and

$$F(x;\alpha,\mu)=\frac{\exp\left\{\alpha\left(\frac{\ln[u(e^{\alpha}-1)+1]}{\alpha}\right)^{\left(\frac{\ln x}{\ln \mu}\right)^2}\right\}-1}{e^{\alpha}-1}, \tag{4}$$

respectively. In this formulation, $\alpha>0$ serves as the shape parameter, $\mu\in(0,1)$ represents the quantile parameter corresponding to the probability u (which is assumed to be known), and a random variable X following this distribution is denoted by $X\sim ZUR(\alpha,\mu,u)$.

We now develop a quantile regression model based on the re-parameterized ZUR distribution defined in Eq. (3). Consider a sample $y_1,y_2,\ldots,y_n$, where each observation y_i is assumed to follow a $ZUR(\alpha,\mu_i,u)$ distribution for $i=1,\ldots,n$. Here, α and the quantile parameters μ_i are unknown, while the probability level u is fixed and known.

The quantile regression model links the conditional quantile parameter μ_i to a set of covariates via the linear predictor:

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

where $\boldsymbol{\delta}=(\delta_0,\delta_1,\ldots,\delta_p)^{\top}$ is the vector of unknown regression coefficients, and $\mathbf{x}_i=(1,x_{i1},x_{i2},\ldots,x_{ip})$ is the known vector of covariates for the i-th observation. The function $g(\cdot)$ serves as the link function connecting the covariates to the conditional quantile of the response variable. For example, setting $u=0.5$ allows the model to target the conditional median. Since the ZUR distribution is supported on the unit interval $(0,1)$, we employ the logit link function:

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

Alternatively, other link functions suitable for bounded responses, such as the probit or log-log links, may also be used to relate the covariates to the conditional quantiles.

5.1. **Parameter estimation.** We now focus on estimating the unknown parameters of the ZUR quantile regression model using the maximum likelihood estimation (MLE) method. Specifically, we consider the logit link function:

$$g(\mu_i)=\log\left(\frac{\mu_i}{1-\mu_i}\right)=\mathbf{x}_i\boldsymbol{\delta}^{\top}. \tag{5}$$

From Equation (5), the conditional quantile parameter μ_i can be expressed as:

$$\mu_i=\frac{\exp\left(\mathbf{x}_i\boldsymbol{\delta}^{\top}\right)}{1+\exp\left(\mathbf{x}_i\boldsymbol{\delta}^{\top}\right)}. \tag{6}$$

Let $Y_1,Y_2,\ldots,Y_n$ be a random sample of size n, where each $Y_i\sim ZUR(\alpha,\mu_i,u)$ with corresponding observed values $y_1,y_2,\ldots,y_n$. Here, μ_i is determined by Equation (6) for $i=1,\ldots,n$. Substituting these into the pdf given in Equation (3), the log-likelihood function for the sample

is:

$$\ell(\boldsymbol{\Omega}) = n\ln(2\alpha) + n\ln\left[\ln\left(\frac{\alpha}{\ln[u(e^{\alpha}-1)+1]}\right)\right] - 2\sum_{i=1}^{n}\ln|\ln\mu_i| - n\ln(e^{\alpha}-1)$$
$$+\sum_{i=1}^{n}\ln|\ln y_i| - \sum_{i=1}^{n}\ln y_i - \ln\left(\frac{\alpha}{\ln[u(e^{\alpha}-1)+1]}\right)\sum_{i=1}^{n}\left(\frac{\ln y_i}{\ln\mu_i}\right)^2$$
$$+\alpha\sum_{i=1}^{n}\exp\left\{-\ln\left(\frac{\alpha}{\ln[u(e^{\alpha}-1)+1]}\right)\left(\frac{\ln y_i}{\ln\mu_i}\right)^2\right\}, \tag{7}$$

where $\boldsymbol{\Omega} = (\alpha, \boldsymbol{\delta}^{\top})^{\top}$ is the vector of unknown parameters. The maximum likelihood estimates (MLEs) of $\boldsymbol{\Omega}$, denoted by $\widehat{\boldsymbol{\Omega}} = (\widehat{\alpha}, \widehat{\boldsymbol{\delta}}^{\top})^{\top}$, are obtained by maximizing $\ell(\boldsymbol{\Omega})$. Because Equation (7) is highly nonlinear with respect to the model parameters, the maximization must be performed numerically using statistical software such as R, S-Plus, or Mathematica. Note that setting $u = 0.5$ allows the model to specifically target the conditional median of the response variable. Under standard regularity conditions, the asymptotic distribution of the estimator $\widehat{\boldsymbol{\Omega}}$ is multivariate normal, specifically $\widehat{\boldsymbol{\Omega}} \sim N_{p+1}(\boldsymbol{\Omega}, \mathcal{I}^{-1})$, where $\mathcal{I}$ is the expected Fisher information matrix. In practice, the covariance matrix $\mathcal{I}^{-1}$ is approximated by the inverse of the $(p+1)\times(p+1)$ observed information matrix. We perform the numerical maximization of Equation (7) using the `maxLik` function in R [13], which also computes the asymptotic standard errors based on the observed information matrix.

5.2. **Residual analysis.** Residual analysis is essential to assess the adequacy of the fitted regression model. To this end, we employ the randomized quantile residuals [14] and the Cox-Snell residuals [15]. For $i = 1, \ldots, n$, the i-th randomized quantile residual is defined as

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

where $F(\cdot; \alpha, \mu)$ denotes the CDF of the re-parameterized ZUR distribution given in Equation (4), $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution, and $\hat{\mu}_i$ is obtained from Equation (6) by replacing $\boldsymbol{\delta}$ with its maximum likelihood estimate $\hat{\boldsymbol{\delta}}$. If the model provides an adequate fit to the data, these randomized quantile residuals should approximately follow a standard normal distribution. Alternatively, the i-th Cox-Snell residual is given by

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

Under a correctly specified model, these residuals are expected to follow a standard exponential distribution (i.e., an exponential distribution with a scale parameter of 1).

6. Data Analysis

This section demonstrates the practical utility and modeling flexibility of the ZUR distribution through two real-world applications: univariate data modeling and quantile regression. The first dataset records the number of consecutive air conditioning failures observed for each of 13 Boeing 720 aircraft. To scale the data to the (0, 1) interval, the original values were divided

by 1,000. Al-Olaimat et al. (2026) [16] used this dataset to illustrate univariate fitting. The second application employs a quantile regression framework to analyze educational attainment values (EAV, measured in percentages) across OECD countries. Following Altun (2021) [2], we model the EAV using the Better Life Index, labor market insecurity (LMI), and homicide rate (HR) as covariates. All statistical computations were conducted using the R software environment. Specifically, the `maxLik` package [13] was utilized for maximum likelihood estimation. The detailed analyses and empirical results for both datasets are presented in the following subsections.

6.1. **Univariate modeling for air conditioning failures dataset.** Here, we analyze the air conditioning failures dataset. Summary statistics and a box plot visualization are provided in Table 1 and Figure 4, respectively. The distribution is right-skewed and contains some outliers. Based on this dataset, we evaluate the fitting performance of the proposed ZUR distribution, estimated via the maximum likelihood estimation (MLE) method, against several well-established unit distributions available in the literature. The probability density functions of these competing distributions are given as follows:

TABLE 1. Summary statistics of the first data set

Minimum	Mean	Median	Maximum	Variance	Skewness	Kurtosis	n
0.001	0.096	0.056	0.603	0.012	2.009	7.217	181

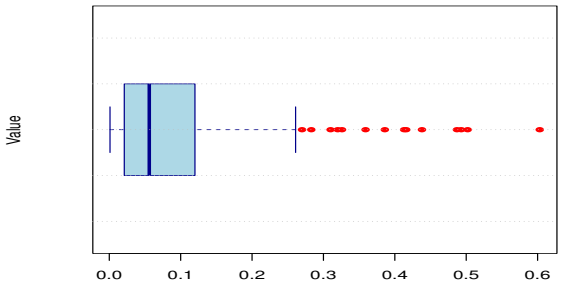


FIGURE 4. Box plot for first dataset.

- **Beta distribution:**

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

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

- **Modified Kies Topp Leone (MKTL) distribution** [3]:

$$f(x;\eta,\beta) = \frac{2\beta\eta x^{\eta-1}(1-x)(2-x)^{\eta-1}\left[1-(1-x)^2\right]^{\eta\beta-\eta}}{\left[1-(1-(1-x)^2)^{\eta}\right]^{\beta+1}} \exp\left(-\left[\left(1-(1-x)^2\right)^{-\eta}-1\right]^{-\beta}\right),$$

for $0 < x < 1$ and $\eta, \beta > 0$.

- **Unit Log-Logistic Exponential (ULE) distribution** [4]:

$$f_{\text{ULE}}(x;\alpha,\theta) = \frac{\alpha\theta \exp\left(\theta\left(\frac{1}{x}-1\right)\right)\left[\exp\left(\theta\left(\frac{1}{x}-1\right)\right)-1\right]^{\alpha-1}}{x^2\left\{\left[\exp\left(\theta\left(\frac{1}{x}-1\right)\right)-1\right]^{\alpha}+1\right\}^2},$$

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

- **Log-Lindley distribution** [1]:

$$f_{\text{LL}}(x;\sigma,\lambda) = \frac{\sigma^2}{1+\lambda\sigma}(\lambda - \ln x)x^{\sigma-1},$$

where $x \in (0,1)$, $\sigma > 0$, and $\lambda \geq 0$.

- **Log-Weighted Exponential (LWE) distribution** [2]:

$$f_{\text{LWE}}(x;\alpha,\lambda) = \frac{\alpha+1}{\alpha}\lambda x^{\lambda-1}\left(1-x^{\alpha\lambda}\right),$$

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

To identify the optimal model, we employ several goodness-of-fit measures: the maximized log-likelihood ($\hat{\ell}$), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC), and Hannan-Quinn Information Criterion (HQIC) statistics. The best-fitting model is characterized by lower values of AIC, BIC, CAIC, and HQIC, alongside higher values of $\hat{\ell}$.

The parameter estimates and goodness-of-fit statistics are presented in Table 2. As shown in this table, the ZUR distribution yields the lowest AIC, BIC, CAIC and HQIC values. These findings demonstrate that the ZUR distribution provides the superior fit for the data set under consideration. Figure 5 plots the estimated pdfs to visually assess their suitability for the dataset. As the figure shows, the ZUR model provides an excellent fit by accurately capturing the data's skewness and kurtosis.

TABLE 2. MLEs, standard errors (SE) of the estimates (in parentheses), $\hat{\ell}$ and goodness-of-fits statistics for the data set

Model	parameter 1	parameter2	$\hat{\ell}$	*AIC*	*BIC*	*CAIC*	*HQIC*
ZUR	0.010(0.263)	0.096(0.076)	239.444	-474.888	-468.491	-474.821	-472.295
Beta	0.814(0.074)	7.401(0.871)	239.105	-474.210	-467.813	-474.143	-471.617
MKTL	0.487(0.027)	0.969(0.048)	218.214	-432.428	-426.031	-432.361	-429.835
ULE	0.803(0.052)	0.025(0.002)	223.252	-442.504	-436.107	-434.107	-439.910
LL	0.669(0.017)	0.010(1.338)	211.646	-419.293	-412.896	-410.896	-416.699
LWE	0.010(0.244)	0.668(0.042)	212.091	-420.183	-413.786	-411.786	-417.590

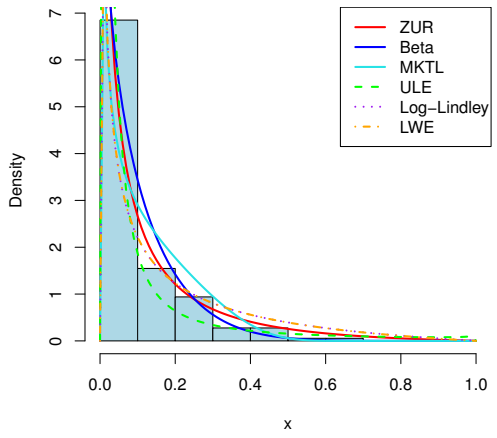


FIGURE 5. The fitted pdfs on the first dataset.

6.2. **Regression modeling for Educational attainment dataset.** To demonstrate the practical utility of the ZUR regression model, this section presents a real-data application using the dataset described above. The objective is to model the relationship between the response variable (y) and a set of covariates. The variables are defined as follows:

- y: Educational Attainment Values (EAV) for OECD countries;
- x_1: Labor Market Insecurity (LMI);
- x_2: Homicide Rate (HR).

This dataset was previously analyzed by Mazucheli et al. (2023) [17]. Given the presence of observed outliers in the response variable, quantile regression provides a more robust framework for inference. To benchmark the performance of the proposed ZUR quantile regression model, we compare it against two well-established competing models from the literature: the beta regression model [18], the new bound unit weibull quantile regression model [8], the unit log-log quantile regression model [7]. Their respective pdfs are given by:

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

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

$$f(y;\beta,\mu) = \beta\left[-\left(\frac{1}{\mu}-1\right)^{-\beta}\log(q)\right]\left(\frac{1-y}{y}\right)^{\beta-1} y^{-2} \times \exp\left[\left(\frac{1}{\mu}-1\right)^{-\beta}\log(q)\left(\frac{1-y}{y}\right)^{\beta}\right], \quad y \in (0,1).$$

where q represents the quantile parameter (supposed to be known, typically $q = 0.5$), $\mu \in (0, 1)$ is the quantile parameter, and $\beta > 0$, and

$$f(y, \alpha, \mu) = \frac{\alpha}{y}(-\log y)^{\alpha-1}(-\log \mu)^{-\alpha} \log(1 - \log u) \times \left[(1 - \log u)^{(-\log \mu)^{-\alpha}}\right]^{(-\log y)^{\alpha}} \times \exp\left\{1 - \left[(1 - \log u)^{(-\log \mu)^{-\alpha}}\right]^{(-\log y)^{\alpha}}\right\}, \quad y \in (0, 1).$$

where $\alpha > 0$ is the shape parameter, $\mu \in (0, 1)$ represents the quantile parameter, and u is known (typically $u = 0.5$). The regression model for the median parameter μ_i is specified as:

$$\text{logit}(\mu_i) = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2}, \quad i = 1, 2, \ldots, 38.$$

For the ZUR, NBUW and ULL quantile regression models, the quantile parameter u is fixed at 0.5. The results of the regression analysis are presented in Table 3. As shown in the table, all parameter estimates across the models are non-zero. This indicates a statistically significant relationship between educational attainment and the covariates.

Table 3 presents the parameter estimates and goodness-of-fit statistics. Notably, the ZUR quantile regression yields the lowest AIC, BIC and HQIC values among all fitted models. Figure 6 presents Q-Q plots for the response variable under the ZUR model to graphically assess model suitability. The plots display randomized quantile residuals against the standard normal distribution and Cox-Snell residuals against the exponential distribution. The close alignment of the residuals with their theoretical distributions indicates an excellent model fit. This suggests that the ZUR model successfully captures the skewness and kurtosis of the dataset.

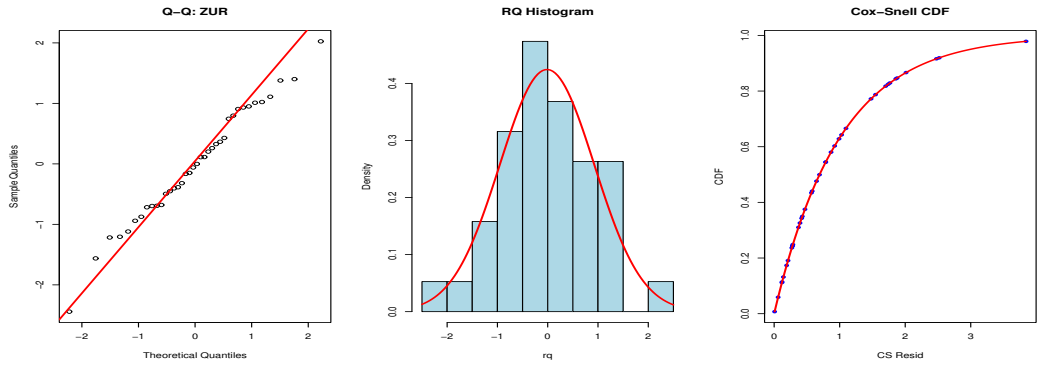


FIGURE 6. left: QQ-plot of response variable vs ZUR, middle: randomized quantile residuals vs fitted normal distribution, right: cox snell residual vs fitted exponential

7. CONCLUSION

A new unit model, termed the Zubair-Unit Rayleigh distribution, is defined for the modeling of percentages, proportions, and rates. The general structural properties of the distribution are investigated, and its parameters are estimated via three distinct methods. Simulation studies

TABLE 3. Parameter Estimates for Beta, ZUR, NBUW, and QULL Regression Models

Param.	**Beta**			**ZUR**			**NBUW**			**QULL**		
	Est.	**SE**	p	**Est.**	**SE**	p	**Est.**	**SE**	p	**Est.**	**SE**	p
δ_0	1.821	0.161	<.001	4.165	0.392	<.001	2.109	0.201	<.001	2.306	0.220	<.001
δ_1	−.078	0.018	<.001	−.243	0.078	.002	−.113	0.029	<.001	−.168	0.044	<.001
δ_2	−.046	0.017	.006	−.111	0.041	.007	−.063	0.019	<.001	−.061	0.017	<.001
α	13.418	3.027	<.001	8.969	2.069	<.001	1.557	0.188	<.001	1.331	0.178	<.001
$\hat{\ell}$		33.803			36.851			34.033			33.864	
AIC		-65.701			-59.605			-60.066			-59.728	
BIC		-59.151			-53.054			-53.515			-53.178	
HQIC		-63.371			-57.274			-57.735			-57.398	

are performed to assess the performance of these estimators. The empirical findings indicate that the proposed model yields a better fit than established unit probability distributions in the literature, in both univariate and regression contexts. It is expected that the proposed distribution will attract significant interest across other disciplines.

Competing interests: The author declares that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] E. Gómez-Déniz, M.A. Sordo, E. Calderín-Ojeda, The Log–Lindley Distribution as an Alternative to the Beta Regression Model with Applications in Insurance, Insur. Math. Econ. 54 (2014), 49–57. https://doi.org/10.1016/j.insmatheco.2013.10.017.

[2] E. Altun, The Log-Weighted Exponential Regression Model: Alternative to the Beta Regression Model, Commun. Stat. Theory Methods 50 (2021), 2306–2321. https://doi.org/10.1080/03610926.2019.1664586.

[3] S.M. Alghamdi, O. Albalawi, S.M. Almarzouki, V.B.V. Nagarjuna, S. Nasiru, et al., Different Estimation Methods of the Modified Kies Topp-Leone Model with Applications and Quantile Regression, PLoS One 19 (2024), e0307391. https://doi.org/10.1371/journal.pone.0307391.

[4] A. Saboor, F. Jamal, S. Shafq, R. Mumtaz, On the Versatility of the Unit Logistic Exponential Distribution: Capturing Bathtub, Upside-Down Bathtub, and Monotonic Hazard Rates, Innov. Stat. Probab. 1 (2025), 28–46. https://doi.org/10.64389/isp.2025.01102.

[5] R.A.R. Bantan, C. Chesneau, F. Jamal, M. Elgarhy, M.H. Tahir, et al., Some New Facts about the Unit-Rayleigh Distribution with Applications, Mathematics 8 (2020), 1954. https://doi.org/10.3390/math8111954.

[6] Z. Ahmad, The Zubair-G Family of Distributions: Properties and Applications, Ann. Data Sci. 7 (2020), 195–208. https://doi.org/10.1007/s40745-018-0169-9.

[7] M.Ç. Korkmaz, Z.S. Korkmaz, The Unit Log–Log Distribution: A New Unit Distribution with Alternative Quantile Regression Modeling and Educational Measurements Applications, J. Appl. Stat. 50 (2023), 889–908. https://doi.org/10.1080/02664763.2021.2001442.

[8] L.P. Sapkota, N. Bam, V. Kumar, New Bounded Unit Weibull Model: Applications with Quantile Regression, PLoS One 20 (2025), e0323888. https://doi.org/10.1371/journal.pone.0323888.

[9] M. Santos-Neto, D.I. Gallardo, E. Costa, C. Marchant, I. Renan, A New Parametric Quantile Regression Model Based on an Owen Distribution, J. Stat. Comput. Simul. 95 (2025), 156–185. https://doi.org/10.1080/00949655.2024.2422921.

[10] M. Shaked, J.G. Shanthikumar, Stochastic Orders, Springer New York, 2007. https://doi.org/10.1007/978-0-387-34675-5.

[11] J.M. Amigó, S.G. Balogh, S. Hernández, A Brief Review of Generalized Entropies, Entropy 20 (2018), 813. https://doi.org/10.3390/e20110813.

[12] J.G. Surles, W.J. Padgett, Inference for Reliability and Stress-Strength for a Scaled Burr Type X Distribution, Lifetime Data Anal. 7 (2001), 187–200. https://doi.org/10.1023/A:1011352923990.

[13] A. Henningsen, O. Toomet, maxLik: a Package for Maximum Likelihood Estimation in R, Comput. Stat. 26 (2011), 443–458. https://doi.org/10.1007/s00180-010-0217-1.

[14] P.K. Dunn, G.K. Smyth, Randomized Quantile Residuals, J. Comput. Graph. Stat. 5 (1996), 236–244. https://doi.org/10.1080/10618600.1996.10474708.

[15] D.R. Cox, E.J. Snell, A General Definition of Residuals, J. R. Stat. Soc. Ser. B 30 (1968), 248–265. https://doi.org/10.1111/j.2517-6161.1968.tb00724.x.

[16] N.M. Al-Olaimat, H.S. Migdadi, R. Saadeh, M. Hafez, On the Two-Parameter Unit Power Rayleigh Distribution: Properties, Inference and Applications, Results Control Optim. 23 (2026), 100734. https://doi.org/10.1016/j.rico.2026.100734.

[17] J. Mazucheli, M.Ç. Korkmaz, A.F.B. Menezes, V. Leiva, The Unit Generalized Half-Normal Quantile Regression Model: Formulation, Estimation, Diagnostics, and Numerical Applications, Soft Comput. 27 (2023), 279–295. https://doi.org/10.1007/s00500-022-07278-3.

[18] S. Ferrari, F. Cribari-Neto, Beta Regression for Modelling Rates and Proportions, J. Appl. Stat. 31 (2004), 799–815. https://doi.org/10.1080/0266476042000214501.