

Article

Generalized Unit Half-Logistic Geometric Distribution: Properties and Regression with Applications to Insurance

Suleman Nasiru ¹, Christophe Chesneau ^{2,*}, Abdul Ghaniyyu Abubakari ¹ and Irene Dekomwine Angbing ¹¹ Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo 03821, Ghana² Department of Mathematics, LMNO, University of Caen, 14032 Caen, France

* Correspondence: christophe.chesneau@unicaen.fr

Abstract: The use of distributions to model and quantify risk is essential in risk assessment and management. In this study, the generalized unit half-logistic geometric (GUHLG) distribution is developed to model bounded insurance data on the unit interval. The corresponding probability density function plots indicate that the related distribution can handle data that exhibit left-skewed, right-skewed, symmetric, reversed-J, and bathtub shapes. The hazard rate function also suggests that the distribution can be applied to analyze data with bathtubs, N-shapes, and increasing failure rates. Subsequently, the inferential aspects of the proposed model are investigated. In particular, Monte Carlo simulation exercises are carried out to examine the performance of the estimation method by using an algorithm to generate random observations from the quantile function. The results of the simulation suggest that the considered estimation method is efficient. The univariate application of the distribution and the multivariate application of the associated regression using risk survey data reveal that the model provides a better fit than the other existing distributions and regression models. Under the multivariate application, we estimate the parameters of the regression model using both maximum likelihood and Bayesian estimations. The estimates of the parameters for the two methods are very close. Diagnostic plots of the Bayesian method using the trace, ergodic, and autocorrelation plots reveal that the chains converge to a stationary distribution.

Keywords: unit distribution; statistical algorithms; power transformation; risk survey data; quantile regression; Bayesian estimation



Citation: Nasiru, S.; Chesneau, C.; Abubakari, A.G.; Angbing, I.D. Generalized Unit Half-Logistic Geometric Distribution: Properties and Regression with Applications to Insurance. *Analytics* **2023**, *2*, 438–462. <https://doi.org/10.3390/analytics2020025>

Academic Editors: Jose Luis Vicente Villardón and Fabio Postiglione

Received: 13 February 2023

Revised: 6 April 2023

Accepted: 4 May 2023

Published: 16 May 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

Risk assessment and management form an integral part of the responsibilities of managers in financial and insurance institutions. Thus, when risk is properly assessed and managed, financial and insurance companies can better manage the risk of financial losses. In order to achieve this, quantitative risk analysis is required. This entails probabilistic methods of handling risk, where the risk is considered random and then quantified using a distribution. These distributions are utilized by investors to predict asset returns and hedge their risks. In this regard, the precision of the analysis is heavily centered on the identification of an appropriate distribution to model the uncertainty. Failure to do so may result in the use of an incorrect distribution to quantify the risk, leading to incorrect decisions.

The selection of the correct distribution for risk analysis is not only essential but also a necessary approach to quantifying and managing risk. This has necessitated the development of new distributions for risk modelling and assessment. Some of the distributions that have been developed and used to model financial and insurance data include: unit half-logistic geometric (UHLG) distribution (see [1]), unit exponentiated Fréchet distribution (see [2]), new modified Kumaraswamy distribution (see [3]), new beta power transformed Weibull distribution (see [4]), unit Weibull distribution (see [5]), WT-XW distribution (see [6]), extended exponential geometric distribution (see [7]), Weibull loss

distribution (see [8]), unit Gompertz distribution (see [9]) and log-Lindley distribution (see [10]).

However, every data-generating process comes with its own feature (such as heavy-tailed, symmetric, asymmetric, or bathtub shape) that characterizes the data generated. This has made it difficult to use a single distribution in all situations. Hence, the development of new distributions to quantify uncertainties with minimal loss of information is very important. Based on this assertion, we are motivated to create a new unit distribution in the following ways: Develop a unit distribution capable of fitting data that exhibit a symmetric, left-skewed, right-skewed, increasing, or bathtub probability density function (PDF); and formulate a quantile regression model to model a bounded response variable that is symmetric, skewed, or heavy-tailed. Based on this information, our study's objectives are fourfold: Develop the generalized UHLG (GUHLG) distribution to model a bounded response variable; study the statistical properties of the new distribution; formulate a quantile regression to model relationships between endogenous and exogenous variables; and demonstrate the application of our models using risk survey data.

The remainder of this paper is organized as follows: The development of the GUHLG distribution is given in Section 2. Its statistical properties are presented in Section 3. In Section 4, the maximum likelihood (ML) method is used to estimate the parameters of the distribution, and Monte Carlo simulations are performed to examine the suitability of the method. The quantile regression is formulated in Section 5, and simulation studies are carried out to investigate how well the considered estimates correspond to the parameters of the regression model. The univariate and multivariate applications of the models are presented in Section 6. The conclusion of the study is given in Section 7.

2. GUHLG Distribution

The UHLG distribution was recently introduced in [1]. The authors defined a random variable X as following the UHLG distribution if its cumulative distribution function (CDF) and PDF are, respectively, given by

$$F_X(x; \alpha) = 1 - \frac{\alpha(1-x)}{\alpha + (2-\alpha)x}, \alpha > 0, x \in (0, 1) \quad (1)$$

and

$$f_X(x; \alpha) = \frac{2\alpha}{(\alpha + (2-\alpha)x)^2}, x \in (0, 1). \quad (2)$$

Ramadan et al. [1] demonstrated that the PDF exhibits decreasing, increasing, and constant shapes for $0 < \alpha < 2$, $\alpha > 2$ and $\alpha = 2$, respectively. In this paper, we present the GUHLG distribution, a new generalization of the UHLG distribution based on the power transformation of X . The power transformation is known to improve the flexibility of the new distribution by enhancing its tail behavior and making it a suitable choice for modelling data with monotonic and non-monotonic hazard rate functions (HRFs) (see [11–15]). As a result, if $Y = X^{1/\gamma}$ with $\gamma > 0$, Y is said to follow the GUHLG distribution if its CDF is defined by

$$\begin{aligned} F_Y(y; \alpha, \gamma) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^{1/\gamma} \leq y) = \mathbb{P}(X \leq y^\gamma) \\ &= 1 - \frac{\alpha(1-y^\gamma)}{\alpha + (2-\alpha)y^\gamma}, \alpha > 0, \gamma > 0, y \in (0, 1). \end{aligned} \quad (3)$$

Thus defined, the GUHLG distribution appears to be a special case of the very general five-parameter Marshall–Olkin beta distribution created by Jose et al. [16]; it corresponds to the so-called $\text{MOBeta}(\gamma, 1, 0, 1, \frac{\alpha}{2})$ distribution. Thanks to this, several theoretical points established in [16] can be transposed to our paper. However, thanks to the simplicity of our case, we are able to provide more precise details for some theoretical results, crucial in our future statistical work (expression of the median, mode, quantiles, etc.). Thus, the GUHLG

distribution can be viewed as a motivated extension of the work of Ramadan et al. [1] and a highlight of a special case in the work of Jose et al. [16], with much more on the inferential aspect.

The corresponding PDF and HRF of the GUHLG distribution are, respectively, given by

$$f_Y(y; \alpha, \gamma) = \frac{2\alpha\gamma y^{\gamma-1}}{(\alpha + (2-\alpha)y^\gamma)^2}, y \in (0, 1) \quad (4)$$

and

$$h_Y(y; \alpha, \gamma) = \frac{2\gamma y^{\gamma-1}}{(1-y^\gamma)(\alpha + (2-\alpha)y^\gamma)}, y \in (0, 1). \quad (5)$$

It can be observed that as $y \rightarrow 0$, $f_Y(y; \alpha, \gamma) \sim h_Y(y; \alpha, \gamma) \sim \frac{2\gamma y^{\gamma-1}}{\alpha}$. This implies that

$$\lim_{y \rightarrow 0} f_Y(y; \alpha, \gamma) = \lim_{y \rightarrow 0} h_Y(y; \alpha, \gamma) = \begin{cases} \infty & \text{if } \gamma < 1 \\ \frac{2}{\alpha} & \text{if } \gamma = 1 \\ 0 & \text{if } \gamma > 1 \end{cases}$$

When $y \rightarrow 1$, we also have $f_Y(y; \alpha, \gamma) = \frac{\alpha\gamma}{2}$ and $h_Y(y; \alpha, \gamma) \rightarrow \infty$. The limiting behavior of the PDF shows that it can exhibit unimodal, reversed-J, bathtub, symmetric, right-skewed, and left-skewed shapes, as shown in Figure 1. The limiting behavior of the HRF also suggests that it can have various shapes, such as bathtub, increasing, and N-shaped, as also shown in Figure 1. It is worth indicating that the PDF does not exhibit the symmetric, left-skewed, right-skewed, or bathtub shapes. Furthermore, the HRF lacks an N-shape.

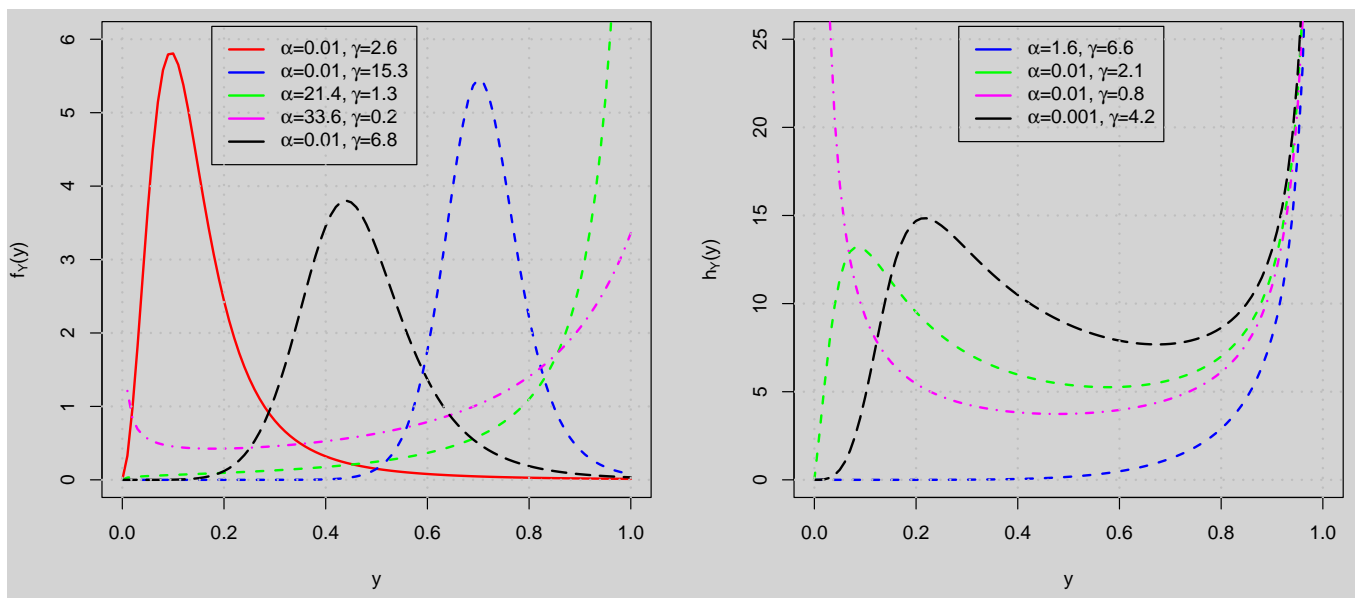


Figure 1. Plots of the PDF (left) and HRF (right).

3. Some Statistical Properties

This section presents the statistical properties of the GUHLG distribution.

3.1. Distributional Inequalities

Distributional inequalities are relevant in the study of first-order stochastic dominance, which is useful in the study of decision theory and analysis (see [17]).

Proposition 1. *The CDF of the GUHLG distribution is a decreasing function of the parameters α and γ .*

Proof. For $y \in (0, 1)$, since $y^\gamma - 1 \leq 0$, we have

$$\frac{\partial F_Y(y; \alpha, \gamma)}{\partial \alpha} = \frac{2y^\gamma(y^\gamma - 1)}{(\alpha + (2 - \alpha)y^\gamma)^2} \leq 0.$$

This implies that $F_Y(y; \alpha, \gamma)$ is decreasing with respect to the parameter α . Furthermore, for $y \in (0, 1)$, since $\log(y) \leq 0$, we have

$$\frac{\partial F_Y(y; \alpha, \gamma)}{\partial \gamma} = \frac{2\alpha y^\gamma \log(y)}{(\alpha + (2 - \alpha)y^\gamma)^2} \leq 0.$$

This means that $F_Y(y; \alpha, \gamma)$ is decreasing with respect to γ . Hence, the proof of the proposition is complete. \square

The following first-order stochastic dominance property follows immediately from the proposition. If $\alpha_1 \leq \alpha_2$, then we have $F_Y(y; \alpha_2, \gamma) \leq F_Y(y; \alpha_1, \gamma)$. Again, if $\gamma_1 \leq \gamma_2$, then we have $F_Y(y; \alpha, \gamma_2) \leq F_Y(y; \alpha, \gamma_1)$.

3.2. Quantile Function

The quantile function (QF) is used in the computation of measures of shapes and dispersion when the classical moments do not exist and also for the generation of random observations from a distribution.

Proposition 2. *The QF of the GUHLG distribution is given by*

$$Q(p; \alpha, \gamma) = \left(\frac{\alpha p}{2 - 2p + \alpha p} \right)^{1/\gamma}, \quad p \in (0, 1). \quad (6)$$

Proof. The QF is obtained by solving the equation $F_Y(y; \alpha, \gamma) = p$ with respect to y . Hence, after some manipulations, solving for y in

$$1 - \frac{\alpha(1 - y^\gamma)}{\alpha + (2 - \alpha)y^\gamma} = p,$$

yields the QF of the GUHLG distribution. \square

The QF can be used to compute measures of shapes such as the Bowley (B) coefficient of skewness and the Moor (M) coefficient of kurtosis. The B coefficient of skewness is given by

$$B = \frac{Q(0.75; \alpha, \gamma) + Q(0.25; \alpha, \gamma) - 2Q(0.5; \alpha, \gamma)}{Q(0.75; \alpha, \gamma) - Q(0.25; \alpha, \gamma)}$$

and the M coefficient of kurtosis is specified by

$$M = \frac{Q(0.375; \alpha, \gamma) - Q(0.125; \alpha, \gamma) + Q(0.875; \alpha, \gamma) - Q(0.625; \alpha, \gamma)}{Q(0.75; \alpha, \gamma) - Q(0.25; \alpha, \gamma)}.$$

The plots of the B skewness and M kurtosis are shown in Figure 2. The B skewness plot shows that the distribution can be left- or right-skewed. Furthermore, the M kurtosis plot reveals that the distribution can assume platykurtic or leptokurtic shapes.

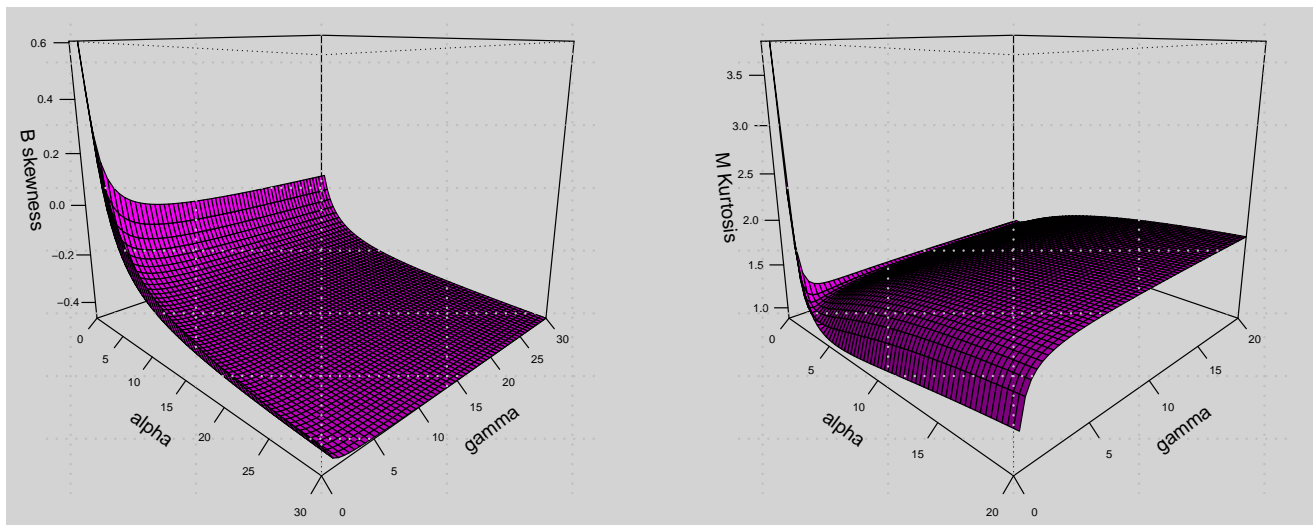


Figure 2. Plots of the B skewness (left) and the M kurtosis (right).

The QF can be used to generate random observations of size n from the GUHLG distribution using the following algorithm:

1. Set the values of the parameters α and γ .
2. Obtain p as a random observation of a random variable that follows the standard uniform distribution, $U(0, 1)$.
3. Estimate $y = \left(\frac{\alpha p}{2 - 2p + \alpha p} \right)^{1/\gamma}$.
4. Repeat steps 2 and 3 n times to obtain n values: y_1, \dots, y_n .

3.3. Moments

The non-central moments of a random variable are useful in estimating measures of central tendency, shape, and dispersion. They always exist for bounded random variables. In full generality, the following integral gives the r th non-central moment of a GUHLG distribution random variable:

$$\mu'_r = \int_0^1 y^r \frac{2\alpha\gamma y^{\gamma-1}}{(\alpha + (2-\alpha)y^\gamma)^2} dy.$$

It is worth noting that μ'_r exists in the mathematical sense, and satisfies $\mu'_r \in (0, 1]$. There is no straightforward analytical expression for μ'_r because of the intricate nature of the integrated function. However, it can always be numerically calculated by setting the parameter values. Basically, the mean is obtained as $\mu = \mu'_1$. The first six moments, standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS) and coefficient of kurtosis (CK) are given in Table 1. The first six moments are estimated numerically using R software. The values for SD, CV, CS, and CK are computed, respectively, using the following standard formulas:

$$\begin{aligned} \text{SD} &= \sqrt{\mu'_2 - \mu^2}, & \text{CV} &= \frac{\text{SD}}{\mu} = \sqrt{\frac{\mu'_2}{\mu^2} - 1}, \\ \text{CS} &= \frac{\mu'_3 - 3\mu\mu'_2 + 2\mu^3}{(\mu'_2 - \mu^2)^{3/2}} & \text{and} & \quad \text{CK} = \frac{\mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4}{(\mu'_2 - \mu^2)^2}. \end{aligned}$$

From Table 1, the values of CS suggest that the distribution can be left- or right-skewed. On the other hand, the values of CK reveal that the distribution can be platykurtic or leptokurtic.

Table 1. Moments, SD, CV, CS and CK.

	$(\alpha, \gamma) = (0.8, 3.5)$	$(\alpha, \gamma) = (0.2, 0.6)$	$(\alpha, \gamma) = (2.5, 0.4)$	$(\alpha, \gamma) = (4.5, 2.3)$
μ'_1	0.681	0.092	0.322	0.790
μ'_2	0.503	0.038	0.195	0.662
μ'_3	0.392	0.024	0.140	0.574
μ'_4	0.318	0.017	0.109	0.508
μ'_5	0.266	0.013	0.090	0.457
μ'_6	0.227	0.011	0.076	0.416
SD	0.196	0.172	0.302	0.195
CV	0.288	1.878	0.936	0.247
CS	−0.413	2.842	0.675	−1.246
CK	2.446	11.341	2.166	4.026

3.4. Order Statistics

The usefulness of order statistics in the areas of finance and insurance is not new in the literature. The applications of order statistics in the study of low or high events, last survivor policies, and exceedances, among others, are essential in finance and insurance. The authors in [18] employed the concept of order statistics to study the expected utility insurance premium principle. The authors of [19] studied ruin and deficit under claim arrivals that exhibit order statistics property. The authors of [20] illustrated how to compute maximum loss using order statistics. On the other hand, the authors of [21] used the properties of order statistics to demonstrate their application in fire protection and insurance problems. The values of the order statistics are obtained when we arrange the observations from the distribution of Y in ascending order. Let $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ be the order statistics of the random sample Y_1, Y_2, \dots, Y_n from the GUHLG distribution. Using the expanded form of the PDF of $Y_{k:n}$ (see [22]), for $k = 1, 2, \dots, n$, the PDF of $Y_{k:n}$ for the GUHLG distribution is

$$f_{k:n}(y; \alpha, \gamma) = D_{k:n} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \frac{2\alpha\gamma y^{\gamma-1} (\alpha(1-y^\gamma))^{n-k+j}}{(\alpha + (2-\alpha)y^\gamma)^{n-k+j+2}}, \quad (7)$$

where

$$D_{k:n} = \frac{n!}{(k-1)!(n-k)!}.$$

The smallest ($Y_{1:n}$) and largest ($Y_{n:n}$) order statistics can be used to predict the minimum and maximum occurrence of extreme events, respectively. Hence, their distributions are of interest for further probabilistic or statistical analysis. Here, the PDF of $Y_{1:n}$ is given by

$$\begin{aligned} f_{1:n}(y; \alpha, \gamma) &= n f_Y(y; \alpha, \gamma) [1 - F_Y(y; \alpha, \gamma)]^{n-1} \\ &= \frac{2n\alpha\gamma y^{\gamma-1} (\alpha(1-y^\gamma))^{n-1}}{(\alpha + (2-\alpha)y^\gamma)^{n+1}}, \end{aligned}$$

and that of $Y_{n:n}$ is

$$\begin{aligned} f_{n:n}(y; \alpha, \gamma) &= n f_Y(y; \alpha, \gamma) [F_Y(y; \alpha, \gamma)]^{n-1} \\ &= \frac{2n\alpha\gamma y^{\gamma-1}}{(\alpha + (2-\alpha)y^\gamma)^2} \left(1 - \frac{\alpha(1-y^\gamma)}{\alpha + (2-\alpha)y^\gamma} \right)^{n-1}. \end{aligned}$$

The possible shapes of the distribution can be investigated using the minimum and maximum (min–max) plots of the order statistics. These plots depend on $\mathbb{E}(Y_{1:n})$ and $\mathbb{E}(Y_{n:n})$. The min–max plot can be used to describe the possible shapes of the distribution. The min–max plots of the GUHLG distribution shown in Figure 3 reveal that the GUHLG distribution can be left- or right-skewed.

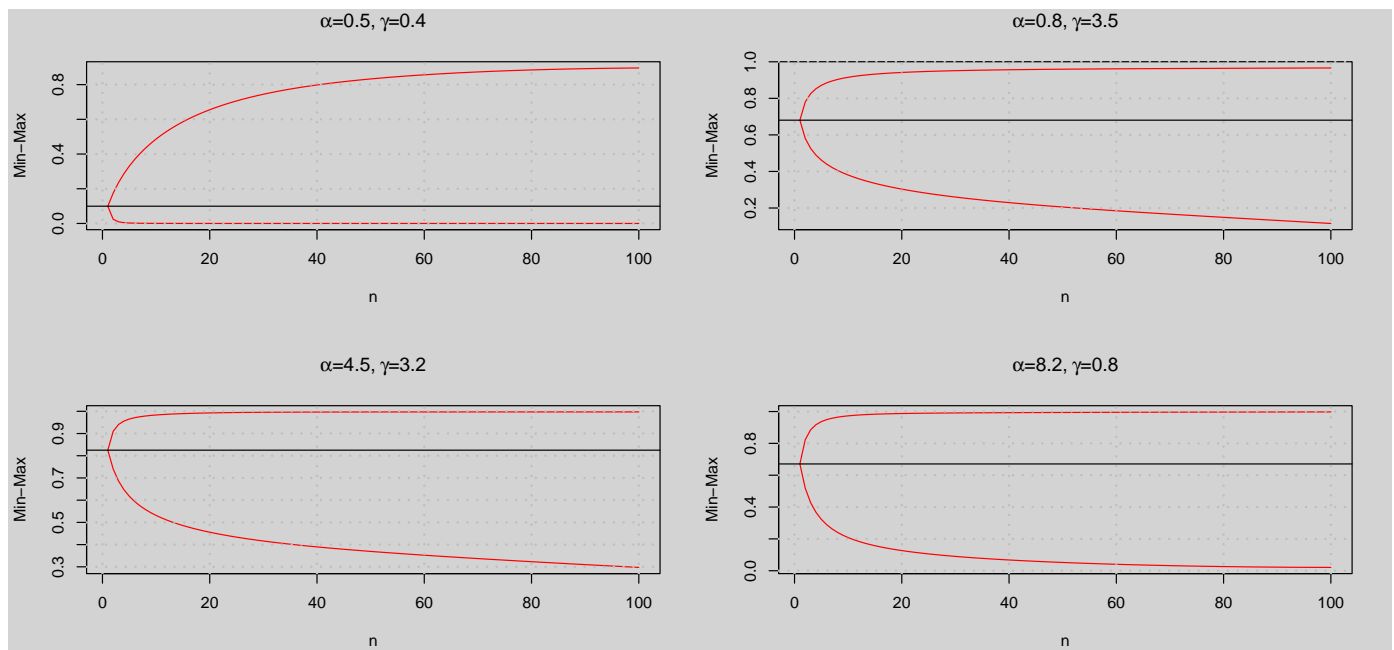


Figure 3. Min–max plots of the GUHLG distribution.

4. Parameter Estimation

In this section, we present how to estimate the parameters of the GUHLG distribution using the ML method. Suppose that y_1, y_2, \dots, y_n are independent and identically distributed random observations of size n from the GUHLG distribution representing the data. The log-likelihood function is then given by

$$\ell = n \log(2\alpha\gamma) + (\gamma - 1) \sum_{i=1}^n \log(y_i) - 2 \sum_{i=1}^n \log(\alpha + (2 - \alpha)y_i^\gamma). \quad (8)$$

The estimates of the parameters can be obtained by directly maximizing the function in Equation (8) according to the parameters. Alternatively, the estimates can be obtained by finding the partial derivatives of the log-likelihood function and solving the resulting system simultaneously. In this case, we need

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} - 2 \sum_{i=1}^n \frac{1 - y_i^\gamma}{\alpha + (2 - \alpha)y_i^\gamma} \quad (9)$$

and

$$\frac{\partial \ell}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^n \log(y_i) - 2 \sum_{i=1}^n \frac{(2 - \alpha) \log(y_i) y_i^\gamma}{\alpha + (2 - \alpha)y_i^\gamma}. \quad (10)$$

Equating Equations (9) and (10) to zero and solving them simultaneously gives the ML estimates of the parameters. However, the solutions of these equations do not have a closed form. Hence, numerical methods are employed. The random version of the vector of the ML estimates of the parameters has an approximate bivariate normal distribution with

zero mean, $\mathbf{0}$, and variance–covariance J^{-1} under the mild regularity conditions (see [23]), where J is the observed information matrix given by

$$J = - \left(\begin{array}{cc} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \gamma} & \frac{\partial^2 \ell}{\partial \gamma^2} \end{array} \right) \bigg|_{(\alpha, \gamma) = (\hat{\alpha}, \hat{\gamma})}.$$

The elements of J are given by

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{n}{\alpha^2} + 2 \sum_{i=1}^n \frac{(1 - y_i^\gamma)^2}{(\alpha + (2 - \alpha)y_i^\gamma)^2}, \quad (11)$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} = -\frac{n}{\gamma^2} - 2 \sum_{i=1}^n \left(\frac{(2 - \alpha)(\log(y_i))^2 y_i^\gamma}{\alpha + (2 - \alpha)y_i^\gamma} - \frac{(2 - \alpha)^2 (\log(y_i))^2 y_i^{2\gamma}}{(\alpha + (2 - \alpha)y_i^\gamma)^2} \right) \quad (12)$$

and

$$\frac{\partial^2 \ell}{\partial \alpha \partial \gamma} = \frac{\partial^2 \ell}{\partial \gamma \partial \alpha} = 2 \sum_{i=1}^n \left(\frac{(2 - \alpha) \log(y_i) y_i^\gamma (1 - y_i^\gamma)}{(\alpha + (2 - \alpha)y_i^\gamma)^2} + \frac{\log(y_i) y_i^\gamma}{\alpha + (2 - \alpha)y_i^\gamma} \right). \quad (13)$$

The variance–covariance matrix can be used to obtain interval estimates of the parameters. The approximate $100(1 - v)\%$ confidence interval for the parameters are given by $\hat{\alpha} \pm z_{v/2} \sqrt{J_{\alpha\alpha}^{-1}}$ and $\hat{\gamma} \pm z_{v/2} \sqrt{J_{\gamma\gamma}^{-1}}$, where $z_{v/2}$ is the upper $(\frac{v}{2})$ th percentile of the standard normal distribution and J_{ii}^{-1} and the diagonal elements of J^{-1} for $i = \alpha$ and γ .

Simulation Studies

In this subsection, Monte Carlo simulation experiments are conducted to investigate the performance of the ML method in estimating the parameters of the distribution. The experiments are carried out using small, moderate, and large sample sizes. Random observations of size $n = 20, 60, 100, 250, 500, 800$ and 1000 are generated from the GUHLG distribution using its QF given in Equation (6). The experiments are replicated 5000 times for each sample size. The following three parameter combinations: I: $\alpha = 0.01, \gamma = 2.6$; II: $\alpha = 0.01, \gamma = 15.3$ and III: $\alpha = 0.01, \gamma = 0.8$ are used during the simulations. The performance of the ML method is assessed using the mean estimate (ME), average bias (AB), average relative bias (ARB), root mean square error (RMSE) and coverage probability (CP) of the ML estimates. The algorithm for the Monte Carlo simulation is as follows:

1. Generate 5000 random samples of size $n = 20, 60, 100, 250, 500, 800$ and 1000 from the GUHLG distribution using the algorithm discussed in Section 3.3.
2. Find the ML estimates of the parameters.
3. Compute the MEs, ABs, ARBs, RMSEs, and CPs of the parameters.
4. Repeat steps 1 to 3 for the three parameter combinations.

The MEs approach the true values of the parameters as the sample size increases. The ABs, ARBs, and RMSEs decrease as the sample size increases, as shown in Table 2. This suggests that the consistency property of the ML method has been attained. The CPs of the parameters are quite high and approach the nominal value of 0.95 as the sample size increases. Hence, it can be concluded that the ML method estimates the parameters well.

Table 2. Simulation results.

Parameter	<i>n</i>	I: $\alpha = 0.01, \gamma = 2.6$					II: $\alpha = 0.01, \gamma = 15.3$					III: $\alpha = 0.01, \gamma = 0.8$				
		ME	AB	ARB	RMSE	CP	ME	AB	ARB	RMSE	CP	ME	AB	ARB	RMSE	CP
α	20	0.013	0.010	0.964	0.017	0.795	0.012	0.009	0.919	0.015	0.782	0.013	0.010	0.964	0.016	0.781
	60	0.011	0.005	0.531	0.007	0.877	0.011	0.005	0.522	0.007	0.881	0.011	0.005	0.535	0.007	0.870
	100	0.011	0.004	0.418	0.006	0.887	0.011	0.004	0.413	0.006	0.893	0.011	0.004	0.404	0.005	0.905
	250	0.010	0.003	0.253	0.003	0.927	0.010	0.003	0.258	0.003	0.926	0.010	0.003	0.253	0.003	0.928
	500	0.010	0.002	0.179	0.002	0.943	0.010	0.002	0.181	0.002	0.938	0.010	0.002	0.181	0.002	0.941
	800	0.010	0.001	0.145	0.002	0.940	0.010	0.001	0.145	0.002	0.942	0.010	0.001	0.147	0.002	0.939
	1000	0.010	0.001	0.130	0.002	0.944	0.010	0.001	0.129	0.002	0.950	0.010	0.001	0.134	0.002	0.942
γ	20	2.769	0.450	0.173	0.599	0.951	16.464	2.742	0.179	3.634	0.952	0.855	0.142	0.177	0.185	0.957
	60	2.654	0.246	0.095	0.316	0.948	15.622	1.433	0.094	1.836	0.953	0.819	0.077	0.096	0.098	0.951
	100	2.636	0.194	0.075	0.245	0.947	15.506	1.104	0.072	1.393	0.949	0.810	0.057	0.071	0.072	0.951
	250	2.613	0.118	0.045	0.147	0.947	15.372	0.690	0.045	0.862	0.952	0.804	0.036	0.045	0.045	0.955
	500	2.606	0.081	0.031	0.102	0.953	15.336	0.485	0.032	0.605	0.952	0.802	0.025	0.032	0.032	0.952
	800	2.606	0.067	0.026	0.083	0.949	15.340	0.390	0.025	0.485	0.951	0.801	0.021	0.026	0.026	0.947
	1000	2.603	0.060	0.023	0.074	0.952	15.313	0.345	0.023	0.430	0.951	0.801	0.019	0.024	0.023	0.949

5. Quantile Regression

The development of parametric quantile regressions has received much attention recently due to their robustness when it comes to modelling asymmetric data or data containing extreme values. The quantile regression is also capable of handling asymmetric and heavy-tailed response variables defined on the interval $(0, 1)$. The development of these regressions requires re-parametrization of the PDFs of the distribution in terms of the quantile to obtain the quantile PDF (see [2,22,24–28]). To formulate the GUHLG distribution quantile regression model, we first make the parameter α the subject in the QF of the GUHLG distribution and then substitute it in the CDF and PDF. These give the quantile CDF and PDF of the GUHLG distribution after simplifications. The quantile CDF and PDF are, respectively, given by

$$F_Y(y; p, \mu, \gamma) = 1 - \frac{2\mu^\gamma(p-1)(1-y^\gamma)}{p(\mu^\gamma-1) \left[\frac{2\mu^\gamma(p-1)}{p(\mu^\gamma-1)} + \left(2 - \frac{2\mu^\gamma(p-1)}{p(\mu^\gamma-1)} \right) y^\gamma \right]}, y \in (0, 1) \quad (14)$$

and

$$f_Y(y; p, \mu, \gamma) = \frac{4\gamma\mu^\gamma(p-1)y^{\gamma-1}}{p(\mu^\gamma-1) \left[\frac{2\mu^\gamma(p-1)}{p(\mu^\gamma-1)} + \left(2 - \frac{2\mu^\gamma(p-1)}{p(\mu^\gamma-1)} \right) y^\gamma \right]^2}, y \in (0, 1), \quad (15)$$

where $\mu \in (0, 1)$ is the quantile parameter and $p \in (0, 1)$. When we substitute $p = 0.10, 0.25, 0.50, 0.75$ and 0.90 , the 10th, 25th, 50th, 75th and 90th percentile PDFs are obtained. Figure 4 shows the quantile PDF plots for different quantiles and parameter values. The quantile PDF shows different shapes such as left-skewed, right-skewed, decreasing, increasing, symmetric, and bathtub. This is an indication that the regression model formulated from this PDF is flexible enough to handle bounded data with such characteristics.

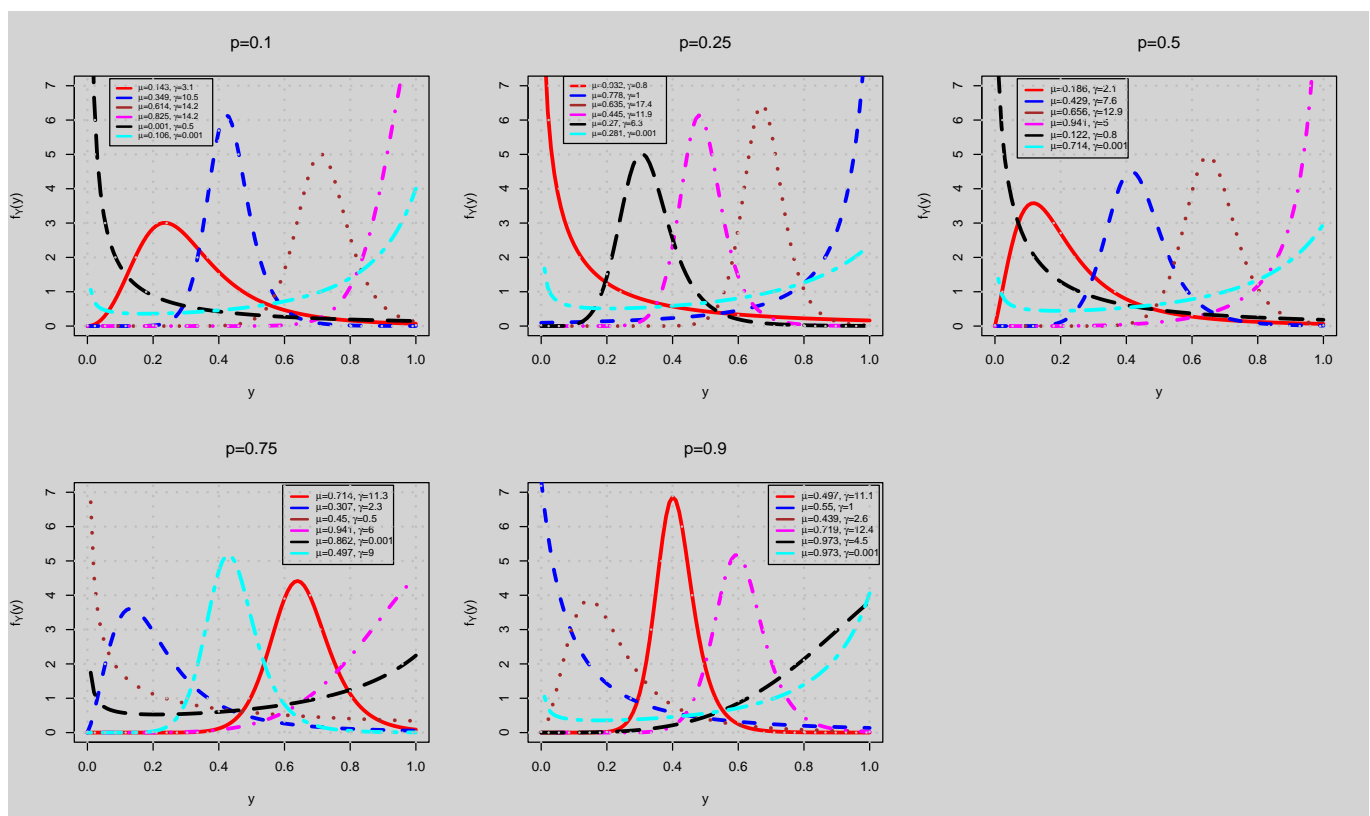


Figure 4. Quantile PDF plots.

Let y_1, y_2, \dots, y_n be random observations from the GUHLG distribution, and x_1, x_2, \dots, x_n be non-random exogenous variables. Then, the GUHLG quantile regression is obtained by relating the conditional quantile of the response variable and the exogenous variables using an appropriate link function in the following manner:

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\eta}, \quad (16)$$

where $\boldsymbol{\eta} = (\eta_0, \eta_1, \dots, \eta_k)^T$ is the vector of the coefficients of the exogenous variables, $\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ik})$, and $g(\cdot)$ is the desired link function. Although different link functions exist, such as the logit, probit, and complementary log-log, among others, the logit link function is used in this study due to the ease of interpretation of the exogenous variable coefficients. Hence, we have the following regression structure:

$$\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\eta}. \quad (17)$$

To obtain the log-likelihood to estimate the parameters of the regression model, we substitute

$$\mu_i = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\eta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\eta})}$$

into the quantile PDF. The log-likelihood function is therefore given by

$$\begin{aligned} \ell = & n \log(4\gamma(1-p)) + \gamma \sum_{i=1}^n \log(\mu_i) + (\gamma-1) \sum_{i=1}^n \log(y_i) - \sum_{i=1}^n \log(p(1-\mu_i^\gamma)) - \\ & 2 \sum_{i=1}^n \log\left(\frac{2\mu_i^\gamma(p-1)}{p(\mu_i^\gamma-1)} + \left(2 - \frac{2\mu_i^\gamma(p-1)}{p(\mu_i^\gamma-1)}\right) y_i^\gamma\right). \end{aligned} \quad (18)$$

The estimates of the parameters are obtained by maximizing Equation (18) with respect to the parameters. Alternatively, we can consider the elements of the score vector obtained by differentiating Equation (18) with respect to the parameters. They are given by

$$\begin{aligned} \frac{\partial \ell}{\partial \gamma} = & \frac{n}{\gamma} + \sum_{i=1}^n \log(\mu_i) + \sum_{i=1}^n \log(y_i) - \sum_{i=1}^n \frac{\mu_i^\gamma \log(\mu_i)}{\mu_i^\gamma - 1} - \\ & 2 \sum_{i=1}^n \left[\frac{\left(\log(\mu_i) - \frac{\mu_i^\gamma \log(\mu_i)}{(\mu_i^\gamma - 1)}\right) (1 - y_i^\gamma) + y_i^\gamma \left(\frac{p(\mu_i^\gamma - 1)}{(p-1)\mu_i^\gamma} - 1\right) \log(y_i)}{1 + y_i^\gamma \left(\frac{p(\mu_i^\gamma - 1)}{(p-1)\mu_i^\gamma} - 1\right)} \right], \\ \frac{\partial \ell}{\partial \eta_r} = & \gamma \left(\sum_{i=1}^n \frac{1}{\mu_i} - \sum_{i=1}^n \frac{\mu_i^{\gamma-1}}{\mu_i^\gamma - 1} \right) \frac{\partial \mu_i}{\partial \eta_r} - 2\gamma \sum_{i=1}^n \left(\frac{(1 - y_i^\gamma) \left(\frac{\mu_i^\gamma}{\mu_i^\gamma - 1} + 1\right)}{1 + y_i^\gamma \left(\frac{p(\mu_i^\gamma - 1)}{(p-1)\mu_i^\gamma} - 1\right)} \right) \frac{\partial \mu_i}{\partial \eta_r}, \end{aligned}$$

for $r = 1, 2, \dots, k$. By taking into account the logit link function, we have

$$\frac{\partial \mu_i}{\partial \eta_r} = \mu_i(1 - \mu_i)x_{ir}, \quad i = 1, 2, \dots, n; \quad r = 1, 2, \dots, k.$$

The estimates of the parameters can be obtained by equating the elements of the score vector to zero and solving the resulting system of equations simultaneously. The median regression is fitted by putting $p = 0.50$ into Equation (18) and then maximizing the resulting log-likelihood function. The estimates of the standard errors of the parameters are obtained based on the large sample property of the ML technique. The authors of [29]

have shown that the observed Fisher information matrix for estimating standard errors of the parameters is

$$I(\hat{\eta}) = -\frac{\partial^2 \ell(\eta|y)}{\partial \eta^T \partial \eta^T} \Big|_{\eta=\hat{\eta}}.$$

5.1. Residual Analysis

After using the regressions to model datasets, it is imperative to examine whether the models provide an adequate fit to the data. This can easily be performed by assessing the behavior of the model's residuals. In this study, the randomized quantile residuals (RQRs) of the models are assessed to see if the model provides a good fit to the data. For any $i = 1, 2, \dots, n$, the i th RQR is given by

$$r_i = \Phi^{-1}(F_Y(y_i; \hat{\eta})),$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF (or QF) of the standard normal distribution and $\hat{\eta}$ is the estimated vector of parameters of the model. The RQRs are expected to follow the standard normal distribution if the model provides good fit to the data (see [30]).

5.2. Monte Carlo Simulations for Regressions

In this subsection, Monte Carlo simulations are performed to investigate how well the ML method estimates the parameters of the quantile regression. The simulation exercise is repeated 5000 times for each sample size $n = 20, 60, 100, 250, 500, 800$ and 1000. The following parameter combinations are used for the quantile regression simulation: $(\eta_0, \eta_1, \eta_2, \gamma) = (0.3, 0.2, 0.7, 1.3)$ and $(\eta_0, \eta_1, \eta_2, \gamma) = (1.3, 0.5, 0.4, 2.5)$. The following regression structure is considered:

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_0 + \eta_1 x_{i1} + \eta_2 x_{i2}$$

for the simulation. The simulation exercise is performed using the median regression by substituting $p = 0.5$. The exogenous variable, x_{i1} , is a binary variable generated from the Bernoulli distribution with probability 0.5, and x_{i2} is a continuous variable generated from the standard normal distribution. These exogenous variables are held as fixed constants during the simulation. The observations for the endogenous variable are random samples generated using the inversion method. The performance of the ML method is assessed using the ME, AB, ARB, RMSE, and CP. The simulation algorithm for the regression is as follows:

1. Generate the exogenous variables x_{i1} and x_{i2} from the Bernoulli and standard normal distributions, respectively.
2. Generate the endogenous variable y_i using

$$y_i = \frac{\alpha_i u_i}{2 - 2u_i + \alpha_i u_i},$$

where u_i is an observation from standard uniform distribution, $\alpha_i = \frac{2\mu_i^\gamma(p-1)}{p(\mu_i^\gamma-1)}$ and $\mu_i = \frac{\exp(\eta_0 + \eta_1 x_{i1} + \eta_2 x_{i2})}{1 + \exp(\eta_0 + \eta_1 x_{i1} + \eta_2 x_{i2})}$.

3. Compute the ML estimates of the parameters of the regression model.
4. Compute the MEs, ABs, ARB, RMSEs and CPs of the parameters.
5. Repeat steps 1 to 4 for the two parameter combinations.

Tables 3 and 4 present the simulation results for the quantile regression for different conditional quantiles. The results show that the MEs approach the true parameter value as the sample size increases. Furthermore, the ABs, ARBs, and RMSEs decrease as the sample size increases. The CPs are quite high and close to the 0.95 value. Hence, the ML approach estimates the parameters of the quantile regression for the different conditional quantiles well.

Table 3. Quantile regression simulation results for $(\eta_0, \eta_1, \eta_2, \gamma) = (0.3, 0.2, 0.7, 1.3)$.

Parameter	n	ME	AB	ARB	RMSE	CP
η_0	20	0.200	0.220	0.734	0.249	0.999
	60	0.275	0.166	0.553	0.200	0.999
	100	0.274	0.120	0.399	0.146	0.999
	250	0.276	0.090	0.301	0.113	0.959
	500	0.302	0.071	0.235	0.087	0.956
	800	0.305	0.056	0.186	0.069	0.952
	1000	0.297	0.055	0.182	0.068	0.931
η_1	20	0.324	0.314	1.571	0.425	0.980
	60	0.248	0.214	1.069	0.265	0.984
	100	0.235	0.184	0.920	0.226	0.978
	250	0.213	0.129	0.647	0.157	0.981
	500	0.196	0.099	0.494	0.119	0.984
	800	0.197	0.075	0.373	0.094	0.962
	1000	0.203	0.073	0.364	0.090	0.955
η_2	20	0.704	0.254	0.363	0.325	0.933
	60	0.696	0.156	0.223	0.197	0.928
	100	0.698	0.115	0.164	0.144	0.946
	250	0.704	0.072	0.102	0.089	0.947
	500	0.699	0.051	0.072	0.063	0.941
	800	0.701	0.040	0.057	0.049	0.962
	1000	0.699	0.036	0.052	0.046	0.940
γ	20	2.468	0.911	0.506	1.177	0.925
	60	1.962	0.385	0.214	0.489	0.935
	100	1.836	0.215	0.120	0.268	0.966
	250	1.851	0.134	0.074	0.169	0.969
	500	1.842	0.113	0.063	0.138	0.965
	800	1.812	0.089	0.050	0.112	0.934
	1000	1.806	0.079	0.044	0.101	0.928

Table 4. Quantile regression simulation results for $(\eta_0, \eta_1, \eta_2, \gamma) = (1.3, 0.5, 0.4, 2.5)$.

Parameter	n	ME	AB	ARB	RMSE	CP
η_0	20	1.323	0.302	0.233	0.381	0.965
	60	1.262	0.191	0.147	0.242	0.963
	100	1.329	0.168	0.129	0.207	0.957
	250	1.277	0.099	0.076	0.125	0.962
	500	1.298	0.076	0.058	0.096	0.932
	800	1.303	0.057	0.044	0.072	0.951
	1000	1.304	0.052	0.040	0.066	0.943
η_1	20	0.630	0.490	0.980	0.601	0.955
	60	0.520	0.295	0.589	0.360	0.962
	100	0.514	0.240	0.480	0.296	0.966
	250	0.503	0.144	0.289	0.183	0.952
	500	0.502	0.108	0.217	0.136	0.942
	800	0.495	0.083	0.166	0.102	0.961
	1000	0.496	0.076	0.153	0.096	0.946
η_2	20	0.433	0.268	0.670	0.325	0.943
	60	0.407	0.160	0.401	0.202	0.936
	100	0.400	0.120	0.299	0.151	0.943
	250	0.399	0.075	0.188	0.095	0.946
	500	0.399	0.049	0.123	0.062	0.955
	800	0.400	0.042	0.104	0.052	0.948
	1000	0.402	0.037	0.093	0.047	0.946
γ	20	3.548	1.848	0.739	2.363	0.843
	60	2.787	0.562	0.225	0.685	0.997
	100	2.653	0.556	0.222	0.711	0.942
	250	2.521	0.302	0.121	0.381	0.929
	500	2.536	0.189	0.076	0.240	0.965
	800	2.488	0.165	0.066	0.203	0.953
	1000	2.519	0.142	0.057	0.180	0.946

6. Application

In this section, the univariate and multivariate applications of the developed distribution are illustrated.

6.1. Univariate Application

In this subsection, the univariate application of the GUHLG distribution is illustrated using insurance data. The data denote the firm cost (firm-specific ratio of premiums plus uninsured losses divided by total assets) reported by 73 managers out of 374 questionnaires sent to managers in large U.S.-based organizations. The data were first reported by Schmit and Roth [31]. Researchers, such as those in [1,2,32], studied the data by dividing it by 100 to rescale it on the unit interval. The GUHLG distribution is fitted to the data, and its performance is compared to that of the UHLG distribution, beta distribution, Kumaraswamy distribution, unit power Weibull (UPW) distribution (see [33]), log-XLindley (LXL) distribution (see [34]), log-Bilal (LB) distribution (see [35]), unit Burr XII (UBXII) distribution

(see [36]), unit Burr III (UBIII) distribution (see [37]), unit Weibull (UW) distribution (see [5]) and exponentiated Topp-Leone (ETL) distribution (see [38]). The comparison benchmarks are the -2ℓ , Akaike information criterion (AIC), AIC difference (ΔAIC), Akaike weights (ω), Bayesian information criterion (BIC) and Kolmogorov–Smirnov (KS) statistic. The ΔAIC is estimated using $\Delta\text{AIC} = \text{AIC}_i - \text{AIC}_{\min}$, $i = 1, 2, \dots, R$, where R is the number of distributions to be compared. The best distribution has $\Delta\text{AIC} = 0$. The difference in the performance of the distributions is considered significant if $\Delta\text{AIC} > 2$. The Akaike weights are computed using the following formula:

$$\omega = \frac{\exp(-\Delta\text{AIC}/2)}{\sum_{i=1}^R \exp(-\Delta\text{AIC}_i/2)}.$$

We recall that the Akaike weight of a distribution is interpreted as the likelihood that the distribution is the best given the data and the other distributions under consideration. The higher the weight, the better the distribution. We consider distributions with $\omega > 0.9$ as the best. Furthermore, the distribution with the smallest values of -2ℓ , AIC, BIC and KS is considered the best. Figure 5 displays the kernel density, boxplot, and violin plots of the data. The plots clearly show that the data are right-skewed and contain some outliers. The ML estimates of the parameters and their standard errors, AIC, ΔAIC , ω , BIC and KS values are given in Table 5. The GUHLG distribution has the smallest values of -2ℓ , AIC, BIC and KS. It has $\Delta\text{AIC} = 0$ and $\omega = 0.9518$. Thus, the GUHLG distribution provides the best fit to the data.

We further explore how well the GUHLG distribution fits the given data using the probability–probability (P-P) plots in Figure 6. These plots further suggest that the GUHLG distribution provides the best fit to the data.

To ascertain whether the ML estimates of the parameters of the GUHLG distribution are unique and represent the true maxima, we plot the profile log-likelihood plots of the parameters in Figure 7. This figure reveals that the estimates are unique and true maxima.

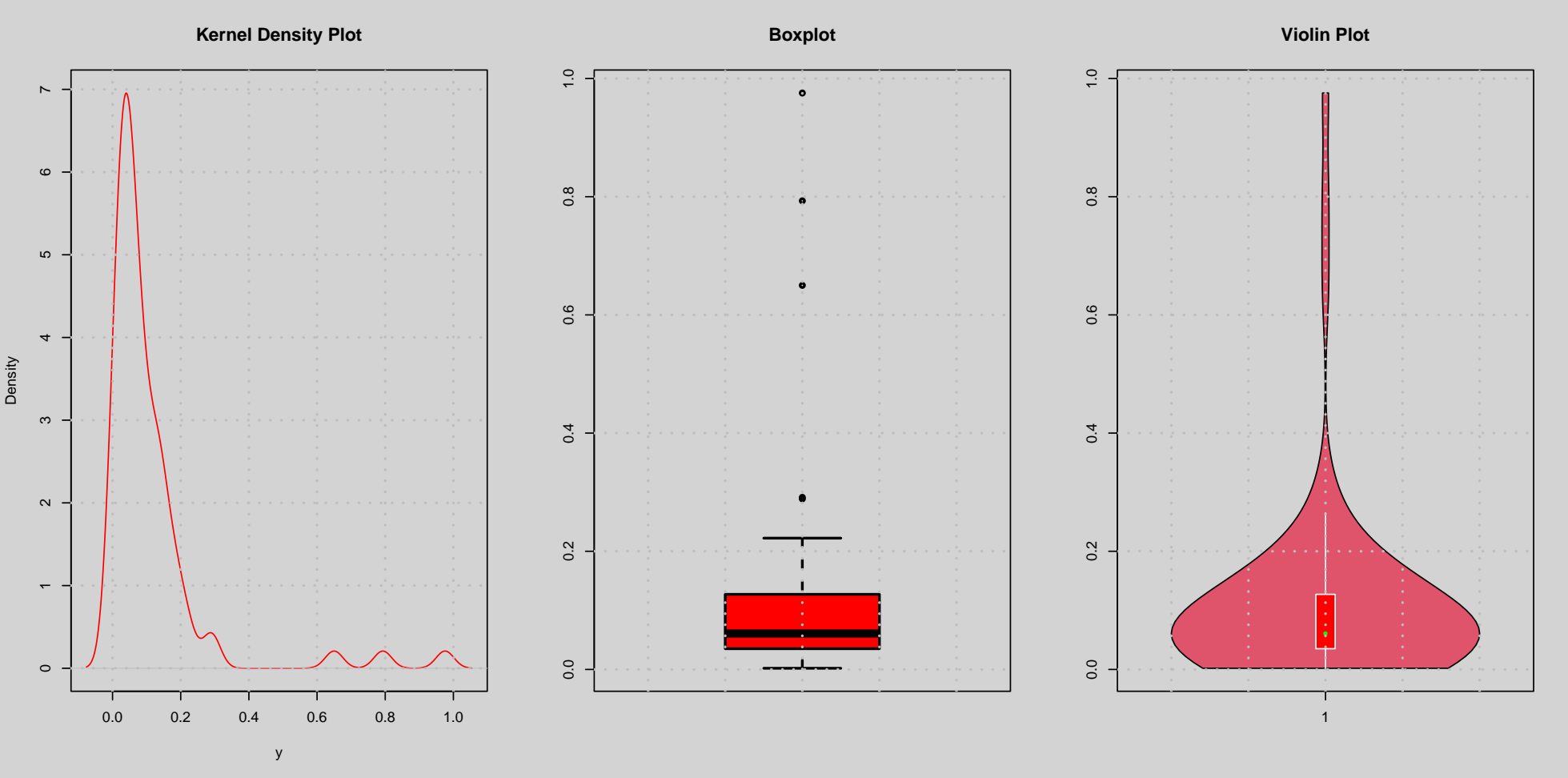


Figure 5. Kernel density, boxplot and violin plots.

Table 5. Estimates, information criteria and goodness-of-fit statistics.

Distribution	α	γ	λ	-2ℓ	AIC	Δ AIC	ω	BIC	KS (<i>p</i> Value)
GUHLG	0.038 (0.019)	1.432 (0.161)		−187.265	−183.265	0.000	0.952	−178.684	0.063 (0.937)
UHLG	0.132 (0.025)			−179.071	−177.071	6.194	0.043	−174.780	0.119 (0.252)
Beta	0.613 (0.086)	3.798 (0.715)		−152.235	−148.235	35.030	<0.001	−143.654	0.181 (0.017)
Kumaraswamy	0.665 (0.072)	3.441 (0.621)		−157.308	−153.308	29.957	<0.001	−148.727	0.154 (0.064)
UBIII	0.234 (0.052)	1.532 (0.297)		−123.663	−119.663	63.602	<0.001	−115.082	0.318 (7.477×10^{-7})
UG	0.150 (0.055)	0.605 (0.076)		−174.298	−170.298	12.967	0.002	−165.717	0.131 (0.162)
UW	0.0655 (0.020)	2.353 (0.214)		−176.201	−172.201	11.064	0.004	−167.620	0.093 (0.552)
ETL	0.654 (0.080)	1.961 (0.322)		−153.906	−149.906	33.358	<0.001	−145.325	0.165 (0.037)
LXL	0.500 (0.044)			−129.518	−127.518	55.746	<0.001	−125.228	0.304 (2.883×10^{-6})
LB	3.164 (0.282)			−149.953	−147.953	35.312	<0.001	−145.662	0.264 (7.515×10^{-5})
UPW	500.000 (9.669×10^{-8})	0.700 (0.054)	0.001 (7.558×10^{-4})	−165.738	−159.738	23.526	<0.001	−152.867	0.126 (0.196)
UBXII	0.348 (0.063)	2.841 (0.421)		−93.013	−89.013	94.252	<0.001	−84.432	0.338 (1.169×10^{-7})

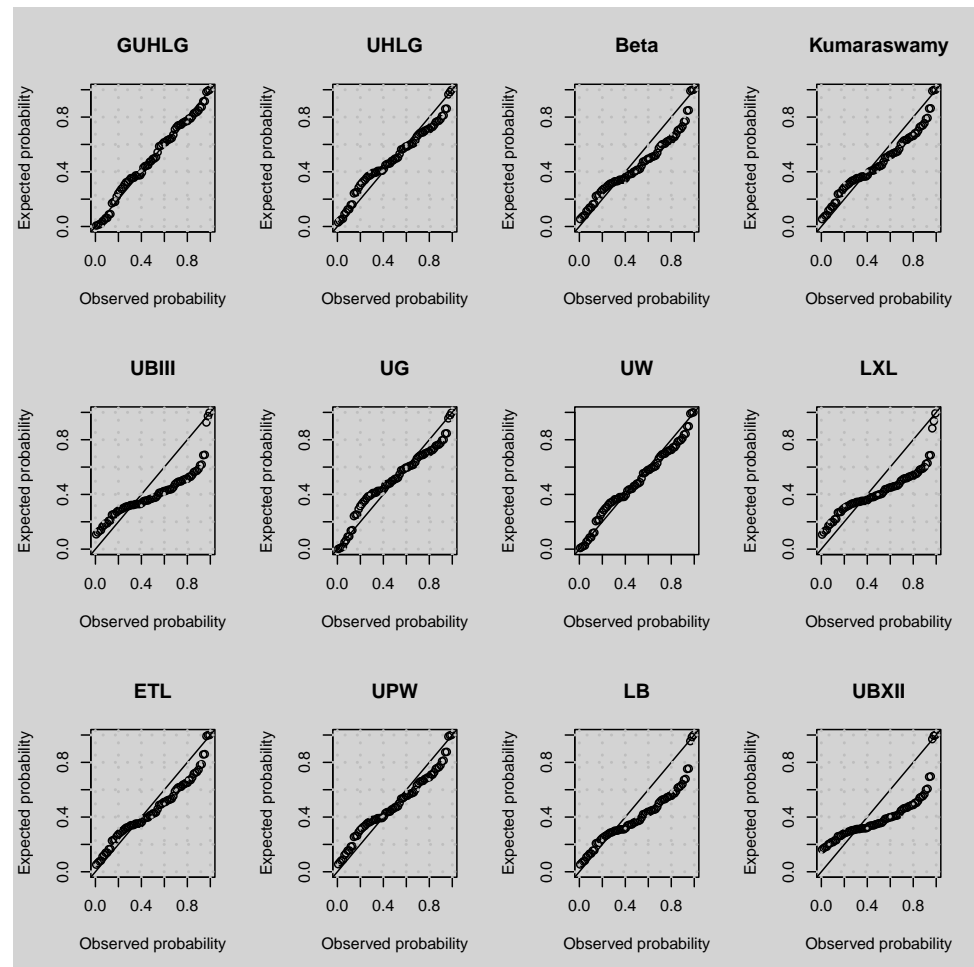


Figure 6. P-P plots of fitted distributions.

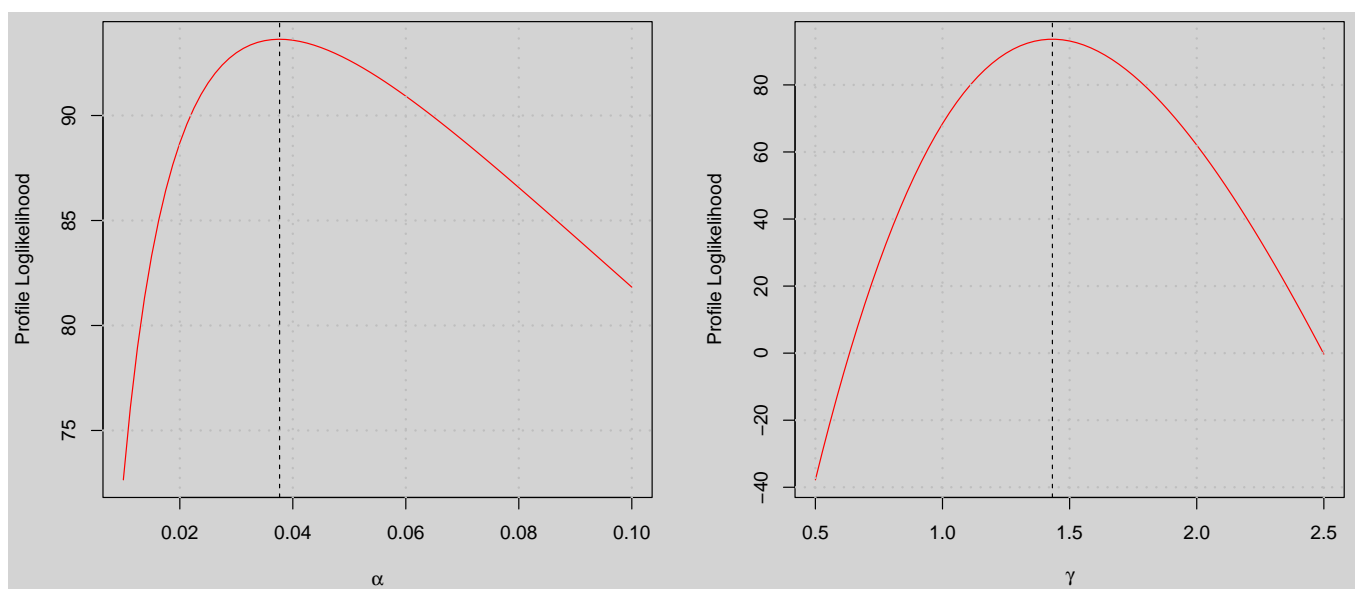


Figure 7. Profile log-likelihood plots.

6.2. Multivariate Application

In this subsection, the multivariate application of the GUHLG quantile regression is illustrated. We demonstrate both the frequentist and Bayesian approaches to fitting the regression model to the given data.

6.2.1. Frequentist Approach

The ML estimation approach is used here to study the effects of the exogenous variables on the conditional median of the endogenous variable. The exogenous variables are:

- ASSUME: Ratio of per occurrence retention levels to total assets.
- CAP: The firm's use of captive (1 if yes and 0 if no).
- SIZELOG: Logarithm of the firm's total asset value.
- INDCOST: Industry average of premiums plus uninsured losses divided by total assets (a measure of the firm's industry risk).
- CENTRAL: Importance of local managers in choosing local retention levels.
- SOPH: Importance of analytical tools in making risk management decisions.

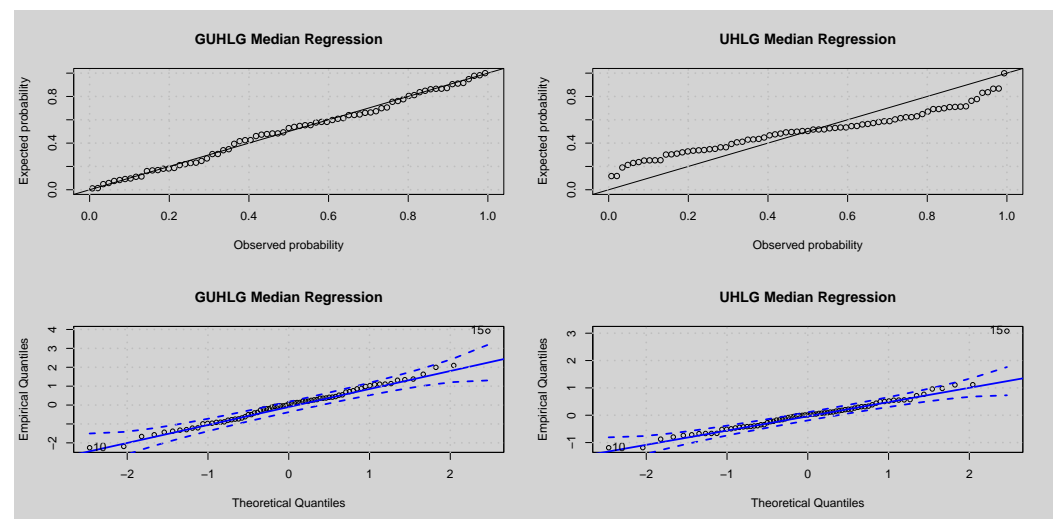
The effects of these exogenous variables on the response variable have been studied by a number of researchers. Recent studies on these variables include: [1,2,32]. More precisely, the authors of [1] fitted the UHLG median regression, the authors of [2] fitted the unit exponentiated Fréchet (UEF) median regression, and the authors of [32] used the unit Weibull (UW) median regression to investigate the relationship between the variables. The authors of [2] recently showed that UEF median regression (AIC = −222.2699, BIC = −201.6400) provided a better fit to the data than the UW median regression (AIC = −206.2200, BIC = −187.9000), Kumaraswamy median regression (AIC = −181.6500, BIC = −163.3300) and beta mean regression (AIC = −159.4500, BIC = −141.3610). The authors of [1] also revealed that the UHLG median regression (AIC = −192.3414, BIC = −176.3081) performs better than the LB median regression (AIC = −151.4600, BIC = −135.4200), Kumaraswamy median regression, and the beta mean regression. Here, we examine the relationship using the following regression structure:

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_0 + \eta_1 \text{ASSUME} + \eta_2 \text{CAP} + \eta_3 \text{SIZELOG} + \eta_4 \text{INDCOST} + \eta_5 \text{CENTRAL} + \eta_6 \text{SOPH}.$$

The exploratory analysis of the response variable shown in Figure 5 suggests that regression models capable of handling extremely skewed data should be used to study the relationship. Hence, our justification for using such a proposed model. Table 6 presents the estimates of the parameters and information criteria for the GUHLG median and UHLG median regressions. The GUHLG median regression outperforms the models studied in [1,2,32]. The GUHLG median regression provided a very good fit to the data. The GUHLG median regression is therefore the best model and provides a significantly better fit to the data compared to the UHLG median regression. We assess the adequacy of the fitted regression models using the P-P (top row) and quantile–quantile (Q-Q) (bottom row) plots of the RQR. The P-P and Q-Q plots in Figure 8 give an indication that the GUHLG median regression provides adequate fit to the data. Although the Q-Q plot shows some outliers, since a larger portion of the residuals are within the simulated envelopes, the model is adequate. Using the best model (GUHLG median), we observe that the variables that significantly influence the firm's cost are the SIZELOG and INDCOST.

Table 6. Estimates of the regression parameters and information criteria.

Parameter	GUHLG Median			UHLG Median		
	Estimates	Standard Error	<i>p</i> Value	Estimates	Standard Error	<i>p</i> -Value
η_0	3.985	1.211	0.001	4.128	2.067	0.046
η_1	−0.012	0.012	0.310	−0.012	0.022	0.580
η_2	−0.053	0.223	0.814	0.018	0.404	0.965
η_3	−0.909	0.125	<0.001	−0.918	0.208	<0.001
η_4	2.343	0.623	<0.001	2.145	0.909	0.018
η_5	−0.137	0.084	0.103	−0.092	0.151	0.544
η_6	0.009	0.020	0.635	0.005	0.036	0.895
γ	2.203	0.227	<0.001			
$-2\ell = -244.962$				$-2\ell = -206.341$		
AIC = −228.962				AIC = −192.341		
BIC = −210.639				BIC = −176.308		

**Figure 8.** P-P and Q-Q plots of the RQR.

6.2.2. Bayesian Approach

In this subsection, we illustrate how to fit the GUHLG median regression using the Bayesian method. To proceed with the Bayesian analysis, we first need to establish the prior distributions of the parameters of the regression model. The prior distribution used for the parameter γ is the non-informative gamma distribution, while that of η is the informative normal distribution. Hence, the prior distributions are:

$$\mathbb{P}(\gamma) \sim \text{Gamma}(a_1, b_1) = \frac{b_1^{a_1} \gamma^{a_1-1}}{\Gamma(a_1)} \exp(-b_1 \gamma), \quad a_1 > 0, b_1 > 0, \gamma > 0$$

and

$$\mathbb{P}(\eta^T) \sim N(a_2, b_2) = \frac{1}{\sqrt{2\pi b_2}} \exp\left(-\frac{(\eta_j - a_2)^2}{2b_2}\right), \quad \eta_j \in \mathbb{R}, a_2 \in \mathbb{R}, b_2 > 0, j = 1, 2, \dots, 6.$$

For more information on the impact of prior distributions on the Bayesian estimates, see [39–41]. The joint PDF of the prior distributions is

$$\mathbb{P}(\gamma, \eta^T) = \mathbb{P}(\gamma) \mathbb{P}(\eta^T).$$

Thus, the joint posterior distribution of the parameters is

$$\mathbb{P}(\gamma, \eta^T | y) \propto \prod_{i=1}^n f_Y(y_i; p, \mu_i, \gamma) \times \mathbb{P}(\gamma, \eta^T),$$

where $\prod_{i=1}^n f_Y(y_i; p, \mu_i, \gamma)$ is the likelihood function of the quantile GUHLG distribution. Since, the joint posterior distribution is not tractable, we employ the Markov chain Monte Carlo (MCMC) algorithm to draw posterior samples from which the marginal distributions are inferred. For the parameter γ , we use the hyperparameter values $a_1 = b_1 = 0.001$ and, for η , we use the hyperparameter values $a_2 = 0$ and $b_2 = 0.001$. The analysis is carried out using three independent chains, each with 600,000 values and a burn-in of 150,000. The thinning interval used is 50, and the sample size per chain is 9000. The R2jags package (see [42]) is used to perform the analysis. The potential reduction scale factor (\hat{R}), the effective sample size (neff), trace plots, ergodic mean plots, and autocorrelation plots are used to examine the chains' convergence to stationary distribution. The Bayesian estimates of the parameters, as well as their standard deviation (SD), naive standard error (SE), \hat{R} , and neff, are shown in Table 7. We observe that the Bayesian estimates are quite close to the ML estimates. The estimated deviance information criterion (DIC) is -227.0000 , which is very close to the estimated AIC value using the ML method. The neff is greater than 400, and \hat{R} is approximately 1. This implies that the MCMC algorithm has converged to a stationary distribution.

Table 7. Estimates of the regression parameters and information criteria.

Parameter	Estimate	SD	Naïve SE	2.5%	25%	50%	75%	97.7%	\hat{R}	neff
η_0	3.968	1.292	0.008	1.428	3.110	3.966	4.835	6.4835	1.001	15,000
η_1	−0.011	0.016	0.000	−0.040	−0.021	−0.011	−0.001	0.022	1.001	27,000
η_2	−0.046	0.240	0.002	−0.509	−0.208	−0.049	0.116	0.430	1.001	27,000
η_3	−0.908	0.132	0.001	−1.168	−0.997	−0.908	−0.819	−0.651	1.001	27,000
η_4	2.373	0.657	0.004	1.111	1.927	2.368	2.806	3.690	1.001	27,000
η_5	−0.130	0.090	0.001	−0.305	−0.190	−0.131	−0.069	0.050	1.001	7000
η_6	0.009	0.021	0.000	−0.033	−0.006	0.009	0.023	0.050	1.001	14,000
γ	2.078	0.225	0.001	1.654	1.924	2.071	2.227	2.537	1.001	27,000

The convergence of the MCMC algorithm is further explored using trace plots. Figure 9 reveals that this algorithm converges with no periodicity.

The ergodic mean plots shown in Figure 10 affirm the convergence of the MCMC algorithm. From these plots, the ergodic mean stabilizes as the chain progresses.

The autocorrelation plots displayed in Figure 11 show a fast decay, which gives an indication that the chains are well mixed and converge to a stationary distribution.

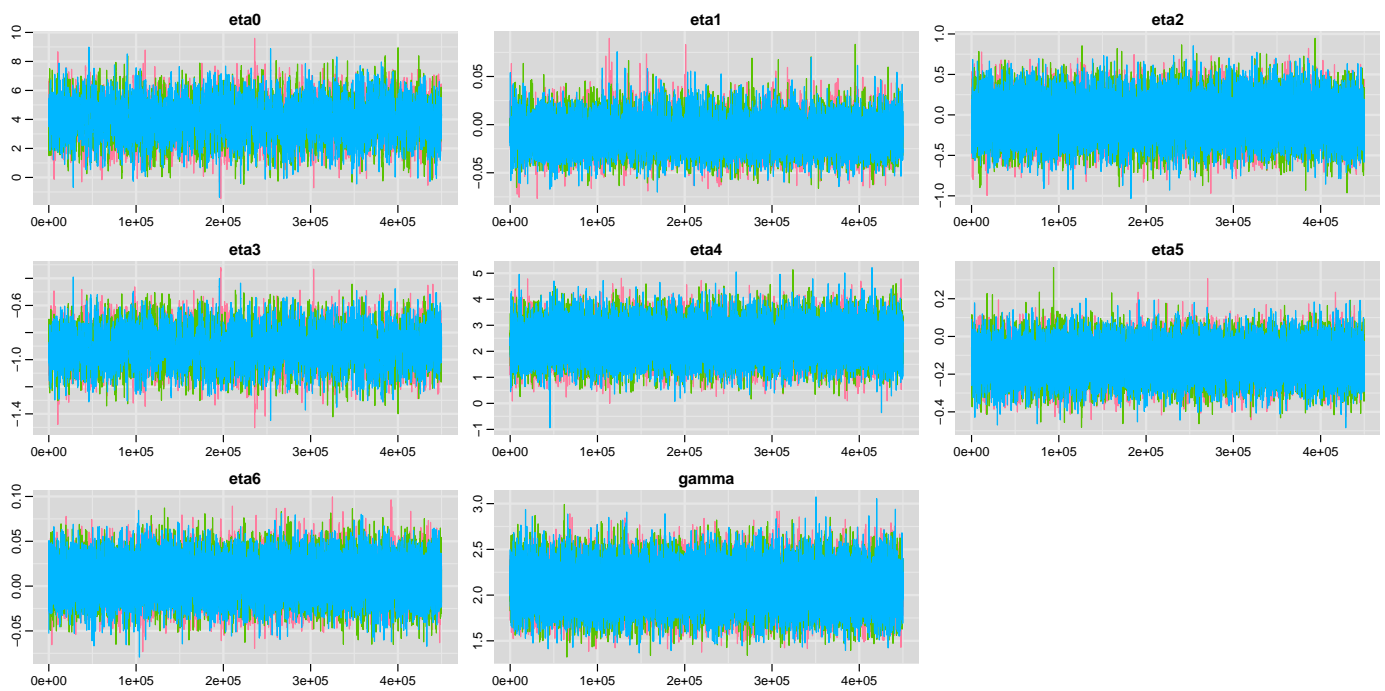


Figure 9. Trace plots of the parameters.

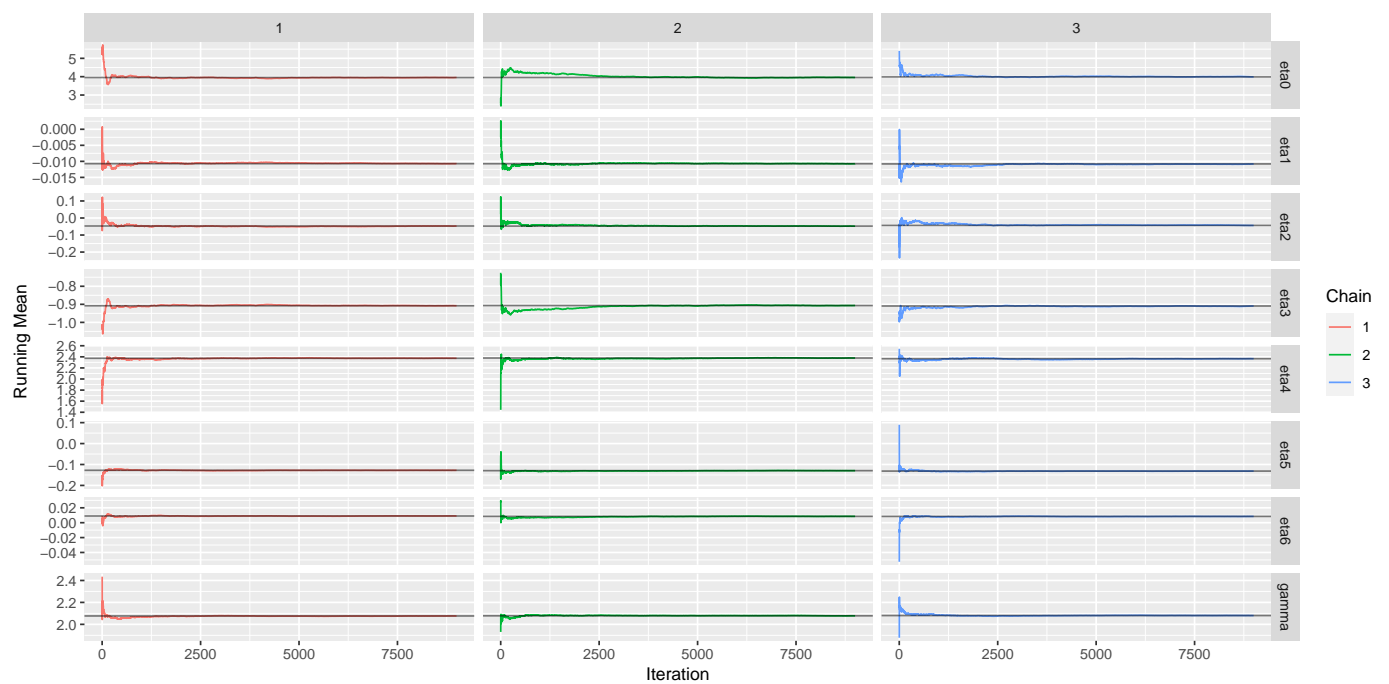


Figure 10. Ergodic mean plots.

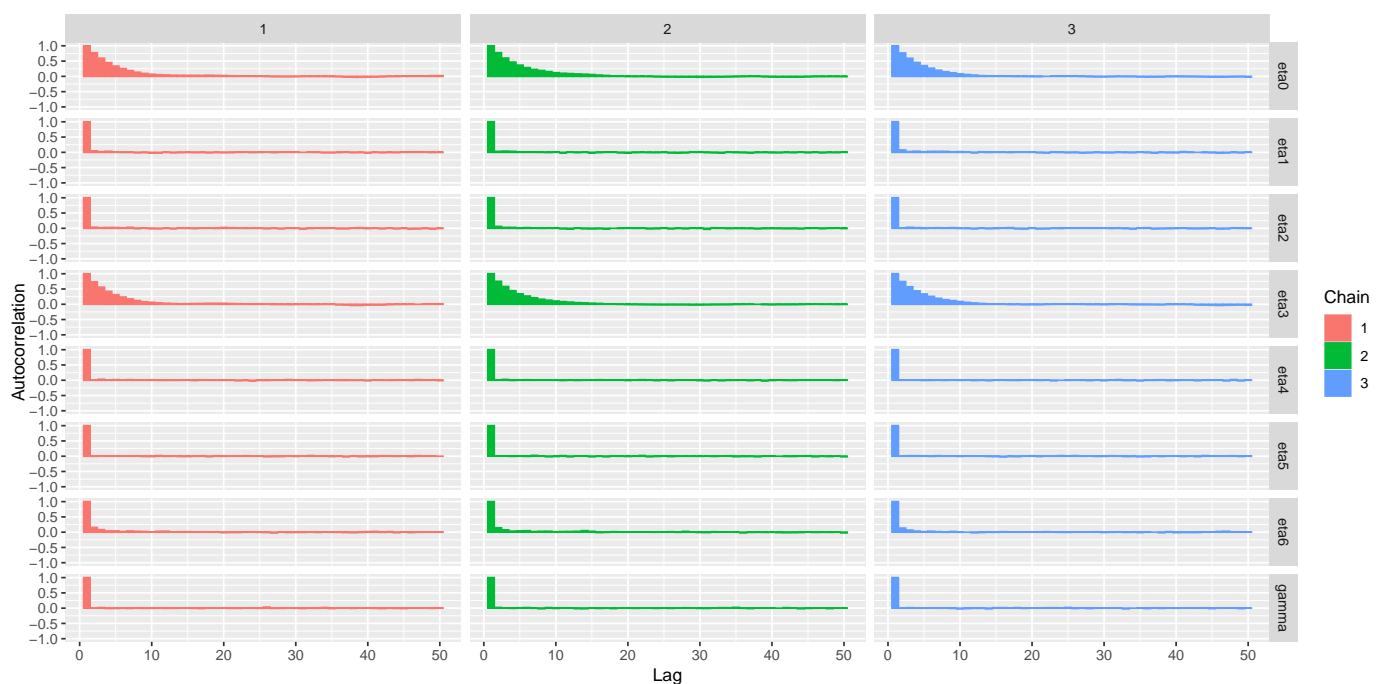


Figure 11. Autocorrelation plots.

7. Conclusions

The GUHLG distribution was studied and then employed to formulate quantile regressions. Some statistical properties, such as distributional inequalities, quantile measures, moments, and order statistics, were derived. The PDF plots suggest that the distribution is capable of modelling data that may have the following characteristics: left-skewed, right-skewed, symmetric, increasing, or bathtub-shaped PDFs. The HRF plots showed that the distribution is capable of modelling data with bathtub, increasing, or N-shaped failure rates. The univariate application of the model using risk survey data revealed that it can provide a better parametric fit than other existing bounded distributions. This is because it has the lowest information criterion and goodness-of-fit statistics. Hence, it offers minimal loss of information compared to the other distributions. The multivariate application using the developed quantile regression model showed that the new regression model provides a better fit to the risk survey data than other regression models that have already been used to model the data. Finally, we illustrated the multivariate application using frequentist and Bayesian methods. The estimates of the parameters from the two methods were quite close. Diagnostic checks of the Bayesian method showed that the MCMC algorithm converges to a stationary distribution. Our future extension of this research is to develop an R package for the univariate and multivariate models.

Author Contributions: Conceptualization, S.N., C.C., A.G.A. and I.D.A.; methodology, S.N., C.C., A.G.A. and I.D.A.; software, S.N., C.C., A.G.A. and I.D.A.; validation, S.N., C.C., A.G.A. and I.D.A.; formal analysis, S.N., C.C., A.G.A. and I.D.A.; investigation, S.N., C.C., A.G.A. and I.D.A.; data curation, S.N., C.C., A.G.A. and I.D.A.; writing—original draft preparation, S.N., C.C., A.G.A. and I.D.A.; writing—review and editing, S.N., C.C., A.G.A. and I.D.A.; visualization, S.N., C.C., A.G.A. and I.D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the five reviewers for their in-depth and helpful comments, which considerably improved the paper.

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

References

1. Ramadan, A.T.; Tolba, A.H.; El-Desouky, B.S. A unit half-logistic geometric distribution and its application in insurance. *Axioms* **2022**, *11*, 676. [\[CrossRef\]](#)
2. Abubakari, A.G.; Luguterah, A.; Nasiru, S. Unit exponentiated Fréchet distribution: actuarial measures, quantile regression and applications. *J. Indian Soc. Probab. Stat.* **2022**, *23*, 387–424. [\[CrossRef\]](#)
3. Alanzi, A.R.A.; Rafique, M.Q.; Tahir, M.H.; Sami, W.; Jamal, F. New modified Kumaraswamy distribution: actuarial measures and applications. *J. Math.* **2022**, 1–18. [\[CrossRef\]](#)
4. Ahmad, Z.; Mahmoudi, E.; Alizadeh, M. Modelling insurance losses using a new beta power transformed family of distributions. *Commun.-Stat.-Simul. Comput.* **2022**, *51*, 4470–4491. [\[CrossRef\]](#)
5. Mazucheli, J.; Menezes, A.F.; Ghitany, M.E. The unit Weibull distribution and associated inference. *J. Appl. Probab. Stat.* **2018**, *13*, 1–22.
6. Ahmad, Z.; Mahmoudi, E.; Dey, S.; Khosa, S.K. Modeling vehicle insurance loss data using a new member of the T-X family of distributions. *J. Stat. Theory Appl.* **2020**, *19*, 133–147. [\[CrossRef\]](#)
7. Jodrá, P.; Jiménez-Gamero, M.D. A quantile regression for bounded responses based on exponential geometric distribution. *Revstat* **2020**, *18*, 415–436.
8. Ahmad, Z.; Mahmoudi, E.; Hamedani, G.G. A family of loss distributions with an application to the vehicle insurance loss data. *Pak. J. Stat. Oper. Res.* **2019**, *15*, 731–744. [\[CrossRef\]](#)
9. Mazucheli, J.; Menezes, A.F.; Dey, S. Unit-Gompertz distribution with applications. *Statistica* **2019**, *79*, 25–43.
10. Gómez-Déniz, E.; Sordo, M.A.; Calderín-Ojeda, E. The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insur. Math. Econ.* **2014**, *54*, 49–57. [\[CrossRef\]](#)
11. Al-Mofleh, H.; Afify, A.Z.; Ibrahim, N.A. A new extended two-parameter distribution: properties, estimation methods and, applications in medicine and geology. *Mathematics* **2020**, *8*, 1578. [\[CrossRef\]](#)
12. Jodrá, P.; Gómez, H.W.; Jiménez-Gamero, M.D.; Alba-Fernández, M.V. The power Muth distribution. *Math. Model. Anal.* **2017**, *22*, 186–201. [\[CrossRef\]](#)
13. Iqbal, Z.; Tahir, M.M.; Riaz, N.; Ali, S.A.; Ahmad, M. Generalized inverted Kumaraswamy distribution: properties and application. *Open J. Stat.* **2017**, *7*, 645–662. [\[CrossRef\]](#)
14. Iqbal, Z.; Hasnain, S.A.; Salman, M.; Ahmad, M.; Hamedani, G.G. Generalized exponentiated moment exponential distribution. *Pak. J. Stat.* **2014**, *30*, 537–554.
15. Cordeiro, G.M.; Brito, R.S. The beta power distribution. *Braz. J. Probab. Stat.* **2012**, *26*, 88–112.
16. Jose, K.K.; Joseph, A.; Ristic, M.M. A Marshall-Olkin beta distribution and its applications. *J. Probab. Stat. Sci.* **2009**, *7*, 173–186.
17. Shaked, M.; Shanthikumar, J.G. *Stochastic Orders*; Wiley: New York, NY, USA, 2007.
18. Mazzocchi, A.; Naldi, M. The expected utility insurance premium principle with fourth-order statistics: Does it make a difference? *Algorithms* **2020**, *13*, 116. [\[CrossRef\]](#)
19. Dimitrova, D.S.; Ignatov, Z.G.; Kaishev, V.K. Ruin and deficit under claim arrivals with order statistics property. *Methodol. Comput. Appl. Probab.* **2019**, *21*, 511–530. [\[CrossRef\]](#)
20. Wilkson, M. Estimating maximum loss with order statistics. In *Casualty Actuarial Society*; CAS: Arlington, VA, USA, 1982; pp. 195–209.
21. Ramachandran, G. Properties of extreme order statistics and their application to fire protection and insurance problems. *Fire Saf. J.* **1982**, *5*, 59–76. [\[CrossRef\]](#)
22. Nasiru, S.; Abubakari, A.G.; Chesneau, C. New lifetime distribution for modeling data on the unit interval: properties, application and quantile regression. *Math. Comput. Appl.* **2022**, *27*, 105. [\[CrossRef\]](#)
23. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
24. Mazucheli, J.; Alves, B.; Korkmaz, M.Ç.; Leiva, V. Vasicek quantile and mean regression models for bounded data: new formulations, mathematical derivations and numerical applications. *Mathematics* **2022**, *10*, 1389. [\[CrossRef\]](#)
25. Mazucheli, J.; Korkmaz, M.C.; Menezes, A.F.B.; Leiva, V. The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics and numerical applications. *Soft Comput.* **2022**, *27*, 279–295. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Mustapha, M.H.B.; Nasiru, S. Unit gamma/Gompertz quantile regression with applications to skewed data. *Sri Lankan J. Appl. Stat.* **2022**, *23*, 49–73. [\[CrossRef\]](#)
27. Korkmaz, M.Ç.; Emrah, A.; Chesneau, C.; Yousof, H.M. On the unit-Chen distribution with associated quantile regression and applications. *Math. Slovaca* **2021**, *72*, 765–786. [\[CrossRef\]](#)
28. Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications. *Symmetry* **2021**, *13*, 56. [\[CrossRef\]](#)
29. Lindsay, B.G.; Li, B. On second-order optimality of the observed Fisher information. *Ann. Stat.* **1997**, *25*, 2172–2199. [\[CrossRef\]](#)
30. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.

31. Schmit, J.T.; Roth, K. Cost effectiveness of risk management practices. *J. Risk Insur.* **1990**, *57*, 455–470. [[CrossRef](#)]
32. 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 modelling of quantiles conditional on covariates. *J. Appl. Stat.* **2020**, *47*, 954–974. [[CrossRef](#)]
33. Bantan, R.A.R.; Shafiq, S.; Tahir, M.H.; Elhassanein, A.; Jamal, F.; Almutiry, W.; Elgarhy, M. Statistical analysis of COVID-19 data: Using a new univariate and bivariate statistical model. *J. Funct. Spaces* **2022**, *2022*, 2851352. [[CrossRef](#)]
34. Eliwa, M.S.; Ahsan-ul-Haq, M.; Al-Bossly, A.; El-Morshedy, M. Properties and estimation techniques with application to model data from SC16 and P3 algorithms. *Math. Probl. Eng.* **2022**, *2022*, 9289721. [[CrossRef](#)]
35. Altun, E.; El-Morshedy, M.; Eliwa, M.S. A new regression model for bounded response variable: an alternative to the beta and unit-Lindley regression models. *PLoS ONE* **2021**, *16*, e0245627. [[CrossRef](#)] [[PubMed](#)]
36. [[CrossRef](#)] Korkmaz, M.Ç.; Chesneau, C. On the unit Burr XII distribution with the quantile regression modeling and applications. *Comput. Appl. Math.* **2021**, *40*, 29. [[CrossRef](#)]
37. Modi, K.; Gill, V. Unit Burr-III distribution with application. *J. Stat. Manag. Syst.* **2019**, *23*, 579–592. [[CrossRef](#)]
38. Pourdarvish, A.; Mirmostafae, S.M.T.K.; Naderi, K. The exponentiated Topp-Leone distribution: properties and application. *J. Appl. Environ. Biol. Sci.* **2015**, *5*, 251–256.
39. Muse, A.H.; Chesneau, C.; Ngesa, O.; Mwalili, S. Flexible parametric accelerated hazard model: simulation and application to censored lifetime data with crossing survival curves. *Math. Comput. Appl.* **2022**, *27*, 104. [[CrossRef](#)]
40. Khan, S.A. Exponentiated Weibull regression for time-to-event data. *Lifetime Data Anal.* **2018**, *24*, 328–354. [[CrossRef](#)]
41. Ali, S. On the Bayesian estimation of the weighted Lindley distribution. *J. Stat. Comput. Simul.* **2015**, *85*, 855–880. [[CrossRef](#)]
42. Su, Y.S.; Yajima, M. R2jags: A Package for Running Jags from R. 2012. Available online: <https://CRAN.Rproject.org/package=R2jags> (accessed on 3 January 2023).

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.