STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, August 2025, pp 956–969.
Published online in International Academic Press (www.IAPress.org)

Quasi Lindley Regression Model Residual Analysis for Biomedical Data

Ahmed Mahdi Salih*, Wafaa Jaafar Hussein

Department of Statistics, Wasit University, Iraq

Abstract The current study proposes and presents a new regression model for the response variable following the Quasi Lindley Regression. The unknown parameters of the regression model are estimated using the maximum likelihood method. A simulation study is conducted to evaluate the performance of the maximum likelihood estimates (MLEs). In addition, a residual analysis is performed for the proposed regression model. The log- Quasi Lindley Regression model is compared to several other models, including Lindley regression and gamma regression, using various statistical criteria. The results show that the suggested model fits the data better than these other models. The model is expected to have applications in fields such as economics, biological studies, mortality and recovery rates, health, hazards, measuring sciences, medicine, and engineering.

Keywords Quasi Lindley distribution; Quasi Lindley Regression Model; Residual analysis; Martingale residual

AMS 2010 subject classifications 62Jxx

DOI: 10.19139/soic-2310-5070-2649

1. Introduction

Several distributions have been used to model data in various fields, including economics, biological studies, mortality, recovery rates, health, risks, measurement sciences, medicine, engineering, insurance, and finance. In recent years, studies have attempted to provide modeling of data based on their distributions. For example, [3] suggested the unit-improved second-degree Lindley distribution for inference and regression modeling. [15] proposed the log-generalized modified Weibull regression model. [12] introduced a new quantile regression for modeling bounded data using the Birnbaum-Saunders distribution.[19] introduced Log-Burr XII regression models. [16]introduced the Log Beta Generalized Weibull Regression Model for lifetime data [13] suggested the quantile regression modeling on the unit Burr-XII , [9] suggested the Exponentiated Weibull regression, [11] suggested the Log-generalized inverse Weibull Regression Model, [8] introduced the Transmuted Weibull Regression Model , [14] Proposed On the Extension of the Burr XII Distribution: Applications and Regression ,[1] suggested the Gumbel-Burr XII Regression Model , [4] Proposed Zografos–. Balakrishnan Burr XII Regression model ,[10] suggested Unit-Chen quantile regression model, [12] suggested The unit generalized half-normal quantile regression model, [19]. Log-Burr XII regression models with censored data Quasi Lindley distribution can be characterized as over-dispersed when $\mu > \sigma^2$, equi-dispersed when $\mu = \sigma^2$, and under-dispersed when $\mu < \sigma^2$. Its hazard rate function increases with x and when β is involved. It is considered superior to Lindley and exponential distributions for modeling lifetime data in medical science and engineering. Shanker introduced a quasi Quasi Lindley distribution to model lifetime data, discussing its statistical properties and potential applications. Additionally, a comparative study of one-parameter Akash, Lindley, and exponential distributions showed that the Akash distribution sometimes provided a better fit for certain datasets. This article is organized as follows. Section 2

*Correspondence to: Ahmed Mahdi Salih (Email: amahdi@uowasit.edu.iq). Department of Statistics, Wasit University. Alhawi / Alkut, Wasit, Iraq (50001).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

introduces the definition of Akash distribution, while Section 3 suggests a log Akash regression model of location-scale. Section 4 employs the maximum likelihood method to estimate the parameters, and Section 5 presents different types of residual analysis. Section 6 Simulation Study, Section 7 Real data, Section 8 Conclusions.

2. Definition of the Lindley and Quasi Lindley Distribution

In general, the Lindley distribution is an elastic probability distribution that was initially proposed in 1957 as a response to the fiducial statistical approach. It does this by combining the exponential distribution with the gamma distribution with parameters of two and θ, which enables it to be versatile enough to represent the range of processes or equipment. Additionally, this statistical model is utilized in a variety of practical applications across a variety of fields, including biology, engineering, and medical science research. The following is a definition of the PDF:

$$f(x,\gamma)=\frac{\gamma^2}{(\gamma+1)}(1+x)e^{\gamma x}\quad x>0,\gamma\geq 0 \tag{1}$$

The cumulative distribution function for the one-parameter function is¿

$$(F(x,\gamma)=1-\frac{(1+\gamma x+\gamma)}{(1+\gamma)}e^{-\gamma x} \tag{2}$$

The survival function S(x) defined as follows

$$S(x,\gamma)=\frac{(1+\gamma x+\gamma)}{(1+\gamma)}e^{-\gamma x} \tag{3}$$

[17] proposed a new distribution that is referred to as the Two-Parameter Quasi Lindley Distribution (New QLD), with the Lindley Distribution (LD) being an example of this distribution in particular. In their investigation, they looked into a number of characteristics of the New QLD, including as its moments, failure rate function, mean residual life function, and stochastic orderings. According to the findings of their investigation, the formulations for the failure rate function, the mean residual life function, and the stochastic orderings of the QLD highlight the versatility of the QLD in comparison to both the Lindley distribution and the exponential distribution. Furthermore, [18] introduced another variant of the Quasi Lindley Distribution (QLD), characterized by parameters γ and α, delineated through its probability density function (PDF).

$$f(x,\gamma,\alpha)=\frac{\gamma(\alpha+\gamma x)}{1+\alpha}e^{-\gamma x}\quad x\geq 0\quad ,\gamma,\alpha\geq 0 \tag{4}$$

The Quasi Lindley distribution is a flexible distribution that can take on a variety of shapes, including a uni-modal shape, a J-shaped shape, and a U-shaped shape, depending on the values of the shape and scale parameters. It has a long right tail, which makes it suitable for modeling data with heavy tails. One of the advantages of the Quasi Lindley distribution is that it has a simple closed-form expression for its cumulative distribution function (CDF), which makes it easy to compute probabilities and quantiles. It also has a simple form for its moments, which can be used to estimate the parameters of the distribution using methods such as the method of moments or the maximum likelihood method.

$$F(x,\gamma,\alpha)=1-\frac{(1+\alpha+\gamma x)}{(1+\alpha)}e^{-\gamma x}\quad x\geq 0\quad ,\gamma,\alpha\geq 0 \tag{5}$$

3. Mathematical Properties of Quasi Lindley Distribution

3.1. The Survival Function

also known as the survivor function, is a function that gives the probability that a system or a component will survive beyond a certain time. It is defined as the complement of the cumulative distribution function (CDF) of a

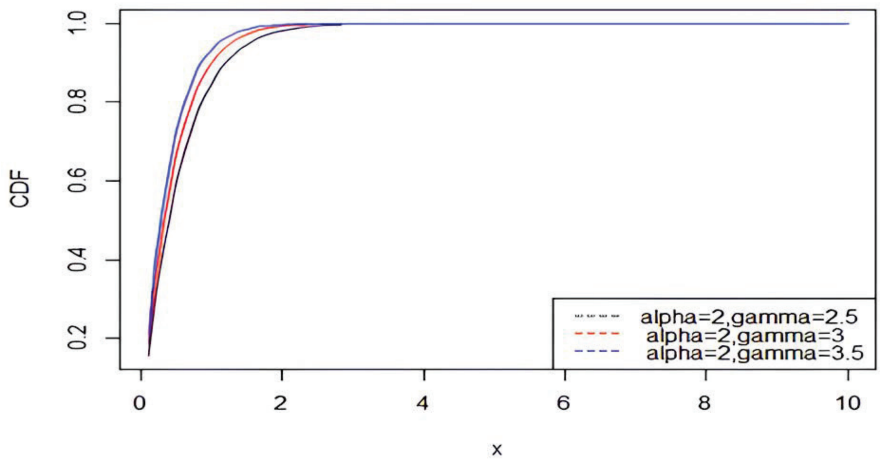

Figure 1. The Cumulative Probability Density Function of Quasi Lindley Distribution

random variable representing the time to failure. Mathematically, the survival function, denoted by S(t), is defined as

$$S(t) = 1 - P(T > t) \tag{6}$$

Where T is the random variable representing the time to failure, $F(t)$ is the cumulative distribution function of T, and $P(T > t)$ is the probability that the time to failure is greater than t.
The survival function of quasi-Lindley is defined as follows [**?**]:

$$S(x) = \frac{1+\alpha+\gamma x}{1+\alpha} e^{-\gamma x} \quad x \geq 0 \quad , \gamma, \alpha \geq 0 \tag{7}$$

3.2. The Hazard Rate Function

also known as the failure rate function, is a function that gives the instantaneous rate of failure at a given time, given that the system or component has survived up to that time. It is defined as the ratio of the probability density function (PDF) to the survival function of a random variable that represents the time to failure. Mathematically, the hazard rate function, denoted by $h(t)$, is defined as:

$$h(t) = \frac{f(t)}{s(t)} \tag{8}$$

Where $f(t)$ is the probability density function of T, $S(t)$ is the survival function of T, and T is the random variable representing the time to failure. The hazard rate function is a non-negative function that can be interpreted as the rate at which the system or component is failing at a given time, given that it has not failed up to that time. It is often used in reliability analysis to model the failure behavior of a system or a component the hazard rate function of Quasi Lindley is defined as follows:

$$h(x) = \frac{\gamma(\alpha+\gamma x)}{(1+\alpha+\gamma x)} \quad x \geq 0 \quad , \gamma, \alpha \geq 0 \tag{9}$$

While the mean and variance as follows:

$$E(x) = \frac{\alpha+2}{\gamma(1+\alpha)} \tag{10}$$

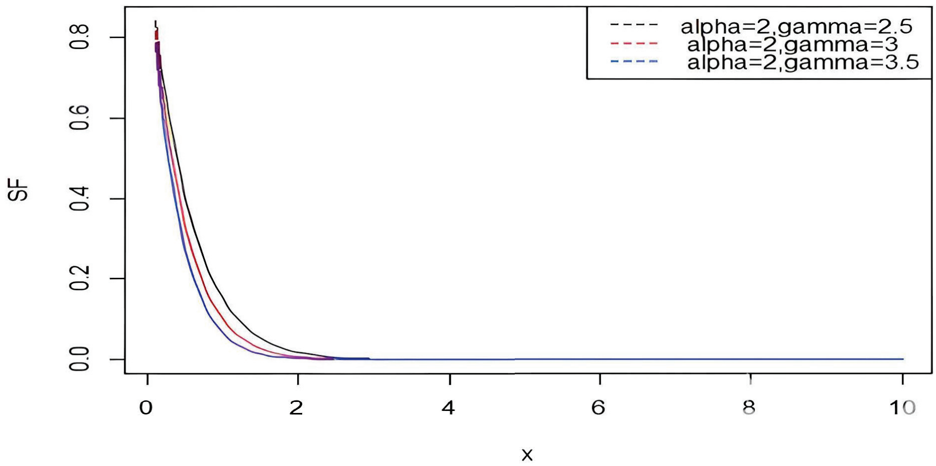

Figure 2. The Survival Function of Quasi Lindley distribution

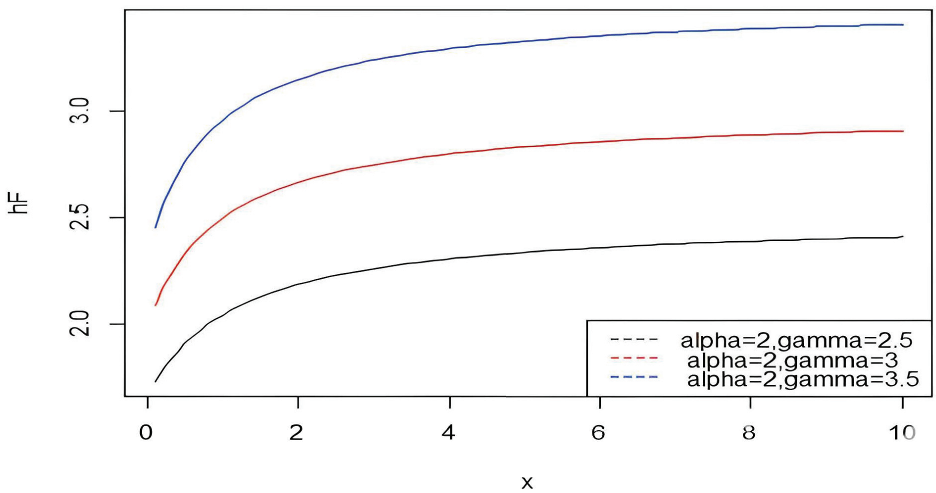

Figure 3. The Hazard Rate Function of Quasi Lindley Distribution

$$V = \frac{(1+\alpha)(2\alpha+6) - (\alpha+2)^2}{\gamma(1+\alpha)^2} \tag{11}$$

3.3. The Construction Regression Model for Quasi Lindley distribution

Let X be a random variable having the Quasi Lindley as in Equation (4) and let $y = \log x$ which leads to $x = exp(y)$, and let $\gamma = \exp(-\mu)$ we can define the density function of Y to be written as [20]:

$$f(y, \mu, \sigma) = f^{-1}(x) \cdot |J| \tag{12}$$

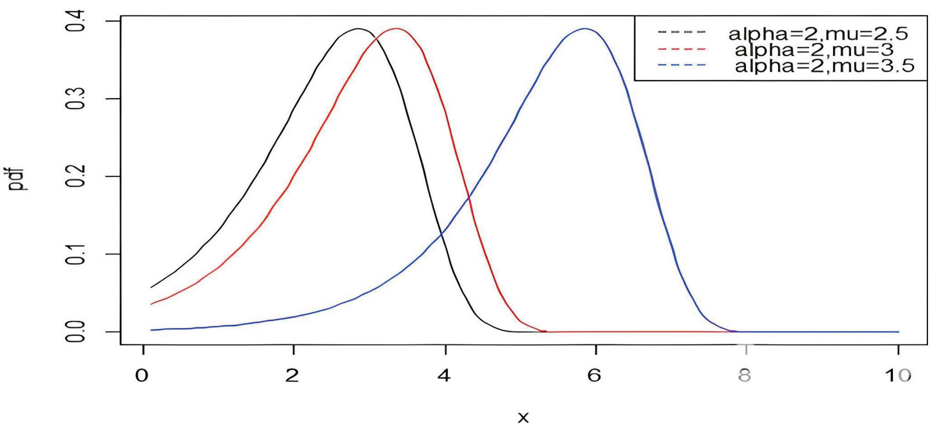

Figure 4. The Probability Density Function of Quasi Lindley distribution

here we can easily calculate that the Jacobian to be $dx/dy = x = exp(y)$ we can rewrite the equation above to be as follows:

$$f(y, \mu, \alpha) = \frac{\exp(y-\mu) * (\alpha + \exp(y-\mu))}{1+\alpha} e^{-(\exp(y-\mu))} \tag{13}$$

Where $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\propto > 0$, y has the log Quasi Lindley, where μ is a location parameter, Log Quasi Lindley distribution (LQL), Further, we defined the standard random variable $Z = y - \mu$ with probability density function.

$$f(z) = \frac{\exp(z) \cdot (\alpha + \exp(z))}{1+\alpha} e^{-\exp(z)} \tag{14}$$

Where $-\infty < z < \infty, \alpha > 0$.

$$S(z) = \frac{(1 + \alpha + \exp(z))}{1+\alpha} e^{-\exp(z)} \tag{15}$$

3.4. Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a statistical method for estimating the parameters of a probability distribution based on observed data. The goal of MLE is to find the values of the parameters that make the observed data most likely.

$$L = \prod_{i=1}^{n} \left[f\left(y, \mu, \sigma\right)\right]^{\sigma_i} * \left[S(y))\right]^{1-\sigma_i}. \tag{16}$$

$$L = \delta_i \sum_{i=1}^{n} (y - \beta_0 - \beta_1 x) + \delta_i \sum_{i=1}^{n} \log[(\alpha + \exp(y - \beta_0 - \beta_1 x))]$$

$$-n\delta_i \log(1+\alpha) - \delta_i \sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x)$$

$$+(1-\delta_i)\left[\sum_{i=1}^{n} \log[1 + \alpha + \exp(y - \beta_0 - \beta_1 x)] - \sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x) - n\log(1+\alpha)\right]$$

To follow this derivation in an easy way, we can express the above equation in the following form:

$$L = K_5 - K_6 + K_7 - K_8 \tag{17}$$

Where

$$K_5 = \delta_i \sum_{i=1}^{n} (y - \beta_0 - \beta_1 x) + \delta_i \sum_{i=1}^{n} \log[(\alpha + \exp(y - \beta_0 - \beta_1 x))]$$

$$K_6 = -n\delta_i \log(1 + \alpha) - \delta_i \sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x)$$

$$K_7 = (1 - \delta_i) \left[\sum_{i=1}^{n} \log[1 + \alpha + \exp(y - \beta_0 - \beta_1 x)] \right]$$

$$K_8 = -(1 - \delta_i) \left[\sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x) - n \log(1 + \alpha) \right]$$

Derivation of the above equation with respect to α as follows:

$$\frac{\partial L}{\partial \propto} = \frac{\partial K_5}{\partial \propto} - \frac{\partial K_6}{\partial \propto} + \frac{\partial K_7}{\partial \propto} - \frac{\partial K_8}{\partial \propto} = 0 \tag{18}$$

$$\frac{\partial K_5}{\partial \alpha} = \sum_{i=1}^{n} \frac{\delta_i}{\alpha + \exp(y - \beta_0 - \beta_1 x)}$$

$$\frac{\partial K_6}{\partial \alpha} = -\frac{n\delta_i}{1 + \alpha}$$

$$\frac{\partial K_7}{\partial \alpha} = \left[\sum_{i=1}^{n} \frac{1 - \delta_i}{1 + \alpha + \exp(y - \beta_0 - \beta_1 x)} \right]$$

$$\frac{\partial K_8}{\partial \alpha} = -\frac{n}{1 + \alpha}$$

Derivation of the above equation with respect to β_0 as follows:

$$\frac{\partial L}{\partial \propto} = \frac{\partial K_5}{\partial \propto} - \frac{\partial K_6}{\partial \propto} + \frac{\partial K_7}{\partial \propto} - \frac{\partial K_8}{\partial \propto} = 0 \tag{19}$$

$$\frac{\partial K_5}{\partial \beta_0} = \sum_{i=1}^{n} (-\delta_i) + \sum_{i=1}^{n} \frac{-\delta_i}{(\alpha + \exp(y - \beta_0 - \beta_1 x))}$$

$$\frac{\partial K_6}{\partial \beta_0} = +\delta_i \sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x)$$

$$\frac{\partial K_7}{\partial \beta_0} = \left[\ \textstyle\sum_{i=1}^{n} \frac{(-1+\delta_i)}{1+\alpha+\exp(y-\beta_0-\beta_1 x)} \ \right]$$

$$\frac{\partial K_8}{\partial \beta_0} = +\sum_{i=1}^{n} \exp(y - \beta_0 - \beta_1 x)$$

Derivation of the above equation with respect to β_1 as follows:

$$\frac{\partial L}{\partial \beta_1} = \frac{\partial K_5}{\partial \beta_1} - \frac{\partial K_6}{\partial \beta_1} + \frac{\partial K_7}{\partial \beta_1} - \frac{\partial K_8}{\partial \beta_1} = 0 \tag{20}$$

$$\frac{\partial K_5}{\partial \beta_1} = \sum_{i=1}^{n} (-\delta_i) + \sum_{i=1}^{n} \frac{x\delta_i}{(\alpha + \exp(y - \beta_0 - \beta_1 x))}$$

$$\frac{\partial K_6}{\partial \beta_1} = +\delta_i \sum_{i=1}^{n} (-x) \exp(y - \beta_0 - \beta_1 x)$$

$$\frac{\partial K_7}{\partial \beta_1} = \left[\sum_{i=1}^{n} \frac{(-1 + \delta_i)\, x}{1 + \alpha + \exp(y - \beta_0 - \beta_1 x)} \right]$$

$$\frac{\partial K_8}{\partial \beta_1} = + \sum_{i=1}^{n} (x) \exp(y - \beta_0 - \beta_1 x)$$

We use the Newton-Raphson numerical method to solve equations (18-20)

4. Residual Analysis

Residual analysis is an essential tool in statistical modeling and analysis. It refers to the examination of the residuals, which are the differences between the observed values and the predicted values from a statistical model. The primary purpose of residual analysis is to assess the goodness-of-fit of the model and identify any patterns or outliers that may indicate a lack of fit or violation of assumptions. By examining the residuals, analysts can evaluate the appropriateness of the model, detect anomalies, and refine the model to improve its accuracy and predictive power. Residual analysis is a crucial step in model building, validation, and diagnosis, and it is widely used in various fields, including finance, engineering, biology, and social sciences. In this introduction, we will discuss the concept of residual analysis, its importance, and the techniques used to perform residual analysis.

4.1. Definition of Residual Analysis

Residual analysis is the process of evaluating the residuals, which are the differences between the observed values and the predicted values from a statistical model. Mathematically, the residual for a given observation is calculated as:

$$\textbf{Residual} = \textbf{observed value} - \textbf{predicted value} \tag{21}$$

The predicted value is obtained from the statistical model, which is a mathematical representation of the relationship between the dependent variable and the independent variables. The residuals provide a measure of how well the model fits the data. If the model is a good fit, the residuals should be small and randomly distributed around zero. If the residuals show any patterns or trends, it suggests that the model may not be capturing the underlying relationship between the variables adequately.

4.2. Martingale Residuals

The martingale residual is a type of residual used in time series analysis and survival analysis. It is defined as the difference between the observed value and the expected value of the response variable at a given time point, given the history of the process up to that time point. In other words, the martingale residual represents the unexpected component of the response variable that cannot be explained by the history of the process up to that time point. , the martingale residual can be expressed as [22] and has been used by researchers such as [21], [5], and [7].

$$r_M = \delta_i + \int_0^y h(u)\, du. \qquad i = 1, 2, 3 \tag{22}$$

and as we know

$$\int_0^y h(u)\, du = \log\left[S(y_i)\right]$$

Where $\delta_i = 1$ or 0, one when observation is censored and 0 when observation is uncensored, the r_M reduce to

$$r_M = \delta_i + \log(S(y)) \tag{23}$$

$$r_M = \begin{cases} \delta_i + \log(S(y)) & \delta_i = 1 \\ \log(S(y)) & \delta_i = 0 \end{cases} \tag{24}$$

$$r_M = \begin{cases} \delta_i + \begin{cases} \sum_{i=1}^{n} \log(1+\alpha+\exp(y-\beta_0-\beta_1 x)) \\ -\sum_{i=1}^{n} \exp(y-\beta_0-\beta_1 x) - n\log(1+\alpha) \end{cases} & : \delta_i = 1 \\ \begin{cases} \sum_{i=1}^{n} \log(1+\alpha+\exp(y-\beta_0-\beta_1 x)) \\ -\sum_{i=1}^{n} \exp(y-\beta_0-\beta_1 x) - n\log(1+\alpha) \end{cases} & : \delta_i = 0 \end{cases} \tag{25}$$

4.3. Deviance Component Residual

The deviance component residual is a type of residual used in generalized linear models (GLMs) to assess the fit of the model and identify any systematic patterns or outliers in the data. It is defined as the square root of the contribution of each observation to the overall deviance of the model. This residue was suggested to make the martingale residual more symmetric around zero. The deviance component for the parametric regression model is given:

$$\hat{r}_{Di} = \text{sign}(\hat{r}_{Mi})\{-2[\hat{r}_{Mi} + \log(1-\hat{r}_{Mi})]\}^{1/2} \tag{26}$$

Where r_{Mi} is the martingale residual. sign() function is a function that drives the (+1) values if the argument is positive and (-1) is negative,

4.4. Pearson Residuals

Pearson residuals are a type of residual used in generalized linear models(GLMs) to assess the fit of the model and identify any systematic patterns or outliers in the data. They are defined as the standardized difference between the observed value and the predicted value for each observation. Pearson residuals are calculated as

$$r_i = \frac{y_i - \hat{\mu}}{\sqrt{\widehat{\text{var}(y)}}}$$

4.5. Cox and Snell Residual

is used to evaluate the fit of the model and identify any systematic patterns or outliers in the data. [6] residual defined as follows

$$e_i = -\ln[1 - F(y_i, \beta)]$$

5. Simulation Study

In this section, a simulation study is given to evaluate the performance of coefficients for the proposed regression model, According to and following steps using Monte Carlo Simulation.

1-Taking initial value $(a = 2, b = 2, \beta_0 = 0.6, \beta_1 = 0.6, \sigma = 1)$
2-Generate y = *QL*(μ, σ) where $\mu = \beta_0 + x\beta_1$
3-Generate x $\sim U(0,1)$
4-Generate z $\sim U(0,1)$
5-Taking samples (20, 50, 100)
6-The simulation replication is N = 1 000
7-For each generated sample size, evaluate (AS)and (MSE) at the three levels

Table 1. Simulation Study Results of Quasi Lindley Regression Model Censoring Rate 0.20

Censoring rate 0.20	n=20			n=50			n=100		
Parameter	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
β_0	1.255	0.0127	0.0032	1.252	0.0050	0.0012	1.252	0.0025	0.0006
β_1	3.125	0.0439	0.0692	3.304	0.0260	0.0340	3.269	0.0126	0.0161
σ	2.242	0.0421	0.0430	2.141	0.0242	0.0211	9.342	0.012	0.001

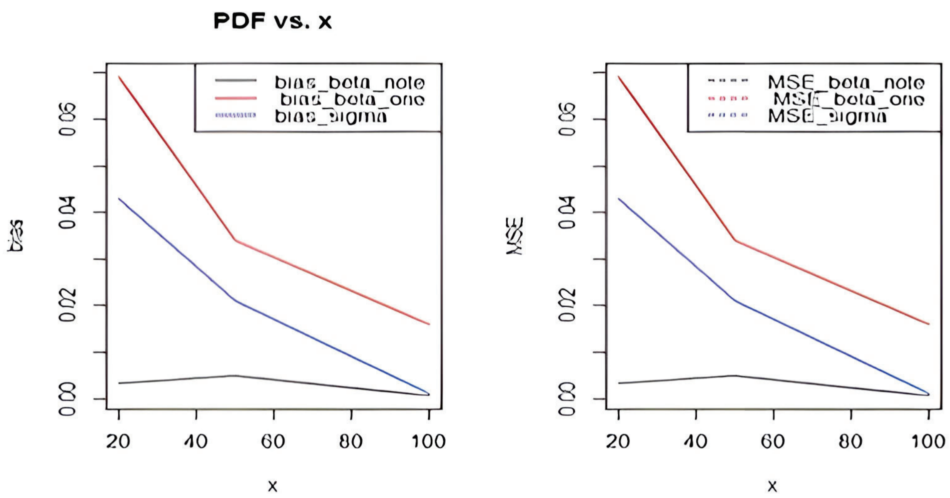

Figure 5. Bias and MSE at different value for Quasi Lindley Regression Model

Censoring rates $(20\%, 30\%, 40\%)$ The simulation results are reported in Table (1),Table(2) and Table (3) . As seen from the results, the estimated biases, average of estimates (AS), and mean square error (MSE) are near the desired value, zero.

Figure 5 contains two panels. The left panel shows the bias of three estimators—labeled "beta_note," "beta_one," and "sigma"—plotted against increasing sample sizes (the horizontal axis). The right panel shows the corresponding mean squared error (MSE) for the same three estimators, again across different sample sizes. In both panels, the curves for all three estimators generally slope downward, indicating that as the sample size grows, bias and MSE both decrease. Among the three, the estimator "beta_note" (black line) consistently has the smallest bias and MSE, "beta_one" (red line) has the largest, and the estimator for "sigma" (blue line) lies in between those two. The fact that the lines converge closer to zero at larger sample sizes demonstrates.

Table 2. Simulation Study Results of Quasi Lindley Regression Model Censoring Rate 0.30

Censoring rate 0.30	n=20			n=50			n=100		
Parameter	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
β_0	1.143	0.012	0.004	1.252	0.004	0.002	1.252	0.002	0.0001
β_1	2.314	0.043	0.061	3.304	0.014	0.022	2.465	0.013	0.012
$\sigma 2$	242	0.043	0.062	2.141	0.024	0.021	9.342	0.012	0.001

Figure 6 compares three estimators—labeled "beta_note," "beta_one," and "sigma"—in terms of bias (left panel) and mean squared error (MSE, right panel) across increasing values of x. In the bias plot, the estimator "beta_note" (black line) remains closest to zero over all x, indicating the smallest bias, whereas "beta_one" (red line) displays the highest bias at smaller x values before decreasing, and "sigma" (blue line) falls between these two. In the MSE plot, a similar trend emerges: "beta_note" consistently shows the lowest overall MSE, "beta_one" the highest for

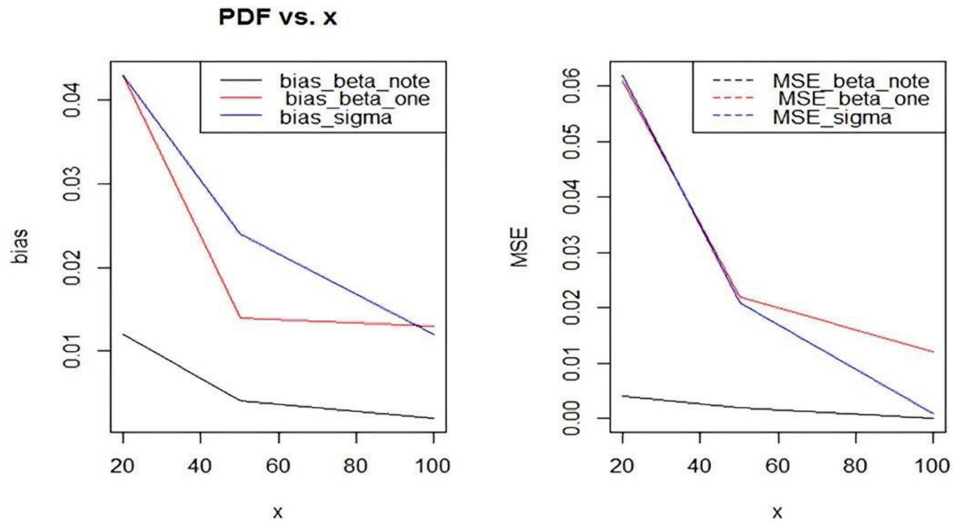

Figure 6. Bias and MSE at different value for Quasi Lindley Regression Model

small x, and "sigma" in between. As x grows larger, the gap between the estimators narrows, reflecting that all estimators improve in both bias and MSE as more data (or larger sample sizes) become available.

Table 3. Simulation Study Results of Quasi Lindley Regression Model Censoring Rate 0.40

Censoring rate 0.40	n=20			n=50			n=100		
Parameter	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
β_0	1.232	0.011	0.003	1.212	0.003	0.001	1.341	0.005	0.001
β_1	3.111	0.054	0.079	3.304	0.013	0.032	3.269	0.011	0.012
σ	1.243	0.031	0.041	2.141	0.014	0.020	9.342	0.011	0.001

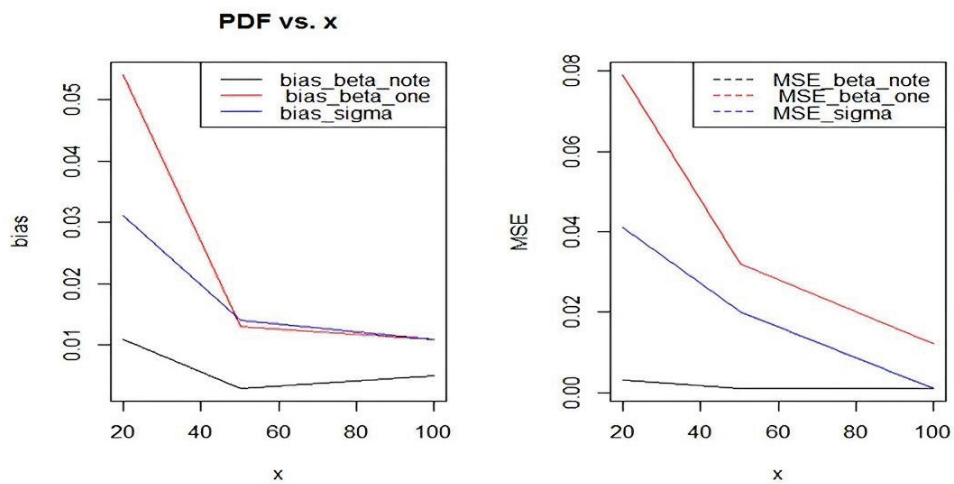

Figure 7. Bias and MSE at different value for Quasi Lindley Regression Model

Figure 7 compares three estimators—“beta_note,” “beta_one,” and “sigma”—in terms of bias (left panel) and mean squared error (MSE, right panel) as x increases. The “beta_note” estimator (black) consistently shows the lowest bias and MSE, with both measures staying close to zero across all x. In contrast, “beta_one” (red) begins with a notably higher bias and MSE at smaller x before decreasing, while “sigma” (blue) lies somewhere in between. As x grows, the performance of all three estimators improves—both their bias and MSE values diminish—indicating that with more data (larger x), each estimator becomes more accurate.

6. Martingale Residuals for Quasi Lindley Regression Model

In survival analysis with a theoretical normal distribution. In this plot, the X-axis represents the quintiles of a normal distribution, while the Y-axis shows the observed quantiles from the data. If the residuals perfectly follow a normal distribution, the points would lie along the dashed red diagonal line. In this case, most points align fairly well with the line, indicating that the residuals are close to being normally distributed. However, there are some deviations at both ends (the tails), where the points drift slightly above or below the line. This suggests that the extreme values (tails) of the residuals deviate from what we’d expect in a perfect normal distribution, possibly indicating heavier tails than normal. Overall, the plot suggests that the residuals are approximately normal but may have minor departures at the extreme.

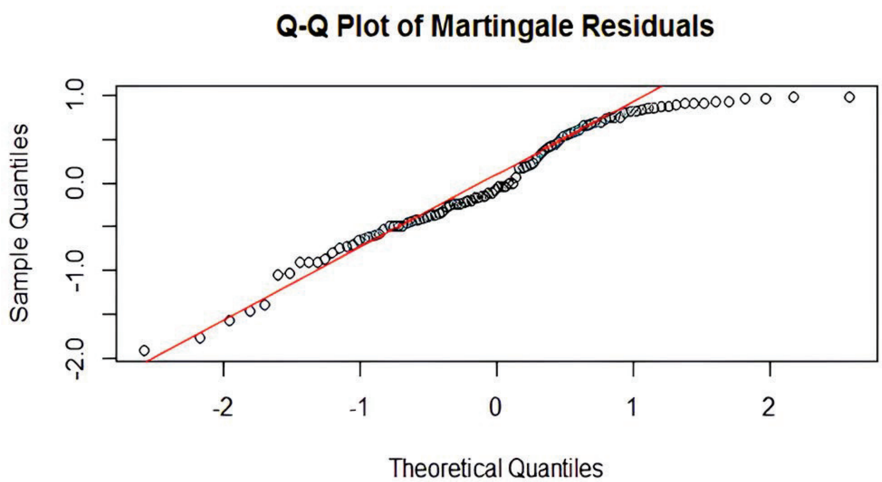

Figure 8. Martingale Residuals for Quasi Lindley Regression Model

7. Real Data

The plot you provided is a scatterplot matrix, also known as a pairs plot. This type of plot is useful for visualizing the relationships between multiple variables in a data set. In your specific plot, the variables are labeled as y (TESTOSTERONE ng/ml), x1 (Years of work-x1), and x2 (age) as follows: As age, their testosterone levels naturally diminish due to physiological changes such testes function and brain hormone signaling. Leydig cells that make testosterone decrease with age, and the hypothalamus and pituitary gland become less effective at hormone regulation. Sex hormone-binding globulin (SHBG) increases, lowering blood testosterone. After 30, this drop averages 1-2 percent every year. Work-related factors can also affect testosterone. Chronic stress from difficult work conditions raises cortisol, which suppresses testosterone. Poor sleep owing to long hours or shift work, poor

Table 4. Descriptive Statistics for Data

Statistic	y	x1	x2
Min	0.700	1.00	20.00
1st Qu.	2.800	4.00	28.00
Median	4.000	8.00	30.00
Mean	3.889	10.56	33.95
3rd Qu.	5.200	13.00	40.00
Max	6.200	39.00	55.00

food, lack of exercise, and smoking can also suppress testosterone levels. Thus, while age is key, job and lifestyle might accelerate the decline. Table 4 presents a concise summary of the minimum, first quartile, median, mean, third quartile, and maximum values for each of the three.

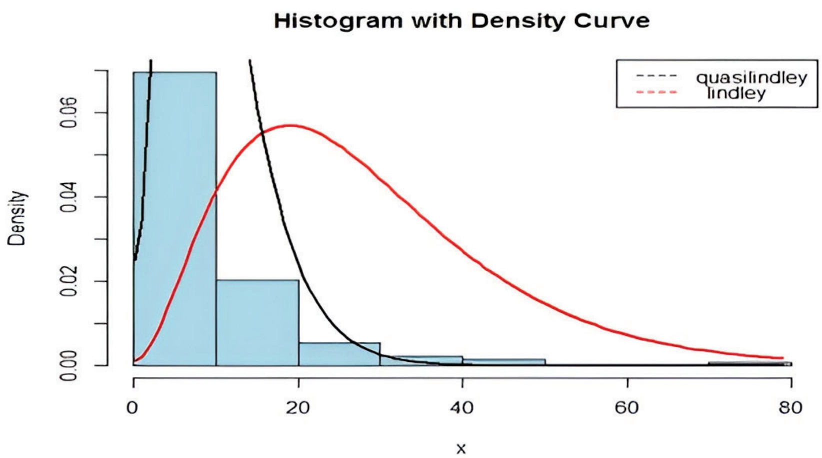

Figure 9. Prabability Density Function for Lindley and Quasi Lindley Distribution

Fiqure 9 displays a histogram of observed data alongside two density curves: the Quasi-Lindley (black dashed line) and Lindley distributions (red dotted line). The histogram represents the frequency distribution of the data, which is heavily skewed to the left, with most values concentrated near zero. Both density curves attempt to model this skewness, but the Quasi-Lindley distribution provides a better fit overall, particularly in the peak region around $x = 0$, while the Lindley distribution slightly overestimates the density at higher values of x. Table 5 compares the performance of two statistical models, Lindley and Quasi Lindley , using four information criteria: AIC, CAIC, BIC, and HQIC. Lower values for these criteria indicate a better-fitting model. Across all four metrics, the Quasi Lindley model consistently outperforms the Lindley model, as evidenced by its lower values (e.g., 723.675 vs. 832.684 for AIC, and 726.527 vs. 838.3881 for BIC). This suggests that the Quasi Lindley distribution provides a better fit to the data and is more efficient in modeling the underlying patterns.

Table 5. Criterion of Comparison Between Lindley and Quasi Lindley Distributions

Criterion	Lindley	Quasi Lindley
AIC	832.684	723.675
CAIC	832.78	723.7067
BIC	838.3881	726.527
HQIC	835.0016	724.8338

8. Fit Data

Table 6. Estimates of Parameter Lindley and Quasi Lindley Regressions

Regression	Beta note	Beta one	AIC	BIC	HQIC
Lindley	0.9755213	0.9965693	2888.863	2894.992	2891.245
Quasi Lindley	0.9761118	0.6298699	2886.864	2890.95	2888.495

Table 6 compares Lindley Regression and Quasi Lindley Regression models, showing parameter estimates (β_0, β_1) and model fit metrics (AIC, BIC, HQIC). The Quasi Lindley model has slightly lower values for all criteria (e.g., AIC $= 2886.864$ vs. 2888.863 for Lindley), indicating a better fit to the data. Furthermore, while the intercept (β_0) is similar for both models, the slope (β_1) is smaller in the Quasi Lindley model, suggesting that it provides a more efficient representation of the relationship.

9. Conclusions

This study introduced and evaluated the Quasi Lindley (Q-L) regression model through extensive simulation studies and real data analysis. The simulation results demonstrated the robustness of the Q-L model, with parameter estimates achieving minimal bias and mean squared error (MSE) across different sample sizes and censoring rates. Notably, the bias and MSE values consistently decreased as sample sizes increased, underscoring the efficiency of the proposed model. Among the estimated parameters, ”beta_note” consistently exhibited the lowest bias and MSE, further validating the model’s stability and accuracy.

The real data analysis provided additional evidence for the superiority of the Q-L model over the traditional Lindley model. Across all statistical criteria, including AIC, BIC, CAIC, and HQIC, the Q-L model consistently outperformed the Lindley model, indicating its improved capacity to fit the data and capture underlying patterns. Additionally, the Q-L model’s probability density function provided a better representation of the observed data, particularly in cases of pronounced skewness.

Overall, the Quasi Lindley regression model proves to be a valuable and efficient tool for analyzing censored data with skewed distributions. Its ability to achieve lower bias, MSE, and better fit metrics makes it a promising alternative for statistical modeling in various applications. Future research may explore extending this model to multivariate scenarios or incorporating additional covariates to enhance its applicability further.

Acknowledgments

We are deeply appreciative of the opportunities provided to explore this fascinating field and for the collaborative spirit that defined this research journey.

REFERENCES

1. Al-Aqtash, R., Mallick, A., Hamedani, G. & Aldeni, M. On the gumbel-burr XII distribution: regression and application. *International Journal of Statistics and Probability* (2021).
2. Altun, E. A new model for over-dispersed count data: Poisson quasi-Lindley regression model. *Mathematical Sciences* **13**, 241–247 (2019).
3. Altun, E. & Cordeiro, G. M. The unit-improved second-degree Lindley distribution: inference and regression modeling. *Computational Statistics* **35**, 259–279 (2020).
4. Altun, E., Yousof, H. M., Chakraborty, S. & Handique, L. Zografos-Balakrishnan Burr XII distribution: regression modeling and applications. *International Journal of Mathematics and Statistics* **19**, 46–70 (2018).
5. Commenges, D. & Rondeau, V. Standardized martingale residuals applied to grouped left truncated observations of dementia cases. *Lifetime Data Analysis* **6**, 229–235 (2000).
6. Cox, D. R. & Snell, E. J. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)* **30**, 248–265 (1968).
7. Elgmati, E., Farewell, D. & Henderson, R. A martingale residual diagnostic for longitudinal and recurrent event data. *Lifetime Data Analysis* **16**, 118–135 (2010).

8. Granzotto, D., Santos, C. & Louzada, F. The transmuted weibull regression model: an application to type 2 diabetes mellitus data. *International Journal of Statistics and Probability* **7**, 1 (2018).
9. Khan, S. A. Exponentiated Weibull regression for time-to-event data. *Lifetime Data Analysis* **24**, 328–354 (2018).
10. Korkmaz, M. C., Altun, E., Chesneau, C. & Yousof, H. M. On the unit-Chen distribution with associated quantile regression and applications. *Mathematica Slovaca* **72**, 765–786 (2022).
11. Kumar, C. S. & Nair, S. R. A generalization to the log-inverse Weibull distribution and its applications in cancer research. *Journal of Statistical Distributions and Applications* **8**, 14 (2021).
12. Mazucheli, J., Korkmaz, M. Ç., Menezes, A. F. & Leiva, V. The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics, and numerical applications. *Soft Computing* **27**, 279–295 (2023).
13. Mazucheli, J., Leiva, V., Alves, B. & Menezes, A. F. A new quantile regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics. *Symmetry* **13**, 682 (2021).
14. Ocloo, S. K., Brew, L., Nasiru, S. & Odoi, B. On the extension of the Burr XII distribution: applications and regression. *Computational Journal of Mathematical and Statistical Sciences* **2**, 1–30 (2023).
15. Ortega, E. M., Cordeiro, G. M. & Carrasco, J. M. The log-generalized modified Weibull regression model (2011).
16. Ortega, E. M., Cordeiro, G. M. & Kattan, M. W. The log-beta Weibull regression model with application to predict recurrence of prostate cancer. *Statistical Papers* **54**, 113–132 (2013).
17. Shanker, R. & Amanuel, A. A new quasi Lindley distribution. *International Journal of Statistics and Systems* **8**, 143–156 (2013).
18. Shanker, R. & Mishra, A. A quasi Lindley distribution. *African Journal of Mathematics and Computer Science Research* **6**, 64–71 (2013).
19. Silva, G. O., Ortega, E. M., Cancho, V. G. & Barreto, M. L. Log-Burr XII regression models with censored data. *Computational Statistics & Data Analysis* **52**, 3820–3842 (2008).
20. Tharshan, R. & Wijekoon, P. Poisson-Modification of Quasi Lindley regression model for over-dispersed count responses. *Communications in Statistics - Simulation and Computation* **53**, 1258–1273 (2024).
21. Therneau, T. M., Grambsch, P. M. & Fleming, T. R. Martingale-based residuals for survival models. *Biometrika* **77**, 147–160 (1990).
22. Verweij, P. J., van Houwelingen, H. C. & Stijnen, T. A goodness-of-fit test for Cox's proportional hazards model based on martingale residuals. *Biometrics*, 1517–1526 (1998).