Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 2, 2025 (New Series), pp 221–236
http://www.ssca.org.in/journal

On Log Odd Burr III Weibull Regression Model and its Application in Survival Analysis

Deepthy G. S., Lakshmi R. and Nicy Sebastian

Department of Statistics
St.Thomas College (Autonomous), Thrissur, affiliated to the University of Calicut, 680 001, Kerala, India

Received: 01 September 2024; Revised: 18 February 2025; Accepted: 18 March 2025

Abstract

This paper focuses on parametric forms, particularly the location-scale regression model using the odd Burr III Weibull distribution. This model serves as an alternative to the log odd log-logistic Weibull and log-Weibull regression model for analyzing data with decreasing, increasing, unimodal, and bathtub-shaped failure rate functions. Several mathematical properties of the log-transformed distribution have been established. The suggested distribution has the benefit that it encompasses various classical distributions as submodels found in the literature. Maximum likelihood and jackknife estimation techniques are used to assess the model parameters. We conduct simulations under different parameter settings to guarantee model robustness. In our proposed model, we employ diagnostic techniques that employ case deletion known as global influence to determine the influential observations. Additionally, we present martingale and modified deviance residuals to figure out the outliers and assess the model assumptions. The performance of the newly developed regression model is exhibited using a real life dataset.

Key words: Regression model; Censored data; Residual analysis; Sensitivity analysis.

AMS Subject Classifications: 62F10, 60E05, 62N01

1. Introduction

In order to develop more reliable and significant models, the trend in recent statistics literature has changed towards introducing novel methods that incorporate additional parameters into the extensive range of continuous univariate distributions. A distribution that has faced a major hurdle is the Weibull distribution, which is well-known for usage in lifetime data analysis. The drawback of this distribution is its ineffectiveness to consider the non-monotone hazard rates, such as bathtub-shaped hazard rates. To address this, researchers are seeking for generalizations of Weibull distribution that might fit the datasets better than the traditional two parameter Weibull models. For instance, exponentiated Weibull distribution proposed by Mudholkar *et al.* (1995), the generalized modified Weibull distribution

Corresponding Author: Deepthy G. S.
Email: deepthygs@gmail.com

presented by Carrasco *et al.* (2008a), and the beta Weibull distribution explored by Lee *et al.* (2007) represent some recent advancements in Weibull generalizations, demonstrating the continued relevance of the Weibull distribution in lifetime data modeling.

In everyday scenarios, a number of factors might influence survival time, which can explain the variability in the time of survival. To assess how these factors influence survival time, it is important to employ proper regression model for censored and time-to-failure data. Thus, describing a probabilistic model for survival time is crucial to design regression models. Different types of regression models exist for this purpose. The location-scale regression model (LSRM), as highlighted by Lawless (2011) stands out among these and is often utilized in clinical trials. In these models, applying a log transformation serves to linearize multiplicative relationships, making the model well-suited for time-to-event analysis. A linear combination is assumed for the log-lifetimes models preserving the characteristics of standard regression models and also incorporating the advantage of handling censored data, effectively managing skewed distributions, and stabilizing variance. Numerous research studies have been carried out employing log-LSRM. Among these, regression analysis using bivariate location-scale models was discussed by He and Lawless (2005) with applications to lifetime data, Silva *et al.* (2010) suggested the log-Weibull extended regression model for censored data, whereas Carrasco *et al.* (2008b) applied log-modified Weibull regression models to evaluate censored data. Cruz *et al.* (2016) investigated the log-odd log-logistic Weibull regression model, whereas Elbatal *et al.* (2022) introduced a new log LSRM based on the odd Perks-Weibull distribution. Furthermore, Zamanah *et al.* (2022) used survival time data from hypertensive patients to illustrate the use of the log-harmonic mixed Weibull-Weibull distribution, a log-modified alpha power-transformed Burr XII regression model was proposed by Anafo (2024) and then analyze a stock market liquidity dataset. The relevance of these distributions also extends to their potential to incorporate non-monotonic failure rate functions in lifetime analysis, a regular occurrence in both survival analysis and reliability studies.

In this study, we discuss a new log-location regression model based on the logarithm of the odd Burr III Weibull distribution, which we call the log odd Burr III Weibull (LOBIIIW) distribution. The modification in the current distribution results in a LSRM that is apt for modelling censored survival times exhibiting bathtub shaped hazard rates. The logarithm of the lifetimes is carried out to regulate variance and establish a more linear relationship with the covariates. This model provide an adequate alternative for the log-logistic regression model. Our suggested distribution should therefore become beneficial in a variety of applications for evaluating survival and reliability data.

Following model execution, it becomes necessary to illustrate the model's assumptions and perform out robustness checks in order to detect any significant observations that may influence the results of the analysis. A study conducted by Cook (1977) established an influence diagnostic approach that uses case deletion technique to evaluate the significance of each observation on parameter estimates. Our proposal suggests using case deletion based diagnostic measures to identify influential data points in the analysis of censored data using LOBIIIW regression models.

The OBIIIW and LOBIIIW distributions are described in Section 2, along with the structural characteristics of the LOBIIIW distribution. In Section 3, we present an location-scale LOBIIIW regression model. Section 4 presents estimation methods like maximum likelihood estimation and jackknife method and discusses the findings from the simulation studies. Two different kinds of residuals are defined in Section 5 for evaluating deviations from the error assumptions and the existence of outliers. An evaluation of a real-data set in Section 6 highlights the significance of the new model. Conclusions are provided at the end of Section 7.

2. Log Odd Burr III Weibull distribution

Jamal *et al.* (2017) defines the cumulative distribution function (CDF) for the odd Burr III Generating (OBIII-G) family as follows:

$$F(t;c,k,\gamma)=\left\{1+\left[\frac{1-G(t,\gamma)}{G(t,\gamma)}\right]^{c}\right\}^{-k},$$

where $c>0$ and $k>0$ denotes the shape parameters and γ represents the parameters of the baseline distribution.
A novel extension of the Weibull distribution has been introduced using the OBIII-G family. This novel distribution named the odd Burr III Weibull (OBIIIW) distribution is established by incorporating the Weibull distribution as base distribution in the OBIII-G family.
Consider the random variable T following OBIIIW distribution, characterized by the parameters c,k,a and b. Then, the survival function is given as follows:

$$S(t;c,k,a,b)=1-\left\{1+\left[\frac{e^{-(at)^{b}}}{1-e^{-(at)^{b}}}\right]^{c}\right\}^{-k}, \tag{1}$$

and the corresponding probability density function (PDF) defined by,

$$f(t;c,k,a,b) \;=\; ckba^{b}t^{b-1}\frac{[e^{-(at)^{b}}]^{c}}{[1-e^{-(at)^{b}}]^{c+1}}\left\{1+\left[\frac{e^{-(at)^{b}}}{1-e^{-(at)^{b}}}\right]^{c}\right\}^{-k-1}, c,k,a,b>0, \tag{2}$$

where c, k and b are the shape parameters, and a is the scale parameter.

Usman and Haq (2019) undertook an extensive investigation of the characteristics and applications of the above proposed model. Their conclusions revealed that the model offers the benefit of accommodating various hazard rate patterns including constant, decreasing, increasing, unimodal and bathtub shaped hazard rates (see Figure 1). This flexibility holds particular significance in reliability analysis, where non-monotone hazard rates are frequently encountered in real world data sets.

If the random variable T has the OBIIIW density function (2), then the random variable $Y=\log(T)$ has the log odd Burr III Weibull (LOBIIIW) distribution. Its density function, parameterized by $a=e^{-\mu}$ and $b=\sigma^{-1}$, can be formulated as follows:

$$f(y;c,k,\mu,\sigma)=\frac{ck}{\sigma}\frac{\exp\left[\left(\frac{y-\mu}{\sigma}\right)-c\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{\left[1-\exp(-\exp(\frac{y-\mu}{\sigma}))\right]^{c+1}}\left\{1+\left(\frac{\exp[-\exp(\frac{y-\mu}{\sigma})])}{1-\exp(-\exp(\frac{y-\mu}{\sigma}))}\right)^{c}\right\}^{-k-1}, \tag{3}$$

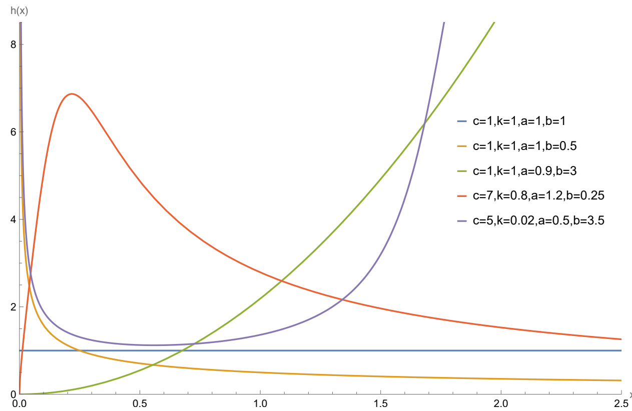

Figure 1: Hazard plot of OBIIIW distribution

where $-\infty < y < \infty$, $c > 0$, $k > 0$, $\sigma > 0$ and $-\infty < \mu < \infty$. We refer to Equation (3) as representing the distribution of $Y \sim LOBIIIW(c, k, \mu, \sigma)$, where μ indicates the location parameter, σ is scale parameter, and c and k are the shape parameters. Figure 2 displays graphs depicting the density function (3) for various parameter values. These findings suggest that this distribution serves as a suitable choice for effectively modeling both left-skewed, right-skewed, and symmetric datasets.
The survival function associated with Equation (3) can be expressed as,

$$S(y; c, k, \mu, \sigma) = 1 - \left\{1 + \left[\frac{\exp[-\exp(\frac{y-\mu}{\sigma})]}{1 - \exp[-\exp(\frac{y-\mu}{\sigma})]}\right]^c\right\}^{-k}. \tag{4}$$

Then the standardized random variable $Z = \frac{Y-\mu}{\sigma}$, which is characterized by the density function is given as follows:,

$$\pi(z; c, k) = ck\frac{\exp\left[z - c\exp\left(z\right)\right]}{\left[1 - \exp(-\exp(z))\right]^{c+1}}\left\{1 + \left(\frac{\exp[-\exp(z)])}{1 - \exp(-\exp(z))}\right)^c\right\}^{-k-1}, \; -\infty < z < \infty. \tag{5}$$

3. Log Odd Burr III Weibull regression model

Researchers and statisticians have been investigating flexible regression models aimed at effectively capturing non-monotone failure rates, frequently encountered in fields such as reliability and biology. To overcome the limitations of the Weibull distribution in modelling such patterns, a LSRM utilizing the LOBIIIW distribution can be used instead. This model is particularly beneficial for analyzing data characterized by increasing, decreasing and bathtub shaped hazard failures. The importance for the newly developed regression model is its capability to extend the commonly used Weibull distribution to more complex scenarios through continuous extension. This generalization opens up new possibilities for applications that require more sophisticated statistical analysis. Moreover, it is crucial to examine the impact of various explanatory variables on lifetimes such as cholesterol levels, weight, blood pressure *etc*. Regression models, in particular location scale models, give researchers important insight regarding the relationship between lifetime and the explanatory factors. We have now established a linear regression model to analyze the relationship between the

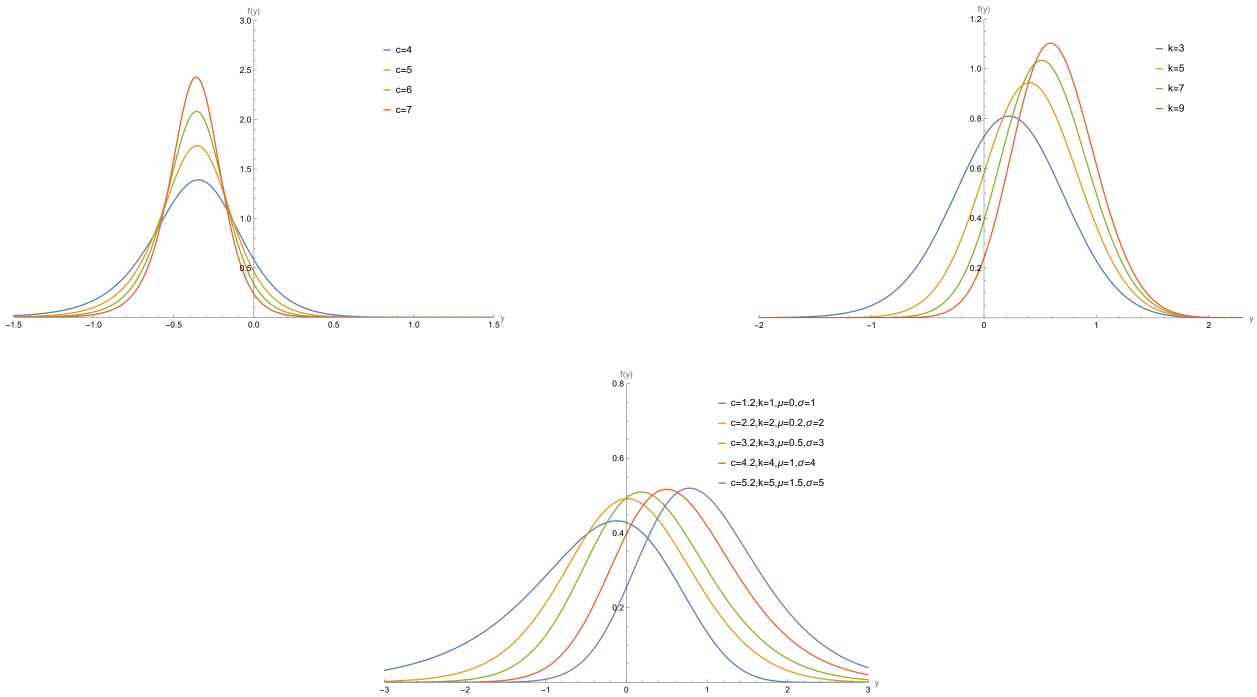

Figure 2: Plots of the LOBIIIW pdf for parameter values (a) $k = 1$, $\mu = 0$, $\sigma = 1$ (b) $c = 1.5$, $\mu = 0$, $\sigma = 1$

response variable y_i and the explanatory variable vector $\mathbf{x_i} = (x_1, x_2, \ldots, x_p)^T$ using the LOBIIIW distribution, in the following way:

$$y_i = \mathbf{x_i^T}\boldsymbol{\beta} + \sigma z_i, i = 1, 2, \ldots, n. \tag{6}$$

where the error term z_i follows the density function (5), $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$, $c > 0$ and $k > 0$ are the unknown parameters. The vector $\mathbf{x_i^T} = (x_{i1}, x_{i2}, \ldots, x_{ip})$ are linked with the explanatory variables. The vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ of LOBIIIW regression model represents the location parameters and is indicated as linear model $\boldsymbol{\mu} = \boldsymbol{X\beta}$ where $\boldsymbol{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is a pre-defined model matrix. New opportunities for fitting several kinds of data are made possible by the LOBIIIW model (6).
Using the log linear model given in Equation (6), the survival function $Y_i|\mathbf{x}$ simplifies to,

$$S(y_i|\mathbf{x}) = 1 - \left\{1 + \left[\frac{\exp[-\exp(\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma})]}{1 - \exp[-\exp(\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma})]}\right]^c\right\}^{-k}. \tag{7}$$

The regression model (6) reduces to the log-odd logistic regression model when $k = 1$, and to the log-Weibull regression model when $c = k = 1$.

4. Estimation of the LOBIIIW regression model

4.1. Maximum likelihood estimation method

Let the sample consists of n independent observations $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)$, where $y_i = \min\{\log(T_i), \log(C_i)\}$, and $\mathbf{x}_i$ denotes the vector of explanatory variables related to the i^{th} individual. We consider that the observed lifetimes ($\log(T_i)$) and censoring times (C_i) are independent and the censorship is non-informative. The log-likelihood function for the vector of parameters $\boldsymbol{\eta} = (c, k, \sigma, \boldsymbol{\beta^T})^T$ is given by,

$$l(\boldsymbol{\eta}) = \sum_{i \in F} l_1(c, k, z_i) + \sum_{i \in C} l_2(c, k, z_i) \tag{8}$$

where,

$$\begin{aligned} l_1(c, k, z_i) &= \log \left[\frac{ck}{\sigma} \frac{\exp\left[z_i - c\exp\left(z_i\right)\right]}{\left[1 - \exp(-\exp(z_i))\right]^{c+1}} \left\{ 1 + \left(\frac{\exp[-\exp(z_i)])}{1 - \exp(-\exp(z_i))} \right)^c \right\}^{-k-1} \right], \\ l_2(c, k, z_i) &= \log \left[1 - \left\{ 1 + \left[\frac{\exp[-\exp(z_i)]}{1 - \exp[-\exp(z_i)]} \right]^c \right\}^{-k} \right]. \end{aligned}$$

and $z_i = \frac{y_i - \mathbf{x_i}^T \boldsymbol{\beta}}{\sigma}$. The set F represents individuals for whom y_i corresponds to log-lifetime, whereas C designates individuals subject to right censoring.

Maximizing the likelihood function (8) yields the maximum likelihood estimates (MLEs) for the parameter vector $\boldsymbol{\eta}$. The estimates are obtained by using the R software, specifically utilizing the MAXLIK package.

Furthermore, the LOBIIIW model can be compared to particular sub-models using the Likelihood Ratio (LR) test. The maximum log-likelihood values for the unconstrained and constrained models can be used to build the LR statistic for testing specific models within the LOBIIIW regression model.

4.2. Simulation

We implement a Monte-Carlo simulation study for examining the behavior of maximum likelihood estimators (MLEs) of c, k, σ, β_0, and β_1 in a finite sample ($n = 50$, 100, and 300). The samples are obtained using the true parameter values: $c = 2$, $k = 4$, $\sigma = 1.5$, $\beta_0 = 3$ and $\beta_1 = 2$ with varying levels of censoring generally at censoring percentages, cp= 0%, 10% and 30%. The log-lifetimes $\log(T_1)$, ... , $\log(T_n)$ are simulated from the LOBIIIW regression model (6), where $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_i$ and x_i are sampled from a uniform distribution within the interval (0, 1). The censored times C_1, C_2, ... , C_n are chosen from a uniform distribution over the interval $(0, \tau)$, where τ has been altered until the required censoring percentages are attained. Every combination of n, c, k, σ, β_0, β_1, and censoring percentages is then used to generate 1000 samples. Each dataset generated was fitted with the LOBIIIW regression model (6), where $\mu_i = \beta_0 + \beta_1 x_i$. The mean squared errors (MSEs) and average estimates (AEs) for the MLEs of c, k, σ, β_0, β_1 computed from simulation results are shown in Table 1. We can see from the data in Table 1 that MSEs increase along with an increase in the censoring percentage. Additionally as predicted, the MSEs decline with sample size increases. These facts indicate that the asymptotic normal distribution accurately approximates the distribution of the estimates from a finite sample.

5. Residual analysis

A key step after fitting a statistical model to the dataset is residual analysis. It serves for various purposes, including validating the data, evaluating the data to unveil valuable

Table 1: AEs and MSEs (in parentheses) of MLEs of $\boldsymbol{c}, \boldsymbol{k}, \boldsymbol{\sigma}, \boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$

Percentage	Parameter	Sample Size		
		50	100	300
	c	2.0489(0.0479)	2.0195(0.0210)	2.0086(0.0072)
	k	4.1593(0.5196)	4.0676(0.2080)	4.0195(0.0682)
0%	σ	1.3882(0.1946)	1.4689(0.0856)	1.4981(0.0384)
	β_0	3.1711(0.1313)	3.2008(0.0937)	3.2220(0.0688)
	β_1	2.3465(0.2200)	2.3294(0.1647)	2.3203(0.1273)
	c	2.0174(0.0494)	1.9894(0.0228)	1.9824(0.0078)
10%	k	4.3935(0.7953)	4.2926(0.3593)	4.2337(0.1380)
	σ	1.3016(0.3028)	1.3745(0.1242)	1.4457(0.0507)
	β_0	3.1143(0.1505)	3.1500(0.0777)	3.1571(0.0497)
	β_1	2.3528(0.2171)	2.3168(0.1409)	2.2673(0.0913)
	c	1.8797(0.0593)	1.8464(0.0437)	1.8328(0.0348)
	k	4.9652(1.8927)	4.8063(0.9988)	4.7374(0.6627)
30%	σ	1.0040(0.8286)	1.0962(0.4321)	1.0828(0.3557)
	β_0	3.0172(0.2901)	3.0948(0.1148)	3.1736(0.0844)
	β_1	2.4716(0.3867)	2.4157(0.2480)	2.3849(0.1784)

information, and confirming the assumptions of the model. Additionally, residual analysis can assist in identifying outliers and reviewing any deviations from the error assumption. Several types of residual analysis have been mentioned in the literature to achieve these goals. For instance, Collett (2023), McCullagh (2019), Fleming and Harrington (2013) and Ortega *et al.* (2008) have discussed about various residual analyses. This study assesses two types of residuals: martingale-type residual and deviance component residual. The martingale residual (MR) in parametric lifetime models may be defined as $r_{Mi} = \delta_i + \log[S(y_i, \hat{\boldsymbol{\eta}})]$, where $\delta_i = 0(\delta_i = 1)$ signifies censored (uncensored) observation and $S(y_i, \hat{\boldsymbol{\eta}})$ refers to the estimated survival function as specified by Equation 7. In fact, r_{Mi} spans from a minimum of $-\infty$ to maximum of $+1$. Thus, the MR for the LOBIIIW model assumes the following expression:

$$r_{Mi} = \begin{cases} 1 + \log\left\{1 - \left(1 + \left[\frac{\exp[-\exp(\hat{z}_i)]}{1-\exp[-\exp(\hat{z}_i]}\right]^{\hat{c}}\right)^{-\hat{k}}\right\} & \text{if} \quad \delta_i = 1, \\ \log\left\{1 - \left(1 + \left[\frac{\exp[-\exp(\hat{z}_i)]}{1-\exp[-\exp \hat{z}_i]}\right]^{\hat{c}}\right)^{-\hat{k}}\right\} & \text{if} \quad \delta_i = 0, \end{cases}$$

where $\hat{z}_i = \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}$.

The drawback of the MR is its significant skewness, indicating that it does not closely resemble a normal distribution. To address this issue, Therneau *et al.* (1990) proposed the modified deviance residual (MDR), which transforms the MR to reduce skewness. The advantage of the MDR is that its distribution approximates a normal distribution as closely as possible, facilitating more effective residual analysis, which can be formulated as:

$$r_{Di} = \text{sign}(r_{Mi})\{-2[r_{Mi} + \delta_i \log(\delta_i - r_{Mi})]\}^{1/2},$$

where r_{Mi} is the MR.

5.1. Simulation studies

A study using simulation was executed to examine the empirical distributions of residuals r_{Di} for the values n = 50, 100 and 300 and censoring percentage 0%, 10%, and 30%, following the procedure outlined in Section 4.2. Following conclusions can be drawn by observing Figure 3:

1. The empirical distribution of the MDR closely matches the standard normal distribution, indicating a strong agreement.
2. The empirical distribution of the MDR tends to converge towards the standard normal distribution as the censoring percentage decreases or the sample size increases.

6. Application

The heart transplant dataset, which includes records up to April 1st, 1974, is detailed in the book by Kalbfleisch and Prentice (2011). During this time frame, certain patients passed away before a suitable heart became available. Out of the 103 patients, 69 underwent heart transplant surgeries, with 75 deaths reported. The remaining 28 patients provided censored survival time data. The survival times from patient entry are represented by the response variable y_i (the number of days from acceptance into the transplantation program to both the transplant and death was recorded for each patient). The failure indicator is shown by the variable "survival status" (0 - censored, 1 - observed). Each patient (i = 1, 2, . . . , 103) was related to the following explanatory variables.

- x_{i1} = Age at acceptance(in years).
- x_{i2} = Prior surgery (marked as 1 for yes and 0 for no).
- x_{i3} = Transplant (0 = no; 1 = yes).

In many cases, it is possible to qualitatively evaluate the form of the failure rate function, which is useful when choosing a model. The total time on test (TTT) plot is a useful tool in this context Aarset (1987). Figure 4 demonstrates the TTT plot for this data, which shows a failure rate function has a monotonically increasing pattern. Hence, the LOBIIIW distribution is appropriate for modeling this dataset.

Now, we demonstrate the results obtained from fitting the LOBIIIW regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma z_i.$$

where the errors z_i are independent random variables with a density function (5)

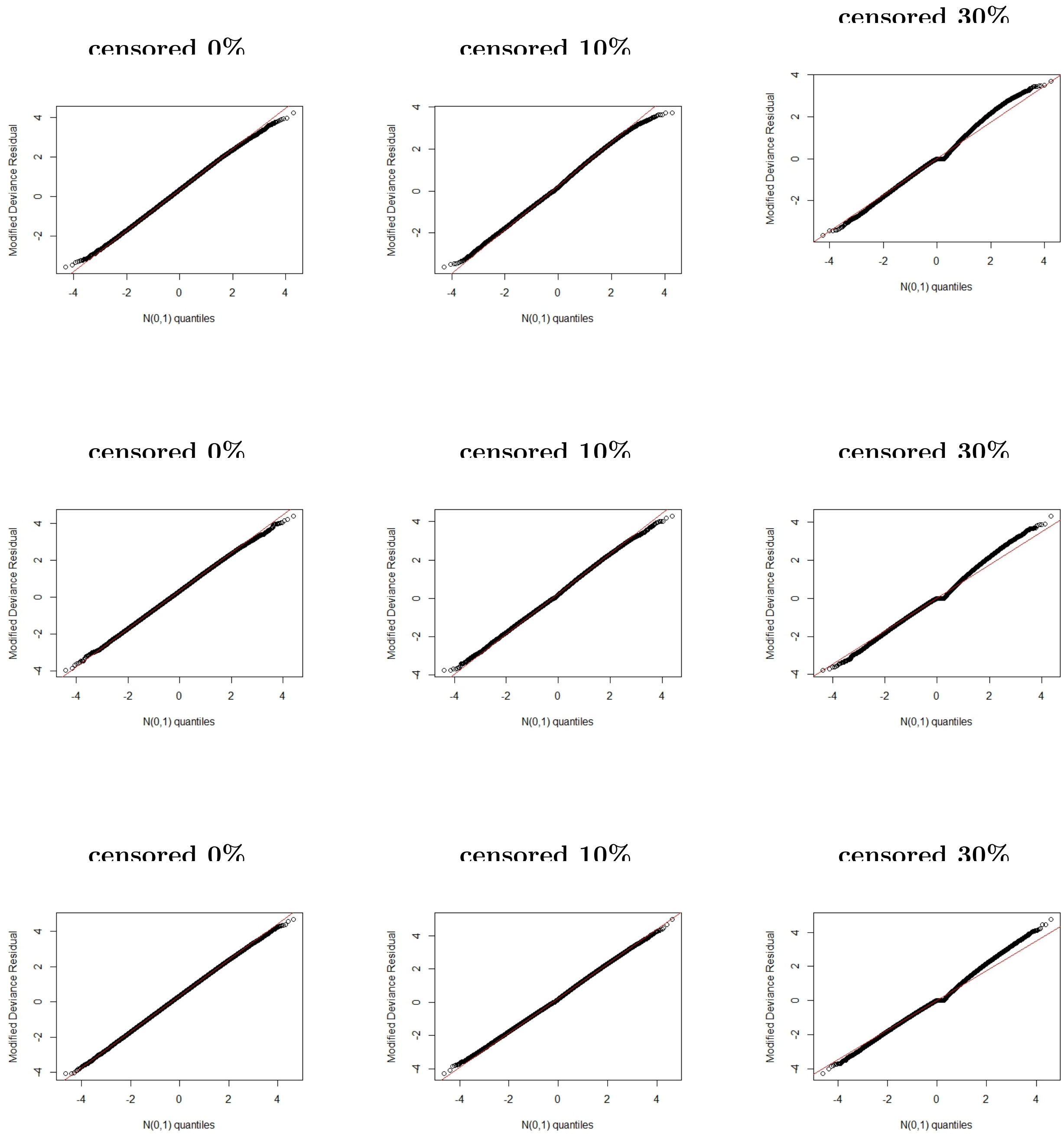

Figure 3: Normal probability plots for the residuals r_{Di} for parameter values $c = 2$, $k = 4$, $\sigma = 1.5$, $\beta_0 = 3$, $\beta_1 = 2$

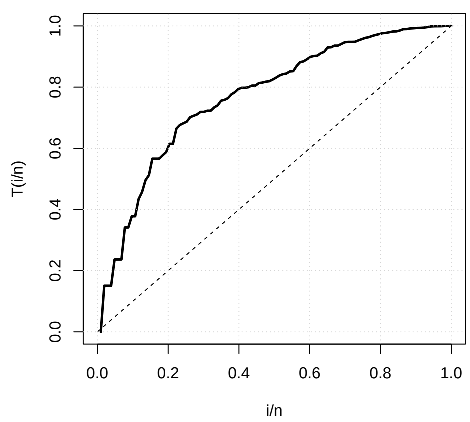

Figure 4: TTT plot of Stanford heart transplant data

6.1. Results of maximum likelihood and jackknife estimation

The statistical program R has been employed for estimating the maximum likelihood estimates of the LOBIIIW, Log Odd Log Logistic Weibull (LOLLW), and Log Weibull (LW) regression models. Table 2 presents the parameter estimates along with their standard errors, p-values, and model selection criteria (AIC, CAIC, and BIC statistics). The smaller standard errors represent more precise parameter estimates and lower p - values reveal that the covariates x_1, x_2, and x_3 have a significant impact. A comparative analysis of these models indicates that the LOBIIIW model consistently yields lower values across the model selection criteria. Also, we can analyze the estimated coefficients in the following way: the expected survival time is expected to decreases by approximately 94.08% ($e^{-0.061} \times 100\%$) when the age at acceptance (x_1) increases by one unit. Variables x_2 and x_3 might also be interpreted in the same way. The jackknife method was also used along with maximum likelihood estimation (MLE) to validate the results. The jackknife serves as a diagnostic tool to verify parameter stability since MLE can be sensitive to influential observations and small sample size. Table 3 provides the jackknife estimates of the model parameters. For both estimation procedures, the LOBIIIW regression model's findings indicate that the explanatory variables x_1, x_2, and x_3 are statistically significant at the 5% level. Both methods presented nearly identical estimates.

The Cox proportional hazards model was employed as a comparative technique since it is a useful regression model for analyzing censored failure times. Further, estimates, standard errors (SE), and p - values for the Cox regression model are displayed in Table 4. Aligning with the findings of the LOBIIIW regression model, the Cox regression model also shows that all explanatory variables are marginally significant at the 5% level.

Furthermore, the LR statistic is employed to compare the LOBIIIW and LW regression models by testing the hypotheses $H_0 : c = k = 1$ against $H_1 : H_0$ is not true. The test yielded a test statistic value of $w = 2(-166.3611 - (-171.7112)) = 10.7002$, with a corresponding p-value of 0.0047. These results signify a favorable indication towards the

LOBIIIW regression model.

Table 2: The parameter estimates, standard errors (given in parentheses) and *p*-values in [.], for the LOBIIIW, LOLLW and LW regression models fitted to the heart transplant data

Models	c	k	σ	β_0	β_1	β_2	β_3
LOBIIIW	1.813 (0.722)	2.378 (1.114)	3.506 (1.153)	5.604 (1.254) [< 0.001]	-0.061 (0.019) [0.001]	1.450 (0.589) [0.013]	2.580 (0.378) [< 0.001]
		AIC = 346.722		BIC = 365.165		CAIC=347.901	
LOLLW	4.628 (3.530)		6.203 (4.685)	8.744 (1.760) [< 0.001]	-0.076 (0.019) [< 0.001]	1.405 (0.574) [0.016]	2.591 (0.388) [< 0.001]
		AIC = 347.595		BIC = 363.404		CAIC = 348.470	
LW			1.465 (0.1314)	7.974 (0.933) [< 0.001]	-0.092 (0.020) [< 0.001]	1.214 (0.647) [0.063]	2.537 (0.373) [< 0.001]
		AIC = 353.420		BIC = 366.594		CAIC = 354.039	

Table 3: Jackknife estimates fitted to the heart transplant data

Parameter	Estimates	SE	95%CI	p-value
c	1.8790	0.7775	(0.3550, 3.4028)	0.0156
k	2.6486	1.2789	(0.1419, 5.1553)	0.0383
σ	3.7520	1.2375	(1.3264, 6.1775)	0.0024
β_0	5.6791	1.3454	(3.0420, 8.3161)	< 0.001
β_1	-0.0649	0.0192	(-0.1297, -0.0271)	< 0.001
β_2	1.5309	0.5878	(0.3788, 2.68290)	0.0091
β_3	2.5511	0.3804	(1.8055, 3.2967)	< 0.001

Table 4: Cox regression model estimates fitted to the heart transplant data

Parameter	Estimate	SE	p-value	95%CI
β_1	0.05919	0.01494	<0.0001	(0.0299, 0.0884)
β_2	-0.74266	0.44225	0.0931	(-1.6094, 0.1241)
β_3	-1.66121	0.27588	<0.0001	(-2.2019, -1.1204)
	AIC	BIC	CAIC	
	554.3522	561.3047	554.6902	

6.2. Results of sensitivity and residual analysis

The global influence measures $GD_i(\eta)$ and $LD_i(\eta)$ have been evaluated to identify the influential observations in the LOBIIIW regression model fitted to the present dataset. Figure 5 shows the index plots of influence measures, which signifies that the possible influential observations are 18 and 99. Another way for detecting potential outliers in the LOBIIIW regression model involves plotting the MDR (r_{Di}) against the adjusted values, as shown in Figure 6(a). Upon examining the MDR plot, a residual analysis indicates that observations 8 and 99 are identified as potential outliers.

Both sensitivity and residual analysis determined observations 99 as the most frequently occurring possible influential points. Observation 99 indicates the longest survival time within the present censored dataset. We modified the model by eliminating the above observations in order determine whether they had a significant effect on the parameter estimates. The Table 5 illustrates the relative change in the estimated parameters, given by the formula $R = \frac{\hat{\eta}_j - \hat{\eta}_{j(i)}}{\hat{\eta}_j}$. Here, $\hat{\eta}_{j(i)}$ represents the MLE estimates calculated excluding the i^{th} record. We evaluate the robustness of the parameters in the LOBIIIW regression model, considering that the covariates x_1, x_2 and x_3 have significance in the fitted model. Thus, the inference remains unchanged when the observation identified as potentially influential in the diagnostic plots is removed.

Also, we display the normal probability plot for MDR with the simulated envelope in Figure 6(b). The plot in Figure 6(b) recommends that the LOBIIIW regression model appears to adequately fit the dataset, with no observations occurring as possible outliers.

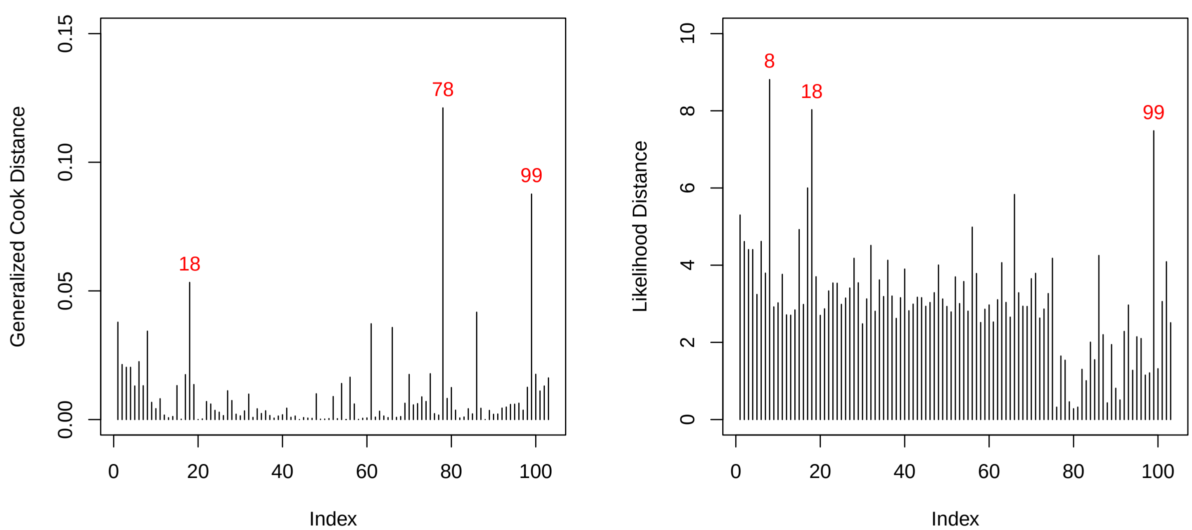

Figure 5: The index plot of (a) $GD_i(\eta)$ and (b) $LD_i(\eta)$ for the LOBIIIW regression model

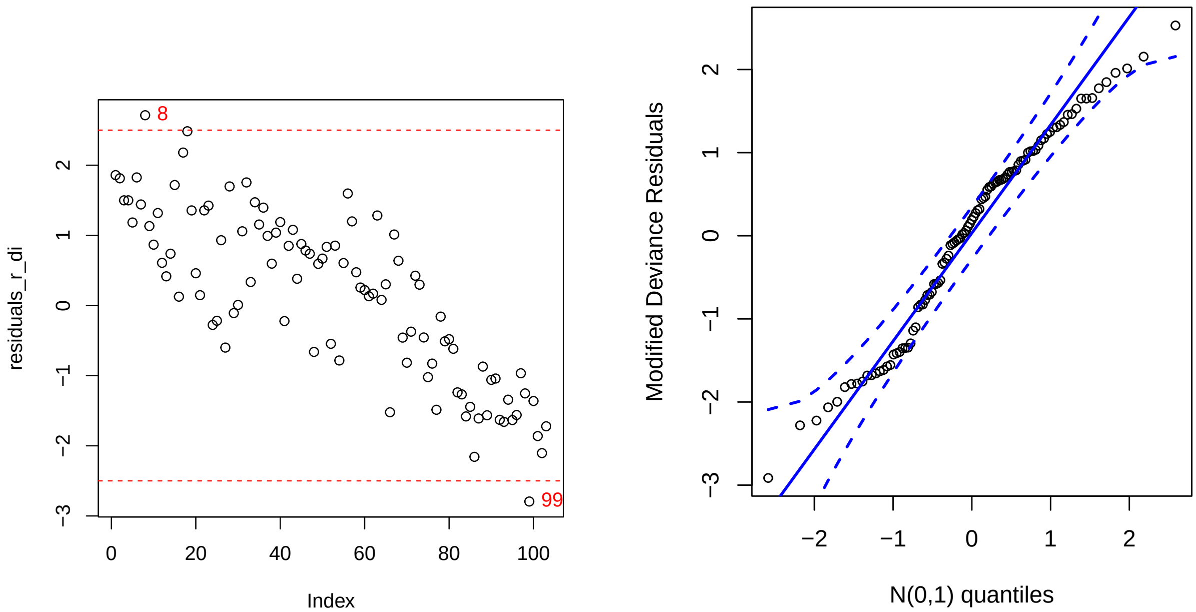

Figure 6: (a) Index plot of the deviance residual and (b) Normal probability plot of r_{Di} with envelopes

Table 5: Relative changes (expressed in percentages) along with the estimates and corresponding p-values for the datasets

Dropped Observation	$\hat{c}$	$\hat{k}$	$\hat{\sigma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
None	-	-	-	-	-	-	-
	1.8136	2.3785	3.5062	5.6040	-0.0610	1.4501	2.5804
				(< 0.001)	(0.0013)	(0.0138)	(< 0.001)
#99	[-2.2618]	[-12.2317]	[-2.6506]	[3.8708]	[0.2569]	[-6.734]	[-1.4363]
	1.8546	2.6694	3.5991	5.3871	-0.0609	1.5478	2.6175
				(< 0.001)	(< 0.001)	(0.0066)	(< 0.001)

7. Conclusion

This study introduces a novel LOBIIIW regression model for analyzing survival data, particularly when dealing with censored observations. The model performs effectively when the hazard function displays a bathtub form. We used two methods to estimate the parameters of the proposed model: the maximum likelihood and jackknife estimator. The assumptions of the model were verified by utilizing MDR. Simulation research has shown that the empirical distribution of MDR approximates a standard normal distribution. In addition, GCD and LD measures were introduced to detect influential observations in the regression model. We also demonstrate the effectiveness of the model by analyzing a real dataset.

Acknowledgements

The authors express their gratitude to the reviewer and the chief editor for their valuable suggestions, which led to considerable improvements in the research paper.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, **36**, 106–108.

Carrasco, J. M., Ortega, E. M., and Cordeiro, G. M. (2008a). A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics and Data Analysis*, **53**, 450–462.

Carrasco, J. M., Ortega, E. M., and Paula, G. A. (2008b). Log-modified Weibull regression models with censored data: Sensitivity and residual analysis. *Computational Statistics and Data Analysis*, **52**, 4021–4039.

Collett, D. (2023). *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cruz, J. N. d., Ortega, E. M., and Cordeiro, G. M. (2016). The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. *Journal of Statistical Computation and Simulation*, **86**, 1516–1538.

Elbatal, I., Alotaibi, N., Almetwally, E. M., Alyami, S. A., and Elgarhy, M. (2022). On odd perks-G class of distributions: properties, regression model, discretization, Bayesian and non-Bayesian estimation, and applications. *Symmetry*, **14**, 883.

Fleming, T. R. and Harrington, D. P. (2013). *Counting Processes and Survival Analysis*, volume 625. John Wiley & Sons.

He, W. and Lawless, J. F. (2005). Bivariate location–scale models for regression analysis, with applications to lifetime data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **67**, 63–78.

Jamal, F., Nasir, M. A., Tahir, M., and Montazeri, N. H. (2017). The odd Burr-III family of distributions. *Journal of Statistics Applications and Probability*, **6**, 105–122.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons.

Lee, C., Famoye, F., and Olumolade, O. (2007). Beta-Weibull distribution: some properties and applications to censored data. *Journal of Modern Applied Statistical Methods*, **6**, 173–186.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1990). Using the jackknife to estimate the variance of regression estimators from repeated measures studies. *Communications in Statistics-Theory and Methods*, **19**, 821–845.

Manly, B. F. (2018). *Randomization, Bootstrap and Monte Carlo Methods in Biology.* chapman and hall/CRC.

McCullagh, P. (2019). *Generalized Linear Models.* Routledge.

Mudholkar, G. S., Srivastava, D. K., and Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, **37**, 436–445.

Ortega, E. M., Paula, G. A., and Bolfarine, H. (2008). Deviance residuals in generalised log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, **78**, 747–764.

Silva, G. O., Ortega, E. M., and Cancho, V. G. (2010). Log-Weibull extended regression model: Estimation, sensitivity and residual analysis. *Statistical Methodology*, **7**, 614–631.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, **77**, 147–160.

Usman, R. M. and Haq, M. A. u. (2019). Some remarks on odd Burr III Weibull distribution. *Annals of Data Science*, **6**, 21–38.

Zamanah, E., Nasiru, S., and Luguterah, A. (2022). Harmonic mixture Weibull-G family of distributions: Properties, regression and applications to medical data. *Computational and Mathematical Methods*, **2022**, 2836545.

ANNEXURE

A. Jackknife method

Jackknifing is a technique that converts the task of estimating a population parameter to that of estimating a population mean. Applying this strategy takes a different way to estimating a mean value. By suggesting another robust estimator for the covariance matrix, Lipsitz *et al.* (1990) made a substantial contribution to the implementation of the jackknife method. This estimator relies on the jackknife method and is valid for analyzing data from repeated measure analyses. This method can be used as an alternative method to estimate population parameters.

Suppose $X_1, X_2, \ldots, X_n$ be a random sample of size n and $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ be the sample mean used to estimate the population mean. Let $\hat{\eta}$ be the estimated parameter vector of η based on each of the n observations and $\hat{\eta}_{-i}$ for $i = 1, 2, \ldots, n$ be the estimated η value after the i^{th} observation was removed from the sample. Hence, the pseudo-values are obtained by;

$$\hat{\eta}_i^* = n\hat{\eta} - (n-1)\hat{\eta}_{-i}, \quad i = 1, 2, \ldots, n.$$

Thus, the jackknife estimator of η is given by,

$$\hat{\eta}^* = \frac{\sum_{i=1}^{n} \hat{\eta}_i^*}{n}.$$

Manly (2018) proposed that an approximate $100(1-\gamma)\%$ confidence interval for each η is provided by $\hat{\eta}^* \pm t_{\frac{\gamma}{2}, n-1} \frac{s}{\sqrt{n}}$, which eliminates bias of order n^{-1}. Here s represents the standard deviation of the pseudo-values and $t_{\frac{\gamma}{2}, n-1}$ is the upper $(1 - \frac{\gamma}{2})$ point of t distribution with $n - 1$ degrees of freedom.

B. Sensitivity analysis: Global influence

The initial technique for performing sensitivity analysis involves global influence through case - deletion as outlined in Cook (1977). Case deletion is a widely utilized technique for assessing the effect of removing the i^{th} record from the dataset. In the context of model (6), the case - deletion process is defined as follows:

$$Y_l = \mathbf{x}_l^T \boldsymbol{\beta} + \sigma Z_l, \quad l = 1, 2, ..., n, \quad l \neq i. \tag{9}$$

The term expressed by "i" describes the actual expression with the exclusion of the i^{th} record. The log-likelihood function for model (9) is represented as $l_{(i)}(\boldsymbol{\eta})$ and let $\hat{\boldsymbol{\eta}}_i = (\hat{c}_i, \hat{k}_i, \hat{\sigma}_i, \hat{\boldsymbol{\beta}}_i^T)^T$represent the corresponding estimate of $\boldsymbol{\eta}$. The basic idea is to evaluate, the difference between $\hat{\boldsymbol{\eta}}_i$ and $\hat{\boldsymbol{\eta}}$ in order to figure out the impact of the i^{th} observation on the maximum likelihood estimate $\hat{\boldsymbol{\eta}} = (\hat{c}, \hat{k}, \hat{\sigma}, \hat{\boldsymbol{\beta}}^{\boldsymbol{T}})^T$. More consideration should be provided to an observation, if its exclusion substantially affects the estimates. Thus, if $\hat{\boldsymbol{\eta}}_i$ is regarded as an influential observation if it differs considerably from $\hat{\boldsymbol{\eta}}$. The generalized Cook distance (GCD), a frequently employed measure to assess the global influence, is defined as,

$$GD_i(\boldsymbol{\eta}) = (\hat{\boldsymbol{\eta}}_i - \hat{\boldsymbol{\eta}})^T \{-\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\eta}})\}(\hat{\boldsymbol{\eta}}_i - \hat{\boldsymbol{\eta}}).$$

Another established measure used to quantify the distinction between $\hat{\boldsymbol{\eta}}_i$ and $\hat{\boldsymbol{\eta}}$ is the likelihood displacement (LD) defined as:

$$LD_i(\boldsymbol{\eta}) = 2\{l(\hat{\boldsymbol{\eta}}) - l(\hat{\boldsymbol{\eta}}_i)\}.$$