RESEARCH ARTICLE

A new one-parameter lifetime distribution and its regression model with applications

M. S. Eliwa[1,2]*, Emrah Altun[3], Ziyad Ali Alhussain[1], Essam A. Ahmed[4,5], Mukhtar M. Salah[1], Hanan Haj Ahmed[6], M. El-Morshedy[7,8]

1 Department of Mathematics, College of Science, Majmaah University, Majmaah, Saudi Arabia, **2** Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt, **3** Department of Mathematics, Bartin University, Bartin, Turkey, **4** Department of Administrative and Financial Sciences, Taibah University, Community College of Khyber, Medina, Saudi Arabia, **5** Department of Mathematics, Sohag University, Sohag, Egypt, **6** Department of Basic Science, Preparatory Year Deanship, King Faisal University, Hofuf, Al-Ahsa, Saudi Arabia, **7** Department of Mathematics and Statistics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia, **8** Department of Mathematics and Statistics, Faculty of Science, Mansoura University, Mansoura, Egypt

* mseliwa@mans.edu.eg

Abstract

Lifetime distributions are an important statistical tools to model the different characteristics of lifetime data sets. The statistical literature contains very sophisticated distributions to analyze these kind of data sets. However, these distributions have many parameters which cause a problem in estimation step. To open a new opportunity in modeling these kind of data sets, we propose a new extension of half-logistic distribution by using the odd Lindley-G family of distributions. The proposed distribution has only one parameter and simple mathematical forms. The statistical properties of the proposed distributions, including complete and incomplete moments, quantile function and Rényi entropy, are studied in detail. The unknown model parameter is estimated by using the different estimation methods, namely, maximum likelihood, least square, weighted least square and Cramer-von Mises. The extensive simulation study is given to compare the finite sample performance of parameter estimation methods based on the complete and progressive Type-II censored samples. Additionally, a new log-location-scale regression model is introduced based on a new distribution. The residual analysis of a new regression model is given comprehensively. To convince the readers in favour of the proposed distribution, three real data sets are analyzed and compared with competitive models. Empirical findings show that the proposed one-parameter lifetime distribution produces better results than the other extensions of half-logistic distribution.

1 Introduction

There are many distributions in the statistics literature. However, each data sings its song. Therefore, finding a suitable probability distribution for each data is an important issue to catch all parts of the song. In the data analysis process, we encounter two types of data. These

are discrete and continuous data sets. It is necessary to decide on the appropriate distribution for each data type. To increase the accuracy in data modeling, the researchers have introduced flexible distributions for both discrete and continuous cases. However, the aim of the presented manuscript complies on the continuous probability distributions. [1–17] have introduced flexible continuous distri-butions based on the T-X family of [18]. In these studies, researchers have generalized the baseline distributions by adding one or more additional shape parameters to increase the flexibility of the baseline distributions. The generalization of the exponential, Weibull, generalized half-normal and Lindley distributions have gained attention by researchers because of their importance in lifetime and reliability modeling. Apart from these distributions, the Half-logistic (HL) is also an important distribution for reliability analysis and increased its popularity in recent years. The detail information on HL distribution can be found in [19]. The probability density function (pdf) of the HL distribution is

$$g(x) = \frac{2e^{-x}}{\left(1 + e^{-x}\right)^2}; \quad x > 0. \tag{1}$$

The corresponding cumulative distribution function (cdf) to (1) is

$$G(x) = \frac{1 - e^{-x}}{1 + e^{-x}}; \quad x > 0. \tag{2}$$

The several generalizations of (1) have been introduced such as exponential half-logistic additive model by [20], extended half-logistic distribution by [21], type-I half-logistic distribution by [22], exponentiated half-logistic-G by [23], transmuted half-logistic distribution by [24] and half-logistic inverse Rayleigh by [25].

In this study, we use the odd Lindley-G family of [8] to introduce a new generalization of HL distribution. The reason for the use of odd Lindley-G (OLi-G) family is that it has only one additional shape parameter. It means that the proposed distribution will also have only one parameter. Increasing the parameter space of the probability distributions causes a problem in the estimation step. The parsimony rule says the best model is a model which requires less assumptions and parameters. Therefore, it is more preferred to study with less parameters and less complexity. However, more complex models are still needed to model the different characteristics of the data sets. The cdf and pdf of the OLi-G family are given, respectively, by

$$F(x; \lambda, \eta) = 1 - \frac{\lambda + \bar{G}(x; \eta)}{(1 + \lambda)\bar{G}(x; \eta)} e^{-\lambda \frac{G(x;\eta)}{\bar{G}(x;\eta)}}; \quad x \geq 0, \tag{3}$$

$$f(x; \lambda, \eta) = \frac{\lambda^2 g(x; \eta)}{(1 + \lambda)\bar{G}(x; \eta)^3} e^{-\lambda \frac{G(x;\eta)}{\bar{G}(x;\eta)}}; \quad x \geq 0, \tag{4}$$

where $\bar{G}(x; \eta) = 1 - G(x; \eta)$ represents the reliability function. As seen from (3) and (4), the OLi-G family has simple forms for its pdf and cdf. This property of the OLi-G family has attracted us to use it for a new generalization of the HL distribution.

The remaining parts of the presented study are arranged as follows. In Section 2, we define the OLiHL distribution and studied its important statistical properties. The parameter estimation problem of the OLiHL distribution is discussed in Section 3 with four different estimation methods. Section 4 deals with a new location-scale regression model based on the OLiHL distribution for modeling extremely left-skewed lifetime variables with covariates. In Section 5, the finite sample performance of the different estimation methods is compared with the simulation study. In Section 6, two real data sets are analyzed with the proposed models. The discussions on the empirical results are given in Section 7. The conclusions are given in Section 8.

2 The OLiHL distribution

Now, we introduce a new generalization of the HL distribution, OLiHL distribution, by inserting (2) in (3). Let the random variable X follows an OLiHL distribution if its cdf is given by

$$F(x; \lambda) = 1 - \frac{\lambda e^x + \lambda + 2}{2(\lambda + 1)} e^{-\frac{\lambda}{2}(e^x - 1)}; \quad x > 0, \tag{5}$$

where $\lambda > 0$. The pdf of the OLiHL distribution is

$$f(x; \lambda) = \frac{\lambda^2 (1 + e^x)}{4(1 + \lambda)} e^{-\frac{\lambda}{2}(e^x - 1) + x}; \quad x > 0. \tag{6}$$

The corresponding hazard rate function (hrf) to (6) is

$$h(x; \lambda) = \frac{\lambda^2 (1 + e^x)}{2[\lambda(1 + e^{-x}) + 2e^{-x}]}; \quad x > 0. \tag{7}$$

The possible shapes of the pdf and hrf are displayed in Fig 1 which shows that the OLiHL distribution could be a proper distribution for right-skewed and unimodal data sets with increasing hrf shapes.

Following the work of [8], the physical interpretation of the OLiHL distribution can be given as follows. Let the random variable X represents lifetimes of individuals following the HL distribution, given in 1. Consider that we are interested to model the odds that an individual dies before a given time X which is given by $F(x)/(1 − F(x))$ where $F(x)$ is the cdf of the HL distribution, given in 2. Let consider that we requires to model the randomness of the odds by the random variable Y, follows the Lindley distribution. Then, we can write

$$\begin{aligned} \Pr(Y < y) &= F_Y\left(\frac{F(x)}{1 - F(x)}\right) = F_Y\left(\frac{1 - e^{-x}}{2e^{-x}}\right) \\ &= 1 - \frac{\lambda e^x + \lambda + 2}{2(\lambda + 1)} e^{-\frac{\lambda}{2}(e^x - 1)} \end{aligned} \tag{8}$$

which is identical to 5.

The statistical properties of the OLiHL distribution such as raw and central moments, incomplete moments, generating functions can be obtained following the results of [8]. [8] introduced the general representation of the OLi-G family for all baseline distributions. Therefore, using the HL distribution as a baseline distribution of the OLi-G family, one can easily obtain the required expansions for the moments and generating functions of the OLiHL distribution. Since these expansions are accessible in the work of [8], we omit them in this study. However, the numerical values of the mean, variance, skewness, and kurtosis measures of the OLiHL distribution are computed and displayed in Fig 2 which shows that the mean and variance of the OLiHL distribution are decreasing function of the parameter λ. The OLiHL distribution can be left or right-skewed. The distribution is nearly symmetric when the parameter $\lambda \cong 0.81$. The OLiHL distribution has also playtokurtic and leptokurtic shapes based on its kurtosis values.

The quantile function (qf) is an important function to generate random variables from a specific continuous probability distribution. The qf is a solution of $F(x) = u$ equation for x,
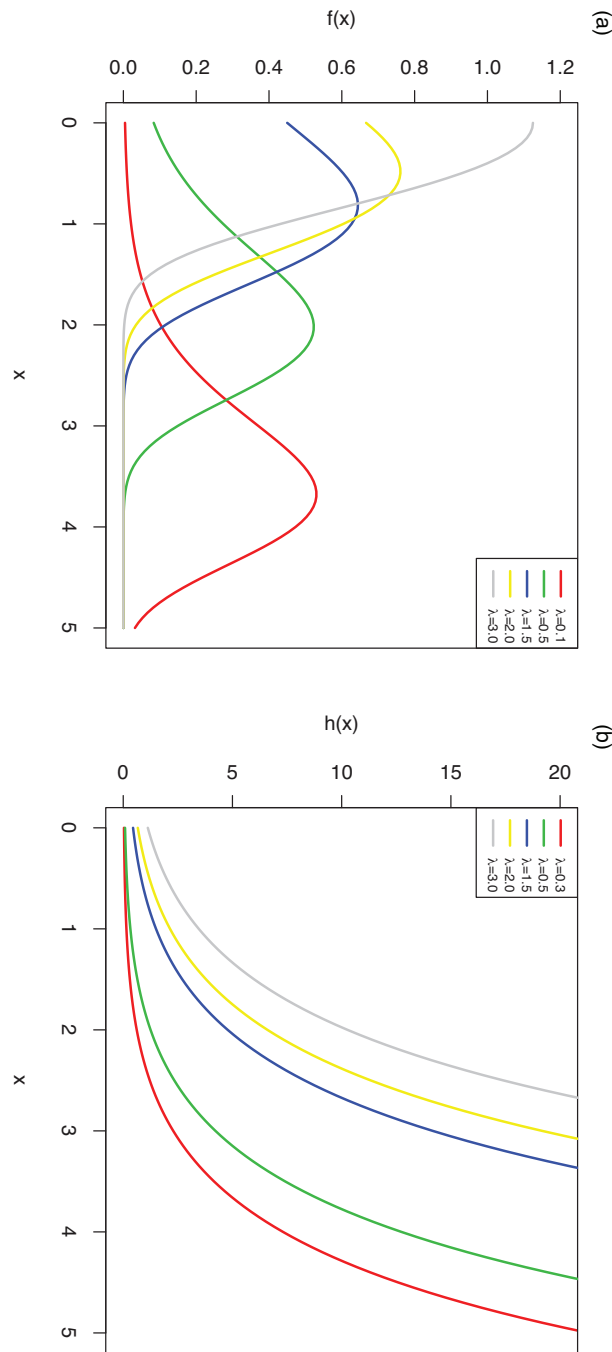
Fig 1. The pdf (left) and hrf (right) plots of the OLiHL distribution.

then $F(u)^{-1} = x$ represents the qf which is denoted as $Q(u)$. The qf of the OLiHL distribution is

$$Q(u) = \ln\left\{\frac{-2}{\lambda}\left[1 + W_{-1}\big((1+\lambda)(u-1)e^{-(1+\lambda)}\big)\right] - 1\right\}, \tag{9}$$

where $u \in (0, 1)$ and W_{-1} is the negative branch of the Lambert-W function. The random variables from OLiHL distribution can be generated by using the qf in (9). To do this, the following algorithm steps can be implemented.
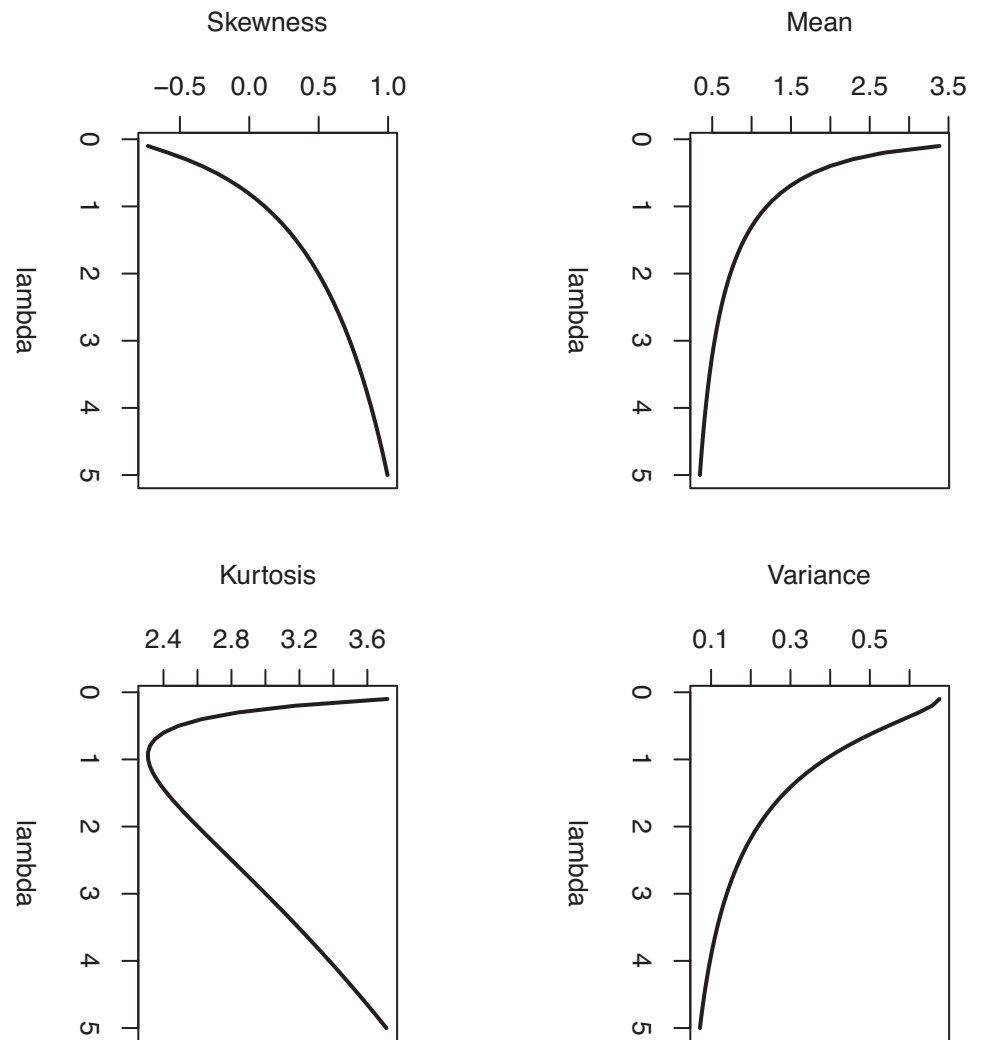
Fig 2. The mean, variance, skewness and kurtosis measures of the OLiHL distribution.

1. Set the parameter λ.

2. Generate $u \sim U(0, 1)$.

3. Using the generated u, calculate

$$Q(u) = \ln\left\{\frac{-2}{\lambda}\left[1 + W_{-1}\big((1 + \lambda)(u - 1)e^{-(1+\lambda)}\big)\right] - 1\right\} \tag{10}$$

4. Repeat the steps 2 and 3 N times.

3 Different estimation methods

The parameter estimation problem of the OLiHL distribution is discussed in detail. The four-parameter estimation methods are used to estimate the unknown parameter of the OLiHL

distribution. The rest of this section is devoted to the mathematical framework of these estimation methods.

3.1 Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ come from the OLiHL distribution unknown parameter λ. The maximum likelihood estimation (MLE) of λ, say $\hat{\lambda}_{MLE}$, is obtained by maximizing the following log-likelihood equation

$$\ell(\lambda) = \sum_{i=1}^{n} \log f(x_i; \lambda)$$

$$= 2n \ln \lambda - n \ln(4 + 4\lambda) + \sum_{i=1}^{n} \ln(1 + e^{x_i}) \tag{11}$$

$$-\frac{\lambda}{2} \sum_{i=1}^{n} (e^{x_i} - 1) + \sum_{i=1}^{n} x_i.$$

Alternatively, the direct maximization of (11) is equal to solve the first derivative of (11) for the parameter λ for zero, which is given by

$$\frac{n(2 + \hat{\lambda})}{\hat{\lambda}(1 + \hat{\lambda})} - \frac{1}{2} \sum_{i=1}^{n} (e^{x_i} - 1) = 0.$$

Since there is no explicit solution of this equation for λ, one should use iterative methods to solve it. More information on MLE method can be found in [26, 27].

3.2 Least square and weighted least square estimations

The least-square estimation (LSE) and weighted LSE (WLSE) methods are based on the minimization of the distance between the empirical and theoretical distribution functions. Assume that $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ represents the ordered sample of a random sample with known cdf, $F(\cdot)$. Let $F(X_{(j)})$ represents the jth order statistics from standard uniform distribution, $U(0, 1)$. It is well-known that the jth order statistics of $U(0, 1)$ is distributed as Beta($j, n - j - 1$) (see [28]). So, we have

$$E\left[F(X_{(j)})\right] = \frac{j}{n + 1}, \tag{12}$$

$$\mathrm{Var}\left[F(X_{(j)})\right] = \frac{j(n - j + 1)}{(n + 1)^2 (n + 2)}. \tag{13}$$

The LSE of the parameter of the OLiHL distribution is obtained by minimizing the following function.

$$\mathrm{LSE}(\lambda) = \sum_{j=1}^{n} \left[F\left(X_{(j)}|\lambda\right) - \frac{j}{n + 1}\right]^2, \tag{14}$$

where $F(X_{(j)}|\lambda)$ is the cdf of the OLiHL distribution, given in (5). Substituting $F(X_{(j)}|\lambda)$ with (5)

in (14), we have

$$\sum_{j=1}^{n} \left(1 - \frac{\lambda e^{x(j)} + \lambda + 2}{2(\lambda + 1)} e^{-\frac{\lambda}{2}\left(e^{x(j)} - 1\right)} - \frac{j}{n+1} \right). \tag{15}$$

The WLSE of the parameter of the OLiHL distribution is obtained by minimizing the following function.

$$\text{WLSE}(\lambda) = \sum_{j=1}^{n} w_j \left[F(X_{(j)}|\lambda) - \frac{j}{n+1} \right]^2. \tag{16}$$

where $w_j = \frac{1}{\text{Var}[F(X_j)]}$. Substituting $F(X_{(j)}|\lambda)$ with (5) in (16), we have

$$\sum_{j=1}^{n} \frac{(n+1)^2 (n+2)}{j(n-j+1)}$$

$$\times \left(1 - \frac{\lambda e^{x(j)} + \lambda + 2}{2(\lambda + 1)} e^{-\frac{\lambda}{2}\left(e^{x(j)} - 1\right)} - \frac{j}{n+1} \right) \tag{17}$$

3.3 Cramér-von Mises minimum distance estimation

Cramér-von Mises estimation (CVME) method is also based on the minimization of the distance between the empirical and theoretical distribution functions. The CVME and WLSE use different weighting functions and the biases CVME is less than those of WLSE. The WLSE of the parameter λ is obtained by minimizing the below function for the parameter λ.

$$\text{CVME}(\lambda) = \frac{1}{12n} + \sum_{j=1}^{n} \left[F\left(x_{(j)}|\lambda \right) - \frac{2j-1}{2n} \right]^2. \tag{18}$$

4 The log-OLiHL regression model

The log-location-scale regression models are popular models to analyze the censored response variable with some covariates. In the last decade, researchers have introduced flexible location-scale regression models to analyze the different characteristics of the data sets. The important papers on location-scale can be cited as follows: log-generalized odd log-logistic-Weibull regression model by [29], log-odd log-logistic Burr XII regression model by [30], log-odd log-logistic generalized half-normal regression model by [31]. These types of regression models were introduced based on the $Y = \log (X)$ transformation and suitable re-parametrization on the baseline distribution.

Now, we adopt this approach to the OLiHL distribution. Let the random variable X follows an OLiHL distribution with the parameter λ. Using the transformation $Y = \log(X)$ and adding location and scale parameters, we have

$$\begin{aligned} f(y; \lambda, \mu, \sigma) &= \frac{\lambda^2}{\sigma(4\lambda + 4)} \left[\exp\left(\exp\left[\frac{y-\mu}{\sigma} \right] \right) + 1 \right] \\ &\times \exp\left[\frac{y-\mu}{\sigma} + \exp\left[\frac{y-\mu}{\sigma} \right] - \frac{\lambda\left(\exp\left(\exp\left[\frac{y-\mu}{\sigma} \right] \right) - 1 \right)}{2} \right], \end{aligned} \tag{19}$$

where $y \in \Re, \mu \in \Re$ and $\sigma > 0$. The parameters μ and σ are the location and scale parameters of the LOLiHL distribution. Hereafter, the pdf in Eq (19) is called as log-OLiHL (LOLiHL)

distribution. The corresponding cdf and survival function (sf) to Eq (19) are given, respectively, by

$$S(y) = \frac{\lambda + \lambda \exp\left(\exp\left[\frac{y-\mu}{\sigma}\right]\right) + 2}{2\lambda + 2}$$

$$\times \exp\left(-\frac{\lambda}{2}\left(\exp\left(\exp\left[\frac{y-\mu}{\sigma}\right]\right) + 1\right)\right). \tag{20}$$

The pdf shapes of the LOLiHL distribution are displayed in Fig 3. As seen from these figures, the LOLiHL distribution has increasing failure rate and can be used to analyze the left-skewed lifetimes.

Consider the following regression model,

$$y_i = \mathbf{x_i^T}\boldsymbol{\beta} + \sigma z_i, \tag{21}$$

where the response variable y_i has the density given in Eq (19). The covariates are linked to location of y_i with identity link function $\mu_i = \mathbf{x_i^T}\beta$ where $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N})^T$ is the model matrix consists of the observations of independent variables and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ is the unknown regression coefficients. Let the random sample $y_1, y_2, \ldots, y_n$ follow a LOLiHL distribution and the response variable is defined as $y_i = \min(x_i, c_i)$ where c_i is the censoring time and x_i is the observed lifetime. Assume that the censoring times and lifetimes are independent.

Let F and C are the sets representing the observed lifetimes and censoring times. The general formulation of the log-likelihood function for the model given in (21) is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i \in F}\log\left[f(y_i)\right] + \sum_{i \in C}\log\left[S(y_i)\right] \tag{22}$$

where $\boldsymbol{\theta}$ is the unknown parameter vector, $\log[f(y_i)]$ and $\log[S(y_i)]$ are given in Eqs (19) and (20), respectively. Inserting Eqs (19) and (20) in Eq (22), we have following the log-likelihood function for the LOLiHL regression model

$$\ell(\boldsymbol{\theta}) = r\left(1 + \ln\left(\frac{\lambda^2}{\sigma(4\lambda+4)}\right)\right)$$

$$+ 2\sum_{i \in F}\left(\exp\left[\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right]\right) + \sum_{i \in F}\left(\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right)$$

$$- \sum_{i \in F} 2^{-1}\lambda\left(\exp\left(\exp\left[\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right]\right) - 1\right)$$

$$+ \sum_{i \in C}\ln\left(\frac{\lambda + \lambda\exp\left(\exp\left[\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right]\right) + 2}{2\lambda + 2}\right)$$

$$- \sum_{i \in C}\frac{\lambda}{2}\left(\exp\left(\exp\left[\frac{y_i - \mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right]\right) + 1\right),$$

where r is the number of uncensored observations and $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}, \sigma)$ is the unknown parameter vector. The unknown parameter vector, $\boldsymbol{\theta}$, is estimated by using the MLE method. The minus of the log-likelihood function is minimized by **optim** function of R software. The observed
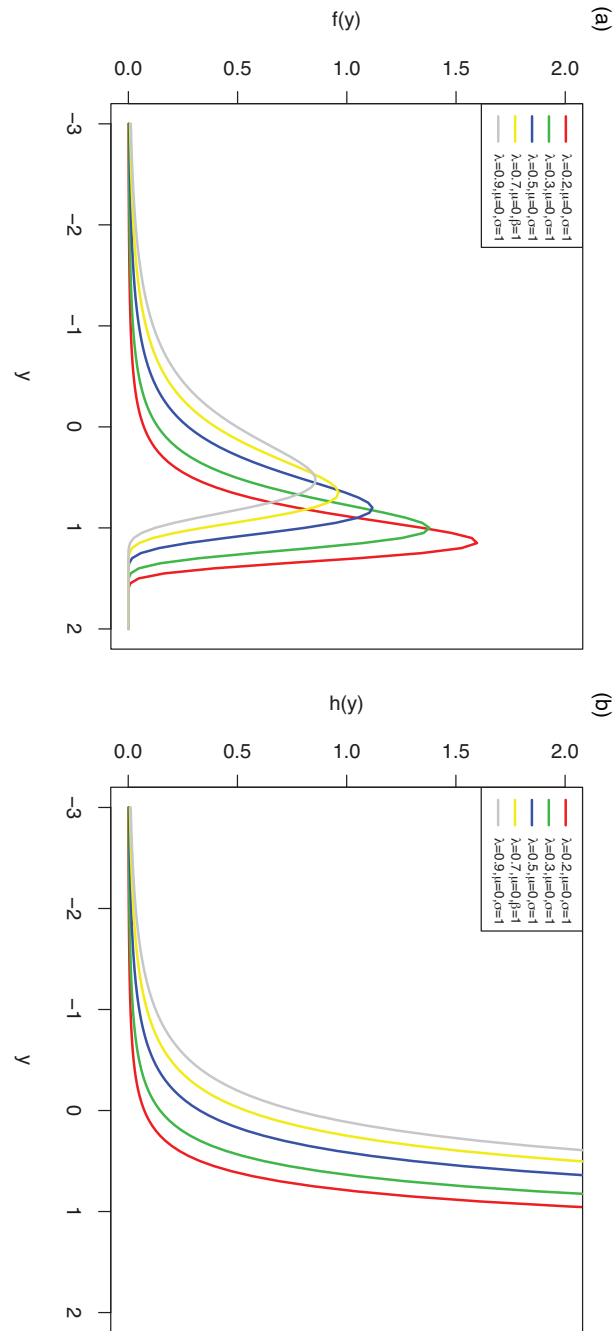
Fig 3. The pdf (left) and hrf (right) plots of LOLiHL distribution.

information matrix evaluated at $\hat{\boldsymbol{\theta}}$ is used to obtain corresponding standard errors to construct asymptotic confidence intervals.

4.1 Residual analysis

Residual analysis is an important step of any regression analysis to check the sufficiency of the fitted model. If the fitted model is accurate for the data used, the residuals have to meet the

distributional assumptions. Here, we used two kinds of residuals: modified deviance and martingale residuals. The martingale residuals under the LOLiHL regression model are given by

$$
r_{M_i} = \begin{cases} 1 + \ln\left(\dfrac{\dfrac{\lambda + \lambda\exp(u_i) + 2}{2\lambda + 2}}{\times \exp\left(-\dfrac{\lambda}{2}(\exp(u_i) + 1)\right)} \right), & \text{if } i \in F \\[2em] \ln\left(\dfrac{\dfrac{\lambda + \lambda\exp(u_i) + 2}{2\lambda + 2}}{\times \exp\left(-\dfrac{\lambda}{2}(\exp(u_i) + 1)\right)} \right), & \text{if } i \in C \end{cases}
$$

where $u_i = \exp\left((y_i - \mathbf{x}_\mathbf{i}^\top \boldsymbol{\beta})/\sigma\right)$. The modified deviance residual of LOLiHL model is

$$
r_{D_i} = \begin{cases} \text{sign}(r_{M_i})\{ -2[r_{M_i} + \log(1 - r_{M_i})]\}^{1/2} & , \text{ if } i \in F \\[1em] \text{sign}(r_{M_i})\{ -2r_{M_i}\}^{1/2} & \text{ if } i \in C \end{cases}
$$

where r_{M_i} is the martingale residual.

5 Simulation results

The relative performance of the MLE, LSE, WLSE, and CVME methods are discussed. The below simulation steps are carried out.

1. Set the simulation replication $N = 1000$ and the parameter value $\lambda = 0.1$

2. Generate random samples from OLiHL with sample sizes $n = 20, 25, 30, \ldots, 100$

3. For each generated sample, obtain the MLE, LSE, WLSE and CVME of the parameter λ, say $\hat{\lambda}_j$ for $j = 1, 2, \ldots, 1000$.

4. Using the estimated value of λ and true parameter value, calculate the biases and means square errors (MSEs) for each parameter estimation methods by using the below equations.

$$
Bias_\lambda(n) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\lambda}_i - \lambda\right)
$$

$$
MSE_\lambda(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \lambda)^2
$$

where $i = 1, 2, 3, \ldots, N$.

The empirical results are graphically summarized in Fig 4. As seen from these results, the estimated biases and MSEs approach to zero for all parameter estimation methods. There are no clear differences between the performance of the estimation methods. However, the LSE method approach to desired values of the biases and MSEs faster than other estimation methods.
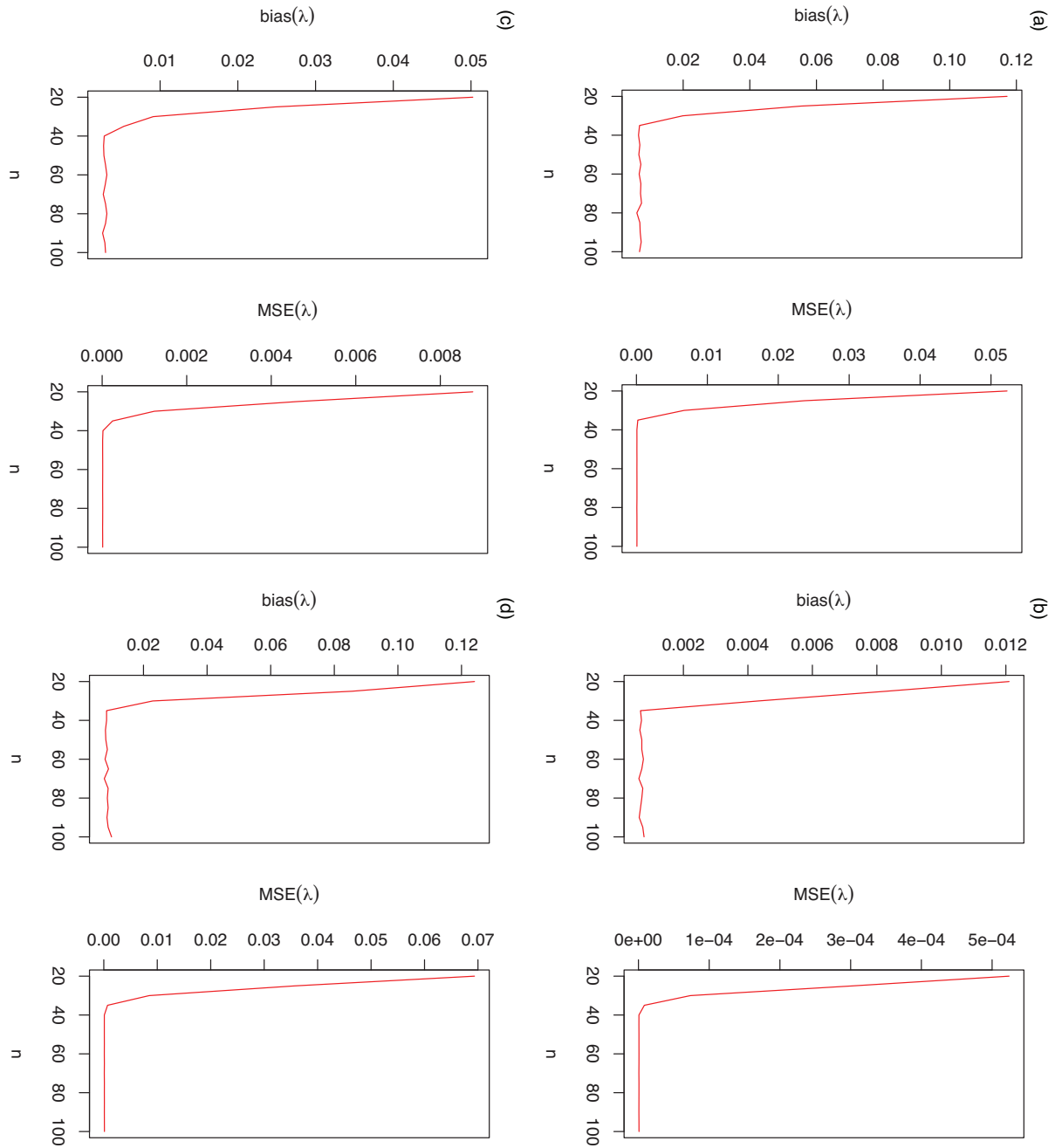
Fig 4. The bias and MSE for λ = 0.1.

6 Empirical results

Two data sets are considered to show the flexibility of the OLiHL distribution against the several competitive models. We make the computational codes available at https://github.com/emrahaltun/Computational-codes-of-OLiHL-distribution.git.

6.1 Carbon fibers data

The fitting performance of the OLiHL distribution is compared with the below competitive models.

1. Exponentiated HL (EHL)

$$F_{\text{EHL}}(x;\lambda) = \left(\frac{1-e^{-x}}{1+e^{-x}}\right)^{\lambda}; \ x, \lambda > 0,$$

2. Generalized HL (GHL)

$$F_{\text{GHL}}(x;\lambda) = 1 - \left(\frac{2e^{-x}}{1+e^{-x}}\right)^{\lambda}; \ x, \lambda > 0,$$

3. Lindley (Li)

$$F_{\text{Li}}(x;\lambda) = 1 - \left(1 + \frac{\lambda x}{1+\lambda}\right)e^{-\lambda x}; \ x, \lambda > 0,$$

4. Inverse Lindley (ILi)

$$F_{\text{ILi}}(x;\lambda) = \left(1 + \frac{\lambda}{(1+\lambda)x}\right)e^{-\frac{\lambda}{x}}; \ x, \lambda > 0,$$

5. Transmuted HL (THL)

$$F_{THL}(x;\lambda) = \frac{(e^x - 1)(1 + 2\lambda + e^x)}{(1+e^x)^2}, x > 0, \tag{23}$$

$$|\lambda| < 1,$$

6. Exponential (Exp)

$$F_{Exp}(x;\lambda) = 1 - e^{-\lambda x}, x > 0, \lambda > 0. \tag{24}$$

We use the results of the Kolmogorov-Smirnov test (KS) with its p-value, Anderson-Darling ($A^{\star}$) and Cramér-von Mises ($W^{\star}$) as well as estimated $-\ell$ to decide the best-fitted model for the data used. The lowest value of the KS, $A^{\star}$ and $W^{\star}$ test statistics and the lowest value of $-\ell$ show the best-fitted model. The data set contains the breaking stress of carbon fibers. The number of observations is $n = 66$ and this data was reported by [32]. Table 1 contains the estimated parameter values of the all fitted model with asymptotic standard errors (SEs) as well as the goodness of fit statistics. The results in Table 1 reveal that the OLiHL could be chosen as

Table 1. The estimated parameters of the fitted models with goodness-of-fit tests.

Distributions	Estimated parameters	SEs	$-\ell$	A^*	W^*	KS	p-value
OLiHL	0.168	0.015	89.925	0.851	0.116	0.110	0.407
EHL	5.067	0.624	93.703	1.534	0.283	0.136	0.172
GHL	0.464	0.057	122.359	1.199	0.223	0.315	<0.001
THL	5.149×10^{-8}	1.046×10^{-5}	147.855	1.086	0.203	0.580	<0.001
Li	0.590	0.053	122.384	1.148	0.214	0.298	<0.001
ILi	2.891	0.296	134.669	4.070	0.721	0.370	<0.001
Exp	0.362	0.045	132.994	1.334	0.246	0.358	<0.001

Table 2. The results of the LSE, WLSE and CVME methods for the breaking stress of carbon fibers.

Statistic ↓ Method →	LSE	WLSE	CVME
λ	0.1909	0.1902	0.1911
KS	0.064	0.064	0.063
p-value	0.950	0.952	0.949

the best model for the data used, since it has the lowest values of the goodness-of-fit statistics. The asymptotic confidence interval of the parameter λ of the OLiHL distribution is (0.139, 197).

Table 2 shows various estimation methods of the OLiHL parameter for the breaking stress of carbonfibers. The test statistics value of KS tests of the LSE, WLSE, and CVME methods are smaller than those of the MLE method for the OLiHL distribution. This result shows the fact that LSE, WLSE, and CVME methods could be more appropriate estimation methods than MLE for this data set. As mentioned in the simulation study, the LSE is a more appropriate method especially for small sample sizes which is consistent with the results obtained in this application.

Fig 5(top) displays the fitted pdfs of the competitive models on the histogram of the data. As seen from these figures, the OLiHL distribution is the best model to describe the characteristics of the modelled data set. Fig 5(bottom) displays the fitted functions of the OLiHL distribution such as fitted pdf, cdf and survival functions with Kaplan-Meier (KM) curve as well as corresponding probability-probability (PP) plot. These figures also reveal that the OLiHL distribution provides superior fit to the modelled data set.

Fig 6 displays the PP plots of the OLiHL distribution obtained under the LSE, WLSE and CVME estimation methods. These estimation methods produce similar results since they are all based on the minimization of the distance between the empirical and theoretical distribution functions.
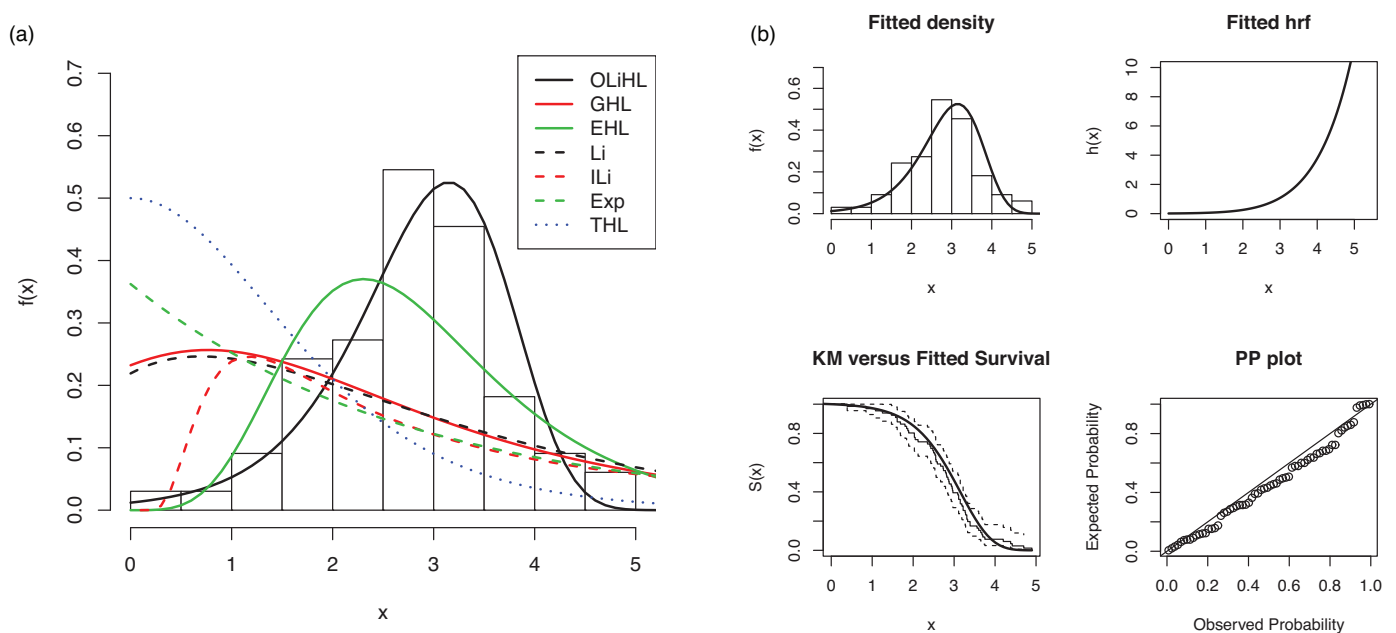

Fig 5. The fitted pdfs of the competitive models (top panel) and estimated functions of the OLiHL (bottom panel) distribution.
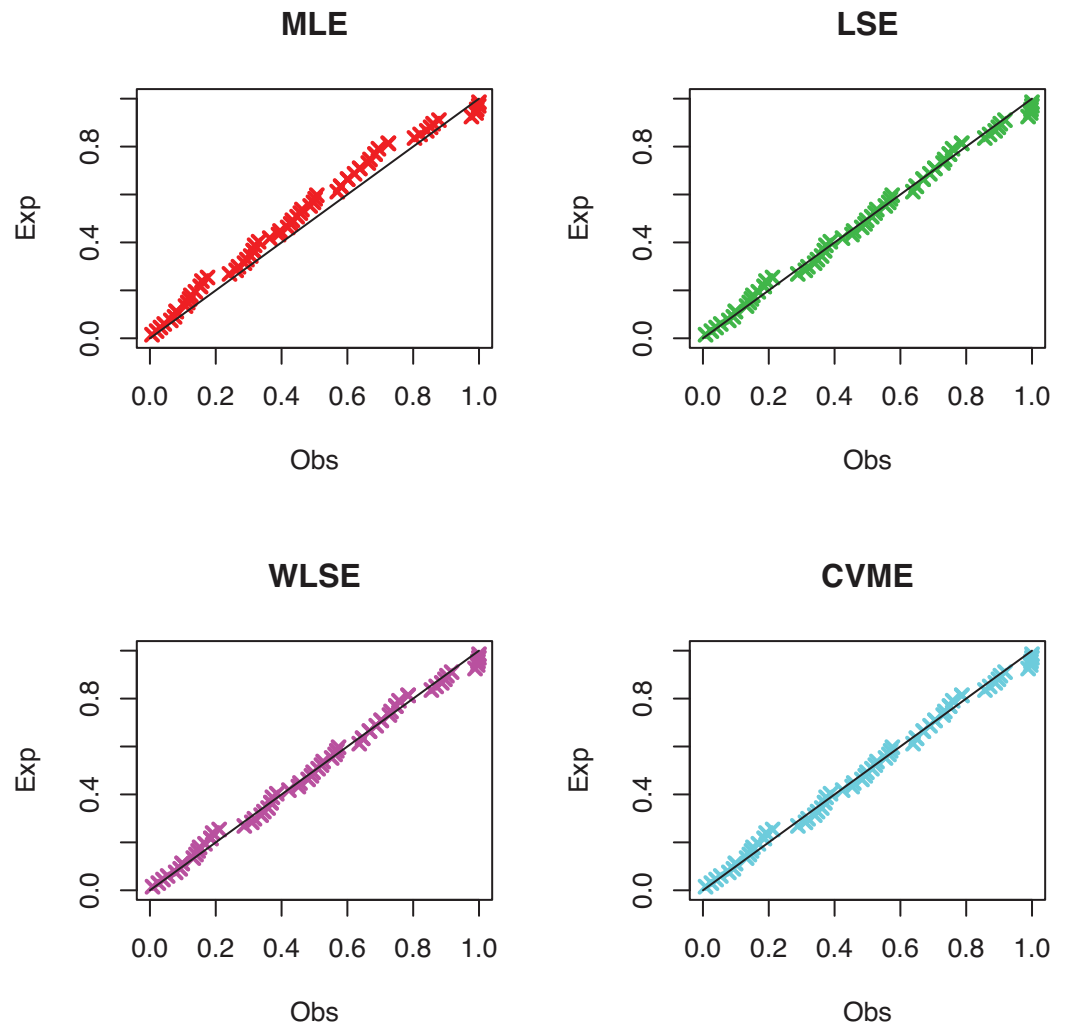
MLE

LSE

WLSE

CVME

Fig 6. The PP plots of OLiHL distribution obtained under LSE, WLSE and CVME methods for the carbon fibers data set.

Finally, the total test time (TTT) plot, introduced by [33], is displayed in Fig 7 to see the empirical behavior of the hrf function. The TTT plot shows that the used data has increasing hrf which means that the OLiHL distribution can be used to model this data.

6.2 HIV+ data

We compare the performance of the LOLiHL regression model with log-exponential (LE) and log-Burr-Hatke-exponential (LBHE) regression models (see, [34], for details of LBHE and LE regression models). The modeled data set consists of 100 individuals having HIV+. The detailed information on the data set can be obtained from **Bolstad2** package of R software. The used data set was modeled by [34]. The depended variable, y_i—observed survival time (in months) with censoring indicator $cens_i$ (0 = alive, 1 = death), is modeled with two covariates: the history of drug use, x_{i1}(1 = yes, 0 = no) and the ages of patients, x_{i2}. The following model is fitted by LOLiHL, LBHE, and LE regression models.

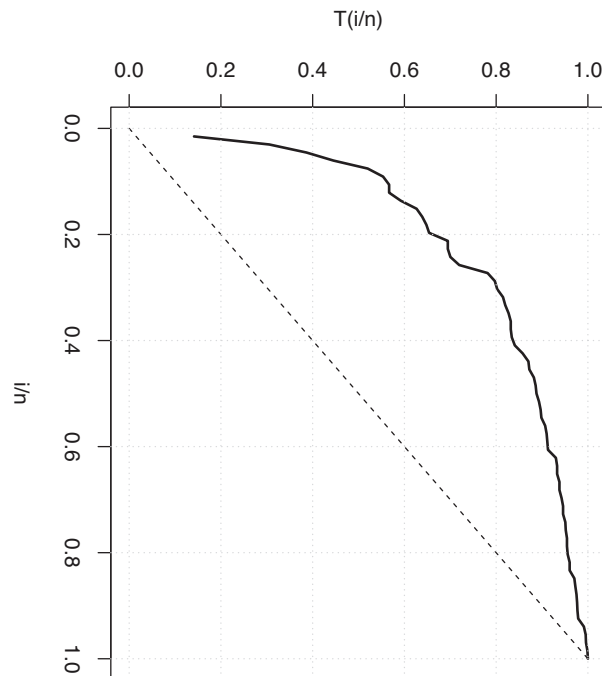$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma z_i.$$

Fig 7. The TTT plots of the carbon fibers data set.

The above regression model is fitted by MLE and estimated parameter value, standard errors, the negative value of log-likelihood, and Akaike Information Criteria (AIC) are reported in Table 3. The AIC statistics is widely used to decide the best model among other competitive models (see [35, 36]). More information on the AIC statistics can be found in [37]. Since the LOLiHL regression model has the lowest value of AIC statistic, we conclude that the LOLiHL regression model produces a better fit than the other two regression model: LBHE and LE regression models. The regression parameters β_0, β_1 and β_2 are found statistically significant 1% level. According to the estimated regression parameters, the individuals having drug use have lower lifetimes than non-drug use individuals. Moreover, when the ages of individuals increase, the lifetimes decrease.

Fig 8 displays the results of the residual analysis for the LOLiHL regression model. As seen from Fig 8, there is no observation to be considered as possible outliers which reveals that the LOLiHL regression model provides an adequate fit to the used data set.

Table 3. The estimated parameters of regression models and AIC statistic.

| Parameters	L-E			LBHE			LOLiHL		
	Estimates	Sth. Errors	p values	Estimates	Sth. Errors	p values	Estimates	Sth. Errors	p values
λ	1.599	13.783	-	1.508	13.659	-	17.490	18.132	-
σ	0.839	0.072	-	0.778	0.067	-	0.991	0.114	-
β_0	6.542	7.256	0.367	6.883	7.064	0.330	2.090	0.158	<0.001
β_1	-0.091	0.014	<0.001	-0.091	0.014	<0.001	-0.015	0.004	<0.001
β_2	-1.049	0.189	<0.001	-1.021	0.193	<0.001	-0.306	0.087	<0.001
$-\ell$	128.502			128.059			125.527		
AIC	267.005			266.117			261.055		
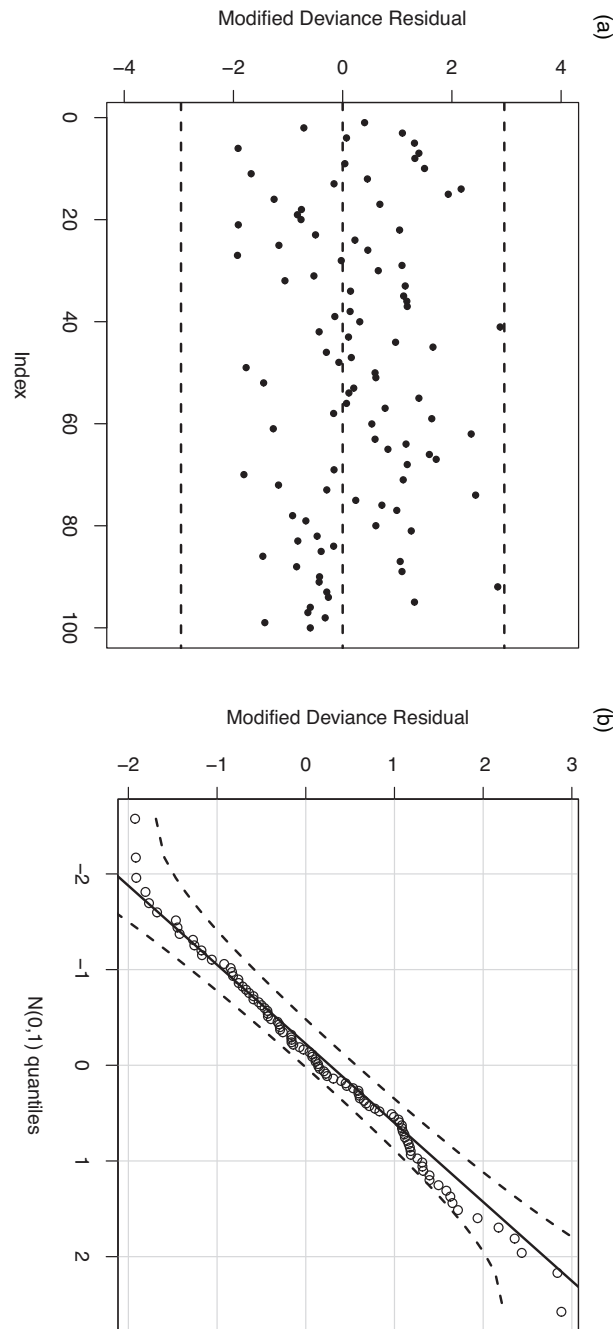
Fig 8. The plot of the modified deviance residual (a) and Q-Q plot of the modified deviance residual.

7 Discussions

The performance of the OLiHL distribution is compared with one-parameter competitive models. The OLiHL distribution has achieved to exhibit better modeling ability than Li, ILi, GHL and EHL distributions. Because the parameters of the Li and ILi distributions treat as location parameters. The location parameters do not affect the flexibility of the distribution. The parameter of the EHL distribution controls the shape of the distribution. However, the

odd Lindley-G family provides more flexibility than the Exponentiated-G family. Therefore, OLiHL distribution works better than other competitive models.

8 Conclusion and future research

This study proposes a new one-parameter lifetime distribution, called as odd Lindley half-logistic distribution, shortly OLiHL distribution. The advantage of the OLiHL distribution is that it has only one parameter and this parameter controls the shape of the distribution which can be left-skewed, right-skewed, or nearly symmetric. This flexibility of the OLiHL distribution provides an opportunity to data scientists in modeling the different types of data sets. Additionally, the LOLiHL regression model will be a useful choice for practitioners studying in the field of survival analysis. As a future work of this study, we plan to develop the heteroscedastic regression model of the OLiHL distribution with its influence diagnostics and residuals analysis. We hope that the OLiHL distribution will find a wider application area in the near future.

Author Contributions

Conceptualization: M. S. Eliwa.

Data curation: Hanan Haj Ahmed.

Formal analysis: Essam A. Ahmed, Hanan Haj Ahmed.

Investigation: Ziyad Ali Alhussain.

Methodology: M. S. Eliwa, M. El-Morshedy.

Project administration: Mukhtar M. Salah.

Resources: Ziyad Ali Alhussain.

Software: Emrah Altun, Essam A. Ahmed, Mukhtar M. Salah, M. El-Morshedy.

Validation: Essam A. Ahmed.

Writing – original draft: M. S. Eliwa, M. El-Morshedy.

Writing – review & editing: Emrah Altun.

References

1. Bourguignon M., Silva R. B., Cordeiro G. M., (2014). The Weibull-G family of probability distributions. Journal of data science, 12, 53–68.

2. Alizadeh M., Emadi M., Doostparast CGBTYDHM., Cordeiro G. M., Ortega E. M.M., and Pescim R. R., (2015). A new family of distributions: the Kumaraswamy odd log-logistic, properties and applications. Hacettepe University Bulletin of natural sciences and engineering series B: Mathematics and statistics. https://doi.org/10.15672/HJMS.2014418153

3. Alizadeh M., Ramires T. G., MirMostafaee S. M. T. K., Samizadeh M. and Ortega E. M. (2019). A new useful four-parameter extension of the Gumbel distribution: Properties, regression model and applications using the GAMLSS framework. Communications in Statistics-Simulation and Computation, 48(6), 1746–1767. https://doi.org/10.1080/03610918.2018.1423691

4. Alizadeh M., Afify A. Z., Eliwa M. S. and Ali S. (2020). The odd log-logistic Lindley-G family of distributions: properties, Bayesian and non-Bayesian estimation with applications. Computational Statistics, 35(1), 281–308. https://doi.org/10.1007/s00180-019-00932-9

5. Tahir M. H., Zubair M., Mansoor M., Cordeiro G. M., Alizadeh M., and Hamedani G. G., (2016). A new Weibull-G family of distributions. Hacettepe journal of mathematics and statistics, 45(2), 629–647.

6. Eliwa M. S., El-Morshedy M., and Ali S. (2020). Exponentiated odd Chen-G family of distributions: statistical properties, Bayesian and non-Bayesian estimation with applications. Journal of Applied Statistics. https://doi.org/10.1080/02664763.2020.1783520

7. Badr M. M. (2019). Beta Generalized Exponentiated Frechet Distribution with Applications. Open Physics, 17(1), 687–697. https://doi.org/10.1515/phys-2019-0071

8. Silva F. S. G., Percontini A., de Brito E., Ramos M. W., Venancio R., and Cordeiro G. M., (2017). The odd Lindley-G family of distributions. Austrian journal of statistics, 46(1), 65–87. https://doi.org/10.17713/ajs.v46i1.222

9. El-Morshedy M., and Eliwa M. S., (2019). The odd flexible Weibull-H family of distributions: properties and estimation with applications to complete and upper record data. Filomat, 33(9), 2635–2652. https://doi.org/10.2298/FIL1909635E

10. Korkmaz M. C., Altun E., Yousof H. M. and Hamedani G. G. (2019a). The odd power Lindley generator of probability distributions: properties, characterizations and regression modeling. International Journal of Statistics and Probability, 8(2), 70–89. https://doi.org/10.5539/ijsp.v8n2p70

11. Korkmaz M. C., Altun E., Alizadeh M. and Yousof H. M. (2019b). A new flexible lifetime model with log-location regression modeling, properties and applications. Journal of Statistics and Management Systems, 22(5), 871–891. https://doi.org/10.1080/09720510.2019.1572980

12. Nofal Z. M. and El Gebaly Y. M. (2017). The Generalized Transmuted Weibull Distribution for Lifetime Data. Pakistan Journal of Statistics and Operation Research, 13(2), 355–378. https://doi.org/10.18187/pjsor.v13i2.1743

13. Ramires T. G., Ortega E. M., Cordeiro G. M., Paula G. A. and Hens N. (2018). New regression model with four regression structures and computational aspects. Communications in Statistics-Simulation and Computation, 47(7), 1940–1962. https://doi.org/10.1080/03610918.2017.1332212

14. Yousof H. M., Altun E., Ramires T. G., Alizadeh M. and Rasekhi M. (2018). A new family of distributions with properties, regression models and applications. Journal of Statistics and Management Systems, 21(1), 163–188. https://doi.org/10.1080/09720510.2017.1411028

15. El-Morshedy M., Eliwa M. S., and Afify A. Z. (2020). The odd Chen generator of distributions: properties and estimation methods with applications in medicine and engineering. Journal of the National Science Foundation of Sri Lanka, 48(2), 113–130. https://doi.org/10.4038/jnsfsr.v48i2.8790

16. Tahir M. H., Hussain M. A., Cordeiro G. M., El-Morshedy M. and Eliwa M. S. (2020). A new Kumaraswamy generalized family of distributions with properties, applications, and bivariate extension. Mathematics, 8, 1989. https://doi.org/10.3390/math8111989

17. Salah M., El-Morshedy M., Eliwa M. S., and Yousof H. M. (2020). Expanded Fréchet model: mathematical properties, copula, different estimation methods, applications and validation Testing. Mathematics, 8, 1949. https://doi.org/10.3390/math8111949

18. Alzaatreh A., Lee C. and Famoye F. (2013). A new method for generating families of continuous distributions. Metron, 71(1), 63–79. https://doi.org/10.1007/s40300-013-0007-y

19. Balakrishnan N., (1992). Handbook of the Logistic Distribution. Statistics: A Series of Textbooks and Monographs, 123, Marcel Dekker, New York, USA.

20. Rao B. S., Nagendram S. and Rosaiah K. (2013). Exponential-half logistic additive failure rate model. International journal of scientific and research publications, 3(5), 1–10.

21. Altun E., Khan M. N., Alizadeh M., Ozel G. and Butt N. S. (2018a). Extended half-logistic distribution with theory and lifetime data application. Pakistan Journal of Statistics and Operation Research, 14(2), 319–331. https://doi.org/10.18187/pjsor.v14i2.2396

22. Olapade A. K., (2014). The type I generalized half-logistic distribution. Journal of Iranian statistical society, 13(1), 69–88.

23. Cordeiro G. M., Alizadeh M., and Ortega E. M. M., (2014). The exponentiated half-logistic family of distributions: Properties and applications. Journal of Probability and Statistics. ID 864396, 1–21. https://doi.org/10.1155/2014/864396

24. Samuel A. F. and Kehinde O. A. (2019). A Study on Transmuted Half Logistic Distribution: Properties and Application. International Journal of Statistical Distributions and Applications, 5(3), 54. https://doi.org/10.11648/j.ijsd.20190503.12

25. Almarashi A. M., Badr M. M., Elgarhy M., Jamal F. and Chesneau C. (2020). Statistical Inference of the Half-Logistic Inverse Rayleigh Distribution. Entropy, 22(4), 449. https://doi.org/10.3390/e22040449 PMID: 33286223

26. Chen F. and Chen S. (2011). Injury severities of truck drivers in single-and multi-vehicle accidents on rural highways. Accident Analysis & Prevention, 43(5), 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026 PMID: 21658494

27. Chen F., Chen S. and Ma X. (2018). Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. Journal of Safety Research, 65, 153–159. https://doi.org/10.1016/j.jsr.2018.02.010 PMID: 29776524

28. Altun E. (2019). The log-weighted exponential regression model: alternative to the beta regression model. Communications in Statistics-Theory and Methods, 1–16. https://doi.org/10.1080/03610926.2019.1664586

29. Cordeiro G. M., Alizadeh M., Ozel G., Hosseini B., Ortega E. M. M. and Altun E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. Journal of Statistical Computation and Simulation, 87(5), 908–932. https://doi.org/10.1080/00949655.2016.1238088

30. Altun E., Yousof H. M. and Hamedani G. G. (2018b). A new log-location regression model with influence diagnostics and residual analysis. Facta Universitatis, Series: Mathematics and Informatics, 33(3), 417–449.

31. Pescim R. R., Ortega E. M., Cordeiro G. M. and Alizadeh M. (2017). A new log-location regression model: estimation, influence diagnostics and residual analysis. Journal of Applied Statistics, 44(2), 233–252. https://doi.org/10.1080/02664763.2016.1168368

32. Nichols M. D., and Padgett W. J., (2006). A Bootstrap Control Chart for Weibull Percentiles. Quality and reliability engineering international, 22, 141–151. https://doi.org/10.1002/qre.691

33. Aarset M. V. (1987). How to identify a bathtub hazard rate. IEEE Transactions on Reliability, 36(1), 106–108. https://doi.org/10.1109/TR.1987.5222310

34. Yadav A. S., Altun E. and Yousof H. M. (2019). Burr-Hatke Exponential Distribution: A Decreasing Failure Rate Model, Statistical Inference and Applications. Annals of Data Science, 1–20.

35. Zeng Q., Wen H., Huang H., Pei X. and Wong S. C. (2017). A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. Accident Analysis & Prevention, 99, 184–191. https://doi.org/10.1016/j.aap.2016.11.018 PMID: 27914307

36. Zeng Q., Hao W., Lee J. and Chen F. (2020). Investigating the impacts of real-time weather conditions on freeway crash severity: a Bayesian spatial analysis. International journal of environmental research and public health, 17(8), 2768. https://doi.org/10.3390/ijerph17082768 PMID: 32316427

37. Hurvich C. M., and Tsai C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. Journal of time series analysis, 14, 271–279. https://doi.org/10.1111/j.1467-9892.1993.tb00144.x