Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Log-Exponentiated Fréchet regression model with censored data

Hanan H. Al-Amoudi [1], *, Salwa A. Mousa [1], [2], Lamya A. Baharith [1]

[1]*Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia*
[2]*Faculty of Commerce (Girls Campus), Al-Azhar University, Nasr city, Cairo, Egypt*

ARTICLE INFO

Article history:
Received 5 April 2016
Received in revised form
27 May 2016
Accepted 27 May 2016

Keywords:
Exponentiated Fréchet distribution
Regression model
Maximum likelihood
Jackknife method
Residual analysis
Global influences

ABSTRACT

In this article, we propose the log-exponentiated Fréchet regression model based on exponentiated Fréchet distribution. Maximum likelihood and jackknife methods are used to estimate the proposed model parameters based on censored data. Model assumptions are checked using martingale and deviance residuals. Moreover, global influence is used to detect influential observations. The behavior of the maximum likelihood and jackknife estimators is investigated through Monte Carlo simulation. The empirical distribution of the martingale and deviance residuals of the proposed model is examined. Leukemia data is analyzed to show the flexibility of the proposed model.

1. Introduction

The construction of flexible distributions has attracted the attention of many statisticians. These distributions can be used to propose a useful parametric model for lifetime data. In survival analysis applications, the study of the effect of covariates on survival time is critical. Different types of regression models can be suggested in survival analysis such as the log-location-scale regression model which has been used in clinical trials, see Lawless (2003).

In recent times, log-location-scale regression model has been conducted frequently in many studies. Among these, the study of log-Burr XII regression model proposed by Silva et al. (2008), the log-modified Weibull regression model proposed by Carrasco et al. (2008), the log-generalized modified Weibull regression model proposed by Ortega et al. (2011), the log-exponentiated Weibull regression model with interval-censored analysis proposed by Hashimoto et al. (2009), the log-Weibull extended regression model proposed by Silva et al. (2010), the log-beta Weibull regression model proposed by Ortega et al. (2011), the log-Burr XII regression model with grouped survival data introduced by Hashimoto et al. (2012), the log-beta log-logistic regression model introduced by Mahmoud et al. (2015), the log-odd log-logistic Weibull regression model introduced by Cruz et al. (2016).

The exponentiated Fréchet (EF) distribution could be used in a wide number of lifetime applications including Fréchet distribution applications such as earthquakes, floods, horse racing, rainfall and wind speeds, see Kotz and Nadarajah (2000). Nadarajah and Kotz (2006) proposed the EF distribution using the exponentiated exponential distribution manner. The cumulative distribution function (cdf) and the corresponding probability density function (pdf) are respectively given by:

$$F(t;\alpha,\lambda,\gamma) = 1 - \left[1 - \exp\left\{-\left(\frac{\gamma}{t}\right)^{\lambda}\right\}\right]^{\alpha}; \; t > 0,$$

$$f(t;\alpha,\lambda,\gamma) = \alpha\lambda\gamma^{\lambda}\left[1 - \exp\left\{-\left(\frac{\gamma}{t}\right)^{\lambda}\right\}\right]^{\alpha-1} t^{-(\lambda+1)}\exp\left\{-\left(\frac{\gamma}{t}\right)^{\lambda}\right\}; \qquad t > 0, \tag{1}$$

where $\alpha > 0$ and $\lambda > 0$ are shape parameters, and $\gamma > 0$ is the scale parameter.

This article aims to investigate the aspects of classical analysis for modeling censored data based on LEF regression model. Asymptotic distribution of the maximum likelihood (ML) estimators is carried out which is useful for small sample size where the normality assumption is hard to justify. Therefore, the use jackknife estimator is explored for the log-exponentiated Fréchet (LEF) regression model for censored data. Moreover, it is essential after modeling to examine the assumptions and perform a diagnosis approach based on case deletion to detect influential and outlying observations, see Cook (1977).

The article is structured as follows: Section 2 presents a brief study on the LEF distribution. The LEF regression model is introduced in Section 3. Section 4 provides estimate of model parameters using ML and jackknife methods based on censored

---

* Corresponding Author.
Email Address: hhalamoudi@kau.edu.sa (H. Al-Amoudi)

data. Global sensitivity analysis is discussed in Section 5. In section 6, a residual analysis is performed to assess the departures from the underlying LEF model and to detect outliers. In section 7, Monte Carlo simulation is conducted to estimate model parameters. Leukemia data set is analyzed to demonstrate the results based on the new regression model in Section 8. Finally, concluding remarks are presented in Section 9.

2. The log-exponentiated Fréchet distribution

If we assume that the random variable T having the EF with density function in (1), then the random variable, $Y = \log(T)$ will have LEF distribution with transformation parameters $\sigma = 1/\lambda$ and $\mu = \log(\gamma)$. Therefore, the pdf and cdf for LEF distribution can be written as:

$$f(y;\alpha,\sigma,\mu) = \frac{\alpha}{\sigma}\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\} \left[1-\exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right]^{\alpha-1} \exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\};\ -\infty < y < \infty, \tag{2}$$

$$F(y;\sigma,\mu) = 1-\left[1-\exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right]^{\alpha}.$$

where $\alpha > 0$, $\sigma > 0$ and $-\infty < \mu < \infty$ are the shape, scale, and location parameters respectively. The survival and hazard functions are defined as:

$$S(y;\alpha,\sigma,\mu) = \left[1-\exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right]^{\alpha}$$

$$h(y;\alpha,\sigma,\mu) = \frac{\frac{\alpha}{\sigma}\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}}{\left[1-\exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right]}$$

It can be noted from Fig. 1 that the LEF distribution has a single mode. The LEF distribution has a monotonic increasing survival function and a monotonic decreasing hazard function.

The standardized random variable $Z = (y-\mu)/\sigma$ has the following density function:

$$f(z) = \alpha\exp(-z)\,[1-\exp\{-\exp(-z)\}]^{\alpha-1} \exp\{-\exp(-z)\};\ -\infty < z < \infty\,. \tag{3}$$

and its survival function is given by:

$$S(z) = [1-\exp\{-\exp(-z)\}]^{\alpha}.$$

For the variable Y, the moment generating function can be written as:

$$M_Y(t) = \int_{-\infty}^{\infty}\exp(ty)\ f(y)dy\,.$$

with setting $u = \exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]$, My(t) can be reduced to:

$$M_Y(t) = \alpha\exp(t\mu)\int_0^{\infty} u^{-t\sigma}\,[1-e^{-u}]^{\alpha-1}\ e^{-u}\,du.$$

For $\alpha > 1$ real non integer, $0 < e^{-u} < 1$ and $u > 0$, using the binomial series expansion:

$$[1-e^{-u}]^{\alpha-1} = \sum_{j=0}^{\infty} C_j^{\alpha-1}\ (-1)^j\,e^{-ju}\ ,$$

then, the $M_Y(t)$ can be rewritten as:

$$M_Y(t) = \alpha\ \exp(t\mu)\ \Gamma(1-t\sigma)\sum_{j=0}^{\infty} C_j^{\alpha-1}\ \frac{(-1)^j}{(1+j)^{1-t\sigma}}\ ,\ t\sigma < 1,$$

where, $C_j^{\alpha-1} = \frac{(\alpha-1)!}{j!(\alpha-1-j)!}$ is the binomial coefficient.

3. The log-Exponentiated Fréchet regression model

For lifetime data, we can use regression model where we can investigate the relations between independent variables and the survival time. Therefore, the model in (2) can be written as linear log- location-scale regression model:

$$y_i = \beta^T x_i + \sigma z_i\,;\quad i = 1,2,3,\dots,n\,, \tag{4}$$

where z_i is the random error with density function in (3), $\beta^T = (\beta_1,\dots,\beta_p)$ and $\sigma > 0$ are unknown parameters, and $x_i = (x_{i1},\dots,x_{ip})^T$ is the vector of independent explanatory variables. The parameter $\mu_i = \beta^T x_i$ is the location of Y_i. The location parameter vector $\mu = (\mu_1,\dots,\mu_n)^T$ can be represented as a linear model such that $\mu = \beta^T X$, where $X = (x_1,\dots,x_n)^T$ is a known model matrix.

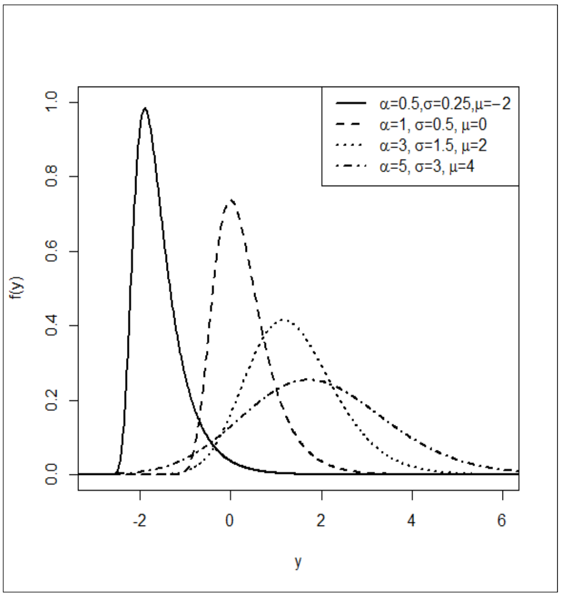


Fig. 1: Plot of the LEF density for some parameter values

In this case, the survival function of $Y|x$ is given by:

$$S(y|x) = \left[1-\exp\left\{-\exp\left[-\left(\frac{y-\beta^T x_i}{\sigma}\right)\right]\right\}\right]^{\alpha}.$$

4. Estimation

4.1. Maximum Likelihood Estimation

For inference, we use ML method to estimate model parameters. If $(y_1,\tau_1,x_1),\dots,(y_n,\tau_n,x_n)$ is a right censored random sample of n observations, where $y_i = \begin{cases}\log(t_i) & \text{if}\ \ \tau_i = 1\\ \log(c_i) & \text{if}\ \ \tau_i = 0\end{cases}$, x_i is the vector of explanatory variables, t_i and c_i are lifetimes and censoring times, respectively. If the censoring times and lifetimes are random and independent then the log-likelihood function is given by:

$$l(\boldsymbol{\theta}) = -r\log(\sigma) + r\log(\alpha) + \sum_{i=1}^{n}\tau_i\begin{Bmatrix}-z_i - \exp(-z_i) + (\alpha-1)\\ \log[1-\exp\{-\exp(-z_i)\}]\end{Bmatrix} + \sum_{i=1}^{n}(1-\tau_i)\alpha\log[1-\exp\{-\exp(-z_i)\}] \tag{5}$$

where r denotes the number of uncensored observations, $\boldsymbol{\theta} = (\alpha, \sigma, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ and $z_i = (y_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)/\sigma$.

The ML estimate for parameter vectors $\boldsymbol{\theta}$ could be obtained by maximizing the log- likelihood in (5).

The distribution of $\hat{\boldsymbol{\theta}}$ is asymptotically normal using some regularity conditions with covariance matrix that represents the inverse of the Fisher information matrix $I(\boldsymbol{\theta})^{-1}$ where, $I(\boldsymbol{\theta}) = E\left[-\left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right]$.

Since the censored observations are present, the computation of the Fisher information matrix is difficult. Therefore, the asymptotic Fisher information (observed matrix) can be obtained by omitting the expectation from $I(\theta)$. Then the asymptotic normal approximation for $\hat{\theta}$ could be defined as $\hat{\theta}^{\mathrm{T}} \sim N_{(p+1)}\{\theta^{\mathrm{T}}, \ddot{L}(\theta)^{-1}\}$, where $\ddot{L}(\theta) = -\left(\frac{\partial^2 \ell(\theta)}{\partial\theta\,\partial\theta^{\mathrm{T}}}\right)$ is the $(p+2) \times (p+2)$ observed matrix such that

$$\ddot{L}(\boldsymbol{\theta}) = \begin{pmatrix} -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\alpha^2}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\alpha\,\partial\sigma}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\alpha\,\partial\beta_j}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} \\ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma\,\partial\alpha}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma^2}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma\,\partial\beta_j}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} \\ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\beta_j\,\partial\alpha}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\beta_j\,\partial\sigma}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} & -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\beta_j\,\partial\beta_k}\Big|_{\hat{\boldsymbol{\beta}},\hat{\alpha},\hat{\sigma}} \end{pmatrix},$$

where $j, k = 1, \dots, p$, and the submatrices can be obtained from authors upon request.

4.2. Jackknife estimation

Jackknife estimation is used to improve the estimate of the parameter by reducing the bias of the estimate. The method is based on “leave one out” procedure. Miller (1974) uses jackknife in linear models to estimate the variance and bias for model parameters, see Tu and Shao (1995).

Suppose that $x_1, x_2, \dots, x_n$ is independent identically distributed random sample and $\hat{\theta}$ is the parameter estimator of the whole sample. Then $\hat{\theta}_{-i}$ is the parameter estimate when we drop the i^{th} observation from the sample. The pseudo-value of the i^{th} observation can be calculated as the difference between parameter estimation from whole sample and parameter estimation obtained without the i^{th} observation. That is,

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}\,.$$

The Jackknife estimate of θ, denoted $\hat{\theta}_J$ is the average of pseudo-values:

$$\hat{\theta}_J = \frac{1}{n}\sum_{i=1}^{n} \tilde{\theta}_i,$$

therefore, the jackknife bias estimator is:

$$b_J = (n-1)\left(\hat{\theta}_J - \hat{\theta}\right).$$

An approximate $100(1-\alpha)\%$ confidence interval for θ is given by:

$$\hat{\theta}_J \pm t_{\alpha/2,n-1}\, {}^{S}\!/\!_{\sqrt{n}}.$$

See, Sahinler and Topuz (2007) and Algamal and Rasheed (2010).

5. Sensitivity analysis: Global Influence

Sensitivity represents the deviation in the model output with respect to the changes in the model's input(s). Global influence is the diagnostic influence depend on case deletion that represents one of the tools to perform sensitivity analysis introduced by Cook (1977). Case deletion is a popular method to investigate the influence of taking out the i^{th}case from the data on the parameter estimate. This method compares between $\hat{\theta}$ and $\hat{\theta}_{-i}$, where $\hat{\theta}_{-i}$ is ML estimate when the i^{th} case is deleted from original data. Then the i^{th} case could be considered as influential observation if $\hat{\theta}_{-i}$ is far from $\hat{\theta}$.

This methodology was conducted in many statistical models, for example, Christensen et al. (1992) and Xie and Wei (2007). The case deletion model for the LEF regression model (4) is given by:

$$Y_J = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + \sigma Z_i; \quad J = 1,2,3, \dots, n, J \neq i. \qquad (6)$$

For model in (6), $\hat{\theta}_{-i} = \left(\hat{\alpha}_{(i)}, \hat{\sigma}_{(i)}, \hat{\beta}^{\mathrm{T}}{}_{(i)}\right)^{\mathrm{T}}$ denotes the ML estimate of θ when i^{th} case is deleted. Then, the generalized cook distance and likelihood distance are used to measure the effect of the i^{th} case on the ML estimate $\hat{\theta} = \left(\hat{\alpha}, \hat{\sigma}, \hat{\beta}^{\mathrm{T}}\right)^{\mathrm{T}}$.

Generalized cook distance is the method that measure global influence defined as the standardized norm of $\hat{\theta}_{-i} - \hat{\theta}$.

$$GD_i(\boldsymbol{\theta}) = \left(\hat{\boldsymbol{\theta}}_{-i} - \hat{\boldsymbol{\theta}}\right)'\left\{\ddot{L}(\hat{\boldsymbol{\theta}})\right\}\left(\hat{\boldsymbol{\theta}}_{-i} - \hat{\boldsymbol{\theta}}\right),$$

where $\ddot{L}(\hat{\theta})$ is the observed information matrix.

The likelihood distance is another method to measure the difference between $\hat{\theta}$ and $\hat{\theta}_{-i}$:

$$LD_i(\boldsymbol{\theta}) = 2\left\{ l\left(\hat{\boldsymbol{\theta}}\right) - l\left(\hat{\boldsymbol{\theta}}_{-i}\right)\right\},$$

where $l\left(\hat{\theta}\right)$ is the log likelihood function of θ from original data and $l\left(\hat{\theta}_{-i}\right)$ is the log likelihood function of θ when i^{th} case is deleted from original data.

6. Analysis of residual

Residual analysis is conducted to examine the assumptions and detect outlying observations. In survival analysis with right censored data, martingale and deviance residuals are performed to assess leverage and goodness of fit, for more details about the martingale and deviance residuals in parametric lifetime model, see Collett (2003).

6.1. Martingale residual

The martingale residual is defined in Barlow and Prentice (1988) as:

$$r_{M_i} = \delta_i + \log S_Y(y_i, \hat{\theta}),$$

where δ_i is the censoring indicator that takes 0 if censored and 1 if a lifetime, and $S_Y(y_i, \hat{\theta})$ is the survival function for LEF regression model. Therefore, the martingale residual for LEF regression model can be written as

$$r_{M_i} = \begin{cases} 1 + \hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\} \\ \qquad \text{if} \quad i \in \text{life time} \\ \hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\} \\ \qquad \text{if} \quad i \in \text{censored} \end{cases},$$

where r_{M_i} is skewed and between $-\infty$ and 1. Thus, we use transform the martingale residual to reduce the skewness.

6.2. Deviance residual

Therneau et al. (1990) proposed deviance residuals for Cox model with no time-dependent explanatory variables and is written as:

$$r_{D_i} = \mathrm{sign}(r_{M_i})\{-2[r_{M_i} + \delta_i \log(\delta_i - r_{M_i})]\}^{\frac{1}{2}},$$

here, r_{D_i} is more symmetrically about zero from martingale residual. Thus, the deviance residual for LEF regression model is:

$$r_{D_i} = \begin{cases} \mathrm{sign}(1 + \hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\}) \\ \left\{-2\begin{bmatrix} 1 + \hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\} + \\ \log(-\hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\}) \end{bmatrix}\right\}^{\frac{1}{2}} \\ \quad \text{if} \quad i \in \text{life time} \\ \\ \mathrm{sign}(\hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\}) \\ \{-2\,\hat{\alpha}\log\{1 - \exp[-\exp(-\hat{z}_i)]\}\}^{\frac{1}{2}} \\ \quad \text{if} \quad i \in \text{censored} \end{cases}.$$

The pots of the martingale and deviance residuals, versus the adjusted values should provide random behavior around zero if the model is adequate.

7. Simulation studies

We performed Monte Carlo simulation to examine the estimates of model parameters α, σ and β_1 when β_0 is fixed and equal to 5, using ML and jackknife methods. This simulation was conducted for different sample sizes, n=30, 50, and 100, generated from EF distribution with the parameters $\alpha = 0.3$, $\lambda = 1.4$ and $\gamma = 50$, using various percentages of censoring (PC), 10, 30, and 50. We generate 1000 samples for each sample size and percentage of censoring. Table 1 displays the ML and jackknife parameter estimates along with the corresponding average of standard error (SE) for LEF regression model given in (4). From these results, it can be noted that the SE of estimates using jackknife method is mostly smaller compared to the corresponding SE of estimates using ML method. Moreover, the SE of estimates increases as censoring percentage increases for the same sample size and the SE of estimates decreases as the sample size increases.

Table 1: Maximum likelihood and jackknife estimates for the LEF regression model

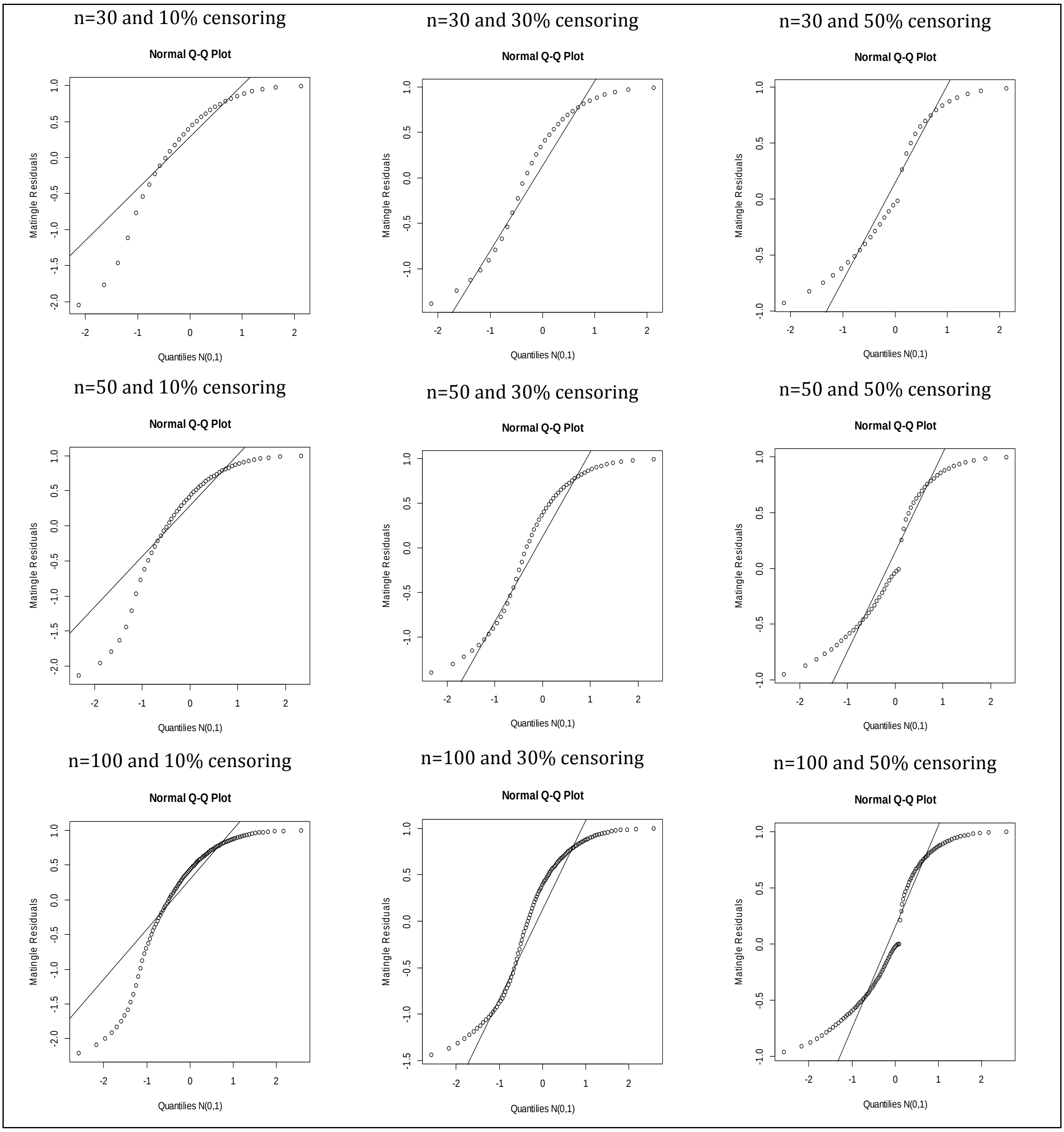
n	PC	θ	ML		Jackknife	
			Estimate	SE	Estimate	SE
30	10%	α	0.3331	0.1243	0.3611	0.0828
		σ	0.8595	0.2211	0.9546	0.2453
		β_1	−1.5961	0.7372	−1.5284	0.5723
	30%	α	0.3371	0.1420	0.3645	0.0849
		σ	0.8670	0.2332	0.9719	0.2618
		β_1	−1.5774	0.7789	−1.5020	0.5803
	50%	α	0.3378	0.1583	0.3624	0.0991
		σ	0.9006	0.2461	0.9949	0.3125
		β_1	−1.5334	0.8310	−1.4879	0.6492
50	10%	α	0.3429	0.0917	0.3533	0.0399
		σ	0.8987	0.1619	0.9336	0.1492
		β_1	−1.5835	0.5582	−1.5790	0.3925
	30%	α	0.3434	0.1013	0.3528	0.0374
		σ	0.9106	0.1689	0.9482	0.1614
		β_1	−1.5646	0.5816	−1.5661	0.3931
	50%	α	0.3424	0.1169	0.3542	0.0407
		σ	0.9294	0.1834	0.9793	0.1883
		β_1	−1.5318	0.6277	−1.5378	0.4079
100	10%	α	0.3451	0.0603	0.3460	0.0070
		σ	0.9081	0.1054	0.9188	0.0890
		β_1	−1.5786	0.3791	−1.5912	0.2553
	30%	α	0.3451	0.0669	0.3460	0.0072
		σ	0.9252	0.1109	0.9363	0.1005
		β_1	−1.5661	0.3984	−1.5792	0.2573
	50%	α	0.3451	0.0776	0.3458	0.0083
		σ	0.9569	0.1226	0.9668	0.1191
		β_1	−1.5291	0.4305	−1.5492	0.2653

In addition, the form of the empirical distribution of r_{M_i} and r_{D_i} residuals is examined by computing r_{M_i} and r_{D_i} residuals based on ML estimates. Then, a plot of the mean ordered residuals versus the expected quantiles of the standard normal distribution (SND) is presented in Fig. 2 and Fig. 3. Appendix A illustrates the steps used to compute the ML estimates and residuals for LEF regression model. From Fig. 2 and Fig. 3, it is clear that the empirical distribution of the deviance residual shows close agreement to the SND compared to the martingale residual. Moreover, the empirical distribution of the deviance residual moves away from the SND when the censoring percentage increases and it approaches the SND as the sample size increases.

8. Leukemia Data

This data set of survival time for 33 patients whose diagnose leukemia described in Lawless (2003) is used to demonstrate the performance of the LEF regression model. This data consists of three variables, survival times in weeks and two explanatory variables: white blood cells count (WBC) at diagnosis and a binary variable (AG) is white blood cells characteristics test indicate positive test (AG=1) or negative test (AG=0). The following variables in the study are:

t_i: Survival time observed (in weeks).
y_i: Log survival time observed (in weeks).
$status_i$: Censoring indicator (0=censoring, 1=lifetime).
x_{i1}: White blood cells characteristics test (0= negative, 1= positive).
x_{i2}: White blood cells count (in thousands).


n=30 and 10% censoring

n=30 and 30% censoring

n=30 and 50% censoring

n=50 and 10% censoring

n=50 and 30% censoring

n=50 and 50% censoring

n=100 and 10% censoring

n=100 and 30% censoring

n=100 and 50% censoring

Fig. 2: Normal probability plots for the martingale residual at sample size n=30, 50 and 100, different censoring percentages =10, 30, and 50, and parameter value $\alpha = 0.3$, $\sigma = 1/1.4$, and fixed $\beta_0 = 5$

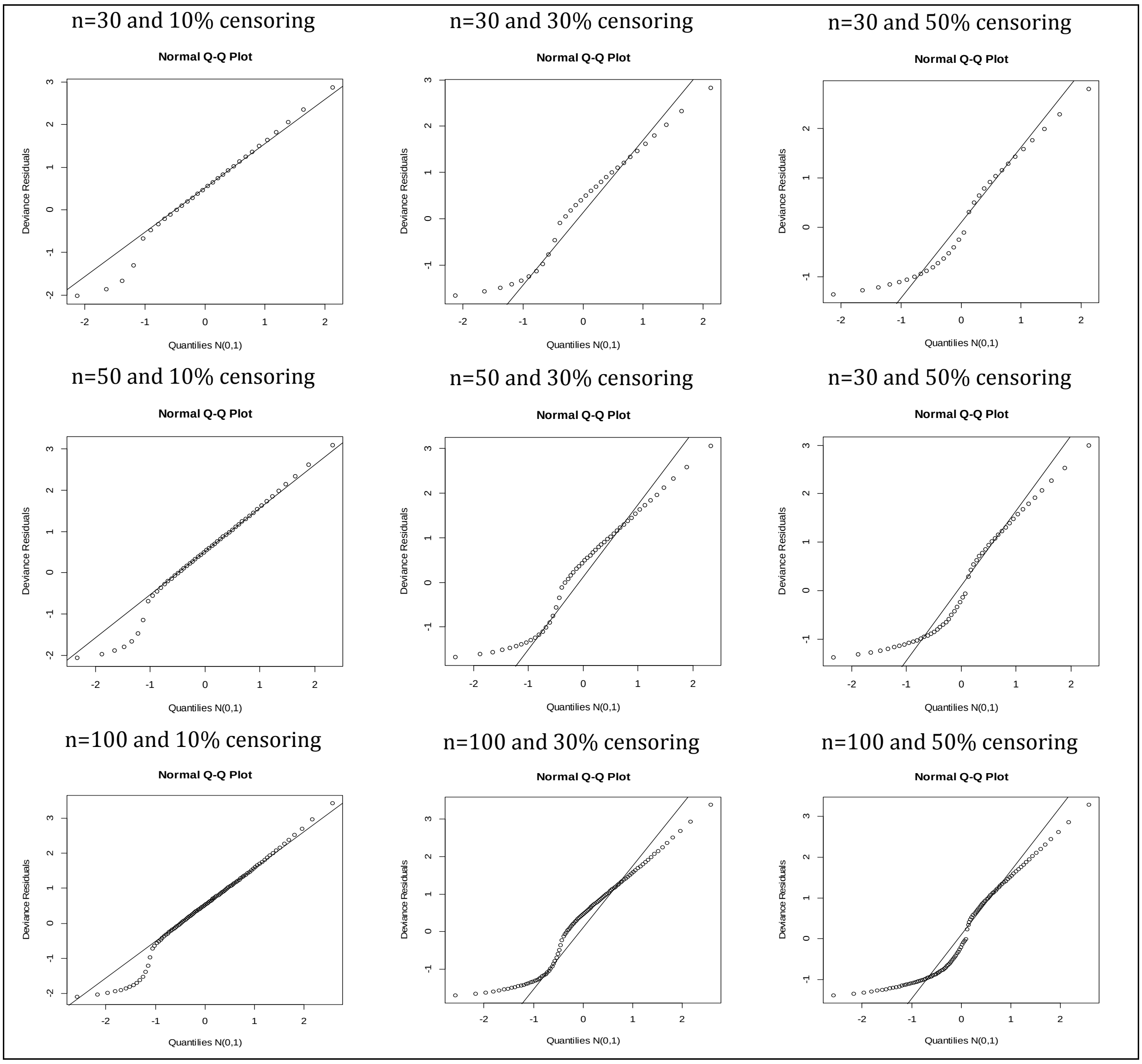


Fig. 3: Normal probability plots for the deviance residual at sample size n=30, 50, and 100, different censoring percentages =10, 30, and 50, and parameter value $\alpha = 0.3$, $\sigma = 1/1.4$, and fixed $\beta_0 = 5$

8.1. Model validation

To examine the suitability of the proposed model, a plot of the empirical survival function using Kaplan-Meier (KM) estimates and the estimated survival function by fitting the EF model is presented in Fig. 4. It is clear that the logarithm of the times to event follows LEF distribution. Therefore, the model fitted can be written as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 \log(x_{i2}) + \sigma z_i; \quad i = 1,2,\ldots,33 \tag{7}$$

where y_i follows the LEF distribution given in (2).

8.2. Maximum likelihood estimation

ML and jackknife methods are used to estimate the model parameters in (7) along with the SE, 95% confidence interval (CI) and p-value as shown in the Table 2. It can be seen that the explanatory variables x_1 and $\log(x_2)$ are significant using both methods. Also, both methods provide very similar estimates except for parameter σ. Moreover, it can be concluded from the parameter estimates that the log of survival time for leukemic patients increases when AG test is positive and white blood cells count are low at diagnosis.

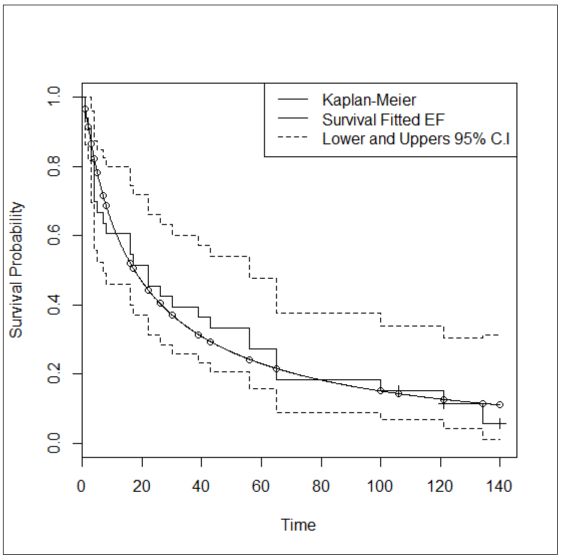


Fig. 4: Plot the survival function by fitting the EF distribution and KM function for leukemia data

Fig. 5 represents the survival function corresponding to explanatory variables for the fitted LEF regression model. From Fig. 5A, it can be seen that $\hat{S}(4|AG = 0) = 0.5760$, which is approximately 58% of the patients who have negative AG test at the diagnosis will be alive at y = 4 (≈ 55 weeks). Furthermore, for patients who have positive AG test at the diagnosis, $\hat{S}(4|AG = 1) = 0.8769$, that is, approximately 88% of patients will be alive at y = 4. Analogously, Fig. 5B shows that $\hat{S}(3|WBC = 2) = 0.7250$, which means that about 73% of the patients who reached 2000 white blood cells at the diagnosis will be alive at y = 3 (≈ 20 weeks). However, for patients who reached 25000 and 100000 white blood cells, $\hat{S}(3|WBC = 25) = 0.1909$ and $\hat{S}(3|WBC = 100) = 0.0682$, this means that the percentages of the patients that will be alive at y = 3 decreases to 19% and 7% respectively.

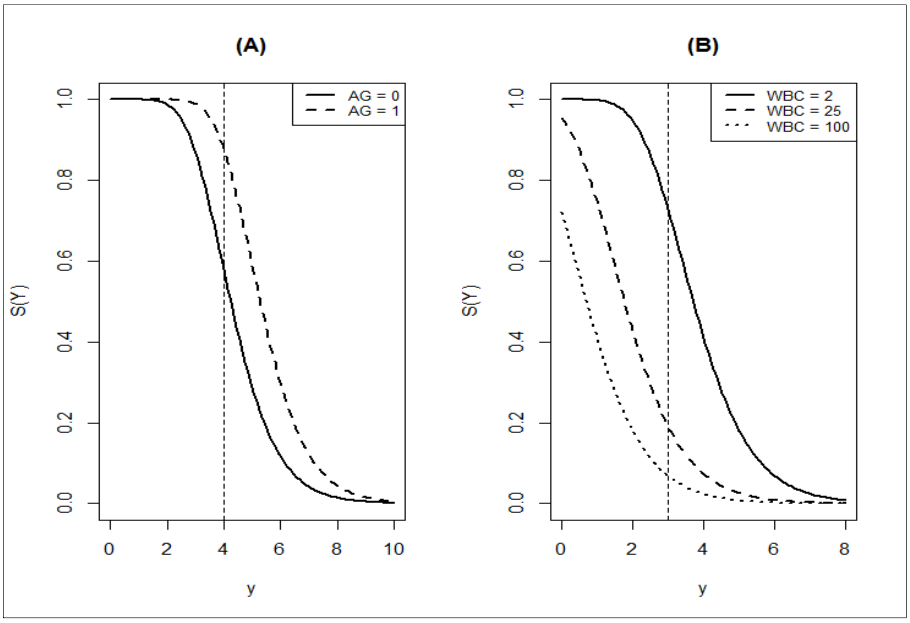


Fig. 5: Fitted survival functions from the LEF regression model for the leukemia. (A) for x_1 = AG test, (B) for x_2 = WBC

8.3. Global influence

The case-deletion measures $GD_i(\boldsymbol{\theta})$ and $LD_i(\boldsymbol{\theta})$ that presented in section 5 are numerically computed. Fig. 6A and Fig. 6B represent the influence measure index plots. It can be seen that the case 21 is a possible influential observation.

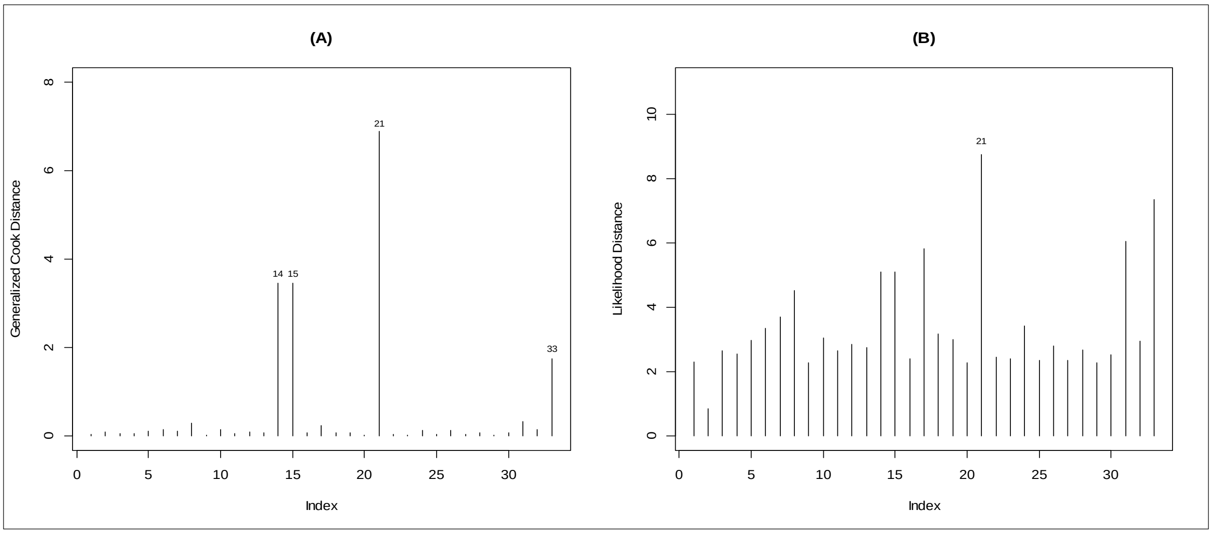


Fig. 6: A) Plot the index of generalized Cook distance for theta. B) Plot the index of likelihood distance for theta

8.4. Analysis of residual

The deviance residual against the index of survival time for the fitted model is shown in Fig. 7A. It is clear that almost all observations fall on the interval (−3, 3) and are distributed randomly about zero. Fig. 7B shows the normal probability plot for the deviance residual with the simulated envelope; see Ortega et al. (2011) and Zhao et al. (2011).

Therefore, LEF regression model provides a suitable fit for leukemia data.

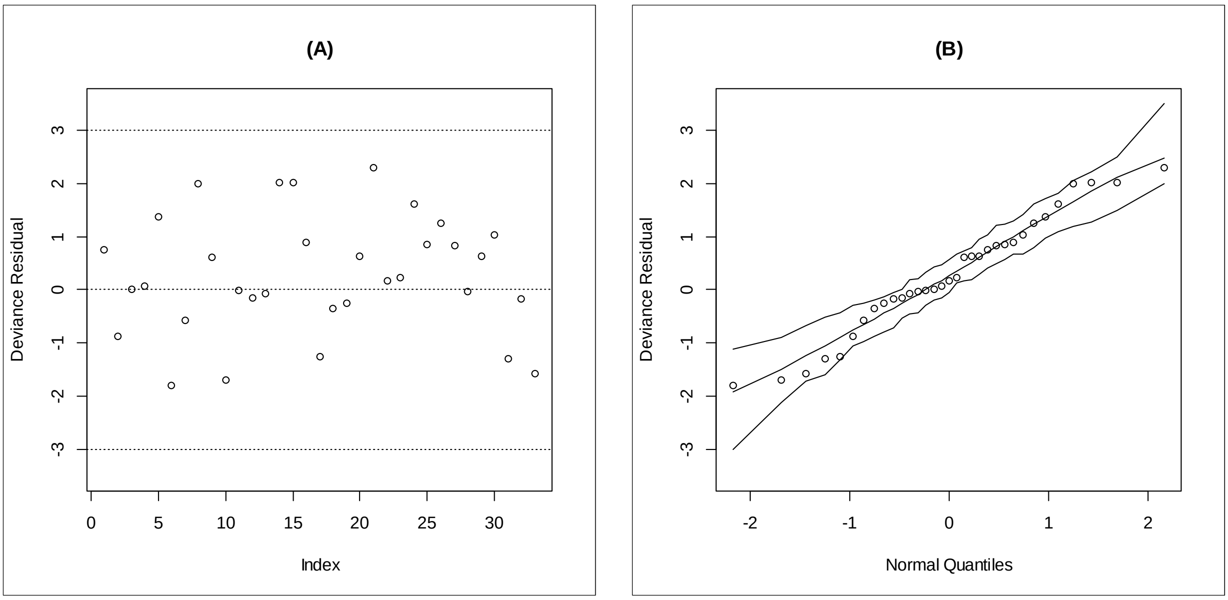


Fig. 7: A) Index plot of deviance residual. B) Normal probability plot for the deviance residual with envelopes

8.5. Comparison between Log-Weibull and LEF Regression Models

In Lawless (2003), analysis has been conducted on the previous leukemia survival time data under log-Weibull regression model. Therefore, we conduct a comparison between log-Weibull and LEF regression model based on AIC and BIC criteria. Table 3 displays the results of these criteria which show that the LEF regression model is more appropriate model compared to the log-Weibull regression model with smaller values of AIC and BIC.

Table 2: The ML and Jackknife estimates for the parameters of the LEF regression model

θ	ML				Jackknife		
	Estimate	SE	95% CI	p-value	Estimate	SE	95% CI
α	1.7128	2.2270	(0.0000 , 6.0777)	-	1.8395	2.1381	(0.0000, 6.0301)
σ	1.4884	0.9754	(0.0000 , 3.4002)	-	2.2491	1.1962	(0.0000 , 4.5937)
β_0	4.3788	1.5209	(1.3979 , 7.3598)	0.0039	4.8536	1.1961	(2.5092 , 7.1979)
β_1	1.0470	0.4262	(0.2116 , 1.8824)	0.0140	1.1399	0.5019	(0.1563 , 2.1236)
β_2	-0.7692	0.1703	$(-1.1031, -0.4354)$	< 0.001	-0.6984	0.3188	$(-1.3232, -0.0736)$

Table 3: Statistics AIC and BIC for comparing the log-Weibull and LEF regression models

Model	AIC	BIC
Log-Weibull	115.6403	121.6263
LEF	113.4414	120.9239

9. Concluding remarks

This article introduced the LEF regression model with right censored lifetime data. ML and jackknife methods were used to estimate model parameters. In addition, Monte Carlo simulation indicated that there is close agreement between the SND with the distribution of the deviance residual. The leukemia data have illustrated the performance and validity of LEF regression model. Moreover, the robustness features of the ML estimator from the fitted LEF regression model are discussed through residuals and sensitivity analysis. The goodness of fit is conducted by constructing normal probability plot with simulated envelope where we observe that almost all observations fall within the envelope. Finally, we use leukemia data to compare the performance of LEF model and log-Weibull regression model using AIC and BIC criteria which indicated that the proposed model is a more appropriate model for the current data set.

Acknowledgement

The authors thank the editors for their comments which improved the original version of the paper.

Appendix A: Algorithm ML estimation and Residual Analysis for the parameters of LEF regression model

(1) For a given values of the parameters $\alpha = 0.3, \lambda = 1.4, \text{and } \gamma = 50$, $T_1, T_2, \dots, T_n$ are lifetimes from EF distribution given (1).
(2) Generate the explanatory variable x_i from a standard uniform distribution.
(3) Generate $C_1, C_2, \dots, C_n$ as a censoring time from uniform distribution $[0, \rho]$, where ρ was

adjusted until obtaining the required censoring percentages.

(4) The log lifetimes considered in each fit were calculated as $y = \min\{\log(T_i), \log(C_i)\}$.
(5) Estimate model parameter given in (4) by ML method using *nlminb* function in R program.
(6) Compute the standard error for each estimate using hessian function in R program.
(7) Compute the standardized, Martingale and deviance residuals.
(8) The above steps are repeated 1000 times.
(9) Plot the martingale and deviance residuals against the expected quantiles of normal distribution.

References

Algamal ZY and Rasheed KB (2010). Re-sampling in Linear Regression Model Using Jackknife and Bootstrap. Iraqi Journal of Statistical Science, 18: 59-73.

Barlow WE and Prentice RL (1988). Residuals for relative risk regression. Biometrika, 75(1): 65-74.

Carrasco JM, Ortega EM and Paula GA (2008). Log-modified Weibull regression models with censored data: Sensitivity and residual analysis. Computational Statistics and Data Analysis, 52(8): 4021-4039.

Christensen R, Pearson LM and Johnson W (1992). Case-deletion diagnostics for mixed models. Technometrics, 34(1): 38-45.

Collett D (2003). Modelling survival data in medical research: CRC press, London.

Cook RD (1977). Detection of influential observation in linear regression. Technometrics, 19(1): 15-18.

Cruz JND, Ortega EM and Cordeiro GM (2015). The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. Journal of Statistical Computation and Simulation, 86(8): 1-23.

Hashimoto EM, Ortega EM, Cancho VG and Cordeiro GM (2010). The log-exponentiated Weibull regression model for interval-censored data. Computational Statistics and Data Analysis, 54(4): 1017-1035.

Hashimoto EM, Ortega EM, Cordeiro GM and Barreto ML (2012). The Log-Burr XII regression model for grouped survival data. Journal of Biopharmaceutical Statistics, 22(1): 141-159.

Kotz S and Nadarajah S (2000). Extreme value distributions: theory and applications. World Scientific. Imperial College Press, London.

Lawless JF (2003). Statistical models and methods for lifetime data: John Wiley and Sons, New Jersey, USA.

Mahmoud MR, El-Sheikhb AA, Morad NA and Ahmad AM (2015). Log-Beta Log-Logistic Regression Model. International Journal of Sciences: Basic and Applied Research (IJSBAR), 22(2): 389-405.

Miller RG (1974). An unbalanced jackknife. The Annals of Statistics, 2(5): 880-891.

Nadarajah S and Kotz S (2006). The Exponentiated Type Distributions. The exponentiated type distributions. Acta Applicandae Mathematica, 92(2): 97-111.

Ortega E MM, Cordeiro GM, and Hashimoto EM (2011). A log-linear regression model for the Beta-Weibull distribution. Communications in Statistics-Simulation and Computation, 40(8): 1206-1235.

Ortega EM, Cordeiro GM and Carrasco JM (2011). The log-generalized modified Weibull regression model. Brazilian Journal of Probability and Statistics, 25(1): 64-89.

Sahinler S and Topuz D (2007). Bootstrap and jackknife resampling algorithms for estimation of regression parameters. Journal of Applied Quantitative Methods, 2(2): 188-199.

Silva GO, Oretga EMM, Cancho VG and Barreto ML (2008). Log-Burr XII regression models with censored data. Computational Statistics and Data Analysis, 52(7): 3820-3842.

Silva GO, Ortega EMM, and Cancho VG(2010). Log-Weibull extended regression model: estimation, sensitivity and residual analysis. Statistical Methodology, 7(6): 614-631.

Therneau TM, Grambsch PM, and Fleming TR (1990). Martingale-based residuals for survival models. Biometrika, 77(1): 147-160.

Tu D, and Shao J (1995). The Jackknife and bootstrap: Springer-Verlag, New York, USA.

Xie FC and Wei BC (2007). Diagnostics analysis for log-Birnbaum–Saunders regression models. Computational Statistics and Data Analysis, 51(9): 4692-4706.

Zhao Y, Lee AH, Yau KK and McLachlan GJ. (2011). Assessing the adequacy of Weibull survival models: a simulated envelope approach. Journal of Applied Statistics, 38(10): 2089-2097.