Afrika Statistika

Afrika Statistika
Vol. 16 (4), 2021, pages 2993 - 3007.
DOI: http://dx.doi.org/10.16929/as/2021.2993.192

ISSN 2316-090X

A Log-Beta Rayleigh Lomax Regression Model

Nofiu Idowu Badmus (1,*), **Mary Idowu Akinyemi** (1) **and Josephine Nneamaka Onyeka-Ubaka** (1)

(1) Department of Mathematics, University of Lagos, Akoka, Nigeria

Received on October 16, 2021; Accepted on November 20, 2021

Abstract. For the first time, a location-scale regression model based on the logarithm of an extended Raleigh Lomax distribution which has the ability to deal and model of any survival data than classical regression model is introduced. We obtain the estimate for the model parameters using the method of maximum likelihood by considering breast cancer data. In addition, normal probability plot of the residual is used to detect the outliers and evaluate model assumptions. We use a real data set to illustrate the performance of the new model, some of its sub-models and classical models consider in the study. Also, we perform the statistics *AIC*, *BIC* and *CAIC* to select the most appropriate model among those regression models considered in the study.

Key words: Breast cancer; location-scale; logarithm; outliers; residual.
AMS 2010 Mathematics Subject Classification Objects : 60E05; 62J12; 62N02.

---

*Corresponding author: Nofiu Idowu Badmus (nibadmus@unilag.edu.ng)
Mary Idowu Akinyemi: makinyemi@unilag.edu.ng
Josephine Nneamaka Onyeka-Ubaka: jonyeka-ubaka@unilag.edu.ng

Resume (Abstract in French) Pour la première fois, un modèle location-scale regression base sur une extension de la distribution de Raleigh Lomaz ayant la capacité de modéliser tous les types de regression (classiques et non-classiques) données de survie est proposé. Les estimations de paramètres sont faites et appliquées aux données sur le cancer du sein. Avec des données réelles, le modèle est evalué ainsi que certains de ses sous-modèles. Entre autres outils de décision, les critères *AIC*, *BIC* ad *CAIC* pour la sélection de modèles. L'étude graphique des résidus est utilisée pour détecter les valeurs aberrantes et pour évaluer les nouvelles méthodes.

The authors.

Nofiu Idowu Badmus, Ph.D., is a Lecturer at University of Lagos, Akoka, Nigeria.

Mary Idowu Akinyemi, Ph.D., is a Senior Lecturer at University of Lagos, Akoka, Nigeria.

Josephine Nneamaka Onyeka-Ubaka, Ph.D., is a Senior Lecturer at University of Lagos, Akoka, Nigeria.

1. Introduction

Classical distributions (Weibull, Rayleigh, Lomax, log-normal etc) do not have the ability to capture the behavior of a lifetime data that exhibits excess skewness, kurtosis and a bathtub-shaped failure rate curve. These classical distributions are only good in modelling monotone or unimodal failure rate functions. Silva *et al.*(2009) stated in their work that distributions that exhibit bathtub-shaped failure rate are usually complex and, at the same time difficult to model. For this reason, there is need to propose/develop that can model this type of excess skewness and kurtosis, and failure rate. Several distributions have been introduced to model survival data, but a few number of regression models have been proposed with this objective, among them, the log-generalized modified Weibull model Ortega *et al.*(2009), the log-beta generalized half-normal regression model Pescim *et al.*(2013), the log-exponentiated-Weibull regression model for interval censored data was introduced by Hashimoto *et al.*(2010) and log-modified Weibull Carrasco *et al.*(2008) regression models.

In this paper, we propose a new regression model using the logarithm of the extended Rayleigh Lomax distribution by Kawsar *et al.*(2018) in a bid to model certain real life phenomena. The modification of the existing distribution leads to a location-scale regression model suitable for fitting censored survival times with bathtub-shaped hazard rates referred to as the log- Beta Rayleigh Lomax (*LBRL*) regression model. In Section 2, we define the *LBRL* distribution and derive its moments. We propose a *LBRL* regression model of location-scale form, obtain the maximum likelihood estimates and provide normal probability plots of the residuals to detect the outliers in section 3. Also, we show in Section 4 that the proposed model is more adequate to fit the breast cancer data analysis than its sub-models and some classical regression models by checking the residual plots for the models and discriminating between the models using three different statistics. Section 5 ends with some concluding remarks.

2. The log- Beta Rayleigh Lomax distribution

In the recent past, many works have been done which extend both Rayleigh and Lomax distribution, for instance: Weibull-Lomax (*WL*) distribution introduced by Tahir *et al.*(2010), Gumbel-Lomax (*GL*) distribution proposed by Tahir *et al.*(2016), Exponential Lomax (EL) distribution by El-Bassiouny *et al.*(2015), Exponentiated Weibull Lomax (*EWL*) distribution was initiated by Hassan and Abd-allah (2018), the Beta-modified weighted Rayleigh (*BMWR*) distribution by Badmus *et al.*(2017), the Gamma-Rayleigh (*GR*) distribution by Akarawak *et al.*(2017). The Rayleigh Lomax (*RL*) distribution with three parameters proposed by Kawsar *et al.*(2018) based on the combination of Rayleigh distribution by Siddiqui (1962) and Lomax distribution by Lomax (2018) which they intend to fit several kinds of survival data.

Here, we are going to present two important classes of distributions. First, we introduce Beta Rayleigh Lomax (*BRL*) distribution using the logit of beta function by Jones (2004) on the Rayleigh Lomax law. The result gives the following non-negative distribution with probability distribution function (*pdf*)

$$\begin{aligned} f(t) \;=\; & \frac{\left\{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\right\}^{a-1}\left\{e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\right\}^{b-1}}{B(a,b)} \\ & \times\; \frac{\alpha\lambda}{\theta}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha+1} e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}} 1_{(t>0)}, \end{aligned} \tag{1}$$

where $B(a,b) = \left(\Gamma(a)\Gamma(b)\right)/\Gamma(a+b)$ is the beta function. We call as a Beta Rayleigh Lomax (*BRL*) distribution any probability law of *pdf* 1.

The parameters a and b govern the weight tails of the distribution, α and λ control the shape of the distribution, while θ take cares of its scale. Furthermore, the important characteristic of the new distribution is that it consists some known special sub-models such as: Exponential Lomax distribution when $a = b = 1$, Rayleigh distribution if $a = b = \theta = \lambda = 1$, Lomax distribution as $a = b = \alpha = 1$ and *RL* distribution if $a = b = 1$. The new distributions emanate from the propose distribution are: Lehmann Rayleigh Lomax (*LRL*) distribution when $a = 1$ and exponentiated Rayleigh Lomax (*ERL*) distribution if $b = 1$.

Secondly, we introduce the log-transform of the *BRL* to get the log-beta Rayleigh Lomax (*LBRL*) distribution $Y = \log T$ with $T \sim$ *BRL*. The paper mainly focuses of the second class, the *LBRL* distribution, which gives the solution of problmes we cited above. Direct computation lead to the following facts about *BRL*'s laws. The reliability function of T is

$$\begin{aligned} R(t) &= 1-\ell_{\left\{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\right\}}(a,b) \\ &= 1-\frac{1}{B(a,b)}\int_0^{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}} k^{a-1}(1-k)^{b-1}dk \end{aligned} \quad (2)$$

where, $\ell_x(a,b) = B_x(a,b)/B(a,b)$ is the incomplete beta function ratio and $B_x(a,b)=\int_0^x k^{(a-1)}(1-k)^{(b-1)}dk$ is the incomplete beta function. Silva *et al.*(2009).

Also, the hazard rate function of T is

$$\begin{aligned} h(t) &= \frac{\alpha\lambda}{\theta}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha+1} e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\left\{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\right\}^{a-1} \\ &\times \frac{\left\{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}\right\}^{b-1}}{B(a,b)\left\{1-l_{1-e^{-\frac{\alpha}{2}\left(\frac{\theta}{t+\theta}\right)^{-2\alpha}}}(a,b)\right\}} \end{aligned} \quad (3)$$

The *pdf* of T can be re-written the distribution in another Rayleigh version as follows

$$f(y)=\lambda\left(\frac{\theta}{\alpha}\right)\frac{\theta}{y}-2+1\left(\frac{y}{\alpha}\right)^{\theta}.exp\left(-\frac{\alpha}{2}\left(\frac{y}{\alpha}\right)^{\theta}\right)^{\frac{2}{\theta}+1} \quad (4)$$

Then, for easy verification on the density function Y can be obtained by replacing $y=log(t)$ that is $t=e^y$, $\theta=\frac{1}{\sigma}$ and $\mu=log(\alpha)$ i.e $\alpha=e^{\mu}$ which becomes, for $y\in\mathbb{R}$,

$$\begin{aligned} f(y;a,b,\lambda,\mu,\sigma) &= \frac{\left\{\frac{\lambda}{\sigma}exp\left(\frac{y-\mu}{\sigma}\right)-2\lambda+1\left(exp\left(\frac{y-\mu}{\sigma}\right)\right)exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}}{B(a,b)} \\ &\times\left\{1-exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right]^{a-1}\left[exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{b-1}, \end{aligned} \quad (5)$$

where $-\infty<\mu<\infty, \sigma>0, \lambda>0, a>0$ and $b>0$. Therefore, Equation (5) is refer to as the *LBRL* distribution, say $Y \approx LBRL(\mu,\sigma,\lambda,a,b)$ where $-\infty<\mu<\infty$ is the location parameter, $\sigma>0$ is the scale parameter and $\lambda>0, a>0$ and $b>0$ are shape parameters. Hence, the new *LGBRL* distribution contains some well-known and unknown distributions as special sub-models. These are: log-Exponential Lomax distribution when $a = b = 1$, log-Beta Rayleigh (*LBR*) distribution if $\lambda = 1$, log-Beta Lomax (*LBL*) distribution as $\mu = 1$, log-Lomax (*LL*) distribution if $a = b = \mu = 1$, log-Rayleigh Lomax (*LRL*) distribution if $a = b = 1$, log-Rayleigh (LR) distribution when $a = b = \lambda = 1$. The new distributions emanate from the propose distribution are: log-Lehmann Rayleigh Lomax (*LLRL*) distribution when $a = 1$ and if $b = 1$ we have log-exponentiated Rayleigh Lomax (*LERL*) distribution.

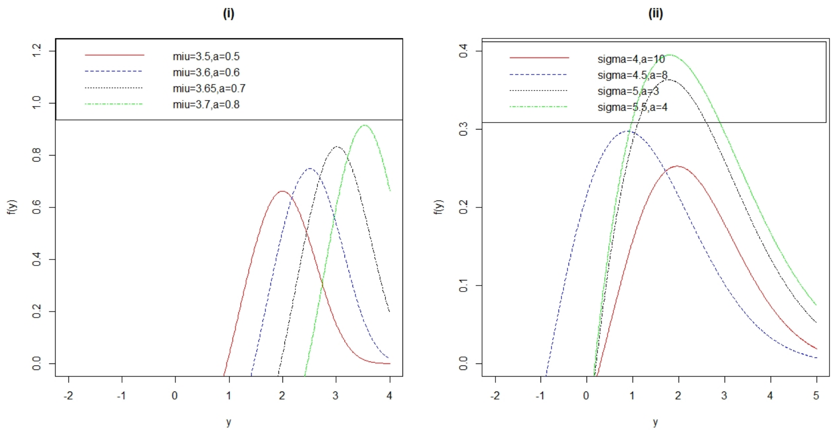

Fig. 1. The *LBRL* density curves: (i) As some values of μ increasing and a increasing with $\sigma = 1, \lambda = 5$ and $b = 1.7$. (ii) Also, for some values of σ increasing and a decreasing with $\mu = 2.5$ and $\lambda = b = 2$

The *LBRL*-reliability function is

$$R(y) = 1 - \ell_{\left\{1-\exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}}(a, b), \tag{6}$$

Furthermore, the random variable $Z = (Y - \mu)/\sigma$ has the following *pdf*

$$\begin{aligned} f(z) &= \frac{\left\{\dfrac{\lambda}{\sigma} exp\,(z) - 2\lambda + 1\,(exp\,(z))\,exp\,(-2\lambda exp\,(z))\right\}}{B(a,b)} \\ &\times \left\{1 - exp\,(-2\lambda exp\,(z))\right\}^{a-1} \left\{(-2\lambda exp\,(z))\right\}^{b-1}, -\infty < y < \infty \end{aligned} \tag{7}$$

We obtain two simple formulae for the distribution function (cdf) and density function (pdf) of the *LBRL* distribution depending if the parameter $b > 0$ is real non-integer or integer.

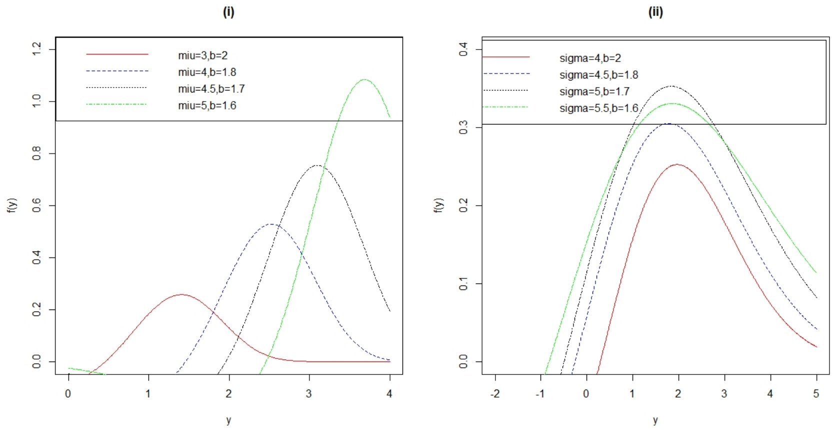

Fig. 2. The *LBRL* density curves: (i) As some values of μ increasing and b decreasing with $\sigma = 1, \lambda = 5$ and $a = 1.7$. (ii) Also, for some values of σ increasing and b decreasing with $\mu = 2.5$ and $\lambda = a = 2$

Theorem 1. *Let $Y \sim$ LBRL$(\mu, \sigma, \lambda, a, b)$ with $a > 0$ and ¿0 being non-integers. Then, we the following facts.*
(1) Its cdf is given by

$$F(y) = \frac{1}{B(a,b)} \lambda \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j! (a+j)} A(y), \; y \in \mathbb{R}, \tag{8}$$

where, $A(y) = \left\{1 - exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{a+j}$ *(as defined in Pescim et al.(2013)).*

If $|z| < 1$ and $b > 0$ is real non-integer, we have

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j!} z^j, \tag{9}$$

where $z^j = \int_0^{A(y)} k^{a+j-1} dk$

(2) Its pdf is given by,

$$\begin{aligned} f(y) \;=\; & \left\{\frac{\lambda}{\sigma} exp\left(\frac{y-\mu}{\sigma}\right) - 2\lambda + 1\left(exp\left(\frac{y-\mu}{\sigma}\right)\right) exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\} \\ & \times \sum_{j,w=0}^{\infty} \sum_{r=0}^{w} K_{j,w,r}(a,b) \left\{1 - exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{r}, \end{aligned} \tag{10}$$

where

$$K_j, w, r(a,b) = \frac{(-1)^{j+w+r}\Gamma(a+b)\Gamma(a+j)\binom{w}{r}}{\Gamma(a)\Gamma(b-j)\Gamma(a+j-w)j!w!}.$$

Proof of Theorem 1. Let us proceed by parts.

Proof of (i). By differentiating (8) under Theorem 1, yields

$$\begin{aligned} f(y) &= \frac{1}{\sigma\Gamma(a)}\left\{\lambda exp\left(\frac{y-\mu}{\sigma}\right) - 2\lambda + 1\left(exp\left(\frac{y-\mu}{\sigma}\right)\right) exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\} \\ &\times \sum_{j,w=0}^{\infty} \frac{(-1)^j\Gamma(a+b)}{\Gamma(b-j)j!}\left\{exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{a+j-1}. \end{aligned} \tag{11}$$

The proof of Part (1) is finished. □

Proof of part (2). By using (9) in (10) for $a > 0$ real non-integer and its expand as:

$$\begin{aligned} f(y) &= \frac{1}{\sigma\Gamma(a)}\left\{\lambda exp\left(\frac{y-\mu}{\sigma}\right) - 2\lambda + 1\left(exp\left(\frac{y-\mu}{\sigma}\right)\right) exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\} \\ &\times \sum_{j,w=0}^{\infty} \frac{(-1)^{j+w}\Gamma(a+b)\Gamma(a+j)}{\Gamma(a)\Gamma(b-j)\Gamma(a+j-w)j!w!}\left\{exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{w}. \end{aligned} \tag{12}$$

After some algebra, we obtain

$$\begin{aligned} f(y) &= \frac{1}{\sigma}\left\{\lambda exp\left(\frac{y-\mu}{\sigma}\right) - 2\lambda + 1\left(exp\left(\frac{y-\mu}{\sigma}\right)\right) exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\} \\ &\times \sum_{j,w=0}^{\infty}\sum_{r=0}^{w} K_{j,w,r}(a,b)\left\{1 - exp\left(-2\lambda exp\left(\frac{y-\mu}{\sigma}\right)\right)\right\}^{r} \end{aligned} \tag{13}$$

where $K_j, w, r(a,b) = \dfrac{(-1)^{j+w+r}\Gamma(a+b)\Gamma(a+j)\binom{w}{r}}{\Gamma(a)\Gamma(b-j)\Gamma(a+j-w)j!w!}$. The proof of Part (2) and of the theorem is now complete. ■

2.1. Properties of the LBRL distribution

Here, we present some properties of the standardized *LBRL* random variable defined by $Z = (Y - \mu)/\sigma$ are studied and expressed in (7). The *pdf* of Z becomes

$$\begin{aligned}\eta(z;\lambda,a,b) &= \frac{\left\{\frac{\lambda}{\sigma} exp\left(z\right) - 2\lambda + 1\left(exp\left(z\right)\right) exp\left(-2\lambda exp\left(z\right)\right)\right\}}{B(a,b)} \\ &\times \left\{1 - exp\left(-2\lambda exp\left(z\right)\right)\right]^{a-1} \left[exp\left(-2\lambda exp\left(z\right)\right)\right\}^{b-1}, -\infty < y < \infty. \end{aligned} \qquad (14)$$

The associated cumulative distribution function (*cdf*) is

$$F_z(Z) = 1 - l_{\{1-exp(-2\lambda exp(z))\}}(a,b).$$

The condition $a = b = 1$ is associated with the standardized Rayleigh Lomax distribution.
Hence, we expand the expression in (14) using binomial expansion; Ortega *et al.*(2009). This gives

$$\begin{aligned}\eta(z;\lambda,a,b) &= \frac{1}{B(a,b)} \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} \\ &\times \left\{\frac{\lambda}{\sigma} exp\left(z\right) - 2\lambda + 1\left(exp\left(z\right)\right) exp\left(-2\lambda exp\left(z\right)\right)\right\} \left\{1 - exp\left(-2\lambda exp\left(z\right)\right)\right\}^{a(i+1)-1} \end{aligned} \qquad (15)$$

Also, the density function $d_a = (a-1)\left[1 - exp\left(-2\lambda exp\left(z\right)\right)\right]$ for greater than zero yields Lehmann type II Rayleigh Lomax; and its corresponding cumulative function is

$$D_a^{(z)} = 1 - \left\{exp\left(-2\lambda exp\left(z\right)\right)\right\}^b$$

Fortunately, $\eta(z;\lambda,a,b) = \sum_{i=1}^{\infty} \varphi_i d_{a(i+1)}(z)$ where the coefficients are

$$\varphi_i = \frac{(-1)^i \binom{b-1}{i}}{B(a,b)a(j+1)}.$$

The *LBRL* distribution function also can be written as a linear combination of LRL densities when equating some parameters to 1 (one). Those distributions have been written immediately after (5).

2.2. Moments and Generating Function

We obtain the s^{th} ordinary moment of the *LBRL* distribution (14) as follows:

$$\begin{aligned} \mu_z^s = E(Z)^s \ &= \frac{1}{B(a,b)} \int_{-\infty}^{\infty} Z^s \left\{ \frac{\lambda}{\sigma} exp\,(z) - 2\lambda + 1\,(exp\,(z))\,exp\,(-2\lambda exp\,(z)) \right\} \\ &\times \ \{1 - exp\,(-2\lambda exp\,(z))\}^{a-1} \{exp\,(-2\lambda exp\,(z))\}^{b-1} dz. \end{aligned}$$

By expanding the binomial term and setting $v = e^s$, gives

$$\begin{aligned} \mu_v^s \ &= \frac{1}{B(a,b)} \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} \int_0^{\infty} logv^s \\ &\times \ \left\{ \left\{ \frac{\lambda}{\sigma} exp\,(v) - 2\lambda + 1\,(exp\,(v))\,exp\,(-2\lambda exp\,(v)) \right\} \{1 - exp\,(-2\lambda exp\,(v))\}^{a(i+1)-1} v \right\} dv. \end{aligned}$$

$$l_{(s,(i+1))} = \frac{\partial}{\partial p}^{s} \left[(a(i+1))^{-p} \Gamma(p) \right] |_{p=1}$$

(see Pascoa *et al.*(2013)). Thus

$$\mu_s^{'} = \frac{1}{B(a,b)} \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} |_{s,(i+1)}. \tag{16}$$

Equation (16) becomes the moments of the *LBRL* distribution. Wherefore, the measures are being controlled by parameters a and b. In the same vein, the moment generating function (mgf) of Z is given by

$$M(t) = E(e^{tz}).$$

It can be written from (14) as

$$\begin{aligned} \mu_v^s \ &= \frac{1}{B(a,b)} \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} \int_0^{\infty} logV^s \\ &\times \ \left\{ \left\{ \frac{\lambda}{\sigma} exp\,(v) - 2\lambda + 1\,(exp\,(v))\,exp\,(-2\lambda exp\,(v)) \right\} \{1 - exp\,(-2\lambda exp\,(v))\}^{a(i+1)-1} v \right\} dv \end{aligned}$$

and

$$M(t)=\frac{\Gamma(t+1)}{B(a,b)}\sum_{i=0}^{\infty}(-1)^{i}\binom{b-1}{i}\left[(a(i+1)-1)\right]^{-(t+1)} \tag{17}$$

Using differentiation method, moment (17)) can be derived from (14).

3. The Log-Beta Rayleigh Lomax Regression Model (*LBRLRM*)

Practically speaking, regression model consists of two different/kinds of variables namely response y and explanatory $x(s)$ variables. The explanatory variables such as blood pressure, cancer, heart problem, weight and others affect lifetimes. Therefore, parametric models have been used for estimating univariate models and censored data regression problems are widely. The advantage is that, the parametric models always provide good fit to the lifetime data set and give more precise estimates of the quantities of interest. Let $x=(x_1,...,x_p)^T$ be the explanatory variable vector associated with the response variable $y_i=log(t_i)$, that is, y_i is the logarithm of the survival time t_i . Now, based on the *LBRL* distribution, a linear regression model linking the response variable y_i and the explanatory variable vector x_i can be defined by

$$y_i=X_i^T\beta+\sigma z_i,\quad i=1,2,...,n \tag{18}$$

where the random error z_i has density function (14) with parameters $\beta=(\beta_1,...,\beta_p)^T$, $\sigma>0$, $a>0$, $b>0$ and $\lambda>0$ are unknown parameters., and $X_i^T=(x_i,\cdots,x_p)$ is the explanatory variable vector modelling the linear predictor $\mu_i=X_i^T\beta$. The linear predictor vector $\mu=(\mu_1,\cdots,\mu_n)^T$ of the *LBRL* regression model is written as $\mu=X\beta$, where $X=(x_1,...,x_n)^T$ is a known model matrix.

Hence, consider a sample $(y_i,x_i),...,(y_n,x_n)$ of n -independent observations, where each random response is defined by $y_i=\min\{\log(t_i),\log(c_i)\}$. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. The log-likelihood function for the model parameters $\gamma=(\lambda,\sigma,\beta^T)^T$ follows from (14) and (18) as

$$\begin{aligned} l(\gamma) &= -rlog\left[log(\sigma)+log\{B(a,b)\}\right]\left\{\sum_{i\in F}\frac{\lambda}{\sigma}exp\left(z_i\right)-2\sum_{i\in F}(\lambda+1)\left(exp\left(z_i\right)\right)exp\left(-2\lambda exp\left(z_i\right)\right)\right\}\\ &+ (a-1)\sum_{i\in F}log\left[1-exp\left(-2\lambda exp\left(z_i\right)\right)\right]+(b-1)\sum_{i\in F}log\left[exp\left(-2\lambda exp\left(z_i\right)\right)\right]\\ &+ \sum_{i\in C}log1-l_{\{1-exp(-2\lambda exp(z_i))\}}(a,b), \end{aligned} \tag{19}$$

where r is the number of uncensored observations (failures) and $z_i = (y - X_i^T \beta)$. The MLE $\hat{\gamma}$ of the parameters vector $\gamma = (a, b, \lambda, \sigma, \beta^T)^T$ of the *LBRL* regression model can be obtained by maximizing the log-likelihood function in (19). The computation process is easy and we used R programming language (maxLik) to estimate $\hat{\gamma}$. After Model (18) is fitted, the reliability function for Y say $P(Y \leq y) = S(y; a, b, \lambda, \sigma, \beta^T)$ can be estimated as

$$S(y; \hat{a}, \hat{b}, \hat{\lambda}, \hat{\sigma}, \hat{\beta}^T) = 1 - l_{\left\{1 - exp\left(-2\hat{\lambda} exp\left(\frac{y - X_i^T \hat{\beta}}{\hat{\sigma}}\right)\right)\right\}}(\hat{a}, \hat{b}). \tag{20}$$

The asymptotic normality is a useful tool for testing goodness of fit of some sub-models (known and unknown) and for comparing some special sub-models using the likelihood ratio (*LR*) statistic. The *LR* statistic for testing the null hypothesis $H_0 : \gamma_1 = \gamma_2$ versus the alternative hypothesis $H_1 : \gamma_1 \neq \gamma_2$ is given by $\psi = 2\{l(\hat{\gamma}) - l(\breve{\gamma})\}$ where $\hat{\gamma}$ and $\breve{\gamma}$ are both the estimate under null and alternative hypotheses, respectively. The statistic ψ is asymptotically as $(n \to \infty)$ distributed as χ^2_q, where q is the dimension of the subset of parameters γ_1 of interest Silva *et al.*(2009).

3.1. Diagnostic Residual Plot

In regression model, it is very important to know the nature of the variables in the model especially the response variable considering in the analysis and the distribution must follow the distribution of the response variable. For example, suppose the response variable is normal it will be normally distributed but in this study, the response variable does not follow normal distribution therefore a more robust distribution is required like *LBRL* distribution. Diagnostic residual plots Figures **3** and 4 show the nature of the response variable consider in the study.

4. Breast Cancer Survival Data

We put in practice the proposed methodology by using breast cancer data set referring to time spent (t) and the explanatory variables : age x_1, occupation x_2, marital status x_3, event status x_4 and type of treatment x_5. The data contain $n = 623$ observations and follow model:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \sigma z_i, i = 1, 2, ..., 623 \tag{21}$$

where, • $y =$ time spent by each patient

- $x_1 =$ Age (1 if 19 – 28, 2 if 29 - 38 8 if 89 – 98)
- $x_2 =$ Occupation (1 if Accountant, 2 if Administration 99 if Unknown)
- $x_3 =$ Marital Status (1 if single, 2 if married and 3 if divorced)
- $x_4 =$ Event (0 if dead, 1 if alive and 9 if unknown)

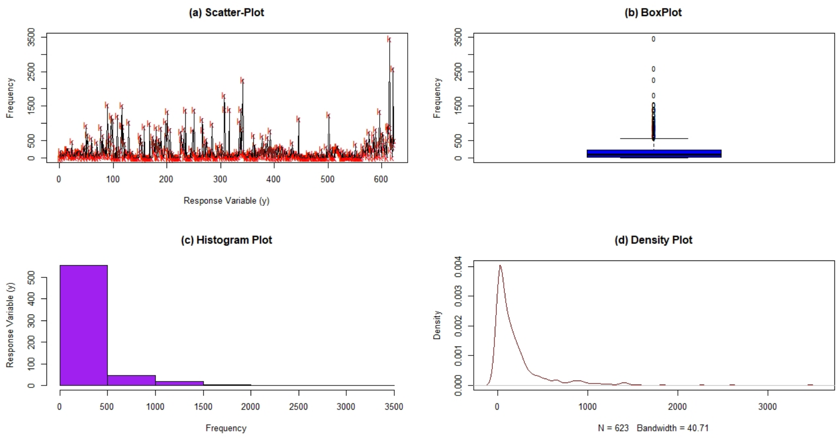

Fig. 3. Diagnostic Plots: The line Scatter, Boxplot, Histogram and Density plot of Response Variable

- x_5 = Types of Treatment (0 if none, 1 if surgery, 2 if radiotherapy 5 if unknown) and βs are regression parameters

4.1. Results

We fitted the *BRLRM*, *LRLRM*, *ERLRM*, *BRRM*. *BLRM* and *RLRM* regression models to breast cancer data. Table 1 contains the estimates, Standard Errors (given in parenthesis) and the probability values [given in braces] of the parameters for regression models. Then, selection of the better model among those considered in the study based on the values of the statistics (model selection criteria) *AIC*, *BIC* and *CAIC*. The statistic gives the values of *AIC*, *BIC* and *CAIC* in Table 2 below. The values of the *BRLRM* are the smallest values compare to other models and these statistics indicates that the *ERL* regression model is more adequate to explain the data set than other models. In addition, we carry out the LR test of non-nested models and the generalized *LR* test statistics gives $TR_{LR,NN} = 206.720$. Therefore, $TR_{LR,NN} > 1.96$, it implies that this fall in the rejection region and we reject at significance level 0.05 the null hypothesis of equivalence of the *BRLRM* and *RLRM* models. Fortunately, the value of this statistic is in line with the previous result and helps in selecting the BRLRM regression model Silva *et al.*(2009).

The estimates of the regression parameters are so close for the models but their standard errors are different., and the conclusions may be different for the models. Also, the residual plots are presented in order to detect possible outliers in the response variable observations. Figure 4 shows the plots of the residuals, contains residual vs fitted, Normal *QQ* plots, scale-location and residual vs leverage of normal plots. Meanwhile, the fitted *BRLRM* model is stated as

Table 1. *MLE*'s of the parameters from the baseline, proposed and special cases of the proposed distributions fitted to the breast cancer patient data set, the corresponding Estimates, Standard Errors (given in parenthesis) and the probability values [given in braces]. **Abbreviation: p/d** (Parameter/distribution)

p/d	*BRLRM*	*LRLRM*	*ERLRM*	*BRRM*	*BLRM*	*RLRM*
a	0.200 (0.009) [<2e-16***]	1	2.500 (0.138) [<2e-16***]	2.500 (0.133) [<2e-16***]	2.500 (0.124) [<2e-16***]	1
b	1.500 (0.118) [<2e-16***]	2.500 (0.127) [<2e-16***]	1	4.500 (0.253) [<2e-16***]	4.500 (0.241) [<2e-16***]	1
θ	0.500 (0.021) [<2e-16***]	1.500 (0.073) [<2e-16***]	1.500 (0.078) [<2e-16***]	1.500 (0.085) [<2e-16***]	1	1.500 (0.073) [<2e-16***]
λ	1.500 (0.002) [<2e-16***]	0.200 (0.008) [<2e-16***]	1.200 (0.009) [<2e-16***]	1	1.200 (0.002) [<2e-16***]	1.200 (0.008) [<2e-16***]
β_0	0.300 (0.172) [0.081]	0.600 (2.112) [0.776]	0.600 (1.113) [0.590]	0.800 (1.337) [0.550]	0.800 (0.134) [<2.2e-09***]	0.600 (2.112) [0.776]
β_1	1.300 (0.013) [<2e-16***]	1.200 (NA) [NA]	1.200 (NA) [NA]	1.200 (1.443) [0.405]	1.200 (0.637) [0.060]	1.200 (NA) [NA]
β_2	1.300 (0.023) [<2e-16***]	1.500 (0.029) [<2e-16***]	1.800 (0.022) [<2e-16***]	1.500 (0.087) [<2e-16***]	1.500 (0.038) [<2e-16***]	1.500 (0.029) [<2e-16***]
β_3	1.200 (3.479) [0.730]	0.200 (2.042) [0.922]	0.200 (NA) [NA]	0.200 (1.127) [0.859]	0.200 (0.911) [0.826]	0.200 (2.042) [0.922]
β_4	0.500 (0.019) [<2e-16***]	0.500 (0.008) [<2e-16***]	0.500 (0.004) [<2e-16***]	0.800 (0.001) [<2e-16***]	0.500 (0.006) [<2e-16***]	0.500 (0.008) [<2e-16***]
β_5	1.500 (0.532) [0.500]	2.500 (0.102) [<2e-16***]	1.500 (0.590) [0.011]	1.500 (0.020) [<2e-16***]	0.500 (0.372) [0.180]	2.500 (0.102) [<2e-16***]

$$\begin{aligned} Timespent(y) &= 0.300 + 1.300(Age) + 1.300(Occupation) \\ &+ 1.200(Marital - Status) + 0.500(Event) + 1.500(Types) \end{aligned} \tag{22}$$

5. Conclusion

In the study, we properly derived a new distribution using the transformation method on Rayleigh Lomax introduced by Kawsar *et al.*(2018) to what we called log-Beta Rayleigh Lomax distribution, which is able to accommodate lifetime data

Table 2. The log-likelihood and model selection criteria for the models

Model	-2Log-lik	AIC	BIC	CAIC
BRLRM	289.110*	600.220*	649.000*	650.000*
LRLRM	386.620	793.240	837.586	838.586
ERLRM	381.290	782.580	826.926	827.926
BRRM	470.640	961.280	1005.625	1006.625
BLRM	359.590	739.180	783.526	784.526
RLRM	392.470	802.940	842.851	843.851

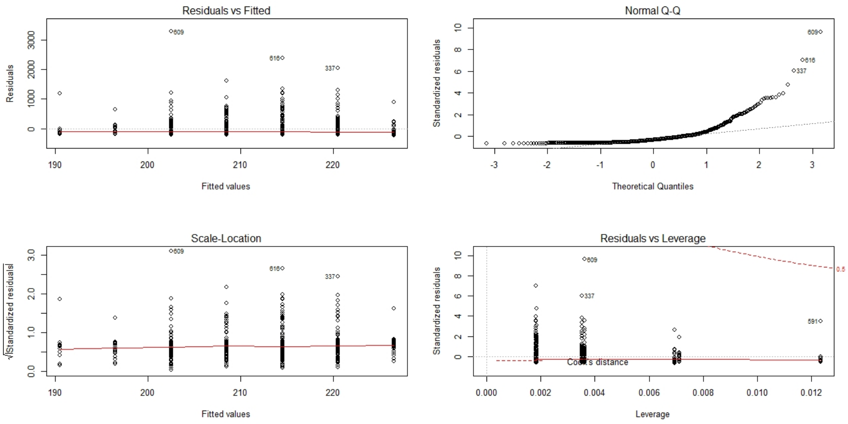

Fig. 4. Residual Plots: Residual vs Fitted, Normal QQ, Scale-Location and Residual vs Leverage of Response Variable

that skewed in nature. Some of its properties are obtained such as reliability function, binomial expansion, moment and moment generating function even transformed to model called the beta Rayleigh Lomax regression model and so on. Based on the new distribution, we develop a *BRL* regression model to compete with known and unknown regression models like *BRLRM*, *ERLRM*, *BRRM*. *BLRM* and (*RLRM*, Silva *et al.*(2009)). A breast cancer real data set is analyzed to show the performance of the proposed regression model. We even show that the *BRL* regression model has better performance than other regression models *ERLRM*, *BRRM*. *BLRM* and *RLRM* for these data.

Acknowledgments. The authors would like to thank the referee and the editor for carefully reading the paper and their valuable comments, which greatly improved the presentation in this paper.

Conflict of interest. The authors declare that no conflict of interests.

References

Akarawak, E., Adeleke, I. A., and Okafor, R. O. (2017). The Gamma-Rayleigh Distribution and Applications to Survival Data. *Nigerian Journal of Basic and Applied Science.* 130-142.

Badmus, N. I., Alakija, T. O., and Olanrewaju, G. O. (2017). The Beta-Modified Weighted Rayleigh Distribution. *International Journal of Mathematical Analysis and Optimization: Theory and Applications.* 154-167.

Kawsar, F., Uzma J. and Ahmad S. P. (2018). Statistical Properties of Rayleigh Lomax distribution with applications in Survival Analysis *Journal of Data Science.* 531-548.

Tahir, M. H., Gauss , M. C., Mansoorx , M., and Zubair, M. (2010). The WeibullL-Lomax Distribution: Properties And Applications. *Bhawalphur, Pakistan.*

Lomax, K. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association.* **49**, 847–852.

Jones, M. C. (2004). Families of distributions arising from distributions of order Statistics. *Test*, 13 (1), 1 - 43.

Siddiqui, M. M. (1962). Some problems connected with Rayleigh distributions. *Res. Nat. Bur. Stand.* 60D, 167–174.

Tahir, M. H., Hussain, M. A., Cordeiro, G. M., Hamedani, G.G., Mansoor, M and Zubair, M. (2016). The Gumbel-Lomax Distribution: Properties and Applications. *Journal of Statistical Theory and Applications.* **15**(1), 61-79.

El-Bassiouny, A. H., Abdo, N. F. and Shahen, H. S. (2015). Exponential Lomax Distribution. *International Journal of Computer Applications.* **121**(13), 24-29.

Hassan, A. S. and Abd-Allah, M. (2018). Exponentiated Weibull-Lomax Distribution: Properties and Estimation. *Journal of Data Science.* **16**(2), 277-298.

Carrasco, J. M. F., Ortega, E. M. M and Paula, G. A (2008). Log-modified Weibull regression models with censored data: Sensitivity and residual analysis. *Computational Statistics and Data Analysis.* 52, 4021 - 4029.

Ortega, E. M, M., Cordeiro, G. M and Carrasco, J. M. F. (2009). The Log-Generalized Modified Weibull Regression Model. *Brazilian Journal of Probability and Statistics.* **25**(1), 64-89.

Pescim, R. R., Demetrio, C. G. B., Cordeiro, G. M., Ortega, E. M. M and Urbano, M. R. (2013). The log-beta generalized half-normal distribution. text *Journal of Statistical Theory and Applications. **12**(4), 330-347.*

Pascoa, M. A. R., De Paiva, C. M. M., Cordeiro, G. M and Ortega, E. M. M. (2013). The Log-Kumara swamy Generalized Gamma Regression Model with Application to Chemical Dependency Data. Journal of Data Science. 11, 781-818.

*Hashimoto, E. M., Ortega, E. M. M., Cancho, V. G and Cordeiro, G. M. (2010). The Log Exponentiated Weibull Model for Interval-Censored Data. Journal of Computational Statistics and Data Analysis. **54**(4), 1017-1035.*

Silva, G. O., Edwin M. M. Ortega and Gauss M. Cordeiro. (2009). A log-extended Weibull regression model. Computational Statistics and Data Analysis. Article in press.