Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 3395-3403
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6100

Comparison between two new censored regression models extended from Burr-XII system with application

Anwaar Dhiaa Abdulkareem[a,*], Sunbul Rasheed Mohammed[b]

[a]College Education for Pure Sciences, University of Kirkuk, Iraq
[b]College of Administration and Economics, University of Kirkuk, Iraq

(Communicated by Madjid Eshaghi Gordji)

Abstract

This article presents a comparison between two new censored regression models which are extended from some continuous distribution with the Burr-XII system. The two models are the Log-BXII Weibull model(LBXIIW) and the Log-BXII Exponentiated Exponential model(LBXIIEE). The results of the comparison showed the LBXIIW model is best than the LBXIIEE model according to the values of the model selection criteria when the Creatine is a dependent variable in the model.

Keywords: Burr-XII System, Log-BXII Exponentiated Exponential model, Log-BXII Weibull model

1. Introduction

Regression models are one of the basic statistical methods in studying and analyzing the relationship between variables for any phenomenon. Regression models show the shape of the relationship to these variables and the extent of the influence of the explanatory variables on the prediction of the dependent variable according to each phenomenon or problem. Recently, many models have been proposed through location regression models. The researcher Cordeiro presented a class of regression models based on the Log-Gamma model extended from the Weibull distribution. The researcher Hashimoto also proposed a log-BurrXII regression model for a set of survival data, among others [4]. Many univariate distributions have been used extensively for data modelling in many fields such as

*Corresponding author
Email addresses: anwaar71@uokirkuk.edu.iq (Anwaar Dhiaa Abdulkareem), sunbulr79@uokirkuk.edu.iq (Sunbul Rasheed Mohammed)

Received: June 2021 *Accepted:* October 2021

economics, engineering, biological studies, and environmental sciences. However, some applied fields that have extended forms of these distributions have important data-related problems that do not follow any of the classic statistical models [3].

It was necessary to model some distributions for the purpose of obtaining new shapes and models that are able to identify new regression models that can be used in studies and analysis of survival and reliability within censored data so that they become censored regression models.

One of the first attempts to analyze a statistical problem that includes censored data was Daniel's analysis, which was reached by Daniel Bernoulli in 1766 for data that shows the prevalence of a particular disease and the deaths related to that disease. An example of this is the data of smallpox disease, as there is an urgent need to prove the statistical importance of vaccination (Bernoulli.& Blower(2004))[2]. **Patricia M. Odell** [8] in 1992 used a regression model based on the Weibull distribution when observing left-handed and time-interval-censored data. Two methods of analysis were considered, the first of which is to calculate the estimates of (MLE) for the observed censored pattern. This method was compared with the estimates where the midpoints were replaced by the left- and time-lapsed censored data (the midpoint estimator MDE), and the simulation studies used here indicate that there are relatively large samples, and there are many cases where MLE is superior to MDE for samples in which it is failure is likely to occur, as the percentage of interval-censored data is small and therefore MDE is appropriate. An example using data from the Framingham Heart Study is discussed. In 2010, **Patricia F.Paranaibe and Edwin M.M** [9]. proposed the ML method and Bayesian analysis for estimating parameters of a new model of BurrXII system which contains some distributions known in the analysis and study of censored data such as logistic distribution and Weibull distribution. And they used simulations to show the results of their own study. In 2019, researchers **Morad Alizadeh and Mahdi Emadi** [1] reviewed a new distribution consisting of two parameters called log-logistic half-logistic and presented the theoretical properties of this model, including the risk function, survival function and skew coefficients. The ML method estimates were compared with different methods through Conducting simulations using real data for the purposes of the study.

The current article aims to review new censored regression models, as well as estimation of parameters of these models which are Log-BurrXII Exponentiated Exponential Model (LBXIIEE model), and Log-BurrXII Weibull Model, as well as estimation of parameters of the two models in the presence of data subject to the Maximum Likelihood Estimation (MLE) method, and comparison between two models in the presence of censored data and choosing the best model between them through the use of some statistical criteria in this field by applying the estimators to real data for patients with renal failure.

2. Log-BXII Weibull (LBXIIW) Model

The Weibull distribution was introduced by the Swedish scientist Waloddi Weibull in 1939. This distribution was named after him as he was the first to present it to represent the distribution of breaking strength of materials, although this distribution was used earlier by the French scientist Maurice in 1927 and was also applied by R. Rosin And E. Rammler in 1933, it was the scientist Weibull who formally defined it as one of the continuous distributions and one of the most common failure models in the past years, and it became of great importance and a wide place in the fields of reliability and life tests [10].

Kao (1958 - 1959) called for the use of this distribution in the field of reliability and quality, and this distribution is often appropriate when the strict randomness conditions of the exponential distribution are not met [4], and thus this distribution adds flexibility, and it has several forms in terms of the number of parameters One parameter, two parameters, and three parameters [11]. The mathematical formula of a two-parameter Weibull distribution is as follows:

Let x be a random variable representing the failure time, then both the probability density function and the cumulative distribution function of the Weibull distribution are expressed as follows [7, 11, 12]:

$$f(x;\alpha,\beta)=\alpha\beta^{-1}(x/\beta)^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right) \tag{2.1}$$

$$F(x;\alpha,\beta)=1-\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right) \tag{2.2}$$

where $x\geq 0, \alpha\geq 0, \beta\geq 0$, x is the random variable, α is the shape parameter and β is the scale parameter.

Table 1: Some characteristics of the Weibull distribution

Properties	Formulas
Mean	$\beta\Gamma(1+\frac{1}{\alpha})$
Variance	$\beta^2\ \Gamma\left(1+\frac{2}{\alpha}\right)-M^2$
$M(t)$	$\sum_{n=0}^{\infty}\frac{t^n\beta^n}{n!}\Gamma(1+\frac{n}{\alpha})$

As we mentioned earlier that there are many newly proposed regressions models during the past years, including the log-BurrXII Weibull model (LBXIIW), which is a model that is constructed using the Location Regression model and the Burr distribution with Weibull. And the mathematical formula in which BXIIW can be expressed is (Cordeiro,G.(2018))[3]:

$$f(x)=ab\alpha\beta^{-1}(x/\beta)^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)^{\alpha}\frac{\left[1-\exp\left(-\frac{x}{\beta}\right)^{\alpha}\right]^{a-1}}{\left[1-\exp\left(-\frac{x}{\beta}\right)^{\alpha}\right]^{a+1}}\left[1+\left[\frac{\left(1+\exp\left(-\frac{x}{\beta}\right)^{\alpha}\right)}{1-(\exp\left(-\frac{x}{\beta}\right)^{\alpha})}\right]^{a}\right]^{-(b+1)} \tag{2.3}$$

$$x\ \sim\ BXIIW\ (\alpha,\beta,a,b)$$

where a, b parameters of the Burr distribution, x is a random variable, and location regression models can be distinguished by that the random variable $y = log(x)$ has a distribution, where position $\mu(\nu)$ depends on the explanatory variable vector (ν), in practice this means that for different values or levels of explanatory variables x has different values for the position parameter. The site regression model can be expressed as follows:

$$y=\mu(\nu)+\sigma Z, \tag{2.4}$$

where Z has a distribution that does not depend on (ν) and the random variable y for each $(y\in R)$ has a probability function and considering that the random variable x has a probability function (2.3) and is distributed BXIIW and that the random variable $y=\log(x)$,

$$\alpha=e^{-\mu}$$

and

$$\beta = \frac{1}{\sigma}.$$

The probability function (2.11) is as follows:

$$f(y;a,b,\mu,\sigma) = \frac{ab}{\sigma}\ \frac{\left[1-\left[\exp\left(-e^{(y-\mu)/\sigma}\right)\right]\right]^{a-1}}{\exp\left[-ae^{(y-\mu)/\sigma} - \left(\frac{y-\mu}{\sigma}\right)\right]}\ \left[1+\left[\exp\left(e^{(y-\mu)/\sigma}\right)\ -1\right]^{a}\right]^{-(b+1)} \tag{2.5}$$

where $a>0, b>0$ are the two parameters of the shape and that $\mu\in R$ is the location parameter and $\sigma>0$ is the scale parameter. Since the above equation represents the LBXIIW distribution and $Y\sim$ LBXIIW (a, b, μ, σ).

Since $Y\sim$ LBXIIW (a, b, μ, σ) and the standard random variable $Z = (Y-\mu)/\sigma$, the probability density function for Z is as follows:

$$f(z;a,b) = ab\frac{[1-exp(-e^{z})]^{a-1}}{\exp\left[-e^{z}-z\right]}[1+[\exp\left(e^{z}\right)\ -1]^{a}]^{-(b+1)}, \tag{2.6}$$

and a linear regression model that links the response variable y_i and the explanatory variable vector $\nu_i{}^T = (\nu_{i1}, \ldots ., \nu_{ip})$,T is a random variable that follows a BXIIW distribution
It will be like this:

$$y_i = \mu_i + \sigma z_i \qquad\qquad i = 1, \ldots, n \tag{2.7}$$

where

z_i : the random error has a probability density function (2.6);
μ_i: the location parameter of y_i;
$\mu_i = \nu_i{}^T\tau$, $\tau = (\tau_1, \ldots, \tau_p)^T$ vector related to explanatory variables;
$a > 0, b > 0$, σ: unknown parameters.

We can use the LBXIIW model to fit different types of data where the explanatory variables have significant effects on the average response to the Y variable.

Let G, H be a set of observations for y_i that represent log-lifetime or log-censoring respectively and by applying the method of MLE and taking the natural logarithm of equation (2.7) to estimate the parameters of the vector $\theta = (a, b, \tau^T, \sigma)$ will be as following:

$$u_i = e^{z_i} \qquad\qquad z_i = \frac{(y_i - \mu_i)}{\sigma}$$

r = number of failed observations.

The equation will be as follows:

$$\begin{aligned} \ln(\theta) = & r\ln(ab) - r\ln\sigma + (a-1)\sum_{i\in G}\ln\left(1-e^{-u_i}\right) + a\sum_{i\epsilon G}(u_i + z_i) \\ & -(b+1)\sum_{i\in G}\ln\left(1+\left(e^{u_i-1}\right)^{a}\right) - b\sum_{i\epsilon H}\ln\left(1+\left(\frac{1-e^{-u_i}}{e^{-u_i}}\right)^{a}\right). \end{aligned} \tag{2.8}$$

The estimations for the parameters vector $\widehat{\theta}$ can be found using some statistical software, including NLMixed in the SAS software.

3. Log-BXII Exponentiated Exponential Model

It is a new continuous distribution of families of the exponential distribution which was discussed by Gupta (1998) and this distribution contains two parameters which are the scaling parameter and the shape parameter similar to the Weibull distribution family and the Gamma distribution. It was noted that the properties of this distribution are somewhat similar to the properties of the Weibull and Gamma distributions, and it can also be used as a possible alternative to them (Gupta,(2001))[5].

The probability density function pdf of the EE distribution can be expressed as follows:

$$f(x;\alpha,\lambda)=\alpha\lambda(1-e^{-\lambda x})^{\alpha-1}e^{-\lambda x}, \tag{3.1}$$

and the cumulative distribution function cdf for the distribution is as follows:

$$F(x;\alpha,\lambda)=(1-e^{-\lambda x})^{\alpha} \tag{3.2}$$

whereas

λ: the scale parameter of the EE distribution;

α: shape parameter of the EE distribution;

Table 2: Some characteristics of the distribution

Properties	Formulas
Mean	$\frac{1}{\lambda}(\psi(\alpha+1)-\psi(1))$
Variance	$\frac{1}{\lambda^2}(\psi'(1)-\psi'(\alpha+1))$
$M(t)$	$\frac{\Gamma(\alpha+1)\Gamma(1-\frac{t}{\lambda})}{\Gamma(\alpha-\frac{t}{\lambda}+1)}$

Another recently proposed model is the Log-BXIIEE model reviewed by Mohamed I. and Haitham M. (2020), which is an extension of the Exponentaited Exponentail distribution with the BXII distribution using the location regression model. And the mathematical formula for BXIIEE is [6]:

$$f_{(x)}= ab\alpha\lambda e^{-\lambda x}\frac{(1-e^{-\lambda x})^{a\alpha-1}}{[1-(1-e^{-\lambda x})^{\alpha}]^{a+1}}\left\{1+\left[\frac{(1-e^{-\lambda x})^{\alpha}}{1-(1-e^{-\lambda x})^{\alpha}}\right]^{a}\right\}^{-b-1} \tag{3.3}$$

$$x \sim BXIIEE(a,b,\alpha,\lambda).$$

In order to obtain the Log-BXIIEE regression model, we have to model the new BXIIEE distribution in a similar manner to the LBXIIW model as follows: when $x\sim$ BXIIEE, and $Y=\log(x)$, the equation (3.3) is as follows:

$$f(y)=\frac{ab\alpha\lambda}{\sigma}\exp\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right)\left(1-\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{\alpha-1}$$
$$\times\frac{\left\{\left(1-\exp\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{\alpha}\right\}^{a-1}}{\left\{1-\left(1-\exp\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{\alpha}\right\}^{a+1}}\left\{1+\left[\frac{\left\{(1-\exp\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right))^{\alpha}\right\}}{1-(1-\exp\left(-\lambda\exp\left(\frac{y-\mu}{\sigma}\right)\right))^{\alpha}}\right]^{a}\right\}^{-b-1} \tag{3.4}$$

whereas

μ: location parameter;

$\sigma>0$: scale parameter;

The above equation represents the LBXIIEE distribution and $Y\sim$ LBXIIEE $(a, b, \alpha,\lambda,\mu,\sigma)$. Using the location-scale regression model that links the explanatory variable vector ${\nu_i}^T=(\nu_{i1},\ldots,\nu_{ip})$ with the average response variable y_i , and the model can be expressed as follows:

$$y_i={\nu_i}^T\beta+\sigma z_i \ \ , \ i=1,\ldots,n \tag{3.5}$$

where y_i follows the LBXIIEE distribution and T is a random variable that follows the BXIIEE distribution.

Assuming that M, N are a set of items for y_i which represents log-lifetime or log-censoring respectively, and assuming lifetimes and censoring times are independent, the LBXIIEE regression model and by applying the MLE method to estimate the parameters of the vector $\theta=\left(a, b,\alpha,\lambda,\beta^T\right)$ and taking the natural logarithm will be as follows [6]:

$$\begin{aligned}\ln(\theta) \ =& r\ln\left(\frac{ab\alpha\lambda}{\sigma}\right)-\lambda\sum_{i\in M}u_i+(\alpha-1)\sum_{i\in M}\ln\left(1-\exp\left(-\lambda u_i\right)\ \right)+(\alpha-1)\sum_{i\in M}\ln\left(1-\exp\left(-\lambda u_i\right)\ \right)^a\\ &+(\alpha+1)\sum_{i\in M}\ln\left(1-\left(1-\exp\left(-\lambda u_i\right)\ \right)^\alpha\right)-(b+1)\sum_{i\in M}\ln\left(1+\left[\frac{\left(1-\exp\left(-\lambda u_i\right)\ \right)^\alpha}{1-\left(1-\exp\left(-\lambda u_i\right)\ \right)^\alpha}\right]^a\right)\\ &+\sum_{i\in N}\ln\left(1-\left[1+\left[\frac{\left(1-\exp\left(-\lambda u_i\right)\ \right)^\alpha}{1-\left(1-\exp\left(-\lambda u_i\right)\ \right)^\alpha}\right]^a\right]^{-b}\right).\end{aligned} \tag{3.6}$$

Whereas

$$u_i=e^{zi}, z_i=\frac{(y_i-{\nu_i}^T\beta)}{\sigma}.$$

Using NLMixed in SAS software, the estimations for the feature vector $\widehat{\theta}$ can be obtained.

4. Data Analysis and Estimation

For the purpose of conducting a practical application to compare the two models of LBXIIW and LBXIIEE, real data was taken regarding patients with renal failure in the Kirkuk General Hospital / Industrial College Unit in the city of Kirkuk - Iraq. A sample of 78 Males was taken from the patient records, and the data were representative of the variables: age, Urea and Creatine. By taking Creatine as a dependent variable for the models, as we mentioned previously in support of the censoring models, the importance of which is one of the aims of this article. The mechanism of analyzing these models includes extracting the standard error values S.E, the T-test values and the P-values, on the basis of which the variables that significantly affect the dependent variable will be known.

5. Results and Discussion

Table 3 that the parameters of the LBXIIW model for male patients were estimated and calculated, given that the dependent variable was Creatine. In the first step, all variables and parameters were estimated and depended on the p-value. In the second step, the constant was excluded, being less influential on the variable. In the third step, the age variable was excluded, as it was less influential and the Urea variable was the most influential on the dependent variable for male patients according to the LBXIIW model.

Table 3: Parameters Estimation by LBXIIW model for the Creatine variable for Males

Steps	Variable's in model	β	S.E	T	P-value
1	Constant	0.2887	9.7907	0.03	0.9765
	Age	0.0053	0.0172	0.31	0.7579
	Urea	0.1032	0.0122	8.44	0.0000
	α	13.4328	35.0925	0.38	0.7029
	β	0.4883	0.3363	1.45	0.1504
	σ	9.7820	25.6741	0.38	0.7042
2	Age	0.0056	0.0167	0.34	0.7362
	Urea	0.1033	0.0119	8.67	0.0000
	α	12.5213	5.6855	2.20	0.0306
	β	0.4819	0.3195	1.51	0.1355
	σ	9.0691	2.7397	3.31	0.0014
3	Urea	0.1027	0.0115	8.95	0.0000
	α	11.6149	4.7557	2.44	0.0168
	β	0.4600	0.2925	1.57	0.1198
	σ	8.3544	1.6463	5.07	0.0000

The following table 4 represents the criteria table of the LBXIIW model for the dependent variable Creatine for male patients, in which we note that the values of the criteria AIC, BIC, H-QIC are changed according to the steps. These criteria reach their lowest value in the last step and each step represents the values of the three criteria for the same step in Table 3.

Table 4: Comparative criteria values for the Creatine variable by LBXIIW model for Males

Steps	AIC	BIC	H-Q
1	63.6	77.8	69.3
2	61.6	73.5	66.4
3	59.7	69.2	63.5

Table 5 shows the estimates of the parameters and variables of the LBXIIEE model for males, considering the Creatine variable as the dependent variable and depending on the P-value in determining the variable most affecting the dependent variable. It was found in the second step that the Urea variable was the most influential on the dependent variable in the first place.

Table 5: Parameters Estimation by LBXIIEE model for the Creatine variable for Males

Steps	Variable's in model	β	S.E	T	P-value
1	Constant	1.3807	21.6993	0.06	0.9494
	Age	3.3474	0.4209	7.95	0.0000
	Urea	-1.1056	0.3733	-2.96	0.0040
	a	1.8441	1.1990	1.54	0.1280
	b	2.1999	0.0001	203375	0.0000
	α	0.8703	0.3202	2.72	0.0081
	β	3.8625	2.3900	1.62	0.1101
	σ	89.5632	25.6268	3.49	0.0008
2	Age	5.2556	0.8731	6.02	0.0000
	Urea	-1.1385	0.5586	-2.04	0.0339
	a	1.0829	1.3838	0.78	0.4362
	b	3.0387	0.0001	192776	0.0000
	α	0.6147	0.6270	0.98	0.3299
	β	3.2736	2.6090	1.25	0.2133
	σ	90.8850	20.3288	4.47	0.0000

The following table 6 represents the criteria table of the LBXIIEE model for the dependent variable Creatine for male patients, in which we note that the values of the criteria AIC, BIC, H-QIC are change according to the steps. These criteria reach their lowest value in the last step, and each step represents the values of the three criteria for the same step in Table (3-25).

Table 6: Comparative criteria values for the Creatine variable by LBXIIEE model for Males

Steps	AIC	BIC	H-Q
1	132.1	151.1	139.7
2	128	144.6	134.7

6. Conclusions

Through the results obtained from the application of some censored regression models for renal failure data, we conclude that the LBXIIW model is the best model compared to the LBXIIEE model according to the model selection criteria values, where the values of these criteria for the LBXIIEE model were great than the values of the same criteria for the LBXIIW model when Creatine as a dependent variable in the model. Also, we note that the criteria values (AIC, BIC, H-QIC) change and decrease according to the steps used to estimate and analyze the two models (LBXIIW, LBXIIEE) according to the dependent variable specified Creatine.

References

[1] M. Alizadeh,, M. Emadi and M. Doostparast, *A new two-parameter lifetime distribution: properties, applications and different method of estimations*, Stat., Optim. Inf. Comput. 7(2) (2019) 291–310.

[2] D. Bernoulli and S. Blower, *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it*, Rev. Med. Virol. 14(5)(2004) 275.

[3] G.M. Cordeiro, H.M. Yousof, T.G. Ramires and E.M. Orte ga, *The Burr XII system of densities: properties, regression model and applications*, J. Stat. Comput. Simul. 88(3) (2018) 432–456.

[4] L.W. Falls, *Estimation of parameters in compound Weibull distributions*, Technometrics 12(2) (1970) 399–407.

[5] R.D. Gupta and D. Kundu, *Exponentiated exponential family: an alternative to gamma and Weibull distributions.* Biom. J. 43(1) (2001) 117–130.

[6] M. Ibrahim, E.A. Ea and H.M. Yousof, *A new distribution for modeling lifetime data with different methods of estimation and censored regression modeling*, Stat. Optim. Inf. Comput. 8(2) (2020) 610–630.

[7] J.F. Lawless, *Statistical Models and Mmethods for Lifetime Data*, John Wiley & Sons, 2011.

[8] P.M. Odell, K.M. Anderson and R.B. D'Agostino, *Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model*, Biometrics 48(3) (1992)951–959.

[9] P.F. Paranaíba, E.M. Ortega, G.M. Cordeiro and R.R. Pescim, *The beta Burr XII distribution with application to lifetime data*, Comput. Stat. Data Anal. 55(2) (2011) 1118–1136.

[10] H. Rinne, *The Weibull distribution: a handbook*, CRC press, 2008.

[11] M.A. Salih , *Parameters Estimation For Pareto Type – II Distribution*, PhD Thesis, College of Administration and Economics, Al-Mustansiriya University, 2015.

[12] N.T. Thomopoulos, *Statistical distributions. Applications and Parameter Estimates.* Cham, Switzerland: Springer International Publishing, 2017.