American Journal of Applied Mathematics and Statistics, 2016, Vol. 4, No. 3, 74-86
Available online at http://pubs.sciepub.com/ajams/4/3/3

DOI:10.12691/ajams-4-3-3

Performance of Log-Beta Log-Logistic Regression Model

Mahmoud Riad Mahmoud[1], Naglaa A. Morad[2], Moshera A. M. Ahmad[2,*]

[1]Department of Mathematical Statistics, Institute of Statistical Studies and Research, Cairo University
[2]Department of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University
*Corresponding author: moshera_ahmad@pg.cu.edu.eg

Abstract For the log-beta log-logistic regression model, we derive the appropriate matrices for assessing the local influence on the parameter estimates under perturbation scheme. Using a set of real data, global and local influences of individual observations on the stated model are considered. Besides, for different parameter settings, sample sizes, and censoring percentages, various simulation studies are performed to the performance of the log-beta log-logistic regression model. In addition, the empirical distribution of the martingale residuals is displayed against the normal distribution for comparison. These studies suggest that the martingale residual has shaped normal form.

Keywords: *likelihood displacement, local influence approach, beta log-logistic distribution, martingale residuals, sensitivity analysis*

Cite This Article: Mahmoud Riad Mahmoud, Naglaa A. Morad, and Moshera A. M. Ahmad, “Performance of Log-Beta Log-Logistic Regression Model.” *American Journal of Applied Mathematics and Statistics*, vol. 4, no. 3 (2016): 74-86. doi: 10.12691/ajams-4-3-3.

1. Introduction

Frequently in parametric regression models in survival analysis, there are covariates whose values are related to the lifetime under study. The main objective in such cases is, usually, to estimate the relationship between the lifetime and the explanatory variables and to test its significance. For example; the length of time it takes an employee to retire from a given job may be affected by variables such as the employee’s age, experience, education,...etc. Another example of this situation has been discussed by [10], in which survival times for 65 multiple myeloma patients were recorded and related to a number of factors, such as the level of hemoglobin in the blood, the white blood cells count at diagnosis, sex, and age. The main problem in this type of studies is to identify which concomitant variables were strongly related to survival time. To examine the relationship between the lifetime and the concomitant variables we will use a regression model in which lifetime has a distribution that depends on concomitant variables. This involves specifying a model for the distribution of T, given x (which may have some censored observation), where T represents lifetime and x is a vector of regressor variables for an individual.

In the last decade, a new class of models has been proposed for use with this type of data, that the covariates whose values are related to the lifetime. [17] proposed a location-scale regression model based on the logarithm of an extended Weibull distribution which has the ability to deal with bathtub-shaped failure rate functions. [9] showed that the log-exponentiated Weibull regression model for interval-censored data represented a parametric family of models that include other regression models that are broadly used in lifetime data analysis. [13] proposed a log-β-Birnbaum–Saunders regression model that can be applied to censored data and be used more effectively in survival analysis. However, there are few practical regression models of this type of failure rate function. A log-beta log-logistic regression model (which we wish has a wide use in the lifetime) is proposed. After modeling, it is important to check assumptions in the model and to conduct a robustness study to detect influential or extreme observations that can distort the results of the analysis. So this paper performs a log-beta log-logistic regression model by various methods.

The paper is organized as follows. In sec. 2 we display the review of regression failure models. Background of the log-beta log-logistic regression model in sec.3. Sensitivity analysis in sec 4. Curvature calculations for log-beta log-logistic regression model are introduced in sec 5. We also discuss some simulations studies and a real data set is analyzed in sec. 6.

2. Review

[7] gave an early discussion of parametric regression models in survival analysis where there are covariates whose values are related to the lifetime; they present a method of estimating survival distribution when the survival times are assumed to follow simple exponential distributions, with a different parameter for each patient. The parameter associated with each patient’s distribution is functionally related to the concomitant variates. [18] generalized the work of [7]. They extended the statistical model to permit maximum likelihood (ML) estimation of the parameters of the linear regression where not all patients in a follow-up study have died by the end of the study. [10,11,15] have discussed this approach and introduced exponential, Weibull, and Gamma regression model in two cases, complete and censored data. Recently,

this approach was discussed by many authors, see for example, [1,5,8,12,14,16]. They introduced exponential Weibull, log-Burr XII, log-modified Weibull, the log-generalized inverse Weibull, the log-beta exponentiated Weibull, and log-beta log-logistic regression model.

3. Background of the Log-beta Log-logistic (LBLLogistic) Regression Model

The beta log-logistic distribution, with positive parameters a, b, α and δ, BLLog(a; b; α; δ), considers that lifetime T has a density function given by

$$f(t)=\frac{(\frac{\delta}{\alpha})(\frac{t}{\alpha})^{a\delta-1}}{B(a,b)[1+(\frac{t}{\alpha})^{\delta}]^{a+b}},t>0. \tag{1}$$

Where a, b, α and $\delta > 0$ is a shape parameter, and $\alpha > 0$ is a scale parameter. The survival function corresponding to random variable T with **BLLog** density is given by

$$S(t)=1-I_{\frac{1}{[1+(\frac{t}{\alpha})^{\delta}]}}(a,b),$$

and the associated hazard rate function takes the form

$$h(t)=\frac{\left(\frac{\delta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{a\delta-1}\left(1+(\frac{t}{\alpha})^{\delta}\right)^{-(a+b)}}{B(a,b)S(t)},t>0. \tag{2}$$

Recently, [12] suggested a regression model based on the **BLLog** distribution described in (1). This model is so-called the **LBLLogistic** regression model. The **LBLLogistic** regression model is represented by

$$y_i=\log(t_i)=x_i^{'}\beta+\sigma z_i, i=1,\ .\ .\ .\ ,\ n, \tag{3}$$

where y_i is the response variable, $\boldsymbol{x}_i^{'}$ = (x_{i1}, x_{i2}, . . ., x_{ip}) is the vector of explanatory variable, $\boldsymbol{\beta}$ = (β_1, . . ., β_p)$'$ $\sigma > 0$. Note that Y = log(T) follows the **BLLog** distribution. The density function of Y can be written as

$$f(y,a,b,\sigma,\mu)=\frac{\sigma^{-1}}{B(a,b)}\left[\exp\left(\frac{y-\mu}{\sigma}\right)\right]^{a}\left[1+\exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-(a+b)}, \tag{4}$$

where a, b, $\sigma > 0$, $-\infty < \mu < \infty$, and $-\infty < y < \infty$.

Further, after suitable transformation, we define the standard random variable $Z = (Y - \mu)/\sigma$ with density function

$$f(z)=\frac{1}{B(a,b)}[\exp(z)]^{a}[1+\exp\exp(z)]^{-(a+b)} \tag{5}$$
$$-\infty<z<\infty.$$

The survival function takes the form

$$S(y)=1-I_{\frac{\exp(\exp((y-\mu)/\sigma)}{[1+\exp(\exp((y-\mu)/\sigma)]}}(a,b), \tag{6}$$

and the associated hazard rate function takes the form

$$h(y)=\frac{\frac{\sigma^{-1}}{B(a,b)}\left\{\begin{array}{c}\left[\exp\left(\frac{y-\mu}{\sigma}\right)\right]^{a}\\ \left[1+\exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-(a+b)}\end{array}\right\}}{1-I_{\frac{\exp(\exp((y-\mu)/\sigma)}{[1+\exp(\exp((y-\mu)/\sigma)]}}(a,b)}, \tag{7}$$

in terms of t, model (3) is referred to as a log-location scale model.

Consider a sample (y_1,x_1),…,(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 noninformative censoring such that the observed lifetimes and censoring times are independent. Let $\mathcal{F}$ and $\mathcal{C}$ be the sets of indices of individuals for which y_i is the log lifetime and log-censoring, respectively. The log-likelihood function for the vector of parameters $\boldsymbol{\theta}=(a,b,\delta,\boldsymbol{\beta}^T)^T$ from model (3) has the form

$$l(\theta)=\sum_{i\epsilon\mathcal{F}}\log\left[f(y_i)\right]+\sum_{i\epsilon\mathcal{C}}\log\left[S(y_i)\right],$$

where $f(y_i)$ is the density function (4) and $S(y_i)$ is the survival function (6) of Y_i. The log-likelihood function for $\boldsymbol{\theta}$ reduces to

$$\begin{aligned}l(\theta)&=-r\log(\sigma)-r\log B(a,b)+a\sum_{i\epsilon\mathcal{F}}(z_i)\\&-(a+b)\sum_{i\epsilon\mathcal{F}}\log\left[1+\exp(z_i)\right]\\&+\sum_{i\epsilon\mathcal{C}}\log\left\{1-I_{\frac{\exp(z_i)}{[1+\exp(z_i)]}}(a,b)\right\}\end{aligned} \tag{8}$$

where r is the number of uncensored observations (failures) and z_i =(y_i - $\mathbf{x}^T_i\,\boldsymbol{\beta}$)/ σ. The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function (8).

Let $I(\boldsymbol{\theta}) = E[\ddot{L}(\boldsymbol{\theta})]$ is the observed information matrix and the asymptotic covariance matrix $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$ of $\hat{\theta}$ can be approximated by the inverse of the (p+3)(p+3) observed information matrix $\ddot{L}(\hat{\theta})=-\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^T}|_{\theta=\hat{\theta}}$,

$$-\ddot{L}(\theta)=\begin{pmatrix}L_{aa}&L_{ab}&L_{a\sigma}&L_{a\beta_j}\\ .&L_{bb}&L_{b\sigma}&L_{b\beta_j}\\ .&.&L_{\sigma\sigma}&L_{\sigma\beta_j}\\ .&.&.&L_{\beta_j\beta_s}\end{pmatrix},$$

where

$$L_{a\beta_j}=[L_{a\beta_1}\ \cdots\ L_{a\beta_p}], L_{b\beta_j}=[L_{b\beta_1}\ \cdots\ L_{b\beta_p}],$$
$$L_{\sigma\beta_j}=[L_{\sigma\beta_1}\ \cdots\ L_{\sigma\beta_p}],$$

and

$$L_{\beta_j\beta_s}=\begin{bmatrix}L_{\beta_1\beta_1}&\cdots&L_{\beta_1\beta_p}\\ \vdots&\ddots&\vdots\\ L_{\beta_p\beta_1}&\cdots&L_{\beta_p\beta_p}\end{bmatrix}.$$

4. Sensitivity Analysis

After fitting a model, it is important to check the assumptions of the model to detect possible extreme or influential observations. So we will discuss the influence diagnostic based on case deletion, in which the influence of the i^{th} observation on the parameter estimates is evaluated by removing it from the analysis. We will display that by the following measures.

4.1. Generalized Cook Distance

[2] has proposed measuring the "distance" between the $\hat{\theta}$ and the corresponding $\hat{\theta}_{(-i)}$ by calculating the F statistic for the "hypothesis" that $\hat{\theta} = \hat{\theta}_{(-i)}$. This statistic is recalculated for each observation i=1,…,n. The resulting values should not literally be interpreted as F tests. The suggested measure of the distance of $\hat{\theta}_{(-i)}$ from $\hat{\theta}$ is

$$GD_i(\theta) \equiv \left(\hat{\theta}_{(-i)} - \hat{\theta}\right)' \left(-\ddot{L}(\theta)\right)\left(\hat{\theta}_{(-i)} - \hat{\theta}\right) \cong F_{(p,n-p,1-\alpha)}. \quad (9)$$

4.2. Likelihood Displacement

Measures of the influence of the i^{th} observation on the ML estimate $\hat{\theta}$, can be based on the sample influence curve $SIC_i \propto \hat{\theta} - \hat{\theta}_{(-i)}$, where $\hat{\theta}_{(-i)}$ denotes the ML estimate of θ computed without the i^{th} observation. While this idea is straightforward, it may be computationally expensive to implement since n+1 ML estimates are needed, each of which may require several iterations. [4] derived the general measure from the use of contours of the log likelihood function to order observations based on influence. If $L(\theta)$ be the log likelihood based on the complete data, a likelihood distance $LD_i(\theta)$ defined as

$$LD_i(\theta) = 2\left[L(\hat{\theta}) - L(\hat{\theta}_{(-i)})\right] \cong \chi^2_{(q)} \quad (10)$$

Where q is the dimension of θ.

4.3. Local Influence Approach

Removing observations from the analysis suddenly lead to all information on a single data point is deleted suddenly, and therefore, it is difficult to determine whether that data point has some influence on a specific aspect of the model. We can find a solution for this problem by using local influence approach in which one can investigate how the results of an analysis change under small perturbations in the model.

The basic idea in influence analysis -as presented by [4]- is " to introduce small perturbations in the problem formulation, and then monitor how the perturbations change the outcome of the analysis. The important questions in designing methods for influence analysis are the choices of the perturbation scheme, the particular aspect of an analysis to monitor, and the method of measurement".

The numerous influence diagnostics that depend on case-deletion can be regarded as global measures since they are designed to measure total change at various corners of Ω =(0,1)^n, where n is the sample size.

[3] indicated that; single case-deletion diagnostics can be computationally intensive and suffer from a form of masking, and group deletion methods are not easily implemented or well understood. So [3] developed a methodology that is relatively easy to use for the identification of groups of observations that may require special attention. It summarized as follows:

Let θ be a p×1 vector of unknown parameters, and $\hat{\theta}$ be the ML estimate of θ obtained by maximizing $L(\theta)$, where $L(\theta)$ represents the log-likelihood for a postulated model and observed data (unperturbed log-likelihood). Introduce some perturbations into the model via the m×1 vector $\omega \in \Omega$, where ω represents the set of relevant perturbations, and let $L(\theta|\omega)$ denotes the log likelihood corresponding to the perturbed model, and let $\hat{\theta}_{\omega}$ denote ML estimate under $L(\theta|\omega)$. Let $\boldsymbol{d}$ be a fixed nonzero direction of unit length in R^q. Let Δ is $(p+3)\times n$ matrix that depends on the perturbation scheme and whose ij^{th} element

$$\Delta_{ij} = \frac{\partial^2 L(\theta\,|\,\omega)}{\partial\theta_i \partial\omega_j}\Big|_{(\theta=\hat{\theta}),\,(\omega=\omega_0)};$$
$$i = 1,\ldots,n \ \text{ and } \ j = 1,2,\ldots,p+3.$$

The normal curvature for θ at the direction $\boldsymbol{d}$ for the postulated model is given by

$$C_d(\theta) = 2\,|\,d'\Delta'\ddot{L}(\theta)^{-1}\Delta d\,|. \quad (11)$$

The extreme $C_{\max}$=max C_d and $C_{\min}$=min C_d are two possible option. max C_d is the largest eigenvalue of the matrix $\mathbf{B} = \Delta'\ddot{L}(\theta)^{-1}\Delta$, and $\boldsymbol{d}_{max}$ is the corresponding eigenvector. The index plot of $\boldsymbol{d}_{max}$ for matrix $\mathbf{B}$ may indicate how to perturb the postulated model to obtain the greatest local change in the likelihood displacement; if the i^{th} element of $\boldsymbol{d}_{max}$ is found to be relatively large, this indicates that perturbations in the weight ω_i of the i^{th} observation may lead to substantial changes in the results of the analysis and thus that ω_i is relatively influential. It is important to investigate the i^{th} observation to find the specific observation of the sensitivity.

5. Curvature Calculations for Log-beta Log-logistic Regression Model

As we mentioned, [3] proposed a general framework to detect the influence of observations to evaluate how sensitive the analysis is to small perturbations that are agitated within the model. Some authors have investigated the evaluation of local influences in survival analysis models: for example [6] adapted the method of local influence to regression analysis with censoring; [9] studied the problem of evaluating local influences in the log- exponentiated Weibull regression model with censored data. We introduce a similar methodology to detect influential data points in the **LBLLogistic** regression model for interval-censored data.

Next, the perturbation scheme will be calculate, the matrix

$$\Delta = \left(\Delta_{ji}\right)_{(p+3)\times n} = \left(\frac{\partial^2 l(\theta \setminus \omega)}{\partial\theta_i \partial\omega_j}\right)_{(p+3)\times n},$$
$$j = 1,2,\ldots,p+3 \text{ and } i = 1,2,\ldots,n,$$

considering the model defined in (3) and its log-likelihood function given by (8). Let the vector of weights ω=(ω_1,

$\omega_2,\ldots,\ \omega_n)'$. Case-weight perturbation for the log-likelihood function takes the form

$$l(\theta\setminus\omega)=-r\left[\log(\sigma)+\log B(a,b)\right]\sum_{i\in\mathcal{F}}\omega_i +a\sum_{i\in\mathcal{F}}\omega_i(z_i)-(a+b)\sum_{i\in\mathcal{F}}\omega_i\left(\log\left[1+\exp(z_i)\right]\right) +\sum_{i\in\mathcal{C}}\omega_i\log\left\{1-I_{\frac{\exp(z_i)}{[1+\exp(z_i)]}}(a,b)\right\},$$

where $0\le\omega_i\le1$ and $\omega_0=(1,\ldots,1)'$. let us denote $\Delta=(\Delta_1,\ldots,\Delta_{p+3})'$. $\mathcal{F}$ and $\mathcal{C}$ be the sets of indices of individuals for which y_i is the log lifetime and log-censoring respectively. Then, the elements of vector Δ_1 take the form

$$\Delta_{1i}=\begin{cases}-r\left[\psi(\hat a)-\psi(\hat a+\hat b)\right]+\hat z_i-\log\left[1+\exp(\hat z_i)\right], & \text{if } i\in\mathcal{F},\\ [1-I_{G(\hat z_i)}(\hat a,\hat b)]^{-1}\left[\begin{array}{l}\dfrac{G(\hat z_i)\Gamma^2(\hat a)\hat o_i}{B(\hat a,\hat b)}\\ -I_{G(\hat z_i)}(\hat a,\hat b)\left\{\begin{array}{l}\log G(\hat z_i)\\ -\psi(\hat a)+\psi(\hat a+\hat b)\end{array}\right\}\end{array}\right], & \text{if } i\in\mathcal{C},\end{cases}$$

where

$$G(z_i)=\left(\frac{e^{z_i}}{1+e^{z_i}}\right)$$

$$o_i={}_3F_2\left(a,a,1-b;1+a,1+a;G(z_i)\right).$$

On the other hand, the elements of the vector Δ_2 can be shown to be given by

$$\Delta_{2i}=\begin{cases}-r\left[\psi(\hat b)-\psi(\hat a+\hat b)\right]-\log\left[1+\exp\exp(\hat z_i)\right], & \text{if } i\in\mathcal{F},\\ -\dfrac{\left[\begin{array}{l}(1-G(\hat z_i))^{\hat b}\,\Gamma^2(\hat b)\hat v_i\\ +B_{(1-G(\hat z_i))}(\hat a,\hat b)[-\log\log(1-G(\hat z_i))\\ +\psi(\hat b)-\psi(\hat a+\hat b)]\end{array}\right]}{B(\hat a,\hat b)[1-I_{G(z_i)}(\hat a,\hat b)]}, & \text{if } i\in\mathcal{C},\end{cases}$$

where

$$v_i={}_3F_2\left(b,b,(a-1);(b+1),(b+1);(1-G(z_i))\right).$$

The elements of the vector Δ_3 can be shown to be given by

$$\Delta_{3i}=\begin{cases}-\hat\sigma^{-1}\left[r+\hat a\hat z_i-\left[\dfrac{(\hat a+\hat b)z_i\exp(\hat z_i)}{[1+\exp(\hat z_i)]^{-1}}\right]\right] & \text{if } i\in\mathcal{F},\\ -\dfrac{\left\{\begin{array}{l}(G(\hat z_i))^{\hat a-1}(1-G(\hat z_i))^{\hat b-1}\\ \left[\left[\dfrac{\hat z_iG(\hat z_i)}{-\hat\sigma}\left(\dfrac{1}{1-\exp(\hat z_i)}-1\right)\right]\right]\end{array}\right\}}{B(\hat a,\hat b)[1-I_{G(\hat z_i)}(\hat a,\hat b)]} & \text{if } i\in\mathcal{C}.\end{cases}$$

The elements of the vector Δ_j, for j=4,…,p+3, may be expressed as

$$\Delta_{ji}=\begin{cases}-\hat\sigma^{-1}\left[\hat a\mathbf{x}_i-\left[\dfrac{(\hat a+\hat b)\mathbf{x}_i\exp(\hat z_i)}{[1+\exp(\hat z_i)]^{-1}}\right]\right] & \text{if } i\in\mathcal{F},\\ -\dfrac{\left\{\begin{array}{l}(G(\hat z_i))^{\hat a-1}(1-G(\hat z_i))^{\hat b-1}\\ \left[\left[\dfrac{\mathbf{x}_iG(\hat z_i)}{\hat\sigma}(G(\hat z_i)-1)\right]\right]\end{array}\right\}}{B(\hat a,\hat b)[1-I_{G(\hat z_i)}(\hat a,\hat b)]} & \text{if } i\in\mathcal{C}.\end{cases}$$

6. Application

An application of the result will be provided by using simulated and real data. The required numerical evaluations were applied using Mathcad and Mathematica programs.

6.1. Simulations Study (1)

In order to assess the performance of estimating the parameters of the **LBLLogistic** regression model, various simulation studies are performed for different settings of sample sizes, censoring percentages, and parameter values. The lifetimes denoted by $T_1,\ldots,T_n$ were generated from the **BLLog** distribution given in (1). The following transformation was made $y_i=\log(T_i)$ $\sigma=\frac{1}{\delta}$, and $\mu=\log(\alpha)$ where $y_i=\mu_i+\sigma z_i$ and $\mu_i=\beta_0+\beta_1x_i$. The values $\beta_0=0.3$ and $\beta_1=-0.23$ were chosen for this study. The fixed components μ_i were generated using x_i from gamma distribution with parameters (0.25, 35). The stochastic components z_i representing the errors in the model was generated from (5) for different values of a and b. The censoring times denoted by $C_1,\ldots,C_n$ were generated from a uniform distribution(0, ϑ), where ϑ was chosen to achieve censoring percentages of 0 or 0.10 or 0.30. The lifetimes considered in each fit were calculated as $\min\{C_i,T_i\}$. We generated the model for different values of n= 20, 50 and 100, different values of σ=0.8, 1.8, and 5, different values of a= 0.5, 1.08, 1.8, 3, 5, 8, and 10, and different values of b=0.5, 0.8, 1, 1.1, 1.8, 2, 4, and 5. For each sample a **LBLLogistic** regression model is fitted by estimating the corresponding parameter value using maximum likelihood method. From these one thousand, parameter estimates, bias, the mean square error (MSE), and relative root MSE were calculated. The results are summrize in Table 1.

It may be noticed that:-

- we have small bias and MSE when $a>b$.
- we have very big bias and MSE when $a<b$.
- The bias and MSE decreased when n increased.
- The bias and MSE increased when the censoring percentage increased.

6.2. Simulations Study (2)

In order to investigate the form of the empirical distribution of the martingale residuals for different settings of n and censoring percentages, several simulation studies are performed for which the results are displayed graphically in Figure 1 – Figure 6. We assumed sample sizes 30, 50, and 100. The log-lifetime denoted by $\log(T_1),...,\log(T_n)$ were generated from the **LBLLogistic** regression model given in (3), for a=10, b=1, different value of σ=0.8, 1.8 and 5, β_0=0.3 and β_1=-0.23 with x_i generated from a gamma distribution with parameters (35, 0.25). The censoring times denoted by $C_1,...,C_n$ were generated from uniform distribution(0, ϑ), where ϑ was adjusted until the censoing percentages 0, 10, and 30% be reached. The lifetimes considered in each fit were calculated as $min\{C_i, T_i\}$. For each setting of n. we generate 1000 sample under the **LBLLogistic** regression model (3).

For each fit, the martingale residuals were calculated and stored. Then, the residuals were estimated and plotted in probability plots.

Table 1. The estimates, bias, MSE, and relative root MSE for log-beta log-logistic regression model

σ value	a value	b value	n	% Censoring	Estimate	Bias	MSE	Relative root MSE
0.8	**5**	**0.5**	**20**	0	-0.331	-0.101	0.018	-0.663
				10	-0.331	-0.101	0.018	-0.667
				30	-0.335	-0.105	0.024	-0.776
			50	0	-0.323	-0.093	0.015	-0.62
				10	-0.324	-0.094	0.015	-0.622
				30	-0.324	-0.094	0.016	-0.623
			100	0	-0.326	-0.096	0.016	-0.631
				10	-0.326	-0.096	0.016	-0.631
				30	-0.326	-0.096	0.016	-0.633
		1	**20**	0	-1.142	-0.912	188.175	-68.588
				10	-1.232	-1.002	68.853	-41.489
				30	-1.429	-1.199	241.84	-77.756
			50	0	-0.584	-0.354	2.99	-8.645
				10	-0.64	-0.41	8.237	-14.35
				30	-0.538	-0.308	5.709	-11.947
			100	0	-0.574	-0.344	12.618	-17.761
				10	-0.469	-0.239	0.334	-2.891
				30	-0.836	-0.606	9.351	-15.29
	8	**0.5**	**20**	0	-0.312	-0.082	0.013	-0.566
				10	-0.313	-0.083	0.013	-0.568
				30	-0.313	-0.083	0.013	-0.575
			50	0	-0.305	-0.075	0.011	-0.525
				10	-0.305	-0.075	0.011	-0.526
				30	-0.305	-0.075	0.011	-0.527
			100	0	-0.307	-0.077	0.012	-0.537
				10	-0.307	-0.077	0.012	-0.537
				30	-0.307	-0.077	0.012	-0.538
		0.8	**20**	0	-0.343	-0.113	0.022	-0.745
				10	-0.344	-0.114	0.025	-0.788
				30	-0.361	-0.131	0.147	-1.92
			50	0	-0.344	-0.114	0.021	-0.722
				10	-0.344	-0.114	0.021	-0.724
				30	-0.345	-0.115	0.022	-0.742
			100	0	-0.348	-0.118	0.022	-0.742
				10	-0.348	-0.118	0.022	-0.742
				30	-0.348	-0.118	0.022	-0.745
1.8	**5**	**0.5**	**20**	0	-0.257	-0.027	0.003347	-0.289
				10	-0.257	-0.027	0.00338	-0.291
				30	-0.258	-0.028	0.003462	-0.294
			50	0	-0.254	-0.024	0.002978	-0.273
				10	-0.254	-0.024	0.002993	-0.274
				30	-0.255	-0.025	0.003024	-0.275
			100	0	-0.256	-0.026	0.003133	-0.28
				10	-0.256	-0.026	0.003136	-0.28
				30	-0.256	-0.026	0.003145	-0.28

			20	0	-0.429	-0.199	1.046	-5.113
				10	-0.438	-0.208	1.182	-5.437
				30	-0.475	-0.245	1.31	-5.723
			50	0	-0.341	-0.111	0.051	-1.124
				10	-0.342	-0.112	0.344	-2.932
				30	-0.388	-0.158	0.859	-4.633
			100	0	-0.32	-0.09	0.028	-0.839
				10	-0.32	-0.09	0.043	-1.036
				30	-0.319	-0.089	0.042	-1.027
		1.8	**20**	0	-0.284	-0.054	0.035	-0.93
				10	-0.278	-0.048	0.07	-1.327
				30	-0.28	-0.05	0.065	-1.275
			50	0	-0.262	-0.032	0.081	-1.425
				10	-0.282	-0.052	0.016	-0.633
				30	-0.274	-0.044	0.054	-1.159
			100	0	-0.266	-0.036	0.024	-0.782
				10	-0.274	-0.044	0.014	-0.587
				30	-0.271	-0.041	0.033	-0.907
	10	**4**	**20**	0	-0.443	-0.213	0.22	-2.343
				10	-0.471	-0.241	0.325	-2.849
				30	-0.467	-0.237	0.285	-2.669
			50	0	-0.383	-0.153	0.075	-1.365
				10	-0.399	-0.169	0.129	-1.795
				30	-0.405	-0.175	0.146	-1.912
			100	0	-0.358	-0.128	0.039	-0.989
				10	-0.367	-0.137	0.035	-0.94
				30	-0.37	-0.14	0.036	-0.952
5	**5**	**0.5**	**20**	0	-0.22	0.01	0.0004115	-0.101
				10	-0.22	0.01	0.0004152	-0.102
				30	-0.22	0.01	0.0004381	-0.105
			50	0	-0.22	0.01	0.0004107	-0.101
				10	-0.22	0.01	0.0004137	-0.102
				30	-0.22	0.01	0.000418	-0.102
			100	0	-0.22	0.01	0.0004006	-0.1
				10	-0.22	0.01	0.0004008	-0.1
				30	-0.22	0.01	0.0004025	-0.1
		2	**20**	0	-0.259	-0.029	0.011	-0.533
				10	-0.263	-0.033	0.006939	-0.416
				30	-0.259	-0.029	0.011	-0.513
			50	0	-0.255	-0.025	0.007557	-0.435
				10	-0.25	-0.02	0.012	-0.558
				30	-0.255	-0.025	0.006331	-0.398
			100	0	-0.255	-0.025	0.007656	-0.437
				10	-0.251	-0.021	0.008225	-0.453
				30	-0.248	-0.018	0.009205	-0.48
	10	**0.5**	**20**	0	-0.216	0.014	0.0002709	-0.082
				10	-0.216	0.014	0.0002737	-0.083
				30	-0.217	0.013	0.0002791	-0.084
			50	0	-0.216	0.014	0.0002674	-0.082
				10	-0.217	0.013	0.0002816	-0.084
				30	-0.217	0.013	0.0002817	-0.084
			100	0	-0.216	0.014	0.0002632	-0.081
				10	-0.216	0.014	0.0002639	-0.081
				30	-0.216	0.014	0.0002626	-0.081

			20	0	-0.262	-0.032	0.004885	-0.349
				10	-0.263	-0.033	0.004936	-0.351
				30	-0.265	-0.035	0.004906	-0.35
		4	**50**	0	-0.257	-0.027	0.005664	-0.376
				10	-0.259	-0.029	0.00562	-0.375
				30	-0.26	-0.03	0.005478	-0.37
			100	0	-0.248	-0.018	0.00652	-0.404
				10	-0.25	-0.02	0.006257	-0.396
				30	-0.252	-0.022	0.006029	-0.388
0.8	**5**	**1.1**	**20**	0	-0.915	0.685	14.546	-16.582
				10	-1.368	1.138	149.088	-53.088
				30	-1.189	0.959	30.939	-24.184
			50	0	-1.335	1.105	22.942	-20.825
				10	-1.754	1.524	201.208	-61.673
				30	-1.354	1.124	49.203	-30.498
			100	0	-2.828	2.598	95.105	-42.401
				10	-1.603	1.373	28.16	-23.072
				30	-2.112	1.882	285.445	-73.457
0.8	**0.5**	**5**	**20**	0	-8.355	8.125	$7.25*10^3$	-370.205
				10	-5.182	4.952	$3.99*10^3$	-274.633
				30	-10.584	10.354	$8.187*10^3$	-393.404
			50	0	-3.965	3.735	$1.369*10^3$	-160.882
				10	-5.361	5.131	$2.945*10^3$	-235.934
				30	-4.735	4.505	$1.483*10^3$	-167.456
			100	0	-4.428	4.198	$2.119*10^3$	-200.13
				10	-4.45	4.22	$1.311*10^3$	-157.45
				30	-5.624	5.394	$1.604*10^3$	-174.153
1.8	**1.08**	**1.8**	**20**	0	-36.862	-36.632	$1.645*10^4$	-641.3
				10	-40.343	-40.113	$2.394*10^4$	-773.667
				30	-40.281	-40.051	$1.581*10^4$	-628.621
			50	0	-21.115	-20.885	$9.726*10^4$	-493.107
				10	-28.244	-28.014	$1.584*10^4$	-629.241
				30	-25.817	-25.587	$1.195*10^4$	-546.508
			100	0	-10.396	-10.166	$2.084*10^3$	-228.274
				10	-14.602	-14.372	$4.343*10^3$	-329.503
				30	-17.862	-17.632	$8.29*10^3$	-455.256
1.8	**1.8**	**5**	**20**	0	-9.282	9.052	$1.854*10^3$	-187.186
				10	-11.44	11.21	$4.664*10^3$	-296.933
				30	-6.952	6.722	$1.571*10^3$	-172.321
			50	0	-10.037	9.807	$3.212*10^3$	-246.401
				10	-15.685	15.455	$9.225*10^3$	-417.588
				30	-14.92	14.69	$4.778*10^3$	-300.528
			100	0	-13.792	13.562	$4.265*10^3$	-283.956
				10	-11.754	11.524	$3.511*10^3$	-257.626
				30	-15.974	15.744	$7.918*10^3$	-386.878
5	**3**		**20**	0	-29.746	29.516	$2.146*10^4$	-636.909
				10	-28.953	28.723	$1.481*10^4$	-529.057
				30	-22.493	22.263	$5.662*10^3$	-327.168
			50	0	-36.334	36.104	$1.856*10^4$	-592.372
				10	-39.978	39.748	$2.572*10^4$	-697.278
				30	-38.532	38.302	$2.388*10^4$	-671.886
			100	0	-67.894	67.664	$8.567*10^4$	$-1.273*10^3$
				10	-53.165	52.935	$3.637*10^4$	-829.156
				30	-42.381	42.151	$2.305*10^4$	-660.081

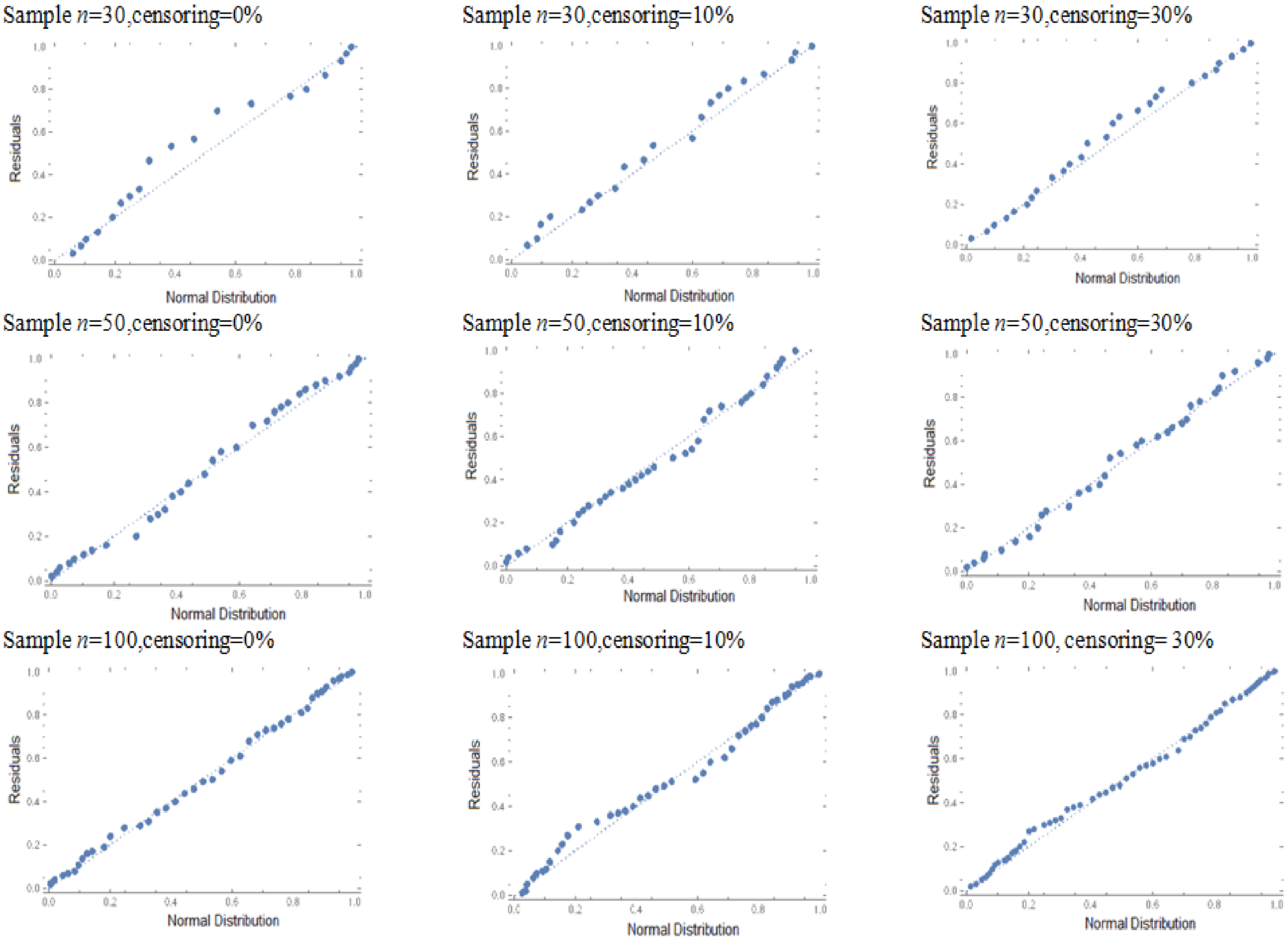

Figure 1. Normal probability plots for martingale residuals (r_{M_i}) . Sample sizes n=30, 50 and 100, percentage of right-censored= 0, 10 and 30, σ=0.8

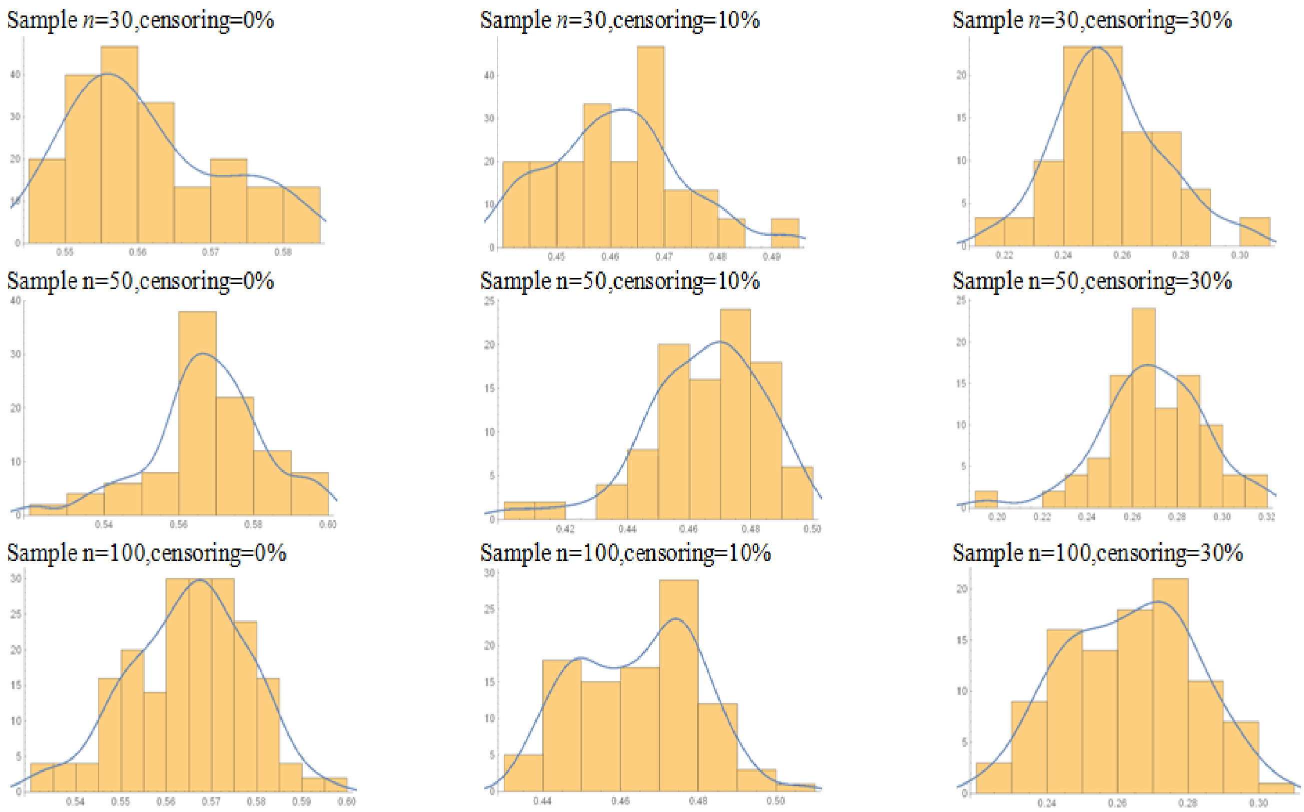

Figure 2. Histogram and Smooth curve plots for martingale residuals$(\mathbf{r_{M_i}})$ .Sample sizes n=30, 50 and 100, percentage of right-censored= 0, 10 and 30, σ=0.8

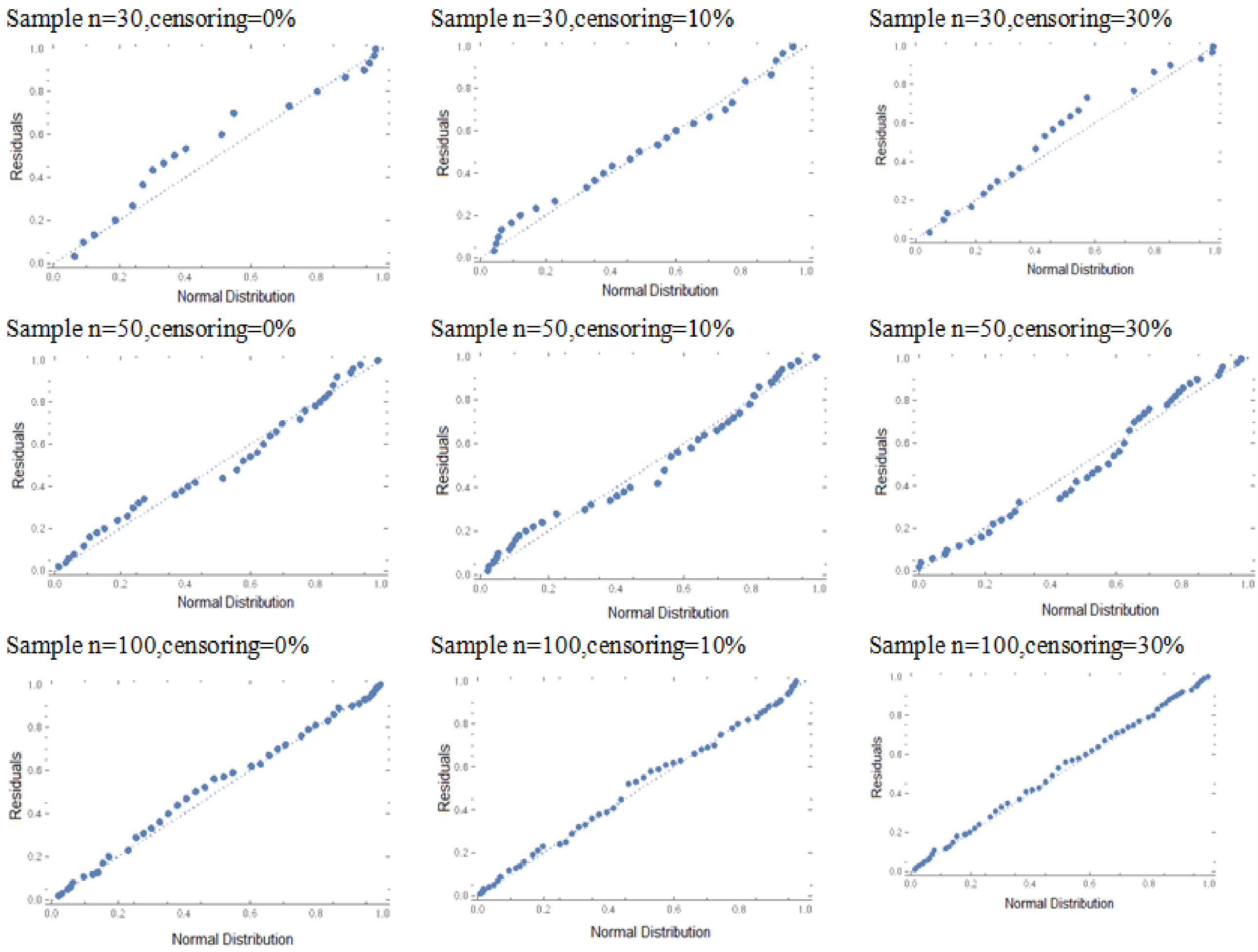

Figure 3. Normal probability plots for martingale residual (r_{M_i}). Sample sizes n=30, 50 and 100, percentage of right-censored=10, 30 and 50, σ=1.8

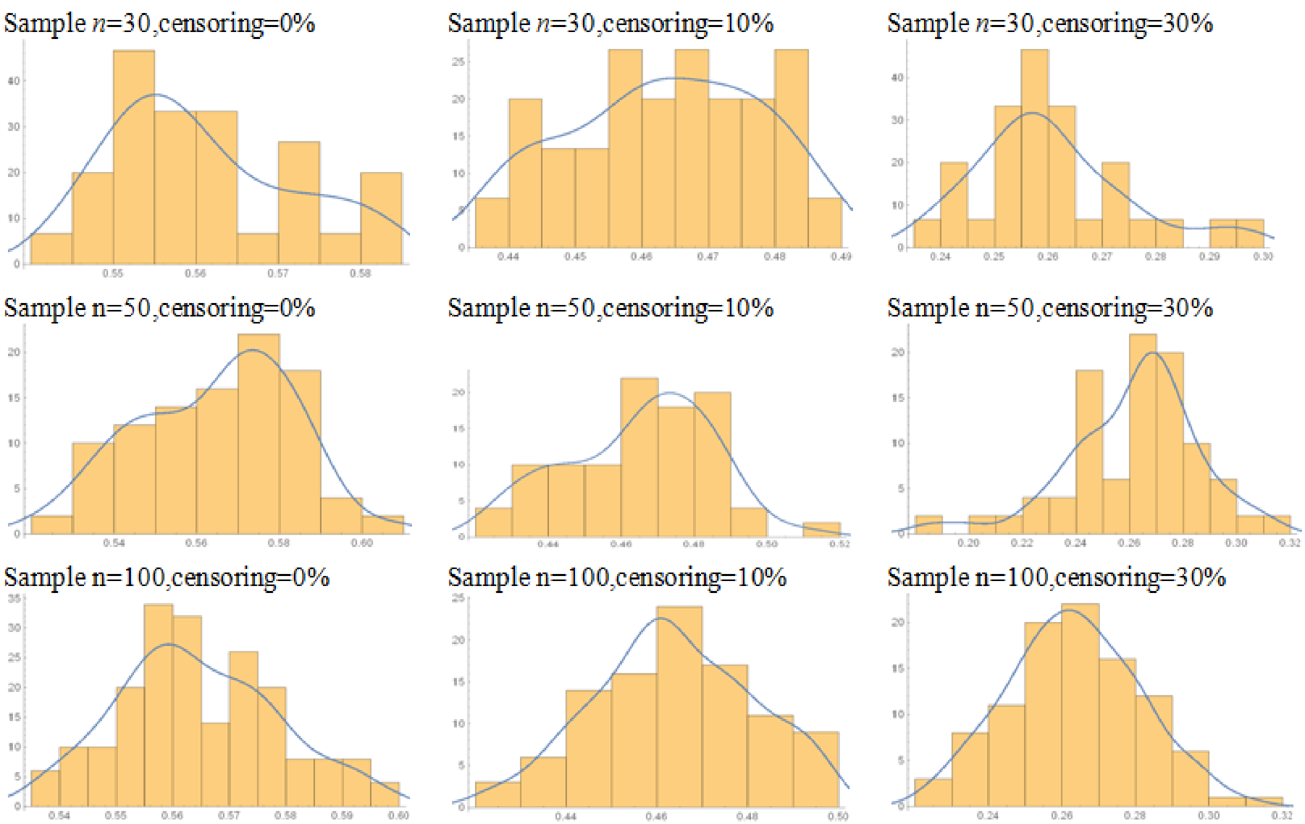

Figure 4. Histogram and Smooth curve plots for martingale residuals (r_{M_i}) . Sample sizes n=30, 50 and 100, percentage of right-censored=10, 30 and 50, σ=1.8

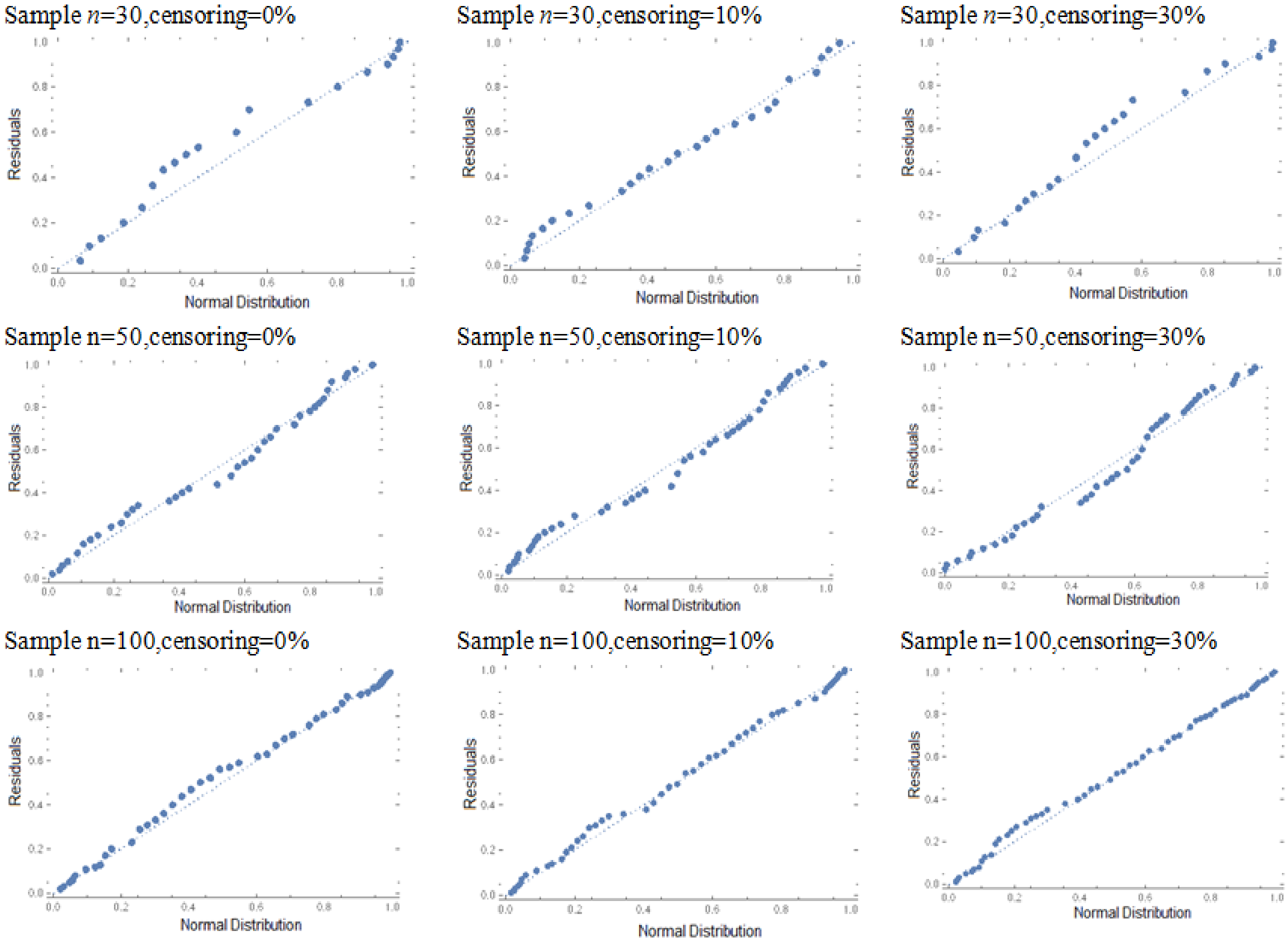

Figure 5. Normal probability plots for martingale residuals (r_{M_i}) . Sample sizes n=30, 50 and 100, percentage of right-censored=10, 30 and 50, σ=5

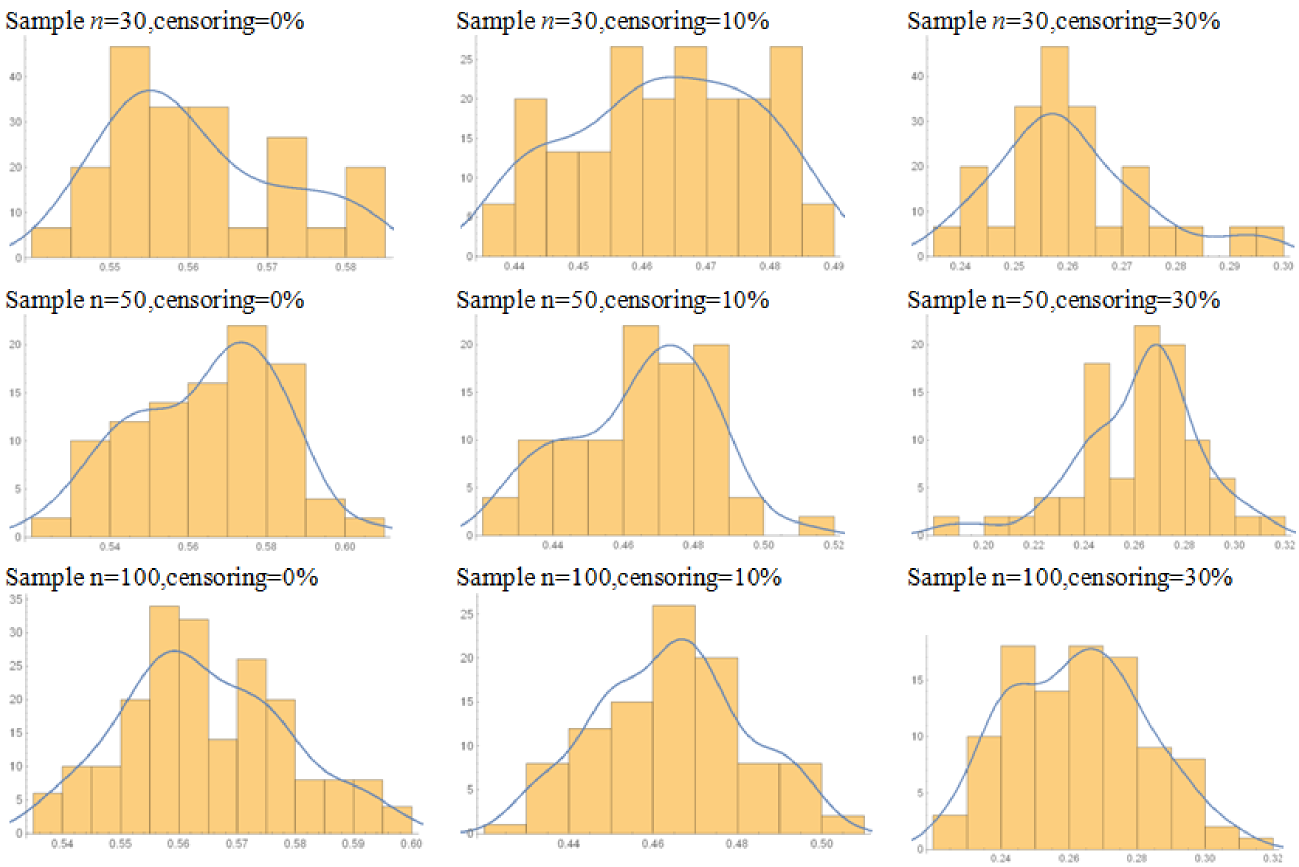

Figure 6. Histogram and Smooth curve plots for martingale residuals (r_{M_i}) . Sample sizes n=30, 50 and 100, percentage of right-censored=10, 30 and 50, σ=5

From Figure 1 – Figure 6 we can extract the following interpretations;

- We can observe that the empirical distribution of the martingale residuals presents agreement with the normal distribution.

- As the sample size increased, the empirical distribution of the martingale residuals seems to present the best agreement with the normal distribution.

- As the $1/\sigma$ increasing, the empirical distribution of the martingale residuals seems to present the best agreement with the normal distribution.

6.3. Application: Myeloma Data

[12] used the data set given in [10] for the fit **LBLLogistic** regression model. The aim of the recent study is to study the performance of the **LBLLogistic** regression model. We use Mathcad and Mathematica to compute Case-deletion measures $GD_i(\theta)$ and $LD_i(\theta)$ defined in (9) and (10). The results of such influence measure index plots are displayed in Figure 7 and Figure 8. These plots show that the cases 29, 40, 44, and 48 are possible influential observations.

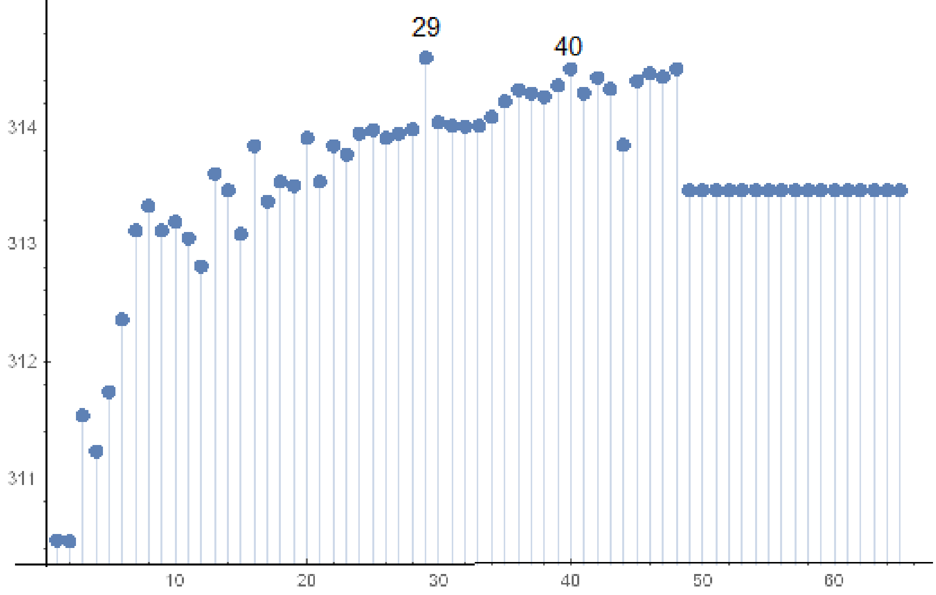

Figure 7. the index plot of $GD_i(\theta)$ on the myeloma data

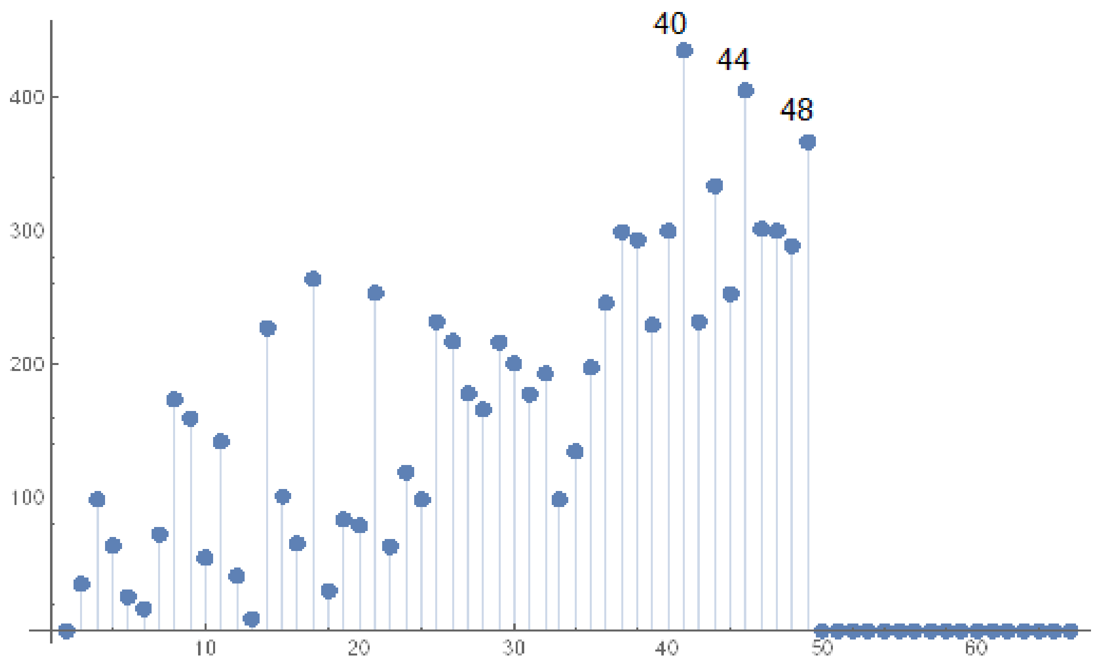

Figure 8. The index plot of $LD_i(\theta)$ on the myeloma data

Impact of the detected influential observations

The diagnostic analysis $GD_i(\theta)$ and $LD_i(\theta)$ detected the four influential observations (cases 29, 40, 44, and 48). The observation 29 corresponds to that one of the largest blood urea nitrogen measurement and age .The observation 40 corresponds to that one of the largest blood urea nitrogen measurement, age, and serum calcium measurement at diagnosis. The observation 44 corresponds to that one of the lowest hemoglobin measurement and serum calcium measurement at diagnosis. The observation 48 corresponds to that one of the largest hemoglobin measurement at diagnosis. In order to reveal the impact of these four observations on the parameter estimates, the model is refitted under some situations. First, each one of these four observations is individually eliminated. Second, we remove from the set "A" (original data set) the totality of potentially influential observations. Table 2 provides the relative change of each estimate (after the "set I" of observations being removed), and the corresponding p-value. Table 2 provides the following sets: I_1={#29}, I_2={#40}, I_3={#44}, I_4={#48}, I_5={#29, #40}, I_6={#29, #44}, I_7={#29, #48}, I_8={#29, #40, #44}, I_9={#29, #40, #48}, I_{10}={ #40, #44, #48}, I_{11}={#29, #40, #44, #48}.

Table 2. Estimates and their p-value and Relative changes, for the corresponding set

Dropping	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
None	0.09	-0.396	0.085	0.00125	0.121	-0.039
P-value	(0.5)	(0.02)	(0.001)	(0.5)	(0.4)	(0.1)
RC	68.694	74.705	46.921	74.933	56.732	74.609
Set I_1	0.09	-0.395	0.085	0.001254	0.122	-0.039
P-value	(0.176)	(0.43)	(0.045)	(0.214)	(0.134)	(0.4)
RC	68.772	74.726	47.425	74.924	56.525	74.49
Set I_2	0.089	-0.394	0.085	0.001255	0.12	-0.039
P-value	(0.179)	(0.429)	(0.046)	(0.214)	(0.135)	(0.399)
RC	69.028	74.783	47.329	74.908	56.988	74.739
Set I_3	0.089	-0.395	0.084	0.001253	0.12	-0.039
P-value	(0.178)	(0.43)	(0.046)	(0.214)	(0.135)	(0.4)
RC	69.059	74.721	47.554	74.933	56.999	74.537
Set I_4	0.089	-0.395	0.086	0.001253	0.121	-0.039
P-value	(0.178)	(0.43)	(0.045)	(0.214)	(0.134)	(0.4)
RC	69.074	74.713	46.629	74.934	56.638	74.637
Set I_5	0.089	-0.394	0.084	0.001255	0.121	-0.039
P-value	(0.179)	(0.429)	(0.046)	(0.214)	(0.135)	(0.399)
RC	69.118	74.808	47.863	74.897	56.779	74.618
Set I_6	0.089	-0.395	0.084	0.001254	0.121	-0.039
P-value	(0.178)	(0.43)	(0.046)	(0.214)	(0.135)	(0.4)
RC	69.149	74.744	48.095	74.922	56.79	74.412
Set I_7	0.089	-0.395	0.085	0.001254	0.122	-0.039
P-value	(0.179)	(0.43)	(0.045)	(0.214)	(0.134)	(0.4)
RC	69.201	74.735	47.141	74.924	56.432	74.513
Set I_8	0.088	-0.394	0.083	0.001255	0.12	-0.039
P-value	(0.181)	(0.429)	(0.047)	(0.214)	(0.136)	(0.399)
RC	69.488	74.83	48.582	74.895	57.052	74.543
Set I_9	0.088	-0.394	0.084	0.001255	0.121	-0.039
P-value	(0.182)	(0.429)	(0.046)	(0.214)	(0.135)	(0.398)
RC	69.47	74.821	47.601	74.897	56.656	74.65
Set I_{10}	0.087	-0.394	0.084	0.001255	0.12	-0.039
P-value	(0.183)	(0.428)	(0.047)	(0.214)	(0.135)	(0.398)
RC	69.802	74.817	47.742	74.906	57.175	74.703
Set I_{11}	0.087	-0.393	0.083	0.001255	0.121	-0.039
P-value	(0.183)	(0.428)	(0.047)	(0.214)	(0.135)	(0.398)
RC	69.922	74.845	48.333	74.895	56.954	74.572

The figures in Table 2 indicate that the estimates of the **LBLLogistic** regression model are not highly sensitive under deletion of the outstanding observations except for $\hat{\beta}_1$ where it became nonsignificant. In general, the significance of the parameter estimates does not change after removing the sets. Hence, we do not have inferential changes after removing the observations handed out in the diagnostic plots.

7. Concluding Remarks

An appropriate matrix for assessing local influence is obtained. We have displayed various simulation studies to assess the performance of estimating the parameters of the **LBLLogistic** regression model and we noticed that; the bias, MSE, and relative root MSE decreased when $a > b$ and when the sample size increased, and the bias, MSE and relative root MSE increased when $a < b$ and when the sample size decreased. We also noticed that when the censoring percentages increase, the bias, MSE, and relative root MSE increase. Also, various simulation studies are performed to investigate the form of the empirical distribution of the martingale residual and we noticed that the martingale residual has shaped normal form. Finally, the authors have analyzed a data set as an application of influence diagnostics in the **LBLLogistic** regression model, although the diagnostic plots detected some possible influential observations, their deletion did not cause substantial changes in the results.

References

[1] Carrasco, J. M. F., Ortega, E.M.M., and Paula, G. A. Log-modified Weibull regression models with censored data:

sensitivity and residual analysis. Computational statistics and data analysis. 52(2008). 4021-4039.

[2] Cook, R.D., (1977) Detection of influential observation in linear regression, Technometrics, Vol., 17, No., 1, 15-18.

[3] Cook, R. D. Assessment of local influence. Journal of the Royal Statistical Society. Series B (Methodological) . 48. 2. (1986). 133-169.

[4] Cook, R. D. and Weisberg, S. Residuals and Influence in Regression. Chapman and Hall. London. (1982).

[5] Cordeiro, G. M., Gomes, A. E., Silva, G. O., and Ortega, E. M. M. The beta exponentiated Weibull distribution. Journal of statistical computation and simulation. 83. 1. (2013). 1141-138.

[6] Escobar, L. A. and Meeker, W. Q. Assessing influence in regression analysis with censored data. Biometrics. 48(1992). 507-528.

[7] Feigl, P., and Zelen, M. Estimation of exponential survival probabilities with concomitant information. Biometrics. 21. 4. (1965). 826-838.

[8] Gusmao, F. R. S., Ortega, E.M.M., and Cordeiro, G. M. The generalized inverse Weibull distribution. Stat Papers. 52(2011). 591-619.

[9] Hashimotoa, E.M., Ortega, E.M.M., Canchoc, V.G., and Cordeiro, G.M. The log-exponentiated Weibull regression model for interval-censored data. Computational Statistics and Data Analysis. 54(2010). 1017–1035.

[10] Krall, J., Uthoff, V. and Harley, J. A step-up procedure for selecting variables associated with survival. Biometrics. 31(1975). 49-57.

[11] Lawless, J. F. Statistical Models and Methods for Lifetime Data. John Wiley & sons. New York. (1982).

[12] Mahmoud, M. R., EL- Sheikh, A. A., Morad, N. A., and Ahmad, M. A. M. Log-beta log-logistic regression model. International Journal of Sciences: Basic and Applied Research. 22. 2 (2015). 389-405.

[13] Ortega, E. M. M., Cordeiro, G. M., and Lemonte, A. J. A log-linear regression model for the β-Birnbaum–Saunders distribution with censored data. Computational Statistics and Data Analysis. 56(2012). 698-718.

[14] Ortega, E.M.M., Cancho, V. G. and Bolfarine, H. Influence diagnostics in exponentiated-Weibull regression models with censored data. SORT. 30. 2. (2006). 171-192.

[15] Prentice, R. L. Exponential survival with censoring and explanatory variables. Biometrika. 60 (1973). 279-288.

[16] Silva, G. O., Ortega, E.M.M, Cancho, V. G and, Barreto, M.L. Log-Burr XII regression models with censored data. Computational statistics and data analysis. 52 (2008). 3820-3842.

[17] Silva, G.O., Ortega, E.M.M., Cordeiro, G.M. A log-extended Weibull regression model. Computational Statisticsand Data Analysis. 53(2009). 4482-4489.

[18] Zippin, C., and Armitage, P. Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. Biometrics. (1966). 665-672.