DIAGNOSTIC TESTS FOR QUANTILE REGRESSION WITH WEIBULL RESPONSE VARIABLES

Besta Okey Onyegbuchulem[1], Chialuka Adline Onyegbuchulem[2], Obiageri Ogwo[3], Emmanuel Oliwe[4]

1,3&4. Department of Mathematics/Statistics, Imo State Polytechnic, Omuma

2. Department Mathematics, Alvan Ikoku University of Education Owerri

Corresponding Author's E-mail: bokey@imopoly.net. +234 8032692379

Abstract

Quantile regression is a valuable tool for modeling the relationship between covariates and the conditional quantiles of a response variable. However, the validity of quantile regression inferences relies heavily on the accuracy of the model assumptions. This study focuses on developing diagnostic tests for quantile regression models with Weibull-distributed response variables. We propose a series of diagnostic tests based on cumulative residual processes to assess the adequacy of the quantile regression model. The proposed tests are evaluated through simulation studies, and their performance is compared to existing methods. The results demonstrate the importance of checking model assumptions in quantile regression analysis, particularly when dealing with Weibull response variables. The proposed diagnostic tests are applied to a real dataset to illustrate their practical utility.

Keywords: Quantile regression, Weibull distribution, diagnostic tests, cumulative residual processes, model checking.

1.0 Introduction

Quantile regression model is a type of regression analysis used in statistics and econometrics. Unlike the traditional linear regression model that predicts the mean of the response variable, quantile regression model predicts a specified quantile of the response variable. This is important to researchers because it is useful to understand how predictors affect different parts of the response variable's distribution.

Ordinary Least Square (OLS) Estimator is one of the estimators of Linear regression model and is also Best Linear Unbiased Estimator (BLUE). However, violation of one or more of the OLS Estimator may result in a situation where the OLS estimator can no longer be said to be BLUE (Nwabueze, 2022). This led Researchers to the concept of Generalized Least Squares. The concept of Generalized Least Square metamorphosized into the median regression. The median is one of the many quantiles and a special quantile which describes the central position of a distribution. Hence, the Conditional-median regression is a type of quantile regression where the conditional 50th quantile is modeled as a function of dependent variable, other quantiles can obviously be used to determine the none central positions of a distribution. Onyegbuchulem (2019) stated that the quantile notion however generalizes some terms like the percentile, the decile, the quintile and the quartile. As for the p^{th} quantile, it is used to denote that value of the dependent variable which its proportion is below the part of population that is p. Thus, quantiles can specify and understudy any position of a distribution. For example, 7.5% of the population lies below the 0.075th quantile.

Koenker & Bassett (1978) introduced the first order Quantile Regression model which has the form

$$Q_{y_i}\left(\tau/x\right) = \beta_0 + \beta_i x_i + F_u^{-1}(\tau) \qquad (1)$$

were

Q_{yi} is the conditional value of the dependent variable given τ in the i^{th} trial,

β_0 is the intercept parameter,

β_i is the slop parameter,

τ denotes the quantile (e.g., $\tau = 0.5$ for the median),

X_i is the value of the independent variable in the i^{th} trial,

F_u is the common distribution function of the error τ,

$E\left(F_u^{-1}(\tau)\right) = 0,\ for\ i = 1, - - -, n,\ e.g.\ F^{-1}(0.50)$ is the median or 0.50 Quantile.

Onyegbuchulem et al. (2019) introduced Cauchit quantile regression. In cauchit regression analysis, the dependent variable is ordinal (statistically, it is polytomous ordinal) and the independent variables can be ordinal or continuous. Onyegbuchulem et al. (2019) introduced Cauchit quantile regression model for two major reasons which include causal analysis and forecasting an effect.

In this work, it may not be out of place to extend the work of Onyegbuchulem et al., (2019). This work therefore used Weibull as the Response Variables to conduct a diagnostic test on quantile regression model. The choice of Weibull as the response variable is because of the need to make the response variable a left skewed continuous variable.

2.0 Methodology

2.1 Quantile Regression

In order to estimate other conditional quantile functions, the conditional quantiles are considered and the absolute value is replaced by $\left(y_i - \xi\left(x_i^t\beta\right)\right)$ to obtain equation (2):

$$\hat{\beta}^{(\tau)} = \min_{\beta \in \Re^n} \sum_{i=1}^{n} \rho_\tau\left(y_i - \xi\left(x_i^t\beta\right)\right)\left(Huo\, et.\, al., 2013\right) \quad \textbf{(2)}$$

Where

$\xi\left(x_i\beta\right)$ is the conditional quantile function, $y_i = y_1, y_2, \ldots, y$ which are the observed values of y and ξ is the estimator of the $\tau-$ quantile of Y. In case of linear quantile function, $Q_N\left(\tau/\beta\right) = x_i^t\beta^\tau$ is obtained by solving equation (2) by linear programming hence parameter $\beta^{(\tau)}$ and parameter $\hat{\beta}^{(\tau)}$ both depend on $\tau \in (0,1)$

$$y_i = Q_{yi}\left(\tau/\beta\right) = \min_{\xi \in \Re^n} \sum_{i=1}^{n}\left(\beta_0^{(\tau)} + \beta_1^{(\tau)}x_i\right) \quad \left(Huo\, et.\, al., 2013\right) \quad \textbf{(3)}$$

Minimizing equation (3) results in a quantile regression model. The resulting minimization problem of equation (3), when $\left(x_i\beta(\tau)\right)$ is formulated as a linear function of the parameters can be solved very efficiently by linear programming method. The progression of the idea that led to equation (3) motivated the original quantile regression model presented in Koenker & Bassett (1978).

2.2 Weibull Distribution

The Weibull distribution is a continuous probability named after Swedish mathematician Waloddi Weibull, who described it in detail in 1951, although it was first applied by Rosin & Rammler (1933) to describe a gain size distribution. Liu et al. (2023) used the Weibull accelerated failure time regression model to predict time to health events, demonstrating its application in medical predictions. Li et al. (2025) focus on reliability assessment of series system with Weibull-distributed components based on zero-failure data, Introducing a practical approach for high-reliability components. Additionally, Oketch and Sepehrifar (2026) developed a Bayesian framework using adaptive MCMC to analyze Weibull failure time which was applied to prostate cancer survival data. This research will focus on examining the diagnostic test of Weibull distribution as a response variable of a quantile distribution variable. The probability density function (PDF) of Weibull distribution is therefore given as:

$$f(x) = \frac{\alpha}{\psi}\left(\frac{x}{\psi}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \quad (4)$$

$$\alpha \rightarrow shape\ parameter\ (\alpha > 0)$$
$$\psi \rightarrow shape\ parameter\ (\psi > 0)$$
$$(x \geq 0)$$

Equation (4) is the probability density function of Weibull distribution. Next is to derive the Cumulative density function of the Weibull distribution, which is derived as follows:

$$F(x) = F(X \leq x) = \int_0^u \frac{\alpha}{\psi}\left(\frac{x}{\psi}\right)^{\alpha-1} e^{-\left(\frac{\alpha}{\psi}\right)} \quad (5)$$

Let

$$\left.\frac{x}{\psi}\right|_0^{\alpha} = A$$

$$\rightarrow d\psi\left(\frac{x}{\psi}\right)^{\alpha} = \frac{d(A)}{dx} \quad (6)$$

$$d\psi\left(\frac{x}{\psi}\right)^{\alpha-1}\frac{1}{\psi} = \frac{d\alpha}{dx} \quad (7)$$

$$\Rightarrow d\psi\left(\frac{x}{\psi}\right)^{\alpha-1}\frac{1}{\psi}dx = d\alpha$$

$$\therefore \int_0 F(x) = \int_0 \left(\frac{x}{\psi}\right)^{\alpha} e^{-\alpha} d\alpha$$

$$= \left[\frac{e^{-\alpha}}{-1}\right]^{\left(\frac{x}{\psi}\right)^{\alpha}} = -\left[e^{-\alpha}\right]^{\left(\frac{x}{\psi}\right)^{\alpha}}$$

$$-\left[e^{-\left(\alpha/\psi\right)^{\alpha}}\right] - e^{0} = -e^{-\left(\alpha/\psi\right)^{\alpha}} + 1$$

$$\therefore F(x,\alpha,\psi) = 1 - e^{-\left(\alpha/\psi\right)^{\alpha}} \tag{8}$$

Equation (8) is the cumulative density function of Weibull distribution. To find the quantile function, $F(x)$ is set as:

$$F(x) = p,\ p \in [0,1] \tag{9}$$

$$p = 1 - e^{-\left(x/\psi\right)^{\alpha}}$$

$$\Rightarrow 1 - p = e^{-\left(x/\psi\right)^{\alpha}} \tag{10}$$

Applying the natural logarithm to remove the exponential

$$\Rightarrow In(1-p) = -\left(\frac{x}{\psi}\right)^{\alpha} \tag{11}$$

To remove the negative sign on the right-hand solution

$$\Rightarrow -In(1-p) = -\left(\frac{x}{\psi}\right)^{\alpha} \tag{12}$$

Solve for x by first inversing the power

$$-In(1-p)^{\left(1/\alpha\right)} = -\frac{x}{\psi} \tag{13}$$

To derive the quantile function, solve for x

$$Q(p) = \psi\left(-In(1-p)^{1/\alpha}\right)$$

$$\Rightarrow \psi\left(-In(1-p)^{1/\alpha}\right) = x \ \ where\ x = Q(p)$$

$$Q(p) = \psi\left(-In(1-p)^{1/\alpha}\right) \tag{14}$$

2.3 Weibull Transformed Quantile Regression

Equation (14) is the Weibull quantile function. Weibull transformed quantile regression is from the family of extreme value distribution. The family of extreme value distributions is a group of distributions that model the extreme values (maxima or minima) of a set of random variables. The three main types are (Gumbel (type I), Fréchet (type II), and Weibull (type III)). As stated earlier, Weibull distribution was introduced by Swedish mathematician Waloddi (Weibull, 1951)

This work therefore, introduced Weibull transformed quantile regression model where the quantile function of Weibull distribution is used as the transformer to transform the quantile regression into a new model that can handle extreme value response data, much attention will be on its diagnostic examinations using the t-values, the p-values, the mean square errors, standard deviations, residual mean square errors, psoudo-R, skewness, kurtoses, AICs and comparing residual means with the residual medians. The general probability density function of a Weibull distribution is given in equation (4), the general cdf is given in equation (8) while the cdf inverse or the quantile function of the Weibull distribution that will be used for data simulation is derived from the cdf of Weibull distribution in equation (14). The next step is to equate the cdf inverse $\left(F^{-1}\right)$ of the Weibull function of equation (14) to the quantile regression model and solve simultaneously for the Weibull quantile regression model.

To derive the quantile function, solve for x

$$\varphi = \left(\beta_0^{(\tau)} + \beta_1^{(\tau)} x_i\right) \tag{15}$$

$$Q(p) = \psi\left(-In\left(1 - Q(p)\right)^{1/\alpha}\right) \tag{16}$$

$$-In\left(1 - Q(p)\right)^{1/\alpha} = \frac{\varphi}{\psi}$$

$$-In\left(1 - Q(p)\right) = \left(\frac{\varphi}{\psi}\right)^{\alpha}$$

$$-In\left(1 - Q(p)\right) = -\left(\frac{\varphi}{\psi}\right)^{\alpha}$$

$$1 - Q(p) = e^{-\left(\frac{\varphi}{\psi}\right)^{\alpha}}$$

$$-Q(p) = e^{-\left(\frac{\varphi}{\psi}\right)^{\alpha}} - 1$$

$$Q(p) = 1 - e^{-\left(\frac{\varphi}{\psi}\right)^{\alpha}} \quad (17)$$

$$Q(p) = 1 - e^{-\left(\frac{\left(\beta_0^{(\tau)} + \beta_1^{(\tau)} x_i\right)}{\psi}\right)^{\alpha}} \quad (18)$$

$Q(p)$ = the response variable and the cdf inverse $\left(F^{-1}\right)$ of the distribution to be estimated

x_i = the covariates to be simulated

β_0 = the intercept parameter

β_i = the unknown parameters

$\tau =$ specified quantiles of the model. This research examines the following quantiles: 0.05, 0.25, 0.5, 0.75, 0.95

Equation (18) is the proposed Weibull transformed quantile regression model.

2.4 Data Simulation and Analysis

The simulation experiments were adopted from the Hao and Naiman (2007). The data simulation was performed following Nwakuya et al. (2019) on fuel consumption of some selected cars. The data were simulated using the the R – code as follows:

$$x_1 = -\left(\tan\left(pi * s * \left(1/180\right)\right)\right), \text{ were, } s = \left(\frac{7}{22} * (m)\right), \; m = (t - 0.5)$$

$$x_2 = rnorm(1500, 5, 0.8),$$

$$x_3 = rgama(n, 1.3, 0.8)$$

$$y = rweibull(n, 5, 8)$$

$$n = sample = 1500$$

3.0 Results and discussion

The graph of the Weibull response variable is presented in Figure 1, the estimated coefficient, the standard errors, the t-values and the p values are presented in Table 1, the descriptive analysis for the residuals of the Weibull distributed quantile regression are presented in table 2, the graphs of the residual plot are presented in figure 2 and then the graphs of the estimates of the coefficient plots for the quantiles regression are presented in figure 3.

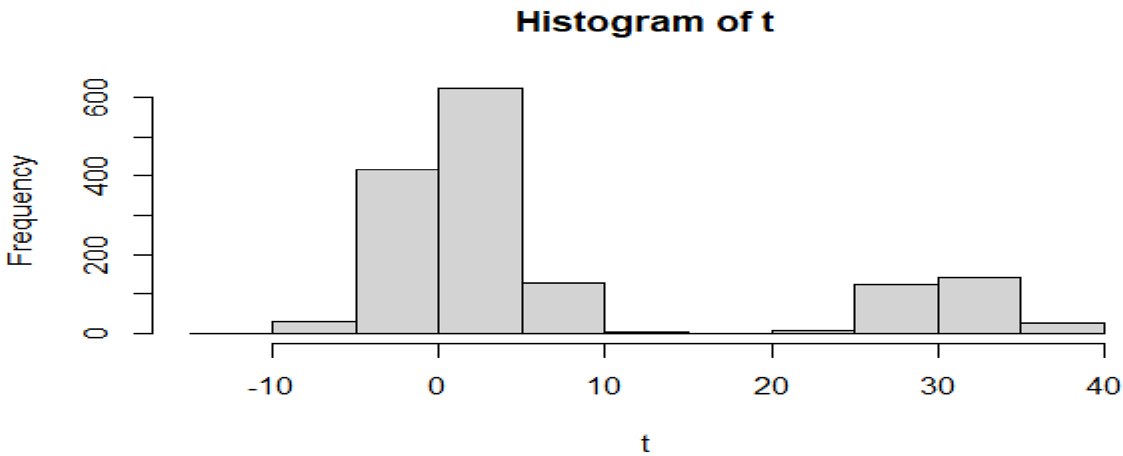


Fig.1: Graph of weibull distributed response variable

Table 1: Estimated Parameters of 1500 Simulated Data for the Weibull transformed QR Model

Quantiles	*parameters*	*coefficient*	*Std error*	*t-value*	*Pr(>|t|)*
0.05	**intercept**	5.5662	**0.5481**	10.1544	**0.0000**
	X_1	0.1446	1.1743	0.1232	0.9019
	X_2	-0.2368	0.1103	-2.1467	0.0319
	X_3	0.0711	0.0536	1.3257	0.1851
0.25	**intercept**	6.45954	0.4050	15.9469	0.0000
	X_1	2.2083	0.9446	2.3377	0.0195
	X_2	-0.0543	0.0777	-0.6991	0.0875
	X_3	0.04753	0.0278	1.7098	0.0875
0.5	**intercept**	7.5777	0.3151	24.0470	0.0000
	X_1	1.87201	0.6260	2.9901	0.0028
	X_2	-0.0146	0.06199	-0.2367	0.8129
	X_3	0.0025	0.0203	0.1258	0.8999
0.75	**intercept**	8.2315	0.2978	27.6353	0.0000
	X_1	0.1645	0.6440	0.2555	0.7983
	X_2	0.0765	0.0587	1.3022	0.1930
	X_3	-0.0238	0.0141	-1.6865	0.0919

0.95	**intercept**	9.3632	0.4527	20.6831	0.0000
	X_1	-0.1231	1.3471	-0.0914	0.9271
	X_2	0.12858	0.0841	1.5273	0.1268
	X_3	0.0147	0.0208	0.7060	0.4802

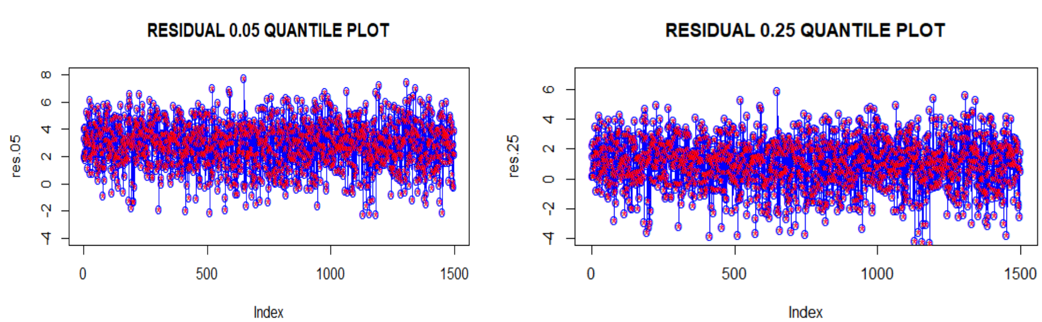
The results of Weibull distributed quantile regression model shows that all the parameters in tables 1 are not significant for all the study quantiles except for the intercept of the 5th, 25th, 50th, 75th and 95thquantiles and x_2 of the 5th quantile, x_1 of the 25th and 50th quantiles. Figure 4, shows that the residual graphs did not cluster around 0 but are uniformly spread around the negative and positive axis in the 50th, 25th and 75th quantile, right skewed in the 5th quantile and left skewed in the 95th quantile.

Table 2: Descriptive analysis for the residuals of the Weibull distributed quantile regression model

Quantiles

Parameters	**0.05**	**0.25**	**0.5**	**0.75**	**0.95**
AIC	7016	6161	5863	5944	6599
Psoudo R	0.995	0.995	0.998	0.999	0.995
skewness	-0.2196	-0.2456	-0.2577	-0.2706	-0.2710
kurtosis	2.9949	3.0508	3.0677	3.0708	3.0612
MSE	11.0505	3.8323	2.7018	3.9087	9.3691
RMSE	3.3242	1.1957	1.6437	1.9770	3.0609
SD	1.6546	1.6430	1.6398	1.6438	1.6443
Mean	2.8835	1.0653	-0.1209	-1.0993	-2.5821
Median	2.9729	1.1823	0.0000	-0.9542	-2.4287

Table 2, show that the skewness for all the study quantiles approximate to zero while the kurtoses for all the study quantiles are within the range of 3, the means are equal to the medians for all the study quantiles, the mean square errors (MSE), and the root mean square errors (RMSE) did not approximate to zero.



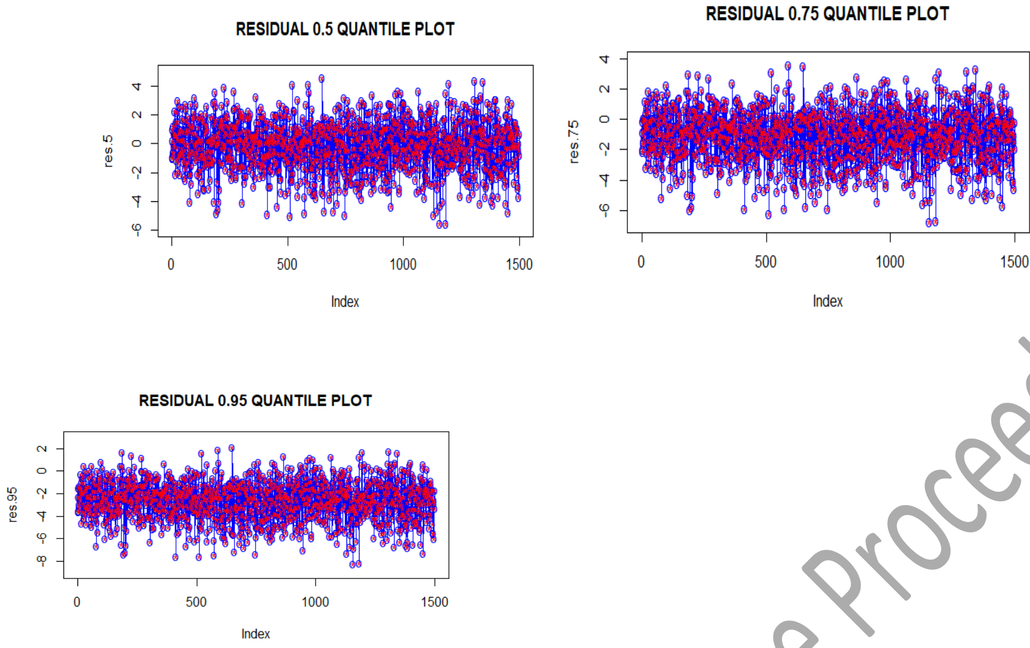


Figure 2: Residual Plot of Quantiles

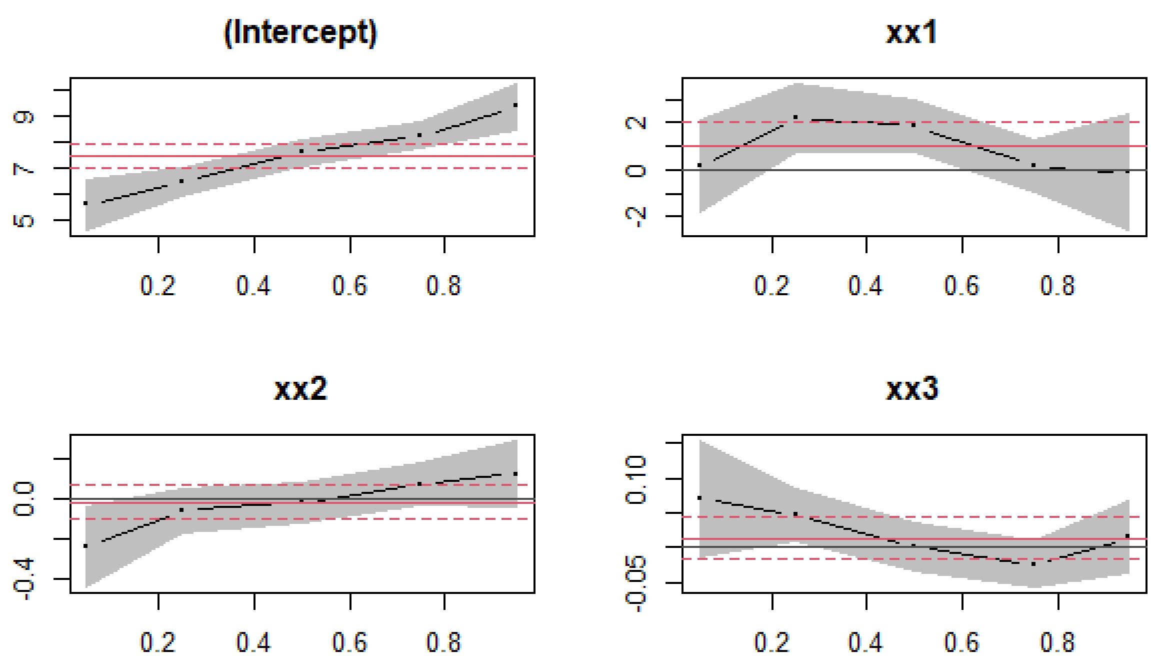


Figure 3: Plot for the Estimate of the Weibull distributed Quantile RM

3.0 Conclusion

Weibull transformed Quantile Regression model shows ability to manage outliers when it is applied to simulated data set. Its residual graphs in Figures 1 shows little presence of outliers especially in the 25th, 50th and 75th quantiles, the summary results of the residuals for the Weibull transformed Quantile Regression model in Tables 2, show that both its skewness and kurtosis are closer to 0 and 3 respectively. Its median and mean are equal; its standard deviation, root mean square error and residual mean square error are smaller compared to the standard deviation that was used to simulate the study data. Based on the above remarks, we therefore conclude that Quantile Regression model can handle data with outliers when transformed with Weibull distribution. Hence it can be recommended for Weibull transformed Quantile Regression model to be used for continuous data with outliers as well data on failure time, data on time, health and other events data.

References

Bosin,P., & Rammler, E. (1933). The laws governing the finesses of powered coal. *Journal of the Institute of fuel. 6(29-36)*

Hao L. &Naiman,D.Q., (2007). *Quantile Regression;* 01-Hao.qxd. 3/13/2007.3.28

Huo, Y., & Wang, Z.(2013). Reliability analysis of consecutive-k-out-of-n:F systems with two types of failures. *Journal of applied mathematics, (1-9)*

Koenker, R., & Bassett, J., G. (1978). Regression quantiles. *Econometrica, 46, 33-50.*

Li, Z., Fu, H., & Guo, J. (2025). Reliability assessment of a series system with Weibull-distributed components based on zero-failure data. *Applied Science 15(5), 2869*

Liu E., Liu, R.Y., & Lim, K. (2023). Using the Weibull accelerated failure time regression model to predict time to health event. *Applied statistics, 13(24)*

Nwabueze, J.C., Onyegbuchulem, B.O. & Nwakuya, M.T. (2022). On the Equivariance of Location Reparameterization of Quantile Regression Model using Cauchy Transformation. *Jouenal of mathematical theory and modelling (12),2*

Nwakuya M. T., Nwabueze J. C., Onyegbuchulem, B. O., & Imoh, J. C. (2019), Response variable transformation for quantile regression model, *International Journal of Scientific & Engineering Research (10)* 8.

Oketch, T. & Sepehrifer, M. (2026). Modeling complex life system: Bayesian inference for Weibull failure time using adaptive MCMC. *Stat paper, 67(2)*

Onyegbuchulem,, B.0, Nwakuya M. T., Nwabueze J. C. & Otu, A. O., (2019), Choice of power transformation of skewed distribution for quantile response model, *Journal of applied mathematics and statistics, vol.7*

Onyegbuchulem,, B.0, Nwakuya M. T., Nwabueze J. C. & Otu, A. O., (2019), Evaluation of the assumptions of linear regression model on the quantile regression model, *International journal of advancement in research & technology, (8)9*

Weibull, W. (1952). Statistical distribution function of wide application, *Journal of applied mechanics, 18(30, 293-297*