

# Choice of Appropriate Power Transformation of Skewed Distribution for Quantile Regression Model

Onyegbuchulem B.O.<sup>1,\*</sup>, Nwakuya M.T.<sup>2</sup>, Nwabueze J.C.<sup>3</sup>, Otu Archibong Otu<sup>4</sup>

<sup>1</sup>Department of Maths/Statistics, Imo State Polytechnic Umuagwo, Nigeria

<sup>2</sup>Department of Maths/Statistics, University of Port Harcourt, River State, Nigeria

<sup>3</sup>Department of Statistics, Federal University of Agriculture Umudike, Nigeria

<sup>4</sup>Department of Research and Statistics, Central Bank of Nigeria, Owerri

\*Corresponding author: bokey@imopoly.net

Received January 18, 2019; Revised March 28, 2019; Accepted May 04, 2019

**Abstract** Quantile Regression (QR) performed better than Ordinary Least Square (OLS) when the Data is skewed. Its best result can be achieved when the Data is transformed. Quantreg package of R software was used to illustrate the various power transformation fitness for quantile regression model. The analysis shows that the best result was obtained from the square root of y transformation with an average error term ( $\epsilon_i$ ) of 0.9539, -0.0494, 0.0238, -0.5309 and -0.7544 for 10th, 25th, 50th, 75th and 90<sup>th</sup> quantile respectively. From the results obtained, it shows that model transformation can greatly improve the result of quantile regression model.

**Keywords:** *Quantile Regression, skewed distribution, power transformation and model selection*

**Cite This Article:** Onyegbuchulem B.O., Nwakuya M.T, Nwabueze J.C, and Otu Archibong Otu, "Choice of Appropriate Power Transformation of Skewed Distribution for Quantile Regression Model." *American Journal of Applied Mathematics and Statistics*, vol. 7, no. 3 (2019): 105-111. doi: 10.12691/ajams-7-3-4.

## 1. Introduction

Conditional-median regression is a special case of quantile regression in which the conditional 50<sup>th</sup> quantile is modeled as a function of covariates. More generally, other quantiles can be used to describe non-central positions of a distribution; the quantile notion generalizes specific terms like quartile, quintile, decile, and percentile. The pth quantile denotes that value of the dependent variable below which the proportion of the population is p. Thus, quantiles can specify any position of a distribution [7].

The first – order Quantile Regression model was introduced by Koenker and Bassett [2]. It has the form

$$Q_y\left(\frac{\tau}{x}\right) = \beta_0 + \beta_i X_i + F^{-1}u(\tau) \quad (1)$$

Where

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

$\beta_0$  is the intercept,

$\beta_i$  is a parameter,

$\tau$  denotes the quantile (eg.,  $\tau = 0.5$  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  $\tau$ ,

$E(F^{-1}_u(\tau)) = 0$ , for  $i = 1, \dots, n$ , eg  $F^{-1}(0.5)$  is the median or 0.5 quantile.

This model conditional quantile is a function of covariates. Therefore, quantile regression model is naturally an

extension of the linear-regression model. While the linear-regression model specifies the change in the conditional mean of the dependent variable associated with a change in the covariates, the quantile- regression model specifies changes in the conditional quantile. Since any quantile can be used, it is possible to model any predetermined position of the distribution. Thus, researchers can choose positions that are tailored to their specific inquiries

However, the expected error term of quantile regression models especially the median regression model which is closely related to the linear regression in term of precision often than not don't approximate to zero. Reference [1] showed that the expected error term of multiple quantile regression can be improved by transforming the response variable using log transformation. Reference [2] uses the relationship between variances and means over several groups to find the appropriate transformation for the study data which makes the variance independent of the mean. Reference [2] shows that procedure for determining the appropriate transformation is to determine the coefficient ( $\beta$ ) of regression of natural logarithm of group standard deviation ( $\hat{\sigma}_i$ ) on the natural logarithm of group average ( $\bar{x}_{i,i=1,2,\dots,m}$ ). He explained that the most popular and common transformations are the power of transformation such as:  $\sqrt{x_t}$ ,  $\log_e X_t$ ,  $1/X_t$ ,  $1/\sqrt{X_t}$ ,  $1/X_t^2$ ,  $X_t^2$ . Reference [1] empirically analysed the monthly earning distribution of Pakistan using logarithm transformation, where the log of monthly earning is taken as a response variable, while education, experience, age, sex, marital status, nature of work, region, and the provinces are used as explanatory variables. Therefore, this study will apply

the five powers of transformation stated by [2] on the response variable to ascertain the appropriate power of transformation that can be used to model quantile regression in the presence of skewed distribution. This study is aimed at investigating the best power transformation of skewed distribution for quantile regression model. The study will specifically:

Assess the best transformation fit of the model based on some selected power transformations, Assess the impact of the covariate on the response variable and Conduct diagnostic tests on the suggested model.

## 2. Methodology

This paper investigates the best power transformation for quantile regression model. The data was generated using Monte Carlo Simulation technique from the data of Weibull distributed data using the sharp and scale parameters of Annual salaries, income and wages of Health workers in Nigeria. The generated data shall be analysed using transformed quantile regression Model. The statistical software to be used in the analysis will be quantreg package of R Software. The hypothesis is therefore, stated as:

$H_0: \beta_i(\tau) = 0$ , Covariate( $x_i$ ) has no significant effect on the response variable

$H_1: \beta_i(\tau) \neq 0$ , Covariate( $x_i$ ) has a significant effect on the response variable

### 2.1. Quantile Regression Model

If we consider the i.i.d sample of  $y_1, \dots, y_n$ , the unconditional sample mean can be defined as the solution to the problem of minimizing sum of squared residual

$$\hat{\mu} = \min_{\mu \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu)^2 \tag{2}$$

Hence the sample median  $\xi$  is the minimizer of the sum of absolute error loss or deviations.

$$\xi = \min_{\xi \in \mathbb{R}} \sum_{i=1}^n |y_i - \xi| \tag{3}$$

To see why median can be define as a minimization problem, it can be written as

$$\begin{aligned} E|y - \xi| &= \int_{-\infty}^{\infty} |y - \xi| f(y) dy \\ &= \int_{y=-\infty}^{\xi} |y - \xi| f(y) dy + \int_{y=\xi}^{+\infty} |y - \xi| f(y) dy \\ &= \int_{y=-\infty}^{\xi} (\xi - y) f(y) dy + \int_{y=\xi}^{+\infty} (y - \xi) f(y) dy \end{aligned} \tag{4}$$

Differentiating with respect to  $\xi$  and setting the partial derivative to zero will lead to the solution for the minimization problem. The partial derivative of the first term is:

$$\begin{aligned} \frac{\partial}{\partial \xi} \int_{y=-\infty}^{\xi} (\xi - y) f(y) dy \\ = (\xi - y) f(y) \Big|_{y=\xi} + \int_{y=-\infty}^{\xi} \frac{\partial}{\partial \xi} (\xi - y) f(y) dy \end{aligned}$$

$$\int_{y=-\infty}^{\xi} f(y) d(y) = F(\xi)$$

And the partial derivative of the second term is:

$$\begin{aligned} \frac{\partial}{\partial \xi} \int_{y=\xi}^{+\infty} (y - \xi) f(y) dy &= \int_{y=\xi}^{+\infty} f(y) d(y) \\ &= -(1 - F(\xi)) \end{aligned}$$

Combining these two partial derivatives lead to:

$$\begin{aligned} \frac{\partial}{\partial \xi} \int_{-\infty}^{+\infty} (y - \xi) f(y) dy &= F(\xi) - (1 - F(\xi)) \\ &= 2F(\xi) - 1. \end{aligned} \tag{5}$$

By setting  $2F(\xi) - 1 = 0$ , we solve for the value of  $F(\xi) = \frac{1}{2}$ , that is, the median, to satisfy the minimization problem.

For the general  $\tau$ th sample quantile  $\xi(\tau)$ , which is the analogue of  $\rho(\tau)$ , may be formulated as the solution of the optimization problem

$$\rho\tau(\xi) = \min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho\tau(y_i - \xi) \tag{6}$$

Repeating the above argument for quantiles, the partial derivative for quantiles corresponding to (6)

$$\begin{aligned} \frac{\partial}{\partial \xi} E[\rho\tau(y, \xi)] &= (1 - \rho)F(\xi) - \rho(1 - F(\xi)) \\ &= F(\xi) - \rho. \end{aligned} \tag{7}$$

We set the partial derivative  $F(\xi) - \rho = 0$  and solve for the value of  $F(\xi) - \rho$  that satisfies the minimization problem. Equation (7) is illustrated thus:

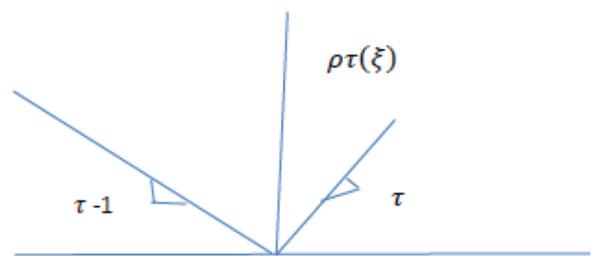


Figure 1. Quantile Regression  $\rho$

Just as the unconditional sample mean in (2) minimizes the sum of square residuals (error lose), the conditional sample mean also minimizes the sum of square residual by replacing the scalar  $\mu$  by  $\mu(x_i, \beta)$ , the estimate of the conditional mean function  $E[Y/x]$  is obtained

$$E(Y/x) = \min_{\beta \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu(x_i, \beta))^2 \tag{8}$$

This can be proceeded in the same way in quantile regression. to obtain an estimate of the conditional median function, the scalar  $\xi$  in (3) is replaced by the parametric function  $\xi(x_i, \beta)$ :

$$\tilde{\beta}(\tau = 1/2) = \min_{\xi \in \mathbb{R}} \sum_{i=1}^n |y_i - \xi(x_i, \beta)| \tag{9}$$

To obtain the estimates of the other conditional quantile function the conditional quantile is considered and the absolute values is replaced by (10):

$$\rho\tau(\tilde{\beta}) = \min_{\xi \in \mathfrak{R}} \sum_{i=1}^n \rho\tau(y_i - \xi(x_i, \beta)). \quad (10)$$

Minimizing (10) results in a quantile regression model. The resulting minimization problem of (10), when  $\xi(x_i, \beta(\tau))$  is formulated as a linear function of the parameters can be solved very efficiently by linear programming method. The progression of ideas that led to (10) motivated the original quantile regression model presented in [8]

### 2.2. Model Specification

Following [8] and [7], our proposed model will take the form:

$$Q(p)^{(\tau)} = \beta_0^{(\tau)} + \beta_i^{(\tau)} X_i + \varepsilon_i^{(\tau)} \quad (11)$$

Where

$Q(p)^{(\tau)}$  = Transformed response variable containing  $n$  observations simulated from the parameters of data on Health Workers Allowances,

$\beta_0$  = Intercept

$\beta_i$  = Unknown Parameters

$\varepsilon$  = a classical error terms

$\tau$  = Specified quantiles of Simulated Data. This research examines the following quantiles: 0.1, 0.25, 0.5, 0.75, 0.9

$X$  = the covariate

$n$  the sample sizes = 3000 was used.

### 2.3. Goodness of Fit of QRM

An analog of  $R^2$  statistics can be readily developed for quantile regression models. Reference [5] stated that “Since linear regression model fits are based on the least square criterion and quantile regression models are based on minimizing a sum of weighted distance  $\sum_{i=1}^n d_p(y_i, \hat{y}_i)$  with different weights used depending on whether  $y_i > \hat{y}_i$  or  $y_i < \hat{y}_i$ . The goodness of fit will be measured in a manner that is consistent with this criterion”, but [9] suggested measuring goodness of fit by comparing the sum of weighted distances for the model of interest with the sum in which only the intercept parameter appears.

Let  $V^1(p)$  be the sum of weighted distance for the full  $p^{th}$  quantile regression model, and  $V^0(p)$  be the sum of weighted distance for the model that includes only a constant term. Therefore, using the one covariate model

$$\begin{aligned} V^1(p) &= \sum_{i=1}^n d_p(y_i, \hat{y}_i) \\ &= p \sum_{y_i \geq \beta_0^{(p)} + \beta_i^{(p)} x_i} \left| y_i - \beta_0^{(p)} - \beta_i^{(p)} x_i \right| \\ &\quad + 1-p \sum_{y_i < \beta_0^{(p)} + \beta_i^{(p)} x_i} \left| y_i - \beta_0^{(p)} - \beta_i^{(p)} x_i \right| \end{aligned} \quad (12)$$

$$\begin{aligned} V^0(p) &= \sum_{i=1}^n d_p(y_i, \hat{Q}^{(p)}) \\ &= \sum_{y_i \geq \bar{y}} p \left| y_i - \hat{Q}^{(p)} \right| + \sum_{y_i < \bar{y}} (1-p) \left| y_i - \hat{Q}^{(p)} \right|. \end{aligned}$$

For the model that only includes a constant term, the fitted constant is the sample  $p^{th}$  quantile  $\hat{Q}^{(p)}$  for the sample  $y_1, \dots, \dots, y_n$  the goodness of fit is then defined as

$$R(p) = 1 - \frac{V^1(p)}{V^0(p)}. \quad (13)$$

Since  $V^0(p)$  and  $V^1(p)$  are nonnegative,  $R(p)$  is at most 1. Also, because the sum of weighted distance is minimized for the full-fitted model,  $V^1(p)$  is never greater than  $V^0(p)$ , so  $R(p)$  is greater than or equal to zero. Thus,  $R(p)$  is within range of  $[0,1]$ , a larger  $R(p)$  indicates a better model fit. The  $R(p)$  defined above allows for comparison of a fitted model with any number of covariates beyond the intercept term to model in which only the intercept term is present. This is the restricted form of a goodness-of-fit introduced by Koenker and Machado (1999) for nested models.

### 3. Data Simulation

For the study data Weibull distribution was found to be left skewed as shown in Figure 2 with sharp and scale parameters 1.78292 and 122560 respectively. To simulate the data, quantile function of Weibull distribution function (probit function) which is simply the inverse of the CDF of Weibull distribution was derive by equating the CDF of  $F(y)$  and theoretically solve for  $y$ .

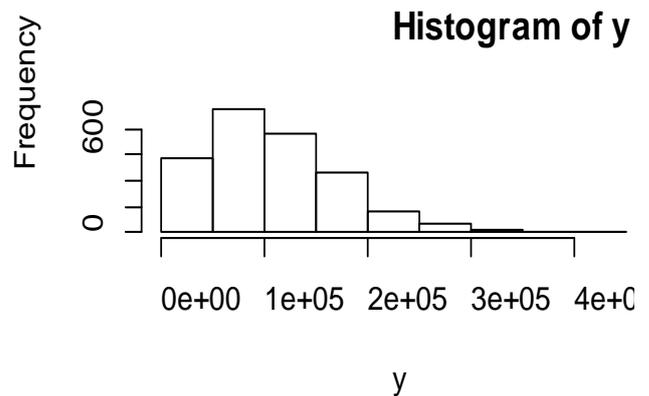


Figure 2. Graph of Original Data Simulated with Weibull Distribution

Probability Density Function of a Weibull distribution function is given as

$$PDF = p(y) = \begin{cases} \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k} & y > 0 \\ 0 & y < 0 \end{cases} \quad (14)$$

Let the Cumulative Density Function (CDF) of a Weibull distribution be denoted by  $F(y)$

$$\therefore F(y) = \int_0^y p(y) dy$$

$$\int_0^y \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} e^{\left(\frac{y}{\lambda}\right)^k} dy$$

$$\frac{1}{\lambda} \int_0^y k \left(\frac{y}{\lambda}\right)^{k-1} e^{\left(\frac{y}{\lambda}\right)^k} dy$$

Let  $u = \left(\frac{y}{\lambda}\right)^k$

$$\frac{du}{dy} = \lambda^{-k} k y^{k-1}$$

$$du = \lambda^{-k} k y^{k-1} dy$$

$$\int_0^y e^u du$$

$$\left[ e^u \right]_0^y$$

$$\left[ e^{\left(\frac{y}{\lambda}\right)^k} \right]_0^y$$

$$\therefore F(y) = 1 - e^{\left(\frac{y}{\lambda}\right)^k} \tag{15}$$

This is proceeded by deriving the probit function theoretically as:

$$1 - e^{\left(-\left(\frac{y}{\lambda}\right)^k\right)} = F(y)$$

$$e^{\left(-\left(\frac{y}{\lambda}\right)^k\right)} = 1 - F(y)$$

$$-\left(\frac{y}{\lambda}\right)^k = \log[1 - F(y)]$$

$$-\left(\frac{y}{\lambda}\right) = \log[1 - F(y)]^{1/k}$$

$$\frac{y}{\lambda} = -\left[\log(1 - F(y))\right]^{1/k}$$

$$y = -\left(\log(1 - F(y))\right)^{1/k} \lambda$$

$$Q(p) = F^{-1}(y)$$

$$Q(p) = -\left(\log(1 - F(y))\right)^{1/k} \lambda \tag{16}$$

Where

$Q(p)$  = probit function

$\lambda$  = scale parameter

$k$  = shape parameter

$F(y)$  = Cumulative Density Function (CDF)

Following the derivation of probity function, Monte Carlo simulation will then be applied on the derived function using both the sharp parameter (1.78292) and the scale parameter (122560) to generate sample size of 3000 for the response variable while the explanatory variable is simulated using the normal distribution with the sharp and scale parameters of 11.27641 and 322999 respectively.

### 3.1. Choice of Data Transformation

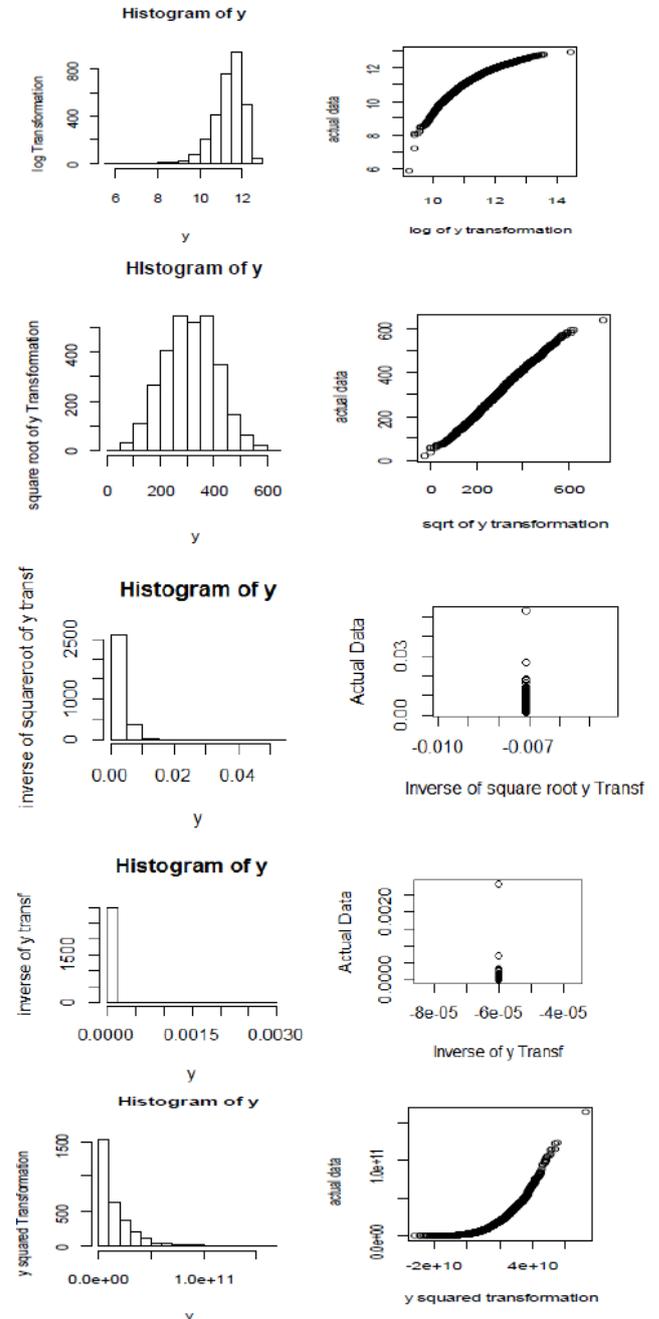


Figure 3. Histograms and the Normality plots

Choice of appropriate data transformation will be based on the Histograms of the transformed data, the normality plots, the values of Skewness and Kurtosis of the transformed values, it will also be based on comparing the values of the mean and the median of the transformed data and assessment of the expected error terms of the quantile regression model for the various transformed series.

Therefore, the following the transformation technique were considered and presented in equation 17 to 21 while their graphs are in Figure 3.

i. Log Transformation:

$$Q(p) = \log \left[ - \left( \log(1 - F(y))^{1/k} \right) \lambda \right] \quad (17)$$

ii. Square root of y transformation

$$Q(p) = \left( - \left( \log(1 - F(y))^{1/k} \right) \lambda \right)^{1/2} \quad (18)$$

iii. Inverse of Square root of y transformation:

$$Q(p) = \left( - \left( \log(1 - F(y))^{1/k} \right) \lambda \right)^{-1/2} \quad (19)$$

iv. Inverse of y transformation:

$$Q(p) = \left( - \left( \log(1 - F(y))^{1/k} \right) \lambda \right)^{-1} \quad (20)$$

v. y-squared transformation:

$$Q(p) = \left[ - \left( \log(1 - F(y))^{1/k} \right) \lambda \right]^2 \quad (21)$$

### 3.2. Results and Discussion

The graph of log of y transformation as well the graph of log of y estimate against the actual log of y show that the data was not transformed, the graph of log of y transformation is right skewed and the graph of log of y estimate against the actual y is partially curved at the center showing that log y estimate is not in full agreement with the actual log of y. Table 1 also shows that the values of the mean and the media are equal but the skewness is

less than 0 while the Kurtosis is greater than 3 which mean that the distribution is not normally distributed. The result of the table two show that despite the fact that the 50<sup>th</sup> quantile is approximately zero but the 25<sup>th</sup> quantile is less than the 50<sup>th</sup> quantile which means that the result may be spurious and cannot be relied upon.

The graph of Inverse of square root of y transformation as well the graph of Inverse of square root of y estimate against the actual Inverse of square root of y show that the data was over transformed. The graph of Inverse of square root of y is left skewed while that of the graph of Inverse of square root of y estimate against the actual Inverse of square root of y shows a vertical straight line meaning that the distribution of Inverse of square root of y estimate is not in agreement with the actual Inverse of square root of y. the result of table One shows that while the median is relatively equal to the mean, the Skewness is greater than zero while kurtosis is far greater than three. The result of Table 2 shows that while the expected error term of the 50<sup>th</sup> quantile is approximately zero that of the 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> quantiles are all less than the 50<sup>th</sup> quantile which suggests that the estimates may be spurious.

The graph of Inverse of y transformation as well the graph of Inverse of y estimate against the actual Inverse of y show that the data was not transformed. The graph of Inverse of y is left skewed with all the distributions grouped into one bar. Also, while the graph of Inverse of y estimate against the actual Inverse of y shows a vertical straight line meaning that the distribution of the inverse of y estimate is never in agreement with that of the actual inverse of y. the result of Table 1, shows that the mean is not equal to the median, the skewness is slightly greater than zero and the kurtosis is far greater than three. The result of Table 2 shows that both the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> quantile all have their expected error term as zero which means that the estimate made with any of the quantile estimate will be the same, which makes the quantile regression model insufficient.

Table 1. Test of Normality of the Transformed Data

Model Trans.	Quantile					
	mean	median	Skewness	kurtosis	sd	AIC
y	107189	97670.	0.4614	3.4456	61889.	63123.
Log y.	11.376	11.489	-0.4734	5.6279	0.7177	-3162.
$y^{1/2}$	312.64	312.52	0.0038	2.6331	97.189	16257.
$y^{-1/2}$	0.0036	0.0032	0.7129	167.711	0.0019	-36675.
$y^{-1}$	1.6969	1.0239	0.3559	2001.45	5.6729	-61720
$y^2$	1.5E10	9.5E9	1.0226	10.8067	1.69E10	143676.

Table 2. Expected Error Term of the Model

Model Transf	Quantile				
	.10	.25	.50	.75	.90
y	8423.	6850.	3824.	-3079.	-11518
Log y	0.2262	0.0298	-0.0691	-0.1176	-0.1307
$y^{1/2}$	6.0081	1.1496	-0.3788	-2.0249	-4.4646
$y^{-1/2}$	0.0110	0.0109	0.089	0.0117	0.0125
$y^{-1}$	0.00008	0.00008	0.00008	0.00009	0.00011
$y^2$	5.496E9	4.759E9	-1.12E11	-1.18E11	-8.24E9

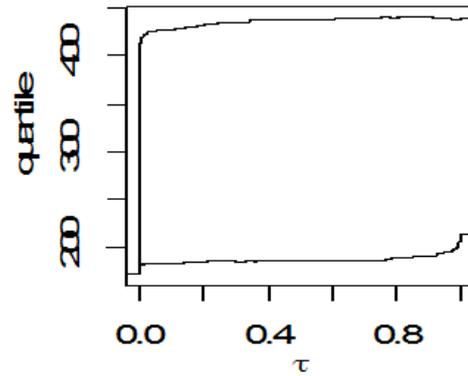
The graph of y-squared transformation as well the graph of y-squared estimate against the actual y-squared show that the data was not transformed. The graph of y-squared transformation is also left skewed, also while the graph of y-squared estimate against the actual y-square transformation shows a curved line meaning that the two distributions are not fully in agreement. The result of Table 1 shows that the mean is not equal to the median and the skewness and the kurtosis are far greater than the zero and 3 respectively meaning that the distribution is not normally distributed. The result of Table 2 shows that the 50<sup>th</sup> quantile as well all the other quantiles are all not equal to zero meaning that the distribution was not actually transformed.

The graph of square root of y transformation as well the graph of square root of y estimate against the actual square root of y show that the data has been transformed. The graph of square root of y transformation is symmetric as well mesokurtic. This can also be seen from the graph of square root of y estimate against the actual square root of y transformation which shows a straight-line curve meaning that the estimated data is in agreement with the actual data. From Table 1, it can be observed that the mean is equal to the median, and the skewness is approximately zero while the kurtosis is approximately three meaning that y-square root transformation is normally distributed. The result of the expected error term of the quantile regression estimate in Table 2 shows that the 50<sup>th</sup> quantile is approximately zero (-0.3788), while the 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> quantiles are distributed around the 50<sup>th</sup> quantile with 6.0081, 1.1496, -2.0249 and -4.4646 respectively meaning that the model has met the assumption that the expected error term must be zero hence can be said to be efficient.

Table 3. shows that the intercept has coefficient values of 25.5475, 25.5516, 25.5583, 25.5560, 25.5556 for the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup> quantile respectively. The explanatory variable ( $x_1$ ) has coefficients value of 0.00004 for all the five quantiles. The results of the p-values show that all the coefficient value have significant effect on the explanatory variable (y). The result of the individual standard errors shows a minimal error in the model hence the model of square root transformation in quantile regression model is efficient at the 50<sup>th</sup> quantile and can be relied on. With the confirmation of the efficiency, as well the reliability of the study model, the model can therefore be used to make some conclusive remark: The graph of Figure 4 shows the existence of wide discrepancy between the upper and lower income earners in the health institutions.

**Table 3. Coefficient and p-value of Square Root Transformed Model**

	.10	.25	.50	.75	.90
intercept	25.54	25.56	25.56	25.56	25.56
Std.Error	0.0022	0.0009	0.0004	0.0001	0.0001
t-value	11420	27371	70285	203497	290428
P(> t )	0.00	0.00	0.00	0.00	0.00
Coef Value	0.00	0.00	0.00	0.00	0.00
Std.Error	0.00	0.00	0.00	0.00	0.00
t-value	63.93	152.03	349.38	969.46	1710.0
P(> t )	0.00	0.00	0.00	0.00	0.00
Psudo - R	0.713	0.757	0.802	0.844	0.870



**Figure 4.** Graph of low and upper class

## 4. Conclusion

Having painstakingly navigated through the transformation of data using different kinds of power transformation which includes: logarithm of y transformation, transpose of y transformation, transpose of square root of y transformation, square root of y transformation, transpose of y-squared transformation and y-squared transformation, the result showed that square root of y transformation is the better transformation fit for Weibull distributed data on quantile regression model based on the plots of the histogram of the transformed data, plot of the estimated data against the actual data, expected error term and normality test using the mean, median, skewness and kurtosis.

## References

- [1] Arshad, I. A., Younas, U., Shaikh,A.W & Chandio,M.S (2016). Quantile Regression Analysis of Monthly Earnings in Pakistan; *Sindh Univ. Res. Jour. (Sci. Ser.)* Vol. 48 (4) 919-924 (2016).
- [2] Bartlett, M.S (1974). The use of Transformation, *Biometrika* 3, 39-52.
- [3] Chaudhuri, P. &Loh, W.-Y. (2002). Nonparametric estimation of conditional quantiles using quantile regression trees, *Bernoulli*, 8, 561-576.
- [4] Frost, J (2012) How to Identify the Distribution of Your Data using Minitab, <http://www.scribd.com/doc/84506538/Body-Fat-Data-for-Identifying-Distribution-in-Minitab>.
- [5] [Hao L. &Naiman, D.Q., (2007). Quantile Regression; *01-Hao.qxd. 3/13/2007.3.28*.
- [6] Iwueze, S.I, Nwogu, E.C., Ohakwe, J. & Ajaraogu, J.C. (2011) Uses of the Buys-Ballot Table in Time Series Analysis, *Applied Mathematics Journal.* (2) 633-645.
- [7] Koenker, R. (2005). *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press. (6)6.
- [8] Koenker,R & Bassett, G. (1978); *Regression Quantiles*, *Econometrica*, Vol. 46, No. 1, pp. 33-50.
- [9] Koenker, R. &D'Orey, V. (1987). Algorithm AS229: Computing regression quantiles, *Applied Statistics*, 36, 383-393.
- [10] Koenker, R. & Machado J.A (1999) Goodness of fit and related inference processes for quantile regression. *Journal of Econometrics*, 93, 327-344
- [11] Lee, B.-J. & Lee, M. J. (2006). Quantile regression analysis of wage determinants in the Korean labor market, *The Journal of the Korean Economy*, 7, 1-31.
- [12] Loh, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection, *Statistica Sinica*, 12, 361-386.
- [13] McMillen, D.P. (2013). *Quantile Regression for Spatial Data*, Springer Briefs in Regional Science.

- [14] Meinshausen, N. (2006); Quantile Regression Forests, *Journal of Machine Learning Research*, (7) 983-99.
- [15] Wen-ShuennDeng, Yi-Chen Lin &JinguoGong (2012) A smooth coefficient quantile regression approach to the social capital–economic growth nexus; *Economic Modelling journal* homepage: [www.elsevier.com/locate/ecmod](http://www.elsevier.com/locate/ecmod).
- [16] Young, T.M., Shaffer, L.B., Guess, F. M., Bensmail, H. &Leon, R.V (2008), A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard; *Forest Products Journal*, 58(4).



© The Author(s) 2019. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).