The International Journal Of Engineering And Science (IJES)
|| Volume || 6 || Issue || 3 || Pages || PP 01-08 || 2017 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Multiple Linear Regression Model with Two Parameter Doubly Truncated New Symmetric Distributed Errors

Bisrat Misganaw Geremew[1],K.Srinivasa Rao[2]
[1]*Department of Statistics, Andhra University,Vishakhapatnam 530 003, India*
[2]*Department of Statistics, Andhra University,Vishakhapatnam 530 003, India*

ABSTRACT
The most commonly used method to describe the relationship between response and independent variables is a linear model with Gaussian distributed errors. In practical components, the variables examined might not be mesokurtic and the populace values probably finitely limited. In this paper, we introduce a multiple linear regression models with two-parameter doubly truncated new symmetric distributed (DTNSD) errors for the first time. To estimate the model parameters we used the method of maximum likelihood (ML) and ordinary least squares (OLS). The model desires criteria such as Akaike information criteria (AIC) and Bayesian information criteria (BIC) for the models are used. A simulation study is performed to analysis the properties of the model parameters. A comparative study of doubly truncated new symmetric linear regression models on the Gaussian model showed that the proposed model gives good fit to the data sets for the error term follow DTNSD.
Keywords:Doubly truncated new symmetric distribution, Maximum likelihood method of estimate, Multiple linear regression model, Regression model, Simulation.

Date of Submission: 09 February 2017 Date of Accepted: 05 March 2017

I. INTRODUCTION

The estimation of parameters in a regression model has got importance in many fields of studies used for realizing practical relationships between variables. The regression model theories and applications are studied by many authors. This method is basically grounded on statistical model wherein the error terms are assumed to be independent and identically distributed Gaussian random variables with zero mean vector and positive covariance matrix (Srivastava, MS, 2002). Recently, there have been an enormous kind of studies on the influences of non-Gaussian in several linear regression analyses. Bell-shaped or roughly bell-shaped distributions are encountered with a big form of applications and, through the Central Limit Theorem, provide the underpinning for the characteristics of sampling distributions upon which statistical inference is primarily based. Regardless of this application, the range of a normally distributed variate $-\infty$ to ∞. But in data sets where the varaite is having finite range may not fit well to regression models with Gaussian error. This problem has stimulated to consider truncated distributions.

The parameter estimates of a singly truncated normal distribution formulae have developed by Pearson and Lee (1908). They used the method of moments. Different papers on this subject include encompassing Cohen (1950), in which cases of doubly truncated and censored samples have been considered. The moment generating function of the lower truncated multivariate normal distribution had in Tallis (1961) and, in principle; it could be used to compute all the product moments for the lower truncated multivariate normal. Tallis (1961) provides explicit expressions of some lower order moments for the $n=2$ and $n=3$cases. Cohen (1949, 1950a, 1950b) has studied the problem of estimating the mean and variance of normal populations from singly truncated samples. Moreover Cohen (1961) found the tables for maximum likelihood estimators of singly truncated and singly censored samples.

J.J. Sharples and J.C.V.Pezzey (2007) stated results of multivariate normal distributions which consider truncation by means of a hyperplane. They presented results by calculating the expected values of the features used in environmental modeling, when the underlying distribution is taken to be multivariate normal. They illustrated the concept of truncated multivariate normal distributions may be employed within the environmental sciences through an example of the economics of climate change control. Karlsson M. *et al*. (2009) derived an estimator from a moment condition. The studied estimators are for semi-parametric linear regression models with left truncated and right censored dependent variables.

The truncated mean and the truncated variance in multivariate normal distribution of arbitrary rectangular double truncation is derived by Manjunath B.G. and Stefan Wilhelm (2012). For most comprehensive account of the theory and applications of the truncated distributions, on can refer the books by N. Balakrishnan and A. Clifford Cohen (1990), A. Clifford Cohen (1991), and Samuel Kotz*et. al*. (1993).

For some instances, even though the shape of the sample frequency curve is symmetric and bell shaped the normal approximation may badly fit the distribution. This may be due to the peakedness of the data which might not be mesokurtc. To resolve this trouble new symmetrical distribution was derived by Srinivasa Rao, et al. (1997) and studied their distributional properties by A. Asrat and Srinivasa Rao (2013). Utilized it for linear model with new symmetric distribution the range of the variate infinite having the values between $-\infty$ to ∞. For instance of double truncation can be determined soundwave frequencies. Frequency of a soundwave can be any non-negative number. However, we cannot listen all sounds. Only those sounds generated from soundwaves with frequencies greater than 20 Hertz and less than 20, 000 Hertz are audible to human ears. Right here, the truncation point at the left is 20 Hertz and that on the right is 20, 000 Hertz.

However, for constrained response variable the usage of such interval my fit the model badly. Hence in this paper we study the multiple regression model with truncated new symmetric distributed errors. Very little work has been reported in literature regarding multiple regression model follows a truncated distributions.

The linear regression model is of the form:

$$\begin{aligned} y_i &= x_i'\beta + u_i \\ &= \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \end{aligned} \qquad (i = 1,2,\ldots,n), a \le y_i \le b \qquad (1)$$

where y_i is an observed response variable of the regression, x_{ik} are observed independent variables, $\beta_1, \beta_2, \ldots, \beta_k$ are unknown regression coefficients to be estimated, and u_i are independently and identically distributed error terms. It is assumed that the error term follows a two parameter doubly truncated new symmetric distribution.

The rest of the paper is organized as follows: In Section 2, the distributional properties of a two- parameter doubly truncated new symmetric distribution are derived. The maximum likelihood estimation of the model parameters is studied in Section 3. In Section 4, simulation studies are performed and the results are discussed. Least square estimation of the model parameters are studied in Section 5. In Section 6, comparison of maximum likelihood estimators and OLS estimators is given. In Section 7, comparison of the suggested model with that of New Symmetric distributed errors and Gaussian model errors are presented. Section 8, the conclusions are given.

II. PROPORTIES OF DOUBLY TRUNCATED NEW SYMMETRIC DISTRIBUTION

The doubly truncated new symmetric distribution was defined by equation (2) specifying its probability density function. There are many properties of doubly truncated new symmetric distribution. Some of the most important properties are:

A random variable Y follows a MDTNS distribution if its probability density function is

$$f(y) = \begin{cases} 0, -\infty < y < a \\ \dfrac{\left[2 + ((y-\mu)/\sigma)^2\right] e^{-(1/2)((y-\mu)/\sigma)^2}}{3\sigma\sqrt{2\pi}\left[F(b) - F(a)\right]}, a \le y \le b \\ 0, b < y < \infty \end{cases} \qquad (2)$$

where $F(b) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{b} e^{-1/2\left(\frac{y-\mu}{\sigma}\right)^2} dy - \left(\dfrac{b-\mu}{3\sigma}\right) e^{-1/2\left(\frac{b-\mu}{\sigma}\right)^2} = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \dfrac{1}{3}\left(\dfrac{b-\mu}{\sigma}\right)\phi\left(\dfrac{b-\mu}{\sigma}\right)$ and

$$F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-1/2\left(\frac{y-\mu}{\sigma}\right)^2} dy - \left(\frac{a-\mu}{3\sigma}\right) e^{-1/2\left(\frac{a-\mu}{\sigma}\right)^2} = \Phi\left(\frac{a-\mu}{\sigma}\right) - \frac{1}{3}\left(\frac{a-\mu}{\sigma}\right)\phi\left(\frac{a-\mu}{\sigma}\right)$$

the parametersa $\le \mu \le$b and $\sigma > 0$ are location and scale parameters, respectively.

Figure 1 shows the frequency curves of the variate under study

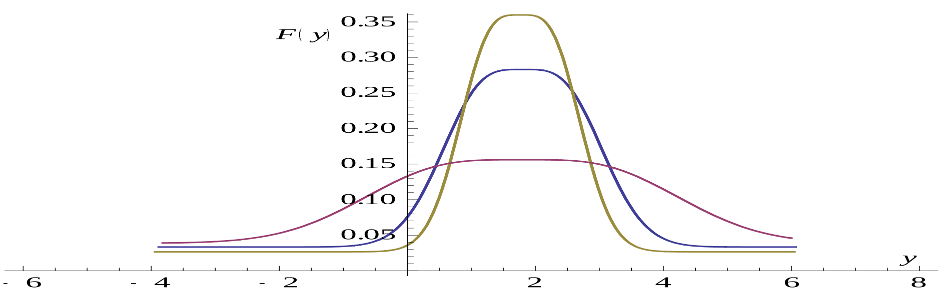

Figure 1. Frequency curves for different values of parameters σ: μ and σ as $(2,0.75) \rightarrow green, (2,1) \rightarrow blue, and (2,2) \rightarrow red$, for $a = -4 \& b = 6.$

The distribution function of the random variable Y is:

$$F_Y(y) = \frac{1}{\sigma\sqrt{2\pi}[F(b)-F(a)]}\int_a^y e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt - \frac{(y-\mu)e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}[F(b)-F(a)]} \tag{3}$$

The Mean of the variable is:

$$E(y) = \mu + \sigma\lambda(\alpha)$$

where
$$\lambda(\alpha) = \left\{\frac{\left(\frac{4}{3}+\frac{1}{3}\left(\frac{a-\mu}{\sigma}\right)^2\right)\phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{4}{3}+\frac{1}{3}\left(\frac{b-\mu}{\sigma}\right)^2\right)\phi\left(\frac{b-\mu}{\sigma}\right)}{F(b)-F(a)}\right\} \tag{4}$$

If $\mu > 0$ and the truncation is both sides, i.e., $\lambda(\alpha) > 0$ the mean of the truncated variable is greater than the original mean. $\lambda(\alpha)$ is the mean of the truncated normal distribution.

The characteristic function $\phi_Y(t)$ of a random variable Y is:

$$\phi_y(t) = E\left(e^{ity}\right) = \int_a^b e^{ity} \frac{\left[2+((y-\mu)/\sigma)^2\right]e^{-(1/2)((y-\mu)/\sigma)^2}}{3\sigma\sqrt{2\pi}\left[F(b)-F(a)\right]} dy$$

$$= \frac{e^{\mu it-\frac{1}{2}\sigma^2t^2}}{F(b)-F(a)}\left\{\begin{array}{l}\left(1+\frac{(\sigma it)^2}{3}\right)\left[\Phi\left(\frac{b-\mu}{\sigma}-\sigma it\right)-\Phi\left(\frac{a-\mu}{\sigma}-\sigma it\right)\right]+ \\ \frac{1}{3}\left[\left(\frac{a-\mu}{\sigma}+\sigma it\right)\phi\left(\frac{a-\mu}{\sigma}-\sigma it\right)-\frac{1}{3}\left(\frac{b-\mu}{\sigma}+\sigma it\right)\phi\left(\frac{b-\mu}{\sigma}-\sigma it\right)\right]\end{array}\right\} \tag{5}$$

The moment generating function is:

$$M(t) = \frac{e^{\mu t+\frac{1}{2}\sigma^2t^2}}{F(b)-F(a)}\left\{\begin{array}{l}\left(1+\frac{(\sigma t)^2}{3}\right)\left[\Phi\left(\frac{b-\mu}{\sigma}-\sigma t\right)-\Phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)\right]+ \\ \frac{1}{3}\left[\left(\frac{a-\mu}{\sigma}+\sigma t\right)\phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)-\frac{1}{3}\left(\frac{b-\mu}{\sigma}+\sigma t\right)\phi\left(\frac{b-\mu}{\sigma}-\sigma t\right)\right]\end{array}\right\} \tag{6}$$

The cumulant generating function is:

$$g(t;\mu,\sigma^2) = \ln M(t;\mu,\sigma^2) = \mu t + \frac{1}{2}\sigma^2t^2 + \ln A - \ln\left[F(b)-F(a)\right] \tag{7}$$

The first two cumulants are

$$\mu + \sigma\left\{\frac{\left(\frac{4}{3}+\frac{1}{3}\left(\frac{a-\mu}{\sigma}\right)^2\right)\phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{4}{3}+\frac{1}{3}\left(\frac{b-\mu}{\sigma}\right)^2\right)\phi\left(\frac{b-\mu}{\sigma}\right)}{F(b)-F(a)}\right\}$$ and

$$\frac{5\sigma^2}{3} - \sigma^2\left\{\frac{\left(\frac{4}{3}+\frac{1}{3}\left(\frac{a-\mu}{\sigma}\right)^2\right)\phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{4}{3}+\frac{1}{3}\left(\frac{b-\mu}{\sigma}\right)^2\right)\phi\left(\frac{b-\mu}{\sigma}\right)}{F(b)-F(a)}\right\}^2 + \tag{8}$$

$$\frac{\sigma^2}{\left[F(b)-F(a)\right]}\left\{\left[\frac{10}{9}\left(\frac{a-\mu}{\sigma}\right)+\frac{1}{3}\left(\frac{a-\mu}{\sigma}\right)^3\right]\phi\left(\frac{a-\mu}{\sigma}\right) - \left[\frac{10}{9}\left(\frac{b-\mu}{\sigma}\right)+\frac{1}{3}\left(\frac{b-\mu}{\sigma}\right)^3\right]\phi\left(\frac{b-\mu}{\sigma}\right)\right\}$$

All higher-order cumulants are equal to zero.

III. MAXIMUM LIKELIHOOD ESTIMATION OF THE MODEL PARAMETERS

In this section, we consider the regression model $(Y, X\beta, \sigma^2 I)$ with $u \sim DTNS(0, \sigma^2 I)$. The unknown parameters of the model $(Y, X\beta, \sigma^2 I)$ are coefficient vector β and the error variance σ^2. The MLEs of β and σ^2 are computed as: the parameter values that maximize the likelihood function:

$$L(y;\beta,a,b,\sigma^2) = \prod_{i=1}^{n} \frac{\left[2+\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}\left[F(b)-F(a)\right]} I_{[a,b]}(y_i)$$

$$= \left[3\sigma\sqrt{2\pi}\left[F(b)-F(a)\right]\right]^{-n} \prod_{i=1}^{n}\left[2+\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] e^{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2} \tag{9}$$

where $F(b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{b} e^{-1/2\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2} dy_i - \left(\frac{b - x_i'\beta}{3\sigma}\right) e^{-1/2\left(\frac{b - x_i'\beta}{\sigma}\right)^2} = \Phi\left(\frac{b - x_i'\beta}{\sigma}\right) - \frac{1}{3}\left(\frac{b - x_i'\beta}{\sigma}\right)\phi\left(\frac{b - x_i'\beta}{\sigma}\right)$ and

$$F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-1/2\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2} dy_i - \left(\frac{a - x_i'\beta}{3\sigma}\right) e^{-1/2\left(\frac{a - x_i'\beta}{\sigma}\right)^2} = \Phi\left(\frac{a - x_i'\beta}{\sigma}\right) - \frac{1}{3}\left(\frac{a - x_i'\beta}{\sigma}\right)\phi\left(\frac{a - x_i'\beta}{\sigma}\right)$$

The log-likelihood function is

$$l = \ln L(y;\beta,a,b,\sigma^2) = -\frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln\left[\frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2} dy_i + \left(\frac{a - x_i'\beta}{3\sigma}\right) e^{-1/2\left(\frac{a - x_i'\beta}{\sigma}\right)^2} - \left(\frac{b - x_i'\beta}{3\sigma}\right) e^{-1/2\left(\frac{b - x_i'\beta}{\sigma}\right)^2}\right] +$$

$$\sum_{i=1}^{n}\ln\left[2+\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right] - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2 \tag{10}$$

The maximum likelihood estimators $\hat{\beta}_{MLE} = (\hat{\beta}, \hat{\sigma}^2)$ are that which maximizes the log-likelihood function. Taking the partial derivatives of the log of the likelihood with respect to the $(k+1)x1$ vector β are nxn matrix of $\sigma^2 I$ and placing the result equal to zero will give (11). The maximum likelihood estimators are the solutions of the equations

$$\frac{\partial l}{\partial \beta} = 0 \text{ and } \frac{\partial l}{\partial \sigma^2} = 0 \tag{11}$$

Since there are no closed form solutions to the likelihood equations, numerical methods such as Fisher Scoring or Newton-Raphson iterative method can be used to obtain the MLEs. The standard procedure for implementing this solution was to use Newton- Raphson iterative method we have,

$$\theta^{(n+1)} = \theta^{(n)} - \left(H^{(n)}\right)^{-1} S^{(n)}, \theta = (\beta, \sigma^2) \tag{12}$$

Here we begin with some starting value, $\theta^{(0)}$ say, and improve it by finding some better approximation $\theta^{(1)}$ to the required root. This procedure can be iterated to go from a current approximation $\theta^{(n)}$ to a better approximation $\theta^{(n+1)}$.

IV. SIMULATION AND RESULTS

The proposed model was evaluated through Monte Carlo experiments in which the data is generated from model (1) using Wolfram Mathematica 10.4. To facilitate exposition of the method of estimation, a several data set with two explanatory variables and one dependent variable are simulated from a model with prespecified parameters for various sample sizes $n = 100, 1000, 3000, 5000,$ and 10000. The dependent variable Y is simulated using doubly truncated new symmetric distribution with mean 2 and variance 1 using the following procedures:

Step 1: Generates the uniform random numbers $d_i \sim U(0,1)$, $i = 1,2,\ldots,n$,

Step 2: Given $\mu = 2, \sigma = 1, a = -1,$ and $b = 5$; solve

$$\frac{1}{\sigma\sqrt{2\pi}[F(b)-F(a)]}\int_a^y e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}dt-\frac{(y-\mu)e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}[F(b)-F(a)]}=d_i \cdot \quad (13)$$

The solution for a random variable y_i , $i = 1,2,..., n$, will have the standard DTNS distribution.

Step 3: We then generate the 2 predictors X_1 and X_2 variables respectively using the simulation protocol and use as explanatory variablesfor the regression model.

$$X_1 = Random\mathrm{Re}\,al[NormalDistribution[1,1]];$$
$$X_2 = Random\mathrm{Re}\,al[LogNormalDistribution[0,1]] \quad (14)$$

MDTNS error regressions were applied to the simulated datasets and the estimated parameters were compared to the true parameters. This process was repeated for sample sizes of $n = 100,1000,3000,5000,$ and 10000 . In the first Monte Carlo experiment we generate the datasets X_1 and X_2 from (14) and Y from the model defined in (1) and 914) . Keeping generality, we took the values of the model parameters to

$$\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}=\begin{pmatrix}2\\ 3\\ 4\end{pmatrix}\text{and}\begin{pmatrix}\mu\\ \sigma\end{pmatrix}=\begin{pmatrix}2\\ 1\end{pmatrix} \quad (15)$$

The error terms generated from a doubly truncated new symmetric distribution, that is $u_i \sim MDTNS(\mu,\sigma)$.

Table 1. Summary of Simulations for the Maximum Likelihood (ML) estimation of the regression model

Sample Size	Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr>ChiSq
				Lower	Upper		
100	β_0	1.6330	0.1612	1.3170	1.9491	102.56	<.0001
	β_1	3.0385	0.0908	2.8606	3.2164	1120.89	<.0001
	β_2	4.1333	0.0460	4.0432	4.2235	8079.28	<.0001
	σ	0.8938	0.0632	0.7781	1.0267		
1000	β_0	2.0082	0.0477	1.9147	2.1017	1772.47	<.0001
	β_1	2.9967	0.0313	2.9354	3.0579	9187.75	<.0001
	β_2	3.9997	0.0096	3.9808	4.0185	172753	<.0001
	σ	0.9776	0.0219	0.9357	1.0214		
3000	β_0	1.9931	0.0299	1.9346	2.0516	4458.15	<.0001
	β_1	3.0231	0.0179	2.9880	3.0583	28448.2	<.0001
	β_2	3.9942	0.0098	3.9750	4.0134	166318	<.0001
	σ	0.9922	0.0128	0.9674	1.0176		
5000	β_0	1.9784	0.0228	1.93337	2.0232	7506.31	<.0001
	β_1	3.0094	0.0140	2.9820	3.0367	46519.3	<.0001
	β_2	4.0069	0.0069	3.9934	4.0205	335462	<.0001
	σ	0.9937	0.0099	0.9744	1.0134		
10000	β_0	1.9792	0.0164	1.9471	2.0112	14650.0	<.0001
	β_1	3.0138	0.0101	2.9940	3.0335	89406.3	<.0001
	β_2	4.0022	0.0048	3.9928	4.0116	692365	<.0001
	σ	1.0080	0.0071	0.9941	1.0220		

Table 1 results suggest that as the size of the sample increases, the estimates of the parameters become more precise. Increasing the sample sizes from 100 to 1000, and then to 10, 000 observations, the estimators all move closer to the true parameter values, and the dispersion of the estimator distributions notably decreases. The fitted linear regression model with MDTNS error terms to the simulated data, based on $n = 10,000$ is,

$$\hat{Y} = 1.9792 + 3.0138X_1 + 4.0022X_2 \quad (16)$$

And their estimated standard errors are:

$$s.e.(\hat{\beta}_0) = 0.0164, s.e.(\hat{\beta}_1) = 0.0101, s.e.(\hat{\beta}_2) = 0.0048 \text{ and } s.e.(\hat{\sigma}) = 0.0071 \quad (17)$$

V. LEAST SQUARES ESTIMATION OF THE MODEL PARAMETERS

The commonly used method for estimating the regression coefficients in a standard linear regression model is the method of ordinary least squares (OLS). The OLS estimator, $\hat{\beta}_{OLS}$, minimizes the sum of squared residuals (18). It can be defined by (19), where β is a $(k+1)x1$ parameter vector. The equation is (19) follows by

i) differentiating equation (13) with respect to β, where $\beta = (\beta_0, \beta_1, \beta_2)'$,
ii) setting the resultant matrix equation equal to zero, and
iii) replacing β by $\hat{\beta}_{OLS}$ and rearranging for $\hat{\beta}_{OLS}$.

Suppose $y_1, y_2, \ldots y_n$ are observations of a random variable y. The estimates of $\beta_0, \beta_1, \ldots, \beta_k$ are the values which reduce

$$SS = \sum_{ii=a1}^{b} (y_i - x_i'\beta) = (Y - X\beta)'(Y - X\beta) \tag{18}$$

The resulting OLS estimator of β is:

$$\hat{\beta} = (X'X)' X'Y \tag{19}$$

Provided that the inversed $(X'X)^{-1}$ exists.

The predicted values of the dependent variable are given by $\hat{Y} = \hat{\beta}X$ and the residuals are calculated using $e = Y - \hat{Y}$. (20)

Properties of the Estimates

Some of the properties of Ordinary Least squares (OLS) estimates are presented as: assuming that $E(u) = 0$, $Var(u) = \sigma^2 I_n$, and $Var(\beta) = \sigma^2 (X'X)^{-1}$ (21)

Table 2 presents the properties of OLS estimations using the data simulated in Section4.

Table 2. OLS estimation output for the simulation data

Equation	DF model	DF error	SSE	MSE	Root MSE	R-Square	Adjusted R-Square
y	3	9997	10149.6	1.0153	1.0076	0.9873	0.9873

Parameter	Estimate	Approximate standard error	t-value	Approximate $Pr > \|t\|$
β_0	1.978464	0.0163	121.03	<.0001
β_1	3.013995	0.0101	298.97	<.0001
β_2	4.002315	0.00481	832.41	<.0001

Table 2 revealed that the Ordinary Least Squares (OLS) estimates vary significantly from the Maximum likelihood (ML) estimates and the ML estimators are closer to the true values of the parameters compared to the OLS estimators.

VI. COMPARSION OF MAXIMUM LIKELIHOOD AND OLS ESTIMATORS

For fitting the multiple linear regression model with two parameter doubly truncated new symmetric error terms comparison of MLEs and LSEs were done. For every estimation methods bias and mean square error (MSE) are computed for $n = 10{,}000$ where

$$MLE(\hat{\beta}) = Var(\hat{\beta}) + [bias(\hat{\beta})]^2$$

The computational result is presented in Table 3.

Table 3. Comparison of the LSEs and MLEs of Doubly Truncated New Symmetric Regression Model

		OLS	ML
$\hat{\beta}_0$	Bias	-0.021536	-0.0208
	MSE	0.000729489	0.0007016
$\hat{\beta}_1$	Bias	0.013995	0.0138
	MSE	0.00029787	0.00029245
$\hat{\beta}_2$	Bias	0.002315	0.0022
	MSE	2.84953E-05	2.788E-05

From Table 3, it is observed that results for comparison criteria approving that deviations from normality causes LSEs to be poor estimators. As the results reported in Table 3 also shown that MLEs have both smaller bias and Mean Square Error (MSE) than the Least Square (LS) estimators. The results confirmed that ML estimation method shows better overall performance than OLS.

VII. COMPARSION OF THE MDTNS-LM WITH THE N-LM

In this paper, the performance of linear regression model with doubly truncated new symmetric distributed error terms with that of normal error terms were examined using simulated data. To choose the best model the Akaike's information criteria (AIC) and the Bayesian information criteria (BIC) with model diagnostics root mean square error (RMSE) were computed. The output of simulation studies using various sample sizes presented in Table 4.

Table 4. Summary for information criteria and model diagnostics for normal and doubly truncated new symmetric error model.

Model	Sample Size (n)	Information Criteria		Model Diagnostics
		AIC	BIC	
MDTNSD	100	-16.4545	-14.2708	0.90752
	1000	-39.24445	-37.2264	0.97910
	3000	-41.0639	-39.0579	0.99268
	5000	-57.0270	-55.0234	0.99402
	10000	-154.5126	-156.5144	0.99576
Normal	100	2.5315	4.7151	0.99789
	1000	-6.0051	-3.9871	0.99551
	3000	-3.7005	-1.6945	0.99888
	5000	-26.9597	-24.9561	0.99701
	10000	-75.2408	-73.2390	0.99610

The model with the smallest AIC or BIC amongst all competing models is deemed to be good model where it can be seen that the MDTNS distribution provides the better fit to the data. That is, both the information criteria techniques (AIC and BIC) and the model diagnostics (RMSE) indicate that linear model with doubly truncated new symmetrically distributed error terms consistently performed better across all the sample sizes of the simulation. This can also be consistently noticed from Figure 2 through Figure 4.

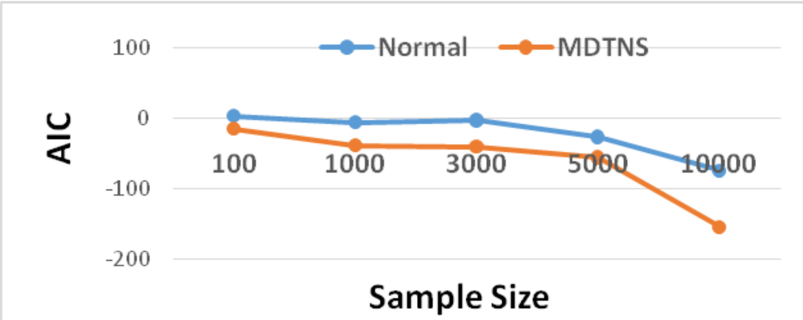

Figure 2. Comparison of Multiple Doubly Truncated New Symmetric (MDTNS) linear model versus Normal Linear Model using AIC.

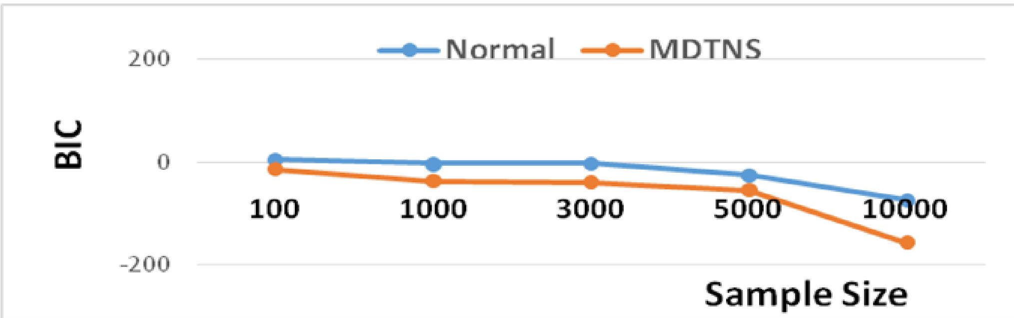

Figure 3. Comparison of Multiple Doubly Truncated New Symmetric (MDTNS) linear model versus Normal Linear Model using BIC.

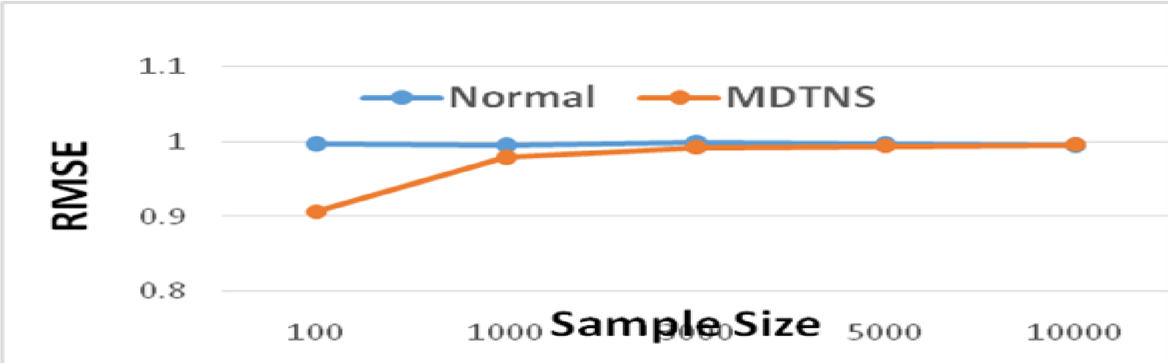

Figure 4. Comparison of Multiple Doubly Truncated New Symmetric (MDTNS) linear model versus Normal Linear Model using RMSE.

Clearly, from Fig. 2 through 4, we note that the MDTNS is better than the normal.

VIII. SUMMARY AND CONCLUSIONS

In this paper, we introduced the multiple regression model with doubly truncated new symmetric distributed errors. The doubly truncated new symmetric distribution serves as an alternative to the new symmetric distribution. The maximum likelihood estimators of the model parameters are derived. Via simulation studies, the properties of these estimators are studied. OLS estimation is carried out in parallel and the results are compared. The simulated results reveal that the ML estimators are more efficient than the OLS. A comparative study of the developed regression model, doubly truncated new symmetric linear regression model, with the Gaussian model showed that this model gives good fit to some data sets. The properties of the maximum likelihood estimators are studied. This regression model is much more useful for analyzing data sets raising that reliability, lifetime data analysis, engineering, survival analysis, and a wide range of other practical problems. This paper can be further extended to the case of non-linear regression with doubly truncated new symmetric distributed errors which will be taken elsewhere.

REFERENCES

[1] AsratAtsedwoyen and K. Srinivasa Rao, Estimation of a linear model with two parameter symmetric platykurtic distributed errors, *Journal of uncertainty analysis and applications*, 1, 2013, 1 -13.

[2] Cohen, A.C.,J, On estimating the mean and variance standard deviation of truncated normal distributions, *Journal of the American Statistical Association, 44*(248), 1949, 518-525.

[3] Cohen, A.C., Estimating mean and variance of normal populations from singly truncated and doubly truncated samples, *the annals of mathematical statistics, 21*(4), 1950, 557-569.

[4] Cohen, A.C., J., Tables for maximum likelihood estimates: Singly truncated and singly censored samples, *Technometrics*. *3*(4), 1961, 535-541.

[5] Cohen, A.C., truncated and censored samples: Theory and applications,*Marcel Dekker, New York*. 1991.

[6] Manjunath B.G. and Stefan Wilhelm, Moments calculation for the doubly truncated multivariate normal density,*http://dx.doi.org/10.2139/ssrn.1472153*, 2012.

[7] Maria Karlsson and Thomas laitila, A semi parametric regression estimator under left truncated and right censoring. *Statistics and probability letters, Elsevier, 78*(16), 2009, 2567-2571.

[8] Pearson, K., Lee, A.,On the generalized probable error in multiple normal correlation, *Biometrica,6*(1), 1908, 59-68.

[9] Sharples. J.J. and .Pezzey, J.C.V., Expectations of linear functions with respect to truncated multinomial distributions with applications for uncertainty analysis in environmental modelling,*Environmental modelling and software, 22*, 2007, 915-923.

[10] Srinivasa Rao, K., Vijay Kumar, C.V.S.R., and Lakshmi Narayana J., On a new symmetric distribution,*Journal of Indian society of agricultural statistics,50*(1), 1997, 95-102.

[11] Srivastava, M.S., Methods of multivariate statistics,*Wiley, New York*. 2002.

[12] Tallis, G.M., The moment generating function of the truncated multi-normal distribution, *Journal of the royal society, 23*(1), 1961, 223-229.

[13] Venkatesh, A. and Mohankumar, S., A mathematical model for the secretion of vasopressin using fuzzy truncated normal distribution,*International journal of pure and applied mathematics, 104*(1), 2015, 69-77.