Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.1, 2014

Linear regression model with generalized new symmetric error distribution

A. Asrat Atsedeweyn (Corresponding author)
Department of Statistics, Andhra University
Visakhapatnam 530003, India
Tel: +918143781411 Email: asrat07@gmail.com

K. Srinivasa Rao
Department of Statistics, Andhra University
Visakhapatnam 530003, India
Tel: +919949659959 Email: keraoau@yahoo.co.in

Abstract
Linear models play a dominant role in analyzing several data sets arising at places like agricultural experiments, space experiments, biological experiments, financial modeling and a wide range other practical problems. One of the major strings in the development of the regression model is the assumption of the error. It is often assumed that the random error of the linear regression model is normally distributed. In numerous situations, however, it is nearly impossible to find a data set that satisfies the normality assumption due to various reasons, such as multivariate skewed and/or heavy-tailed distributions. This problem has been addressed by specifying a different parametric distribution family for the error terms. In this paper, a linear regression model with generalized new symmetric errors is developed and analyzed. The Maximum Likelihood (ML) estimators of the model parameters are derived and their properties with respect to the generalized new symmetric distributed errors are discussed. Simulations were carried out to study the performance of the proposed model with that of Gaussian errors and found that the proposed model perform well when the variables are platykurtic. Some applications of the developed model are also pointed.

Key Words: Generalized new symmetric distribution, Regression model, Simulation

1. Introduction

The linear model is one of the most popular and simplest models in statistics. It has been received significant applications in almost every area of science, engineering and medicine in general and in Statistics and Econometrics in particular. Most of the inferential procedures, however, are based on the basic assumption of the linear regression model that the error terms have normal distribution. For example, Rao (1973), Seber (1977), Drapper and Smith (1981), Atkinson (1985), McCullagh,P., Nilder, J.A.(1989), Montgomery et al. (2001), Grob (2003), Seber and Lee (2003), Sengupta, D. and Jammalamadaka, S.Rao (2003), Weisberg (2005) and Yan and Su (2009) among others, are excellent references covering various aspects of classical linear models.

There have been a wide range of studies on the influences of non-normality on several linear regression analyses during about the last four decades. Zeckhauser and Thompson (1970) on linear regression model with power distributions; Zellner, A. (1976) and Sutradhar, B.C. and Ali, M.M. (1986) on regression model with a multivariate t error variable; Tiku et al. (1999) on linear regression model with symmetric innovations; Tiku et al. (2000) on first-order autoregressive model with symmetric innoviations; Tiku et al. (2001) for the simple linear model with t distribution innovations; Das Gupta, S. and Jammalamadaka, S.R. (2003) on Linear Models; Liu and Bozdogan (2004) for power exponential (PE) multiple regression; Wong and Bian (2005) for the multiple regression coefficients in linear model with underlying Student t distribution; Wong and Bian (2005) extension of the results given in Tiku et al. (1999) where the underlying distribution is a generalized logistic distribution; Liu and Bozdogan (2008) for the multivariate regression models with power exponential random errors; Soffritti, G. and Galimberti, G. (2010) on multivariate linear regression model under the assumption that the error terms follow a finite mixture of normal distributions; Jafari and Hashemi (2011) for simple linear regression with the error term of Skew-Normal distribution; Hettmansperger and McKean (2011) has been developed a complete rank-based inference for linear models based on rank-based estimation analogous to the way that traditional analysis is based on least squares (LS) estimation; S. Jahan and A. Khan (2012) on g-and-k distribution as the underlying assumption for the distribution of error in simple linear regression model; and Guorui Bian, et al. (2013) extension of the results given in Bian and Tiku (1997), Tiku et al. (1999,2000,2001), Wong and Bian (2005), Islam and Tiku (2005, 2010) on the multiple regression model with underlying distribution assumed to be symmetric and Student t are excellent references covering various aspects of linear models with non normal error terms.

Linear model with generalized new symmetric error

For simplicity and clarity here we considered a basic regression model where there is only one independent variable and assume the regression function is linear of the form

$$y_i = \beta_0 + \beta_1 x_i + u_i \,, u \sim GSP(0, \sigma^2 I) \qquad (1)$$

for the conditions in which the distribution of the error terms are assumed to be independent and identically distributed generalized new symmetric random errors with mean *0* and constant variance σ^2.

In this paper, we extend the work of A.Asrat Atsedeweyn and K.Srinivasa Rao (2013) to the case, where the underlying distribution is a generalized new symmetric distribution. The generalized new symmetric distribution family represents very wide platykurtic distributions ranging from mesokurtic to highly flatted platykurtic distributions. It is clear that the generalized new symmetric distribution has been received great interest in many applications such as signal processing, agricultural and biological experiments and financial modeling; see, for example, M.Seshashayee, et.al (2011). But so far, no study has been attempted to investigate a linear regression model with the generalized new symmetric error distribution. Hence, in this paper we considered a linear regression model in which the underlying distribution is a generalized new symmetric.

The rest of the paper is organized as follows. Section 2, briefly introduces the generalized new symmetric distribution and discusses some of its properties. In Section 3, we develop the ML estimators for the generalized new symmetric data, study the asymptotic properties of the proposed estimators and then conduct simulation to the study. We also illustrate the results of Monte Carlo experiments in which datasets were simulated from the proposed model and from models with other distribution for the error terms. In addition, the OLS estimators are also presented. Comparisons of the ML with OLS estimation techniques; and the proposed model with normal regression model are made in Section 4. Finally we state some results in Section 5.

2. Generalized New Symmetric Distribution

On many occasions, even though the shape of the sample frequency curve is symmetric and bell shaped, the normal approximation may badly fit the distribution, indicating some kind of kurtosis. To model such situations a more appropriate distribution of symmetric and platykurtic nature is needed. In recent years, techniques for extending the family of normal distributions have been proposed. A three-parameter distribution, named as Generalized new symmetric (GNS) distribution has been introduced by the authors Srinivasa Rao, et al., (1997) by inserting a new parameter γ to the normal distribution, such that a kurtosis effect will depend on it. It is a family of unimodel, symmetric and a bell-shaped continuous probability distribution, defined on the entire real line, which suits for the situations where the variables are platykurtic.

We say that the random variable y has the generalized new symmetric distribution if y has the density function,

$$f_r(y/\mu,\sigma,\gamma)=\frac{\left[2\gamma+\left(\frac{y-\mu}{\sigma}\right)^2\right]^{\gamma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{\sum_{j=0}^{\gamma}\binom{\gamma}{j}(2\gamma)^{\gamma-j}2^{j+\frac{1}{2}}\Gamma\left(j+\frac{1}{2}\right)\sigma},\quad \begin{matrix}-\infty<\mu<\infty\\ \sigma>0,\gamma=0,1,2,...\\ -\infty<y<\infty\end{matrix} \qquad (2)$$

Here μ is the location parameter, σ is the scale parameter and γ is the shape parameter. We denote the generalized new symmetric distribution with location parameter μ, scale parameter σ and shape parameter γ as $GNS(\mu,\sigma^2,\gamma)$. The factor $k(\sigma,\gamma)=\frac{1}{\sum_{j=0}^{\gamma}\binom{\gamma}{j}(2\gamma)^{\gamma-j}2^{j+\frac{1}{2}}\Gamma\left(j+\frac{1}{2}\right)\sigma}$, which incorporates the well-tabulated gamma function, serves as a normalizing factor to insure that the area under the density curve $f_\gamma(y/\mu,\sigma,\gamma)$ is equal to one.

The distribution has several desirable properties and nice physical interpretations. Because of the shape parameter it has quite a bit of flexibility for analyzing different types of platykurtic data. The attraction of the NS distribution is that from the parent pdf, a large class of distributions can be generated with the parameter γ controlling the kurtosis. Each value of the shape parameter $\gamma(=0,1,2,...)$ gives different bell shaped and symmetric continuous distributions with parameters μ and σ. So, γ can be treated as an index parameter which can be used to determine a larger class of specific distributions. Particularly, if we choose $\gamma=0$ then equation (2) is recognized as the pdf for the univariate normal distribution, i.e. $NS(\mu,\sigma^2,0)=N(\mu,\sigma^2)$. In this sense (2) may be regarded as a generalization of the normal distribution. Moreover, in the limit $\gamma\to\infty$, the pdf tends to the uniform distribution, $U(\mu-\sigma,\mu+\sigma)$. Therefore, this family of distribution has a broad application area in real life.

Figure 1 illustrates several of the possible shapes obtained from (2) for various choices of γ and $\mu=0$, $\sigma=1$

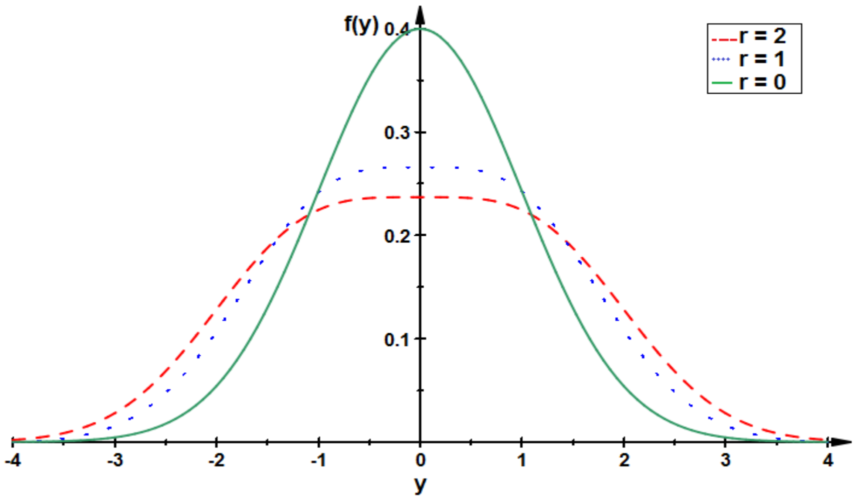

Figure1: The generalized new symmetric probability density for $(\mu, \sigma, \gamma) = (0,1,0)$ (*solid line*); $(\mu, \sigma, \gamma) = (0,1,1)$ (*dotted line*); and $(\mu, \sigma, \gamma) = (0,1,2)$ (*dashed line*)

Figure 1 represents the plot of (2) and shows the effect of varying the index parameter γ for fixed μ and σ. The green (solid) curve is the standard normal distribution while the red (dashed) and blue (dotted) curves are the new symmetric distribution with γ equal to 1 and 2 respectively. It can be seen that the distribution is very versatile and that the value of γ has a substantial effect on the kurtosis of the probability density function. For larger values of γ, we expect heavy fat tails. An advantage of GNS distribution is that it is adaptive to flatness in the data by varying the values of γ. When γ increases, the sharpness diminishes.

When a random variable Y is generalized new symmetric distributed with mean μ, variance σ^2 and index parameter γ or to define *y* as a variate drawn from this distribution, we write

$$Y \sim NS(\mu, \sigma^2, \gamma) \tag{3}$$

The distributions belonging to this family have the following additional characteristics.

i. It is clear from (2) that the mode of the pdf is μ and that it is unimodal and symmetrical about the mode. Therefore, the median and the mean are also equal to μ. These properties coincide with those of the normal distribution.

ii. Since the distribution is symmetric, its skewness is zero.

iii. The Function $f(y;\mu,\sigma,\gamma)$ is log-concave.

iv. Range: Infinity in both directions.

v. The distribution function of the family specified by (2) is given as

$$F(y)=\frac{\sum_{j=1}^{\gamma}\binom{\gamma}{j}(2\gamma)^{\gamma-j}2^{\left(j+\frac{1}{2}\right)}\Gamma\left(j+\frac{1}{2}\right)F_{\gamma}(y)+(2\pi)^{\frac{1}{2}}F_{0}(y)}{\sum_{j=1}^{\gamma}\binom{\gamma}{j}(2\gamma)^{\gamma-j}2^{\left(j+\frac{1}{2}\right)}\Gamma\left(j+\frac{1}{2}\right)+(2\pi)^{\frac{1}{2}}(2\gamma)^{\gamma}} \tag{4}$$

$$\text{where } F_{\gamma}(y)=\phi\left(\frac{y-\mu}{\sigma}\right)-e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^{2}}\sum_{s=1}^{\gamma}\frac{(y-\mu)^{2s-1}}{2^{s+\frac{1}{2}}\Gamma\left(s+\frac{1}{2}\right)\sigma^{2s-1}} \text{ and}$$

$$F_{0}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-t^{2}/2}dt,\quad y\in\Re.$$

$$\text{For } \gamma=2,\ f_{2}(y/\mu,\sigma,2)=\frac{\left[4+\left(\frac{y-\mu}{\sigma}\right)^{2}\right]^{2}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^{2}}}{27\sigma\sqrt{2\pi}},\text{ and}$$

$$F_{2}(y)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^{2}}dt-\frac{11}{27\sqrt{2\pi}}\left(\frac{y-\mu}{\sigma}\right)e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^{2}}-\frac{1}{27\sqrt{2\pi}}\left(\frac{y-\mu}{\sigma}\right)^{3}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^{2}}$$

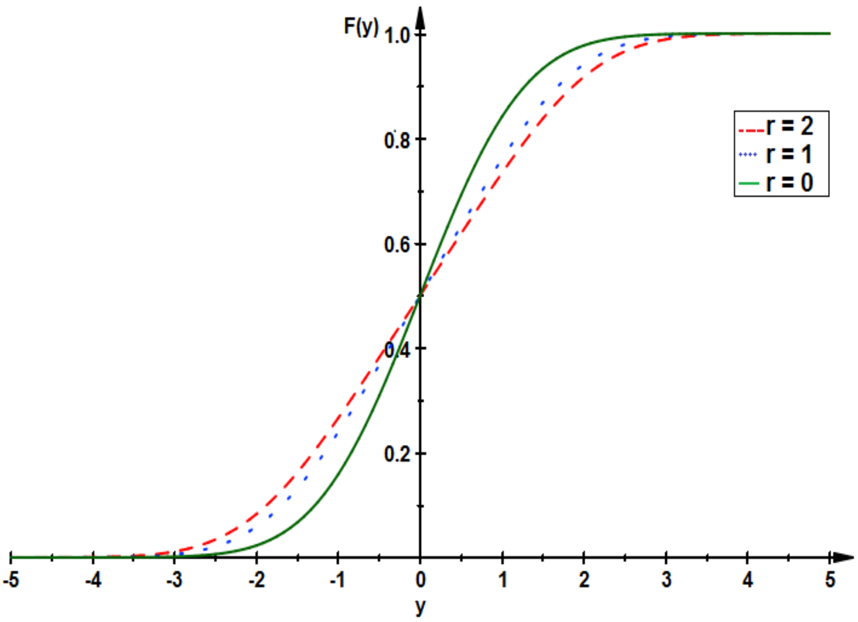

Figure 2: The generalized symmetric distribution for $NS(0,1,0)$, (*solid line*); $NS(0,1,1)$ (*dotted line*); and $NS(0,1,2)$ (*dashed line*)

vi. Central Moments

The GNS has moments of all orders *p*. That is, for new symmetric distributed Y with mean μ, standard deviation σ and index parameter γ, the expectation $E(y-\mu)^{p}$ exists and is finite for all *p*. The odd moments clearly all vanish by symmetry, I.e. $\mu_{2n+1}=0$ and the even moments are

$$\mu_{2n} = \left[\frac{\Gamma\left(n+\frac{1}{2}\right) + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(n+j+\frac{1}{2}\right)}{(\pi)^{\frac{1}{2}} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{1}{2}\right)} \right] 2^n \sigma^{2n} \tag{5}$$

In particular, the first four central moments of this distribution are

$$\mu_1 = 0, \mu_2 = \left[\frac{\frac{\sqrt{\pi}}{2} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{3}{2}\right)}{\sqrt{\pi} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{1}{2}\right)} \right] 2\sigma^2, \mu_3 = 0 \ and$$

$$\mu_4 = \left[\frac{\frac{3\sqrt{\pi}}{4} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{5}{2}\right)}{\sqrt{\pi} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{1}{2}\right)} \right] 4\sigma^4 \tag{6}$$

That is, the variance of the distribution is

$$\mu_2 = \left[\frac{\frac{\sqrt{\pi}}{2} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{3}{2}\right)}{\sqrt{\pi} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{1}{2}\right)} \right] 2\sigma^2 \tag{7}$$

vii. The recurrence relation between the central moments is

$$\mu_{2n} = (2n-1)\sigma^2 \mu_{2n-2} + \frac{2\sigma^2 \sum_{j=1}^{\gamma} j \binom{\gamma}{j} \gamma^{-j} \Gamma\left(n+j-\frac{1}{2}\right)}{\Gamma\left(n+\frac{1}{2}\right) + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j-\frac{1}{2}+n\right)} \mu_{2n-2} \tag{8}$$

viii. The characteristic function of this distribution is

$$\phi_y(t) = e^{it\mu - \frac{t^2\sigma^2}{2}} \left[\frac{\left[\sqrt{\pi} + \sum_{j=0}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \sum_{k=0}^{2j} 2^{\frac{k}{2}} (it\sigma)^{2j-k} \Gamma\{(k+1)/2\} \right]}{\sqrt{\pi} + \sum_{j=0}^{\gamma} \binom{\gamma}{j} \gamma^{-j} 2^j \Gamma\{(j+1)/2\}} \right] \tag{9}$$

This is the product of the normal characteristic function and a polynomial of ‘t’.

ix. The kurtosis of the distribution is

$$\beta_2 = \frac{\left[\frac{3\sqrt{\pi}}{4} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{5}{2}\right) \right] \left[\sqrt{\pi} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{1}{2}\right) \right]}{\left[\frac{\sqrt{\pi}}{2} + \sum_{j=1}^{\gamma} \binom{\gamma}{j} \gamma^{-j} \Gamma\left(j+\frac{3}{2}\right) \right]^2} \tag{10}$$

x. Hazard and reversed hazard functions of the distribution

The hazard function is the ratio of the probability density function to the survival function, $S(y)$. The hazard function $h(y)$ of the GNS distribution proposed in this paper can be utilized to characterize life phenomena and can be written as

$$h(y) = \frac{f(y)}{S(y)} = \frac{f(y)}{1 - F(y)} \tag{11}$$

where f is the *pdf* and F is the CDF of the GNS distribution with mean μ, variance σ and index parameter γ. If $\gamma = 0$, (11) becomes the normal hazard function. Recently, it is observed, Gupta, R. C., Gupta, R. D. (2007), that the reversed hazard function plays an important role in the reliability analysis. The reversed hazard function of the $GNS(\mu, \sigma, \gamma)$ is

$$r(y) = \frac{f(y)}{F(y)} \tag{12}$$

The hazard and reversed hazard functions of the generalized new symmetric distribution for $\mu = 0$, $\sigma = 1$ and some values of $\gamma(= 0,1,2)$ are illustrated in Figure 3.

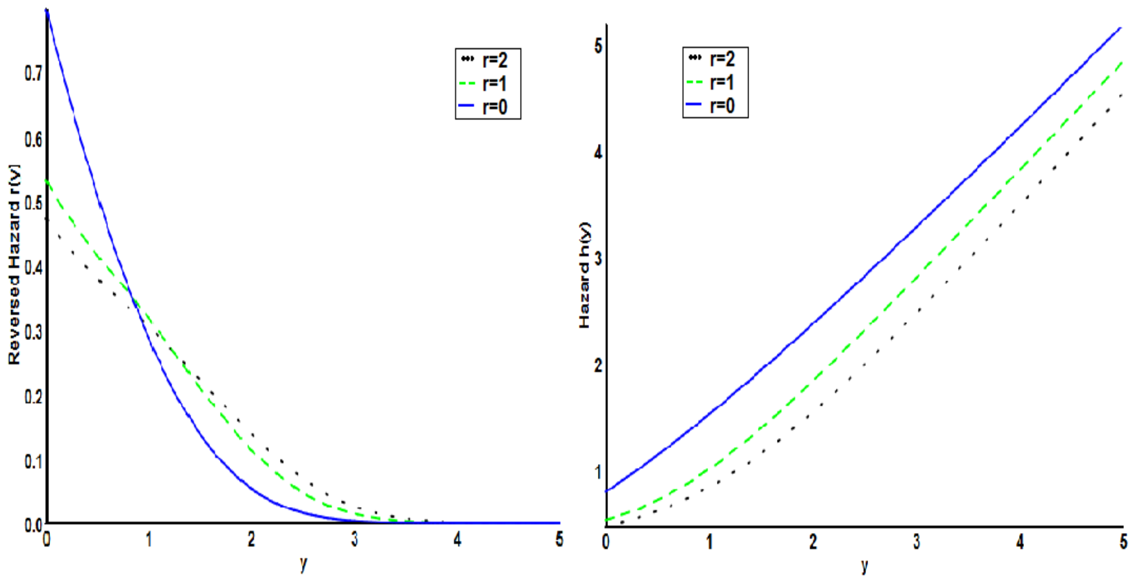

Figure 3: The GNS distribution reversed hazard (left) and hazard (right) functions for $(\mu, \sigma, \gamma) = (0,1,0)$ (solid line), $(\mu, \sigma, \gamma) = (0,1,1)$ (dashed line) and $(\mu, \sigma, \gamma) = (0,1,2)$ (dotted line)

It is well known that the hazard function or the reversed hazard function uniquely determines the corresponding probability density function.

3. Inference

Previously we have studied generalized new symmetric distribution and its distributional properties. Another aspect of any distributional study is to look in to the inferential aspects of the distribution, in particular the estimation of the parameters involved in the distribution under study and their asymptotic properties.

3.1 Estimation of the model parameters

The estimation of parameters in a linear regression model is probably one of the oldest and most widely used problems in many applied areas such as economics, engineering, social, health, and biological sciences. Maximum likelihood (ML) is the principal method of estimation used for all generalized linear models. However, the shape parameter γ is estimated based on the sample kurtosis of the distribution since it takes non-negative integral values. After identifying γ, the ML method is used to estimate the simple linear model parameters (the intercept β_0, and the slope β_1) and the scale parameter (σ).

3.1.1 Estimation of the shape parameter

Since the shape parameter γ takes non-negative integral values, it can be estimated not using the method of ML procedure but through sample kurtosis by solving the equation in which the sample kurtosis β_2 is equated to the theoretical kurtosis β_2. That is, $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{M^4}{M_2^2}$ or

$$\frac{\left[\frac{3\sqrt{\pi}}{4} + \sum_{j=1}^{\gamma}\binom{\gamma}{j}\gamma^{-j}\Gamma\left(j+\frac{5}{2}\right)\right]\left[\sqrt{\pi} + \sum_{j=1}^{\gamma}\binom{\gamma}{j}\gamma^{-j}\Gamma\left(j+\frac{1}{2}\right)\right]}{\left[\frac{\sqrt{\pi}}{2} + \sum_{j=1}^{\gamma}\binom{\gamma}{j}\gamma^{-j}\Gamma\left(j+\frac{3}{2}\right)\right]^2} = \frac{\left[\frac{\sum_{i=1}^{n}(y_i-\bar{y})^4}{n}\right]}{\left[\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\right]^2} \qquad (13)$$

Solving this equation numerically, we can find the value of γ. By Rochies theorem, there is one and only one real root to this equation. We pick up the nearest integer to this real root as an estimator for the shape parameter (γ). For the simulated data of various sample size *n*, we have estimated γ and the results are summarized in Table 1.

Table 1: The estimate of γ for various sample of size *n*

n	$\hat{\gamma}$	
	Calculated	Rounded
100	3.260	3
1000	2.260	2
3000	2.2015	2
5000	2.0240	2
10000	0.9527	1

For each estimated value of γ, the corresponding value of kurtosis β_2 can be found using (13). Some values of β_2 for γ = 0, 1, 2 ..., 21 are given in Table 2.

Table 2: Values of the kurtosis (β_2) for various values of shape parameter (γ)

γ	β_2	γ	β_2
0	3.0000	11	2.29975
1	2.5200	12	2.29509
2	2.43669	13	2.29097
3	2.39539	14	2.28729
4	2.36956	15	2.28398
5	2.35145	16	2.28098
6	2.33786	17	2.27825
7	2.32717	18	2.27574
8	2.31848	19	2.27343
9	2.31123	20	2.27129
10	2.30507	21	2.26931

The table indicates that as γ increases, β_2 decreases. This is also supported graphically in Figure 3.

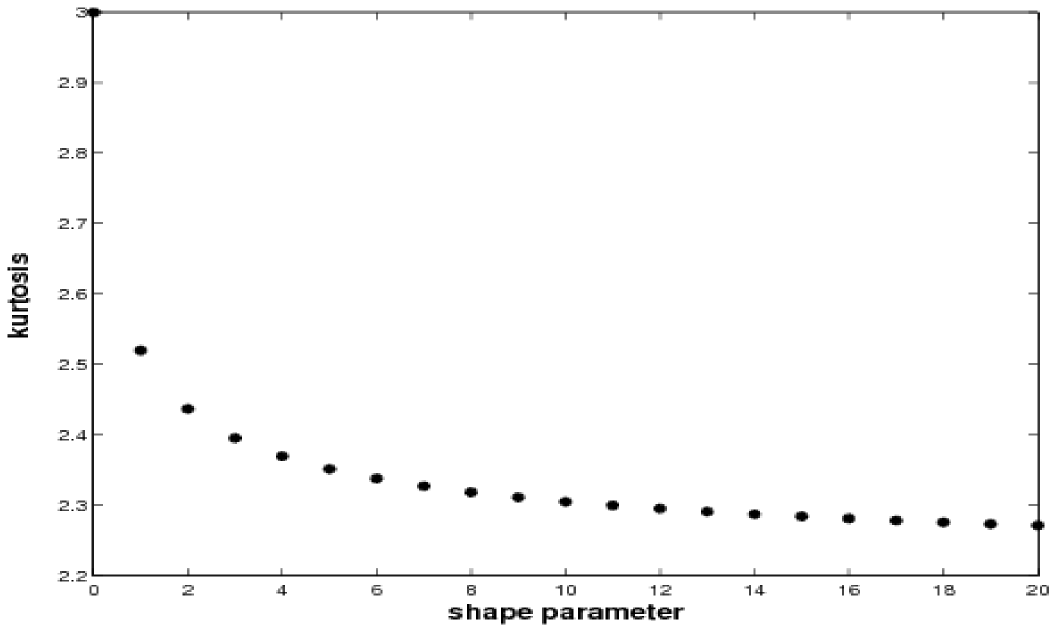

Figure 3: The relationship between kurtosis (β_2) and shape parameter (γ).

3.1.2 Maximum Likelihood Estimation of the Model Parameters

Let $y_1, y_2, ..., y_n$ be a random sample of size *n* drawn from $GNS(\beta_0, \beta_1, \sigma^2, \gamma)$. Since the shape parameter γ is already identified through the sample kurtosis in section 3.1.1, the unknown parameters of the model $(y, X\beta, \sigma^2 I)$ are the regression coefficients β_0 and β_1, and the error variance σ^2. Suppose we take a sample of size 5000, from Section 3.1.1 for the simulated data sets it is found that $\hat{\gamma} = 2.024 \approx 2$. Hence the distribution with $\hat{\gamma} = 2$ viz,

$$f_2(y/\mu,\sigma,2)=\frac{\left[4+\left(\frac{y-\mu}{\sigma}\right)^2\right]^2 e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}}{27\sigma\sqrt{2\pi}} \tag{14}$$

is fitted to the simulated data sets.

The new symmetric linear regression model is

$$y_i=\beta_0+\beta_1 x_i+u_i, u\sim NS(0,\sigma^2 I) \tag{15}$$

Thus, in this section we deal with the problem of estimation of these parameters from the observables **y** and **x** for a sample of size 5000 whose shape parameter estimate is found to be 2.024.

Given a set of *n iid* random vectors $\{Y_i\}_{i=1}^n$, each drawn from the generalized new symmetric distribution, the joint probability or likelihood of a particular realization, $\{y_i\}_{i=1}^n$, is given by

$$L(y;\beta_0,\beta_1,\sigma)=\prod_{i=1}^n f(y_i/\beta_0+\beta_1 x_i,\sigma)$$

$$=\prod_{i=1}^n \frac{\left[4+\frac{1}{\sigma^2}(y_i-\beta_0-\beta_1 x_i)^2\right]^2 e^{-\frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1 x_i)^2}}{27\sigma\sqrt{2\pi}}$$

$$=\left[27\sigma^3\sqrt{2\pi}\right]^{-n}\prod_{i=1}^n\left[4\sigma^2+(y_i-\beta_0-\beta_1 x_i)^2\right]^2 e^{-\frac{1}{2}\left(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma}\right)^2} \tag{16}$$

Taking logarithms of (16) and ignoring additive constants[1], we obtain the commonly used log-likelihood function, $\ell=\ln L,$ for the new symmetric regression model:

$$\ell=-\frac{5n}{2}\ln\sigma^2+2\sum_{i=1}^n\ln\left[4\sigma^2+(y_i-\beta_0-\beta_1 x_i)^2\right]-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\beta_0-\beta_1 x_i)^2 \tag{17}$$

The MLEs of these parameters are provided by taking the derivative with respect to β_0,β_1 and σ^2 and equating to 0, we obtain the normal equations as

$$\frac{\partial\ell}{\partial\beta_0}=-4\sum_{i=1}^n\frac{(y_i-\beta_0-\beta_1 x_i)}{4\sigma^2+(y_i-\beta_0-\beta_1 x_i)^2}+\frac{1}{\sigma^2}\sum_{i=1}^n(y_i-\beta_0-\beta_1 x_i)=0 \tag{18}$$

$$\frac{\partial\ell}{\partial\beta_1}=-4\sum_{i=1}^n\frac{(y_i-\beta_0-\beta_1 x_i)x_i}{4\sigma^2+(y_i-\beta_0-\beta_1 x_i)^2}+\frac{1}{\sigma^2}\sum_{i=1}^n(y_i-\beta_0-\beta_1 x_i)x_i=0 \tag{19}$$

and

$$\frac{\partial\ell}{\partial\sigma^2}=\frac{-5n}{2\sigma^2}+8\sum_{i=1}^n\frac{1}{4\sigma^2+(y_i-\beta_0-\beta_1 x_i)^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n(y_i-\beta_0-\beta_1 x_i)^2=0 \tag{20}$$

That is,

[1] Some terms in the density function have been dropped in the log-likelihood function since they do not affect the estimation of the parameters.

$$S = \begin{bmatrix} -4\sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1x_i)}{4\sigma^2+(y_i-\beta_0-\beta_1x_i)^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i) \\ -4\sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1x_i)x_i}{4\sigma^2+(y_i-\beta_0-\beta_1x_i)^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)x_i \\ \frac{-5n}{2\sigma^2}+8\sum_{i=1}^{n}\frac{1}{4\sigma^2+(y_i-\beta_0-\beta_1x_i)^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad (21)$$

There are no closed form solutions to the likelihood equations. Numerical methods must be applied for simultaneously solving the nonlinear equations to obtain $\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right)$. We need to use either the Scoring algorithm or the Newton-Raphson algorithm. The required numerical evaluations were implemented using MathLab R2012b. Recall that the formula for Newton-Raphson is

$$\theta^{(n+1)} = \theta^{(n)} - H^{-1}S \qquad (22)$$

where H is the Hessian (second derivative) matrix and S is the gradient (first derivative) vector of the log-likelihood function, both evaluated at the current value of the parameter vector.

That is,

$$S = \left[s_j\right] = \left[\frac{\partial \ln L}{\partial \theta_j}\right] \text{ and } H = \left[h_{ij}\right] = \left[\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j{}'}\right] \qquad (23)$$

where $\theta = \beta_0, \beta_1, \sigma^2$

We generate explanatory variable from uniform distribution and the random error from generalized new symmetric distribution with $\beta_0 = -2.45,\ \beta_1 = 4.9$ and $\sigma = 0.2$. Using statistical software Wolfram Mathematica 9, we generate data for sample size n =1000, 2000, 3000, 5000 and 10000. The iteration is to be repeated until the sequence $\left\{\theta^{(n+1)}\right\}$ thus obtained converges to the desired degree of accuracy. The results are reported in Table 3. Essentially, use of this method requires the prior computation of the Hessian matrix H and an initial guess $\theta^{(0)}$ for the model parameters β and scale parameter σ^2. We now need to take the derivative of each of the gradient vector with respect to β_0, β_1 and σ^2 to derive the Hessian matrix. Let's start with β_0.

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -4\sum_{i=1}^{n} \frac{-\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right] + 2(y_i - \beta_0 - \beta_1 x_i)^2}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} + \frac{\sum_{i=1}^{n}(-1)}{\sigma^2},$$

$$= -4\sum_{i=1}^{n} \frac{[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2]}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{n}{\sigma^2} \qquad (24)$$

$$\frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_0} = -4\sum_{i=1}^{n} \frac{-x_i[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2] + 2(y_i - \beta_0 - \beta_1 x_i)x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} + \frac{\sum_{i=1}^{n}(-x_i)}{\sigma^2},$$

$$= -4\sum_{i=1}^{n} \frac{\left[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{n\bar{x}}{\sigma^2} \qquad (25)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_0} = -4\sum_{i=1}^{n} \frac{-4(y_i - \beta_0 - \beta_1 x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{\sigma^4},$$

$$= 16\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{\sigma^4} \qquad (26)$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -4\sum_{i=1}^{n} \frac{-x_i\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right] + 2x_i(y_i - \beta_0 - \beta_1 x_i)^2}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n} x_i}{\sigma^2}$$

$$= -4\sum_{i=1}^{n} \frac{\left[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{n\bar{x}}{\sigma^2}, \qquad (27)$$

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -4\sum_{i=1}^{n} \frac{-x_i^2\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right] - 2x_i^2(y_i - \beta_0 - \beta_1 x_i)^2}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} + \frac{\sum_{i=1}^{n}(-x_i^2)}{\sigma^2},$$

$$= -4\sum_{i=1}^{n} \frac{\left[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]x_i^2}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} \qquad (28)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_1} = -4\sum_{i=1}^{n} \frac{-4(y_i - \beta_0 - \beta_1 x_i)x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i}{\sigma^4},$$

$$= 16\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i}{\sigma^4}, \tag{29}$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma^2} = 8\sum_{i=1}^{n} \frac{2(y_i - \beta_0 - \beta_1 x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{\sigma^4},$$

$$= 16\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{\sigma^4}, \tag{30}$$

$$\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2} = 8\sum_{i=1}^{n} \frac{-2(y_i - \beta_0 - \beta_1 x_i)(-x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} + \frac{-2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i}{2\sigma^4},$$

$$= 16\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i}{\sigma^4}, \tag{31}$$

$$\frac{\partial^2 \ell}{(\partial \sigma^2)^2} = \frac{5n}{2\sigma^4} - 8\sum_{i=1}^{n} \frac{4}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^6},$$

$$= \frac{5n}{2\sigma^4} - 32\sum_{i=1}^{n} \frac{1}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2} - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^6}, \tag{32}$$

$$Thus,\ H = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta_0^2} & \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_0} \\ \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \ell}{\partial \beta_1^2} & \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_1} \\ \frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma^2} & \frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2} & \frac{\partial^2 \ell}{(\partial \sigma^2)^2} \end{pmatrix} \tag{33}$$

where H is the Hessian matrix and each stuff is as derived in equations (24) through (32).

3.1.3 Information Matrix and Standard Errors

The second derivatives of the log-likelihood, given in equations (24) through (32), are complicated nonlinear functions of the data whose exact expected values are unknown. Hence, we use moment approximation as in the case in H. Cramer (1946). The first 4 central moments (6) are used to simplify the computation.

The asymptotic variances and covariances of maximum likelihood estimators are given by the elements of the inverse of the Fisher information matrix:

$$I_{\theta_i\theta_j} = -E\left(\frac{\partial^2 L(\beta)}{\partial\theta_i\partial\theta_j}\right) = \begin{pmatrix} I_{\beta_0\beta_0} & I_{\beta_1\beta_0} & I_{\sigma^2\beta_0} \\ & I_{\beta_1\beta_1} & I_{\sigma^2\beta_1} \\ & & I_{\sigma^2\sigma^2} \end{pmatrix}, \theta' = \left(\beta_0, \beta_1, \sigma^2\right) \tag{34}$$

Unfortunately, the exact mathematical expressions for the above expectations are very difficult to obtain. Therefore, we give the Fisher information matrix for the MLE, which is obtained by using moment approximation, see H.Cramer (1946).

The components of the Fisher information matrix are:

$$I_{\beta_0\beta_0} = -E\left(\frac{\partial^2 \ell}{\partial\beta_0^2}\right) = 4E\left\{\sum_{i=1}^{n}\frac{-4\sigma^2 + \left(y_i - \beta_0 - \beta_1 x_i\right)^2}{\left[4\sigma^2 + \left(y_i - \beta_0 - \beta_1 x_i\right)^2\right]^2}\right\} + E\left(\frac{n}{\sigma^2}\right)$$

$$= 4\left\{\sum_{i=1}^{n}\frac{-4\sigma^2 + E\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{\left[16\sigma^4 + 8\sigma^2 E\left(y_i - \beta_0 - \beta_1 x_i\right)^2 + E\left(y_i - \beta_0 - \beta_1 x_i\right)^4\right]}\right\} + \frac{n}{\sigma^2}$$

$$= 4\sum_{i=1}^{n}\frac{-4\sigma^2 + \mu_2}{\left[16\sigma^4 + 8\sigma^2\mu_2 + \mu_4\right]} + \frac{n}{\sigma^2}$$

$$= 4\left\{\sum_{i=1}^{n}\frac{-4\sigma^2 + \frac{55\sigma^2}{27}}{\left[16\sigma^4 + 8\sigma^2\left(\frac{55\sigma^2}{27}\right) + \frac{91\sigma^4}{9}\right]}\right\} + \frac{n}{\sigma^2}$$

$$= \frac{933n}{1145\sigma^2}, Since\ \mu_2 = \frac{55\sigma^2}{27}, and\ \mu_4 = \frac{91\sigma^4}{9} \tag{35}$$

$$I_{\beta_1\beta_0} = -E\left(\frac{\partial^2 \ell}{\partial\beta_1\partial\beta_0}\right) = 4E\left\{\sum_{i=1}^{n}\frac{\left[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2}\right\} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i = -E\left(\frac{\partial^2 \ell}{\partial\beta_0\partial\beta_1}\right)$$

$$\approx 4\left\{\sum_{i=1}^{n}\frac{x_i(-4\sigma^2 + \mu_2)}{16\sigma^4 + 8\mu_2\sigma^2 + \mu_4}\right\} + \frac{n\bar{x}}{\sigma^2}$$

$$= \frac{-212n\bar{x}}{1145\sigma^2} + \frac{n\bar{x}}{\sigma^2} = \frac{933n\bar{x}}{1145\sigma^2} \qquad (36)$$

$$I_{\sigma^2\beta_0} = -E\left(\frac{\partial^2 \ell}{\partial\sigma^2\partial\beta_0}\right) = E\left\{\sum_{i=1}^{n}\frac{-16(y_i - \beta_0 - \beta_1 x_i)}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2}\right\} + E\left(\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)}{\sigma^4}\right) = I_{\beta_0\sigma^2}$$

$$\approx -16\left\{\sum_{i=1}^{n}\frac{\mu_1}{\left[16\sigma^2 + 8\mu_2\sigma^2 + \mu_4\right]^2}\right\} + \sum_{i=1}^{n}\frac{\mu_1}{\sigma^4} = 0 \qquad (37)$$

$$I_{\beta_1\beta_1} = -E\left(\frac{\partial^2 \ell}{\partial\beta_1^2}\right) = 4E\left\{\sum_{i=1}^{n}\frac{\left[-4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]x_i^2}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2}\right\} + E\left(\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2}\right)$$

$$\approx 4\left\{\sum_{i=1}^{n}\frac{x_i^2\left[-4\sigma^2 + \mu_2\right]}{\left(\left[16\sigma^4 + 8\sigma^2\mu_2 + \mu_4\right]\right)}\right\} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2$$

$$= \frac{933}{1145\sigma^2}\sum_{i=1}^{n} x_i^2 \qquad (38)$$

$$I_{\sigma^2\beta_1} = -E\left(\frac{\partial^2 \ell}{\partial\sigma^2\partial\beta_1}\right) = 4E\left\{\sum_{i=1}^{n}\frac{-4(y_i - \beta_0 - \beta_1 x_i)x_i}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2}\right\} + E\left\{\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i}{\sigma^4}\right\} = I_{\beta_1\sigma^2}$$

$$\approx -16\left\{\sum_{i=1}^{n}\frac{x_i\mu_1}{\left[16\sigma^4 + 8\mu_2\sigma^2 + \mu_4\right]}\right\} + \frac{1}{\sigma^4}\sum_{i=1}^{n} x_i\mu_1 = 0 \qquad (39)$$

$$I_{\sigma^2\sigma^2} = -E\left(\frac{\partial^2 \ell}{(\partial\sigma^2)^2}\right) = -\frac{5n}{2\sigma^4} + 32E\left\{\sum_{i=1}^{n}\frac{1}{\left[4\sigma^2 + (y_i - \beta_0 - \beta_1 x_i)^2\right]^2}\right\} + E\left\{\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^6}\right\}$$

$$\approx -\frac{5n}{2\sigma^4} + 32\left\{\sum_{i=1}^{n}\frac{1}{16\sigma^4 + 8\mu_2\sigma^2 + \mu_4}\right\} + \frac{1}{\sigma^6}\sum_{i=1}^{n}\mu_2 = \frac{18031n}{61830\sigma^4} \qquad (40)$$

The information matrix related to the parameter is thus

$$I = -E\begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta_0^2} & \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_0} \\ \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \ell}{\partial \beta_1^2} & \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta_1} \\ \frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma^2} & \frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2} & \frac{\partial^2 \ell}{(\partial \sigma^2)^2} \end{pmatrix} = \begin{pmatrix} \frac{933n}{1145\sigma^2} & \frac{933n\bar{x}}{1145\sigma^2} & 0 \\ \frac{933n\bar{x}}{1145\sigma^2} & \frac{933\sum_{i=1}^{n} x_i^2}{1145\sigma^2} & 0 \\ 0 & 0 & \frac{18031n}{61830\sigma^4} \end{pmatrix} \quad (41)$$

Consequently, the asymptotic variances and covariances of maximum likelihood estimators are given by the elements of the inverse of the Fisher information matrix (41) and then substituting β_0, β_1 and σ^2 by $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}^2$.

$$Var(\hat{\theta}) = \begin{bmatrix} \frac{1145\sum_{i=1}^{n} x_i^2}{933\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} & \frac{-1145\sum_{i=1}^{n} x_i}{933\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} & 0 \\ \frac{-1145\sum_{i=1}^{n} x_i}{933\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} & \frac{1145n}{933\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} & 0 \\ 0 & 0 & \frac{61830\sigma^2}{18031} \end{bmatrix} \frac{\hat{\sigma}^2}{n} \quad (42)$$

The diagonal elements of $Var(\hat{\theta})$ will give the asymptotic variances of associated with the coefficients. The stuff in the bottom right is the variance of σ^2.

3.2 Least Square Estimation for Model Parameters

The most widely used technique for estimating the unknown regression coefficients in a standard linear regression model is undeniably, the method of ordinary least squares (OLS). The least square estimates of β_0 and β_1 are the values which minimize

$$SS = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \quad (43)$$

And provide the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \equiv \frac{S_{xy}}{S_{xx}}$$

$$and \qquad (44)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$. Table 3 summarizes the results of the OLS estimation of the model to the simulated data sets.

Table 3: OLS Estimation Output for the Simulation Data

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
y	2	9998	430.1	0.0430	0.2074	0.9791	0.9790
Nonlinear OLS Parameter Estimates							
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|			
β_0	-2.45381	0.00413	-593.47	<.0001			
β_1	4.906041	0.00718	683.56	<.0001			
Number of Observations				**Statistics for System**			
	Used	10000		objective	0.0430		
	Missing	0		Objective*N	430.0562		

3.3 Simulation and Results

To better understand, compare and observe their practical performance, the methods developed in this work will be evaluated with simulation studies. This is done by generation of artificial data from an assumed model and application of the estimation methods to the generated data. All computations were carried out using Wolfram Mathematica software (version 9), SAS (version 9), and MathLab (version R2012b). To facilitate exposition of the method of estimation we considered linear model of the form (1) and for illustrative purposes we get a set of simulated data (x_i, y_i) for sample sizes of n=1000, 2000, 3000, 5000, 10000. (The sample size in this case, 10,000, is relatively large, and so finite sample bias is less of an issue.) The dependent y variable is simulated from generalized new symmetric distribution with mean 0, variance 1 and shape parameter 2 based on random variable inversion method, while the predictor x variable is generated uniformly between 0 and 1. Summary statistics of simulations for the regression model using maximum likelihood procedure are presented in Table 4.

Table 4: Summary of ML Estimation of the regression model parameters based on simulations of the new symmetric distribution for $\hat{\beta}_0 = -2.455;$ $\hat{\beta}_1 = 4.9050;$ and $\hat{\sigma}^2 = 0.0430,$ with n=1000, 2000, 3000, 5000 and 10000.

n	Parameter	Estimate	Standard Error	Wald 95% Conf. Limits		Chi-Square	Pr > ChiSq	LL
				Lower	Upper			
1000	β_0	-2.4573	0.0132	-2.4832	-2.4315	34734.5	<.0001	166.2385
	β_1	4.9190	0.0222	4.8755	4.9625	49052.2	<.0001	
	σ	0.2049	0.0046	0.1961	0.2141			
2000	β_0	-2.4525	0.0092	-2.4706	-2.4344	70653.4	<.0001	339.8547
	β_1	4.8866	0.0160	4.8553	4.9179	93687.4	<.0001	
	σ	0.2042	0.0032	0.1979	0.2106			
3000	β_0	-2.4247	0.0065	-2.4373	-2.4120	140901	<.0001	839.5490
	β_1	4.8523	0.0113	4.8301	4.8745	183976	<.0001	
	σ	0.1766	0.0023	0.1722	0.1811			
5000	β_0	-2.4507	0.0057	-2.4620	-2.4395	182311	<.0001	945.3037
	β_1	4.9021	0.0100	4.8825	4.9216	241132	<.0001	
	σ	0.2046	0.0020	0.2006	0.2086			
10000	β_0	-2.4538	0.0041	-2.4619	-2.4457	352273	<.0001	1542.7372
	β_1	4.9060	0.0072	4.8920	4.9201	467353	<.0001	
	σ	0.2074	0.0015	0.2045	0.2103			

From numerical results in Table 4, we conclude that if the sample size n is increased, then

i. The estimated standard errors of the estimators are decreased. This indicates that the maximum likelihood estimates provide asymptotically normally distributed and consistent estimators for the parameters;

ii. The asymptotic variances of the estimators are decreased; and

iii. The confidence interval lengths of the parameters are decreased.

The value of $\hat{\theta}$ that maximizes the likelihood function are thus

$$\hat{\theta} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right) = \left(-2.4538, 4.9060, 0.0430\right)' \quad (45)$$

It can also be verified that the Hessian matrix (33) evaluated at $\hat{\beta}_0 = -2.4538$; $\hat{\beta}_1 = 4.9060$; and $\hat{\sigma}^2 = 0.0430$, is a negative definite matrix. These ensure that this solution does indeed provide a maximum. The fitted simple linear model with new symmetric error terms to the simulated data is:

$$\hat{Y} = -2.4538 + 4.9060X \quad (46)$$

The estimated standard errors of the estimators are just the square roots of the diagonal elements of the variance-covariance matrix given in (42). That is,

$$s.e.(\hat{\beta}_0) = 0.0041,\ s.e.(\hat{\beta}_1) = 0.0072 \text{ and } s.e.(\hat{\sigma}) = 0.0015 \quad (47)$$

Table 5 summarizes the results of the ML estimation based on the sample of size 10,000.

Table 5: Maximum Likelihood Estimation Output for the Simulation Data

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	9998	430.0562	0.0430
Scaled Deviance	9998	10000.000	1.0002
Pearson Chi-Square	9998	430.0562	0.0430
Scaled Pearson X^2	9998	10000.000	1.0002
Log Likelihood		1542.7372	
Algorithm converged.			

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > Chi-Sq
β_0	1	-2.4538	0.0041	-2.4619	-2.4457	352273	<.0001
β	1	4.9060	0.0072	4.8920	4.9201	467353	<.0001
σ	1	0.2074	0.0015	0.2045	0.2103		

NOTE: The scale parameter was estimated by maximum likelihood.

In the "Criteria for Assessing Goodness of Fit" table displayed in Table 5, the value of the deviance divided by its degrees of freedom is less than 1. A p-value is not computed for the scaled deviance; however, a scaled deviance that is approximately equal to its degrees of freedom is a possible indication of a good model fit.

3.4 Properties of the Estimators

In this section, the asymptotic behavior of the estimators obtained in Section 3.2 are studied. The GNS family satisfies all the regularity conditions, and therefore we have the following results. See for more details, Casella and Berger (2002).

The maximum likelihood estimators $\hat{\theta} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right)$ of ($\beta_0, \beta_1, \sigma^2$) are consistent estimators.

Proof: From (42) it is clear that each variance tends to zero as $n \to \infty$ so that we conclude the estimators are consistent since they are composed of *i.i.d*. observations. This is also supported graphically in Figure 4. It is observed that the bias first increase as n increases and then start decreasing.

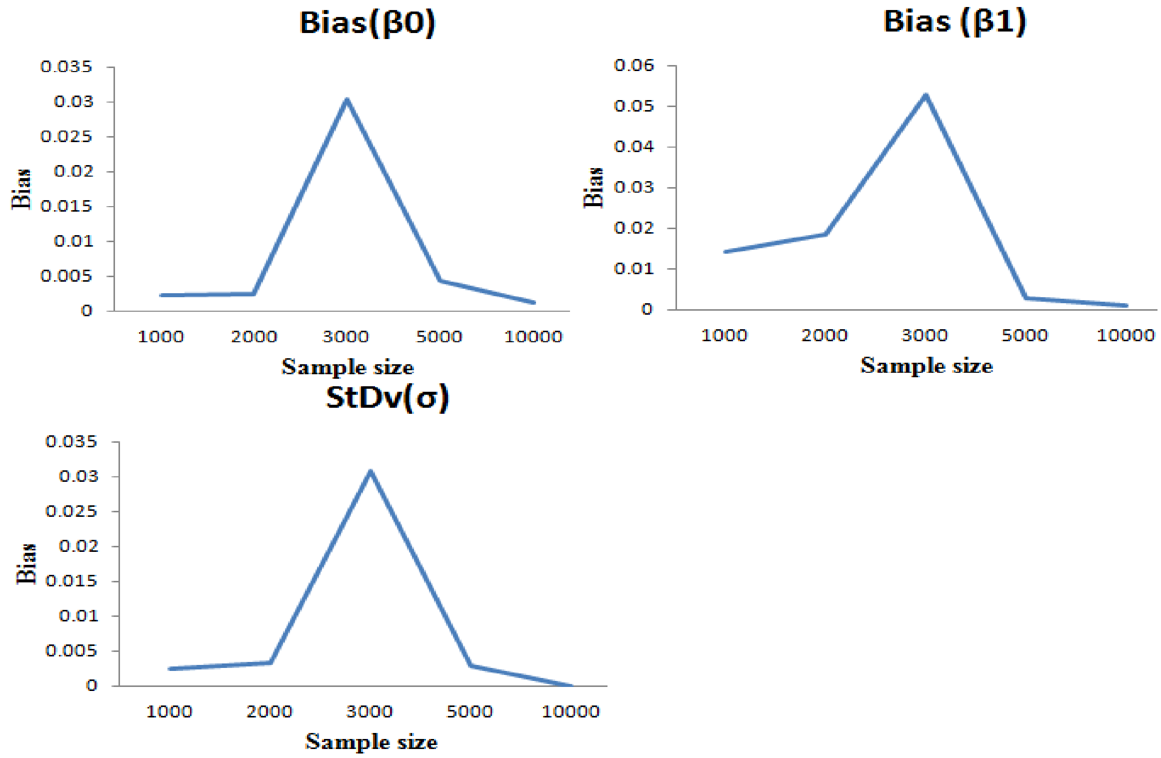

Figure 4: **Bias versus sample size**

$\sqrt{n}\left(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \hat{\sigma}^2 - \sigma^2\right)$ is asymptotically normal with mean vector **0** and the variance covariance matrix given in (42).

Proof: The asymptotic normality of the estimators $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right)$ here follows as in the i.i.d. case (H.Cramer, 1946) and we have that

$$\frac{\sqrt{n}(\hat{\beta}_0 - \beta_0)}{se(\hat{\beta}_0)} \xrightarrow{L} N(0,1), \frac{\sqrt{n}(\hat{\beta}_1 - \beta_1)}{se(\hat{\beta}_1)} \xrightarrow{L} N(0,1)$$
$$and \qquad (48)$$
$$\frac{\sqrt{n}(\hat{\sigma}^2 - \sigma^2)}{se(\hat{\sigma}^2)} \xrightarrow{L} N(0,1)$$

It follows that the MLEs are asymptotically normal.

Maximum likelihood estimators $\hat{\theta} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right)$ of ($\beta_0, \beta_1, \sigma^2$) are efficient estimators.

Proof: An estimator whose variance is as small as Cramer-Rao lower bound when the sample size tends to infinity is called asymptotically efficiency. It can be shown that the Cramer Rao lower bound for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ are respectively

$$Var(\hat{\beta}_0) \geq \frac{1}{-nE\left(\frac{\partial^2 \ln f}{\partial \beta_0^2}\right)} = \frac{1145\sigma^2}{933n},$$

$$Var(\hat{\beta}_1) \geq \frac{1}{-nE\left(\frac{\partial^2 \ln f}{\partial \beta_1^2}\right)} = \frac{1145\sigma^2}{933nx^2} \quad and$$

$$Var(\hat{\sigma}^2) \geq \frac{1}{-nE\left(\frac{\partial^2 \ln f}{(\partial \sigma^2)^2}\right)} = \frac{61830\sigma^4}{18031n} \tag{49}$$

This means that any unbiased estimator that achieves this lower bound is efficient and no better unbiased estimator is possible. Now look back at the variance-covariance matrix in Eq. (3.30). It is interesting to note that the variances of the estimators in the variance-covariance matrix (42) do asymptotically coincide with Cramer-Rao lower bound (49). This means that our MLEs are 100% asymptotically efficient.

4. Comparative Study of the model

4.1 Comparison of ML and OLS Estimators

In this section Maximum Likelihood and Ordinary Least Square estimation methods are compared in fitting the simple linear model with generalized new symmetric error terms with a sample size *n*=10,000. One-step-ahead forecasting is commonly used to compare the performance of different models (Clements and Hendry, 1997; Chiang et al., 2009). For each estimation techniques bias, mean square error (MSE) and relative mean square error (RMSE) are calculated; where

$$MSE(\hat{\theta}) = \text{var}(\hat{\theta}) + \left[bias(\hat{\theta})\right]^2$$
$$and \tag{50}$$
$$RMSE(\hat{\theta}) = \frac{MSE_{OLS} - MSE_{ML}}{MSE_{OLS}}.$$

RMSE is useful to measure the quality of the parameter estimation. Positive values of the RMSE can be expressed as there is a proportional decrease in the MSE of a given estimator with respect to OLS method. The computational result is presented in Table 6.

Table 6: Bias and MSE of Simulation Results for *n*=10,000

Methods	Parameter	Bias	MSE	RMSE
OLS	β_0	-0.00381	3.16E-05	0.01
	β_1	0.006041	8.805E-05	0.002
	σ	0.0074	0.043	0
ML	β_0	-0.0038	3.125E-05	0.01
	β_1	0.0060	8.784E-05	0.002
	σ	0.0074	0.043	0

Table 6 gives summary results for comparison criteria confirming the fact that deviations from normality cause OLS estimators to be poor estimators. The results reported in Table 6 also show that ML estimators have both smaller one-step-ahead forecast bias and MSE; and have positive values of RMSE than the LS estimators. This revealed that ML estimation technique exhibits the stronger performance than OLS.

4.2 Comparison of linear regression models with new symmetric and normally distributed errors

To the simulated data set, we fit both simple linear regression models under normality and new symmetric assumptions. In order to choose the best model among the fitted ones we computed the Akaike's information criterion AIC and the Bayesian information criterion BIC with model diagnostics root mean square error (RMSE). In this case, we expect that the proposed model would be chosen as our best model according to the minimum of AIC or BIC, and that the model parameters would be estimated correctly, since our true model is generated under the NS assumption. The output of simulation study for both the models using various sample sizes are presented in Table 7. Table 7: Summary of Estimation result for normal and new symmetric error Model

	Normal			New symmetric		
n	AIC	BIC	RMSE	AIC	BIC	RMSE
1000	-3112.5252	-3110.5172	0.21071	-3166.3541	-3164.3461	0.20512
2000	-6321.8686	-6319.8646	0.20578	-6351.4636	-6349.4596	0.20426
3000	-10388.294	-10386.291	0.17698	-10400.239	-10398.236	0.17663
5000	-15809.819	-15807.817	0.20573	-15864.483	-15862.482	0.20461
10000	-31306.402	-31304.401	0.20900	-31460.245	-31458.244	0.20740

The model with the smallest AIC or BIC among all competing models is deemed the best model where it can be seen that the SP distribution provides the best fit to the data. Both the information criteria methods (AIC and BIC) and the model diagnostics (RMSE) indicate that linear model with new symmetrically distributed error terms consistently performed best for all of the sample sizes. This can also be consistently noticed from Figure 5 through 7.

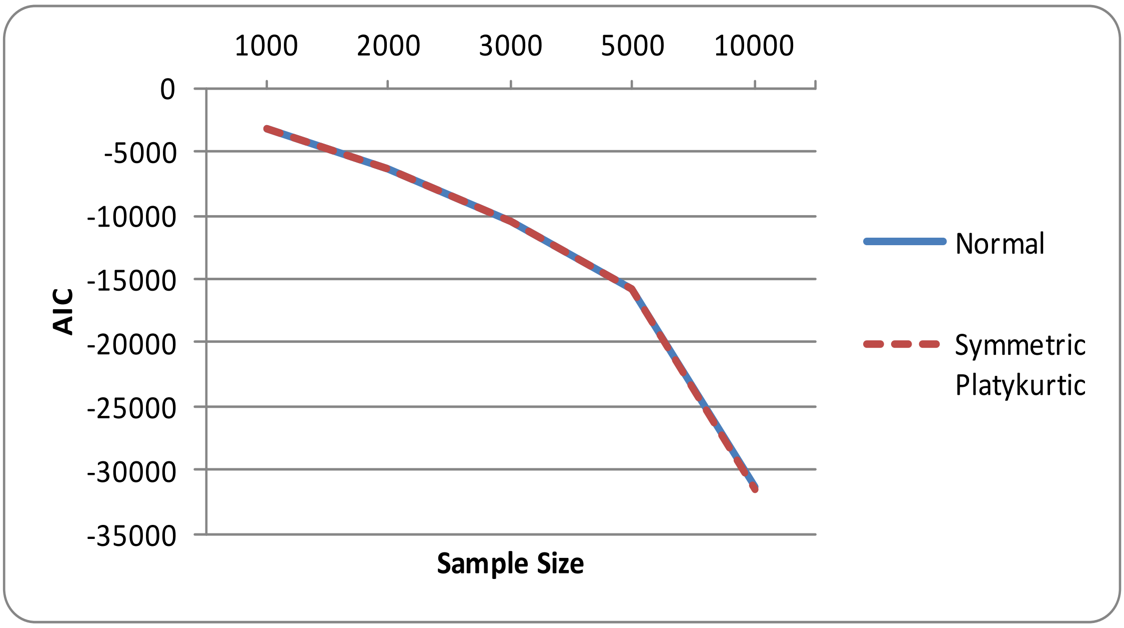

Figure 5: Graphical comparisons of SP and Normal Distributions using AIC

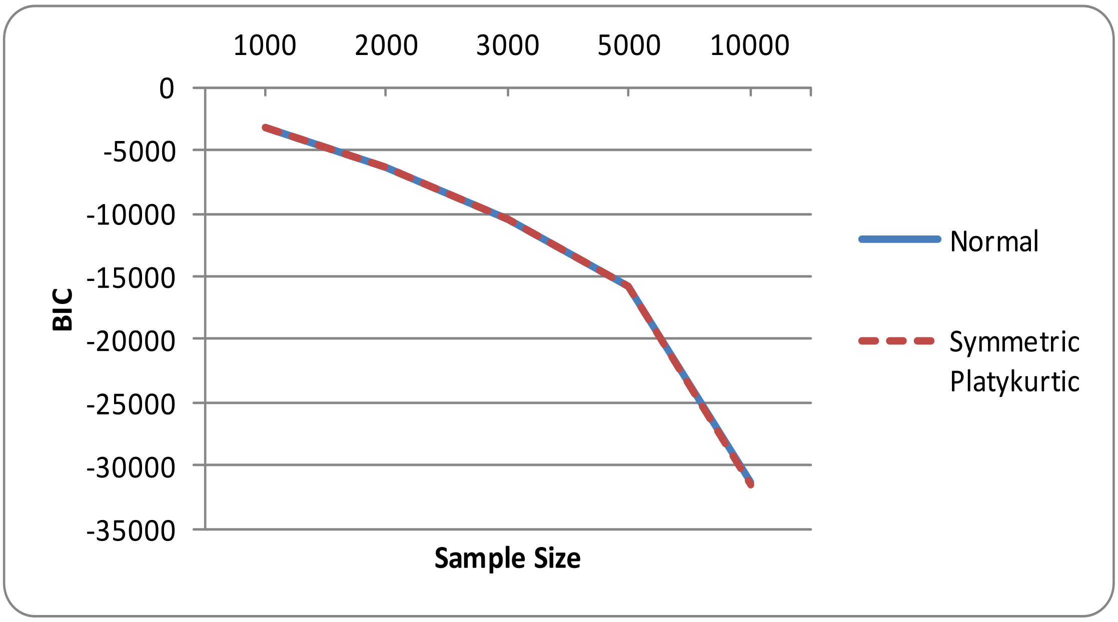

Figure 6: Graphical comparisons of SP and Normal Distributions using BIC

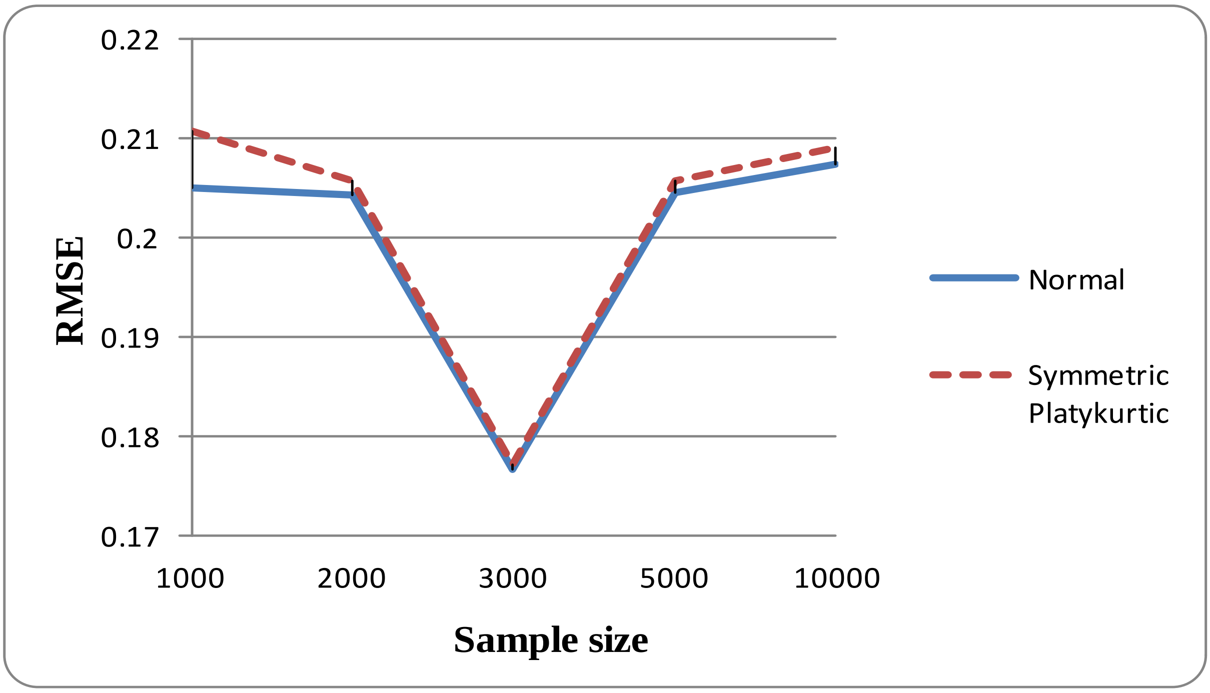

Figure 7: Graphical comparisons of NS and Normal Distributions using RMSE

From Figure 5 and 6, NS and normal are almost indistinguishable. However, from RMSE we note that the NS regression model is better than the normal regression model.

5. Summary and Conclusions

In this paper, we presented a new model to deal with nonnormality in linear models, as possible alternatives to the linear regression model with normal error terms. We developed NS linear model which can be used to model random phenomena whose observations' tails are thicker than those of normal distribution which is used often in the literature. The maximum likelihood estimators of the model parameters are derived and their asymptotic properties are studied. Through simulation studies these estimators are compared with ordinary least square estimators. The simulated results reveal that the ML estimators are more efficient than LSE in terms of the relative efficiency of one-step-ahead forecast bias and mean square error. A comparative study of the developed linear model with that of Gaussian errors revealed that the proposed model gives good fit to the simulated data sets. The proposed model is useful for analyzing data sets arising from agricultural experiments; portfolio management, space experiment and a wide range of other platykurtic nature practical problems.

References

[1] A.Asrat Atsedeweyn and K.Srinivasa Rao (2013). Linear regression model with new symmetric distributed errors. Journal of Applied Statistics, DOI: 10.1080/02664763.2013.839638.

[2] Atkinson, A.C. (1985). Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. New York: *Oxford University Press.*

[3] Bian, G., and Tiku, L.M. (1997). Bayesian inference based on robust priors and MML estimators: Part I, symmetric location-scale distributions. *Statistics*, 29, 317-345.

[4] Casella, G. and Berger, R. L. (2002). Statistical Inference, Second edition, *Duxbury Press.*

[5] C.R. Rao (1973). Linear Statistical Inference and Its Applications, 2nd ed., *Wiley,* New York

[6] Chiang, T.C., Qiao, Z., and Wong, W.K. (2009). New evidence on the relation between return volatility and trading volume. *Journal of Forecasting*, 29(5), 502-515.

[7] Clements, M.P., and Hendry, D.F. (1997). An empirical study of seasonal unit roots in forecasting. *International Journal of Forecasting*, 13, 341-355.

[8] Das Gupta, S. and Jammalamadaka, S.R. (2003). Linear Models: An Integrated Approach. *World Scientific Publishing* Co. Pte. Ltd., 6, 128-133.

[9] Draper and Smith (1981). Applied Regression Analysis, Third Edition, John Wiley and Sons, Inc., New York.

[10] Grob J. (2003). Linear Regression. Lecture notes in statistics, *Springer* Vol. 175.

[11] Guorui Bian, Michael McAleer, and Wing-Keung Wong (2013). Robust Estimation and Forecasting of the Capital Asset Pricing Model. *Annals of Financial Economics* (forthcoming).

[12] H. Cramer (1946). Mathematical Methods of Statistics. Number 9 in Princeton Mathematical Series, Princeton University Press, Princeton.

[13] Habib Jafari and Reza Hashemi (2011). Optimal Designs in a Simple Linear Regression with Skew-Normal Distribution for Error Term. *Journal of Applied Mathematics,* 1(2), 65-68; Scientific & Academic Publishing.

[14] Islam, M.Q., and Tiku, M.L. (2005). Multiple linear regression models under non-normality. *Communications in Statistics-Theory and Methods*, 33(10), 2443-2467.

[15] Islam, M.Q., and Tiku, M.L. (2010). Multiple linear regression models with stochastic design variables. *Journal of Applied Statistics*, 37(6), 923-943.

[16] M.Seshashayee, K. Srinivas Rao, Ch. Satyanarayana and P. Srinivasa Rao (2011). Image segmentation based on a finite generalized new symmetric mixture model with K-Means. *International Journal of Computer Science* Issues, Vol. 8, Issue 3, No. 2.

[17] McCullagh, P. and J.A. Nelder (1989). Generalized Linear Models. 2nd ed. *Chapman & Hall/CRC*, Boca Raton, Florida.

[18] Min-Hui Liu, and H. Bozdogan (2004). PE multiple regression model selection with ICOMP and genetic algorithms. *Working paper*.

[19] Minhui Liu, and Hamparsum Bozdogan (2008). Multivariate Regression Models with Power Exponential Random Errors and Subset Selection Using Genetic Algorithms With Information Complexity. *European Journal of Pure and Applied Mathematics* 1(1), 4-37.

[20] Montgomery D.C., Peck E.A., Vining G.G. (2001). Introduction to Linear Regression Analysis, 3rd edn. *John Wiley and Sons*, New York.

[21] S. Jahan, and A. Khan (2012). Power of t-test for Simple Linear Regression Model with Non-normal Error Distribution: A Quantile Function Distribution Approach. *Journal of Scientific Research Res.* 4 (3), 609-622.

[22] Seber G.A.F. (1977). Linear regression analysis. *Wiley*, New York.

[23] Seber G.A.F. and Lee A.J. (2003). Linear Regression Analysis, 2nd edn. *John Wiley and Sons*, Hoboken.

[24] Soffritti, G. and Galimberti, G. (2011). Multivariate linear regression with non-normal errors: a solution based on mixture models. *Statistics and Computing* 21(4), 523–536

[25] Srinivasa Rao K, C.V.S.R.Vijay Kumar, and J.Lakshmi Narayana (1997). On a New Symmetrical Distribution. *Journal of Indian Society for Agricultural Statistics*, 50(1), 95-102.

[26] Sutradhar, B.C., Ali, M.M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Commun. Stat.Theory* 15, 429–450.

[27] Tiku, M.L., Islam, M.Q., and Selcuk, A.S. (2001). Non-normal regression, II: symmetric distributions. *Communications in Statistics: Theory and Methods* 30, 1021-1045.

[28] Tiku, M.L., Wong, W.K., and Bian, G. (1999). Estimating Parameters in Autoregressive Models in Non-normal Situations: symmetric Innovations. *Communications in Statistics: Theory and Methods* 28(2), 315-341.

[29] Tiku, M.L., Wong, W.K., Vaughan, D.C., and Bian, G. (2000). Time series models with nonnormal innovations: symmetric location-scale distributions. *Journal of Time Series Analysis* 21(5), 571-596.

[30] Weisberg S. (2005). Applied Linear Regression, 3rd edn. *John Wiley and Sons*, Hoboken.

[31] Wing-Keung Wong and Guorui Bian (2005). Estimation of parameters in autoregressive models with asymmetric innovations. *Statistics and Probability Letters* 71, 61-70.

[32] Wing-Keung Wong and Guorui Bian (2005). Robust Estimation of Multiple Regression Model with Non-normal Error: Symmetric Distribution.

[33] Yan X. and Su X.G. (2009). Linear Regression Analysis: Theory and Computing. *World Scientific Publishing Company*, Hackensack.

[34] Zeckhauser, R. and Thompson, M. (1970). Linear Regression with Non-Normal Error Terms. *The Review of Economics and Statistics*, 52(3), 280-286.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with Multivariate student t error terms. *Jour. of American Statistical Association*, 71, 400–405.