TESTING FOR MULTIPLE UPPER OUTLIERS IN DISTRIBUTION SAMPLES: A STUDY OF FOREIGN EXCHANGE DATA

Adesina O. S.
Department of Mathematical Science Olabisi Onabanjo University
Ago Iwoye
NIGERIA
Email: olumidestats@gmail.com
Ayoola F.J.
Department of Statistics, University of Ibadan,
NIGERIA
Email: ayoolafemi@yahoo.com
Dare R.J.
Department of Mathematical Science Olabisi Onabanjo University, Ago Iwoye
NIGERIA
Email:Remd_2005@yahoo.com

Abstract: *In this study, existences of k-upper outliers is investigated in distribution samples of gamma, Normal and exponential by carrying out simulation of ten thousand at different values of* n *using algorithm introduced by Tietjen-Moore, test statistics and critical values were equally estimated from the algorithm. A Normal Q-Q plot was made which aims at distinguishing a data set that follows a normal distribution and one that deviates from normality. The algorithm was applied to Nigeria-US dollars foreign exchange rate, both on raw and logarithmic transformed data. The simulation study reveals the existence upper outliers more in Gamma and exponential sample than the Normal sample. Empirical analysis shows that there are upper outliers in the raw data set but no upper outliers are found in the transformed data. The result in this paper would help researcher in business and economics to take time to explore data before use and properly transform accordingly to avoid error in estimation.*

Keywords: *upper outlier, Gamma, Exponential, Normal, Foreign exchange, discordancy, Tietjen-Moore*

1. INTRODUCTION

The application of Normal, Exponential and Gamma distribution cut across disciplines, the detection study of outlier on the sample of the distribution of utmost importance. Often, some observations from a set of data might appear to be inconsistence with the rest of the observations as a result of outlier Lalitha. S & kurmar. N (2012). Inclusion of the outliers may lead to model misspecification, biased parameter estimation and incorrect results. It is therefore important to identify outliers prior to modeling and analysis Liu et al (2004). Barnett and Lewis (1994) emphasized that in other to identify these outliers *discordancy* test needs to be performed.

Ben-Gal et al (2005), Outlier detection methods have been suggested for numerous applications, such as credit card fraud detection, clinical trials, voting irregularity analysis, data cleansing, network intrusion, severe weather prediction, geographic information systems, athlete performance analysis, and other data-mining tasks (Hawkins, 1980; Barnett and Lewis, 1994; Ruts and Rousseeuw, 1996; Fawcett and Provost, 1997; Johnson et al., 1998; Penny and Jolliffe, 2001; Acuna and Rodriguez, 2004; Lu et al., 2003).

Ben-Gal et al (2005), also stated that most of the earliest univariate methods for outlier detection rely on the assumption of knowingthe underlying distribution of the data. Moreover, many discordance tests for detecting univariate outliers further assume that the distribution parameters and the type of expected outliers are also known (Barnett and Lewis, 1994).

Certain figuresthat are considered as outliers sometimes represent actual values in a given sample, and maybe sending a warning signal to the user of such data. In the study byAdeleke et al (2015) where daily foreign exchange risk of Nigeria naira against nine other foreign currencies were forecasted using 'extreme value theory' where tail area of the distribution was modeled and risk measured accordingly.

In this study, we seek to identify the presence of k-upper outliers in exponential, Gamma and Normal samplesusing Tietjen-Moore test-statistics, also apply the same technique to foreign exchange data andexamine the difference between test statistics for samples that are approximately normally and the one that are not approximately normally distributed.

The underlying distributions are discussed as follows:

1.1 Discordancy Test in Normal distribution

The random variable x is said to have a Normal distribution (Gaussian) if its density function is

$$f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right),\ \sigma>0,\ -\infty<\mu<\infty \tag{1}$$

The normal distribution is an extremely important in many fields; it is applicable in physiological measurements of biological specimens, financial variables like exchange rates, counting problems that follow Binomial or Poisson random variables, light and thermal intensity and many more. It has relationship with distributions like Raleigh, Cauchy and lognormal distributions, Ugwuowo (2009).

According to Ben-Gal et al (2005), A central assumption in statistical-based methods for outlier detection, is a generating model that allows a small number of observations to be randomly sampled from distributions $G_1,........,G_k$, differing from the target distribution F, which is often taken to be a normal distribution $N(\mu,\sigma^2)$. The outlier identification problem is then translated to the problem of identifying those observations that lie in a so-called outlier region. This leads to the following definition (Davies and Gather, 1993): For any confidence coefficient $\alpha, 0<\alpha<1$, the α-outlier region of the $N(\mu,\sigma^2)$ distribution is defined by

$$out(\alpha,\mu,\sigma^2)=\left\{x:|x-\mu|>Z_{1-\alpha}\big/2^{\sigma}\right\} \tag{2}$$

where Z_q is the q quintile of the $N(0,1)$. A number x is an α-outlier with respect to F if $x\in out(\alpha,\mu,\sigma^2)$. Although, traditionally the normal distribution has been used as the target distribution, this definition can be easily extended to any unimodal symmetric distribution with positive density function, including the multivariate case Ben-Gal et al (2005).

To carry out discordancy test in Normal sample, we assume a univariate data set of n observations represented by $x_1,x_2,.................x_{n-2},x_{n-1}$ order statistics where $x_{(1)}$ is the lowest observation and $x_{(n)}$ the highest observation. Surrendra et al (2006) while considering six Dixon type $N7$ and $N9-N13$, identifies that Tests $N7, N9$, and $N13$ are discordance tests for an extreme outlier ($x_{(n)}$ or $x_{(1)}$) in a normal sample with population variance (σ^2) unknown, whereas tests $N11-N13$ are for two extreme observations (either the upper-pair $x_{(n)}$, $x_{(n-1)}$ or the lower-pair $x_{(1)}$, $x_{(2)}$) in a similar normal sample. The corresponding test statistics for example, the test statistic for test $N9$ is

$$TN9_u = {(x_{(2)} - x_{(1)})}\Big/{(x_{(n)} - x_{(2)})} \qquad (3)$$

To perform the discordancy test k upper outliers, we compute $TN9_u$ in equation (3). It is said that the value $x_{(n)}$ is under evaluation, i.e., tested to see if it was drawn from the same normal population as the rest of the sample (null hypothesis H_0), or it came from a different normal sample (with a different mean or a different variance or both), i.e., if it happens to be a discordant outlier (alternate hypothesis H_1). The computed value of test statistic $TN9_u$ is then compared with the critical value (percentage point) for a given number of observations n and at a given significance level (SL), in this study used SL of 5% and 1% was used.

If computed $TN9_u$ is less than the critical value at a given confidence level, H_0 is said to be true at that particular confidence level, i.e., there is no outlier at the chosen confidence level. But if computed $TN9_u$ is greater than the respective critical value at a given confidence level, H_0 is said to be false and, consequently, H_1 is said to be true at that particular confidence level, i.e., the observation tested $x_{(n)}$) by $TN9_u$ is detected as a discordant outlier which can then be discarded, and the test applied consecutively for other extreme values until H_0 is true.

1.2 Discordancy Test in Exponential Distribution

The exponential distribution with scale parameter, θ, is the distribution with probability density Function

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \; x > 0, \theta > 0 \qquad (4)$$

The exponential distribution has many applications in queuing theory life-testing experiment and reliability engineering, other areas of applications of the exponential distribution are stated in Ugwuowo (2009)

According to Lalitha. S &kumar . N (2012), let $X_1, \ldots\ldots\ldots\ldots X_n$ be a random sample from an exponential distribution $f(x|\theta)$ and its corresponding order statistics be $X_{(1)} \le \ldots\ldots\ldots\ldots \le X_{(n)}$. To perform the discordancy test for *k* upper outliers, we set up a null hypothesis H_0 that all the observations are coming from an exponential distribution $f(x|\theta)$ against the slippage alternative, H_k, that *(n − k)* observations are from this population but *k*values are from a $f(x|b\theta)$, *(b >*1*)* population. In fact, the choice of *k* in multiple outlier problems is crucial. They emphasized that improper choice of *k* may give misleading results and the problem of deciding on the number of outliers in a sample has been considered by various authors such as Jain and Pingel(1981), Kale (1976), Rosner (1975) and Tietjen and Moore (1972).

According to Barnett and Lewis (1994), various discordancy tests for single and multiple outliers have been proposed for exponential samples. Likes (1966) proposed a Dixon-type statistic

$$D_k = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(1)}} \qquad (5)$$

A large value of the test statistic D_k signifies the presence of k-upper outliers in the sample.

Zerbet and Nikulin (2003) also proposed test statistic for identifying outliers as follows:

$$T_k = \frac{X_{(n-k)} - X_{(1)}}{\sum_{j=n-k+1}^{n} X_{(j)} - X_{(1)}} \qquad (6)$$

A smaller value of T_k indicates the presence of outlier in the sample.

A popular test statistic which is also used for testing upper outliers is the maximum likelihood ratio test given by

$$L_k = \frac{\sum\limits_{j=n-k+1}^{n} X_{(j)}}{\sum_{j=1}^{n} X_{(j)}} \qquad (7)$$

If L_k is greater than specified value, the test indicates the presence of outliers.

Lalitha. S &Kumar . N (2012), used the test statistics proposed by gap-family to test for multiple outlier in exponential sample

The test statistic for k-upper outliers may be defined as:

$$Z_k = \frac{X_{(n)} - X_{(n-k)}}{S_n} \qquad (8)$$

A larger value of Z_k will indicate the presence of k upper outlier in the sample. Therefore, the null hypothesis is rejected when $Z_k > z_k(\alpha)$, where $z_k(\alpha)$ is the critical value at α level of significance. The exact null distribution Z_k for $k \geq 2$ is rather complex. However, the critical value $z_k(\alpha)$ of test for $k \geq 1$ are found to be very close to:

$$\hat{Z}_{(\alpha)} = 1 - \prod_{j=1}^{K} \left\{ \frac{(j-1+\alpha}{j} \right\}^{\left(\frac{1}{n-1}\right)} \qquad (9)$$

1.3 Discordancy Test in Gamma Distribution

Gamma distribution is also an important distribution, one of its application which is modelling waiting times between Poisson distributed events.

JabbariNooghabi et al. (2010) and Kumar and Lalhita (2012), shows that If F is a Gamma distribution then H_0: $X_1, ...X_n$ are n independent random variables, each following a Gamma distribution with shape parameter $m > 0$ and scale parameter $\sigma > 0$, denoted by $\Gamma(m, \sigma)$,

whose probability density function (pdf) is given by

$$f(x; m, \sigma) = \frac{1}{\Gamma(m)\sigma^m} x^{m-1} \exp\left(\frac{-x}{\sigma}\right), x > 0 \qquad (10)$$

However, Lucini, M.M and Frery A. C (2015) identifies that results presented by JabbariNooghabi et al. (2010) do not hold in all expected cases. With this, the technique proposed by Kumar and Lalhita (2012) for detecting upper outliers in Gamma samples is also not valid. Specifically, the note shows that the probability density functions (pdf) under the null hypothesis of the test statistics therein proposed are not always valid. Both authors proposed tests statistics to detect outliers in Gamma samples using a test of discordancy for outliers framework as defined by Barnnett and Lewis (1994). Lucini, M.M and Frery A. C (2015)pointed out that equation (10) will be assumed that these random variables are distributed according to a $\Gamma(m,1)$ law, that is, with pdf given by

$$f(x;m) = \frac{1}{\Gamma(m)} x^{m-1} \exp(-x), x > 0 \qquad (11)$$

The alternative hypothesis used in JabbariNooghabi et al. (2010) and Kumar and Lalhita (2012) is the slippage alternative. We are interested in detecting $1 \le k \le n$ upper outliers using Z_k, the statistic proposed by Kumar and Lalhita (2012). This statistic, after some computations, can be written as

$$Z_k = \frac{\sum_{j=n-k+1}^{n} (n-j+1)Y_j}{\sum_{j=2}^{n} (n-j+1)Y_j} \qquad (12)$$

Where

$$Y_j = X_{(j)} - X_{(j-1)} \qquad (13)$$

$X_{(j)}$ denotes the j-th order statistics of the ordered sample from $(X_i)_{(1\le i\le n)}$ in non-decreas ing order, that is, $X_{(1)} \le X_{(2)} .... \le X_{(n)}$, and k is the number of observations suspected to be upper outliers.

A strong assumption made in both works is that, under the null hypothesis, each Y_j follows a $\Gamma(m,(n-j+1)^{-1}$ distribution. This is not true when $m \neq 1$.See work done by Lucini, M.M and Frery A. C (2015) for detailedcomments on “Detecting Outliers in Gamma Distribution" by M. JabbariNooghabi et al. (2010).

1.4 Tietjen-Moore Test Statistics

Tietjen-Moore test-statistics reduces to Grubb’s test if k=1, Grubbs' test (Grubbs 1969and Beck 1972), is used to detect a single outlier in a univariate data set that follows an approximately normal distribution. Grubbs' test is also known as the maximum normed residual test.

The Grubbs' test statistic is defined as:

$$G = \frac{|Y_t - \bar{Y}|}{s}$$

with $\bar{Y}$ and $\boldsymbol{s}$ denoting the sample mean and standard deviation, respectively. The Grubbs' test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation.

This is the two-sided version of the test. The Grubbs' test can also be defined as one of the following one-sided tests:

i. test whether the minimum value is an outlier

$$G = \frac{\overline{Y} - Y_{\min}}{s}$$

with Y_{min} denoting the minimum value.

ii. test whether the maximum value is an outlier

$$G = \frac{Y_{max} - \overline{Y}}{s}$$

with Y_{max} denoting the maximum value.

For the two-sided test, the hypothesis of no outliers is rejected if

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{(t_{\alpha}/(2N), N-2)^2}{N-2+(t_{\alpha}/(2N), N-2)^2}}$$

with tα/(2N),N−2 denoting the critical value of the *t* distribution with (*N*-2) degrees of freedom and a significance level of *α*/(2*N*).For one-sided tests, we use a significance level of level of *α*/*N*.

Grubbs' test is defined for the hypothesis:H_0:There are no outliers in the data set and H_a:There is exactly one outlier in the data set.

Tietjen-Moore is a general Grubb’s test when more than one upper outlier is suspected in a given data set. Steps in computing Tietjen-Moore test statistics are considered as follows:

Sort the n data points from smallest to the largest so that y_i denotes the *ith* largest data value. The test statistic for the k largest points is

$$L'_k = \frac{\sum_{i=1}^{n} (y_i - \overline{y}_k)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \qquad (14)$$

with $\overline{y}$ denoting the sample mean for the full sample and $\overline{y}_k$ denoting the sample mean with the largest k points removed. The test statistic for the k smallest points is

$$L''_k = \frac{\sum_{i=k+1}^{n} (y_i - \overline{y}_k)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \qquad (15)$$

with $\overline{y}$ denoting the sample mean for the full sample and $\overline{y}_k$ denoting the sample mean with the smallest k points removed.

To test for outliers in both tails, compute the absolute residuals

$$r_i = | y_i - \overline{y} | \qquad (16)$$

and then let z_i denote the y_i values sorted by their absolute residuals in ascending order.

The test statistics for the case is:

$$E_k = \frac{\sum_{i=k+1}^{n} (z_i - \overline{z}_k)^2}{\sum_{i=1}^{n} (z_i - \overline{z})^2} \qquad (17)$$

with z_i denoting the sample mean for the full data set and $\bar{z}_k$ denoting the sample mean with the largest k points removed. We reject H_0 if $E_k < E_{critical}$

2. METHODOLOGY

Tietjen-Moore test statistics is used for the sample of the three underlying distributions to get the test statistics and the critical value for 10, 000 simulations, k-upper outliers were also examined using Nigeria foreign exchange rate against the US dollars following the same algorithm. The simulation and empirical study was carried out using R 3.2.5 the test statistics and critical value were computed for each value of n .The null hypothesis H_0 for Tietjen-Moore test is defined as there are no outliers in the data set, while the alternative hypothesis H_a , there are exactly k outliers in the data set.

Algorithm for Tietjen-Moore is as follows:
(i) Create a function to compute statistic to (ii) Compute the absolute residuals (iii) Sort data according to size of residual (iv) Create a subset of the data without the largest k values (v) Compute the sums of squares (vi) Compute the test statistic (vii) Call the function and compute value of test statistic for data (viii) Compute critical value based on simulation.

The Tietjen-Moore test is a lower, one-tailed test, so we reject the null hypothesis that there are no outliers when the value of the test statistic is less than the critical value.

3. SIMULATION STUDY

The critical region for the Tietjen-Moore test is determined by simulation. The simulation is performed by generating a standard normal random sample of size n and computing the Tietjen-Moore test statistic. Typically, 10,000 random samples are used. The value of the Tietjen-Moore statistic obtained from the data is compared to this reference distribution. The value of the test statistic is between zero and one. If there are no outliers in the data, the test statistic is close to 1. If there are outliers in the data, the test statistic will be closer to zero. Thus, the test is always a lower, one-tailed test regardless of which test statistics is used, L_k or E_k .

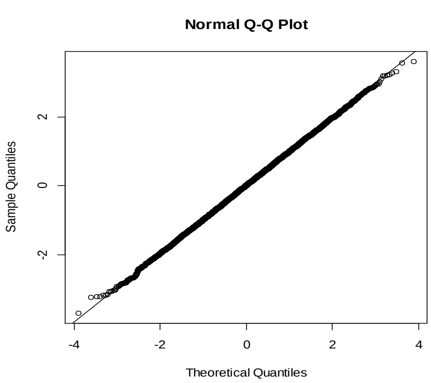

Fig 1: Q-Q fo Normal Sample

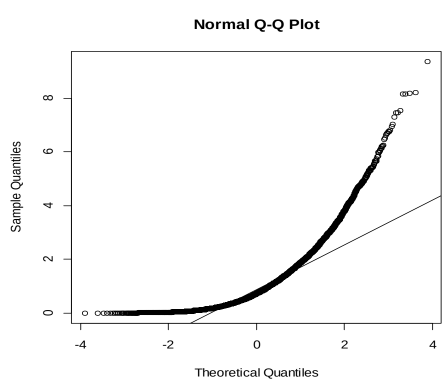

Fig 2: Q-Q for Exponential Sample

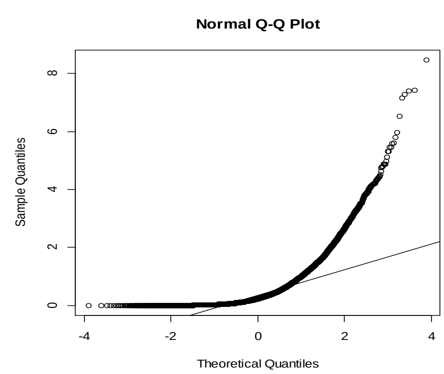

Fig 3: Q-Q for Gamma S Sample

Table 1: Simulated and approximated test statistics when $k = 2,3,4$ for Exponential, Gamma and Normal sample using E_k as test statistics.

n	$E_{E\text{xp}_2}$	E_{Gam_2}	E_{Norm_2}	$E_{E\text{xp}_3}$	E_{Gam3}	E_{Norm3}	$E_{E\text{xp}_4}$	E_{Gam_4}	E_{Norm_4}
10	0.2114	0.1913	0.2177	0.2709	0.1861	0.3485	0.1404	0.1058	0.1868
20	0.4048	0.5039	0.6441	0.3270	0.4684	0.6091	0.3519	0.2856	0.4641
30	0.4026	0.5800	0.5512	0.2824	0.1016	0.6046	0.3481	0.3278	0.4572
40	0.3879	0.4513	0.7662	0.1476	0.3944	0.6599	0.3786	0.4095	0.5240
50	0.4173	0.6162	0.8176	0.5851	0.6492	0.7186	0.5144	0.5282	0.5918
100	0.7095	0.8375	0.8344	0.7131	0.6520	0.8203	0.5412	0.6570	0.7985
200	0.7436	0.8405	0.9191	0.7009	0.8229	0.8895	0.7340	0.7785	0.8739
250	0.8232	0.8567	0.9346	0.8030	0.8231	0.9001	0.7207	0.7643	0.8599
500	0.9194	0.9098	0.9595	0.8575	0.8848	0.9347	0.7647	0.8386	0.9292
1000	0.9423	0.9539	0.9759	0.9215	0.9100	0.9667	0.8597	0.91789	0.9607
5000	0.9654	0.9879	0.9949	0.9651	0.9691	0.9921	0.9514	0.9739	0.9903
10000	0.9855	0.9887	0.9963	0.9831	0.9811	0.9955	0.9789	0.9839	0.9944

Table 2: Simulated and approximated critical values for 5% and 1%,when $k = 2,3,4$ for Exponential, Gamma and Normal sample.

n	$E^*_{E_2}(\alpha)$	$E^*_{G_2}(\alpha)$	$E^*_{N_2}(\alpha)$	$E^*_{E_3}(\alpha)$	$E^*_{G_3}(\alpha)$	$E^*_{N_3}(\alpha)$	$E^*_{E_4}(\alpha)$	$E^*_{G_4}(\alpha)$	$E^*_{N_4}(\alpha)$
10	0.0567 0.0247	0.0939 0.0491	0.1666 0.0984	0.0353 0.0151	0.0561 0.0266	0.0812 0.0446	0.0217 0.0082	0.0296 0.0139	0.0366 0.0164
20	0.1580 0.0923	0.2346 0.1517	0.4206 0.3399	0.1082 0.0664	0.1711 0.1087	0.3028 0.2360	0.0843 0.0515	0.1316 0.0828	0.2226 0.1677
30	0.2482 0.1552	0.3331 0.2414	0.5480 0.4876	0.1688 0.1094	0.2557 0.1865	0.4425 0.3816	0.1347 0.0903	0.2023 0.1417	0.3613 0.3079
40	0.3117 0.2185	0.4018 0.3095	0.6285 0.5669	0.0372 0.0159	0.3219 0.2424	0.5354 0.4794	0.1805 0.1250	0.2684 0.2006	0.4585 0.4061
50	0.3643 0.2620	0.4639 0.3663	0.6826 0.6322	0.2875 0.2023	0.3832 0.3027	0.5996 0.5504	0.2308 0.1684	0.3209 0.2543	0.5262 0.4778
100	0.5327 0.441	0.6225 0.5392	0.8122 0.7803	0.4541 0.3722	0.5494 0.4809	0.7543 0.7260	0.4014 0.3198	0.4924 0.4277	0.7035 0.6709

	9								
200	0.6775	0.7478	0.6079	0.6079	0.6910	0.8550	0.5609	0.6456	0.8236
	0.5964	0.6872	0.5309	0.5309	0.6331	0.8350	0.5000	0.5922	0.8043
250	0.7150	0.7828	0.6580	0.6581	0.7319	0.8783	0.6071	0.6884	0.8507
	0.6444	0.7255	0.5850	0.5850	0.6813	0.8636	0.5507	0.6409	0.8343
500	0.8142	0.8630	0.7716	0.7716	0.8294	0.9310	0.7353	0.7979	0.9148
	0.7674	0.8262	0.7211	0.7211	0.7950	0.9218	0.6877	0.7652	0.9061
1000	0.8845	0.9177	0.8556	0.8555	0.8936	0.9616	0.8292	0.8739	0.9519
	0.8580	0.8996	0.8279	0.8279	0.8726	0.9576	0.7978	0.8532	0.9475
5000	0.9654	0.9761	0.9931	0.9546	0.9681	0.9904	0.9451	0.9777	0.9879
	0.9579	0.9713	0.9923	0.9461	0.9626	0.9896	0.9363	0.9745	0.9868
10000	0.9799	0.9863	0.9732	0.9732	0.9818	0.9948	0.9679	0.9777	0.9934
	0.9754	0.9834	0.9686	0.9686	0.9784	0.9944	0.9632	0.9745	0.9929

Note: $E^{*}_{E_k}(\alpha)$, $E^{*}_{G_k}(\alpha)$, and $E^{*}_{N_k}(\alpha)$ are simulated and approximate critical values for Exponential, Gamma and Normal respectively.
In each two-entry, the first line is the critical values for $\alpha = 0.05$ and the second line is the critical values for $\alpha = 0.01$. We reject H_0 if $E_k < E_{critical}$

4. EMPIRICAL APPLICATION

4.1 Data Description

In this study, daily official central foreign exchange data were collected from Central Bank of Nigeria (CBN) official website, the sample period is from, October 12, 2001 to March 8 2016; this represents 3487. Data set for of three currency exchange rates against Nigeria Naira. Three foreign exchange rates were considered against the NGN which include US Dollars. If S_t be the exchange rate of USD against NGN and let $r_t = \log\left(S_t / S_{t-1}\right)$, then the exchange rate of NGN against USD is $1/S_t$. This will yield the log returns $r_t = \log\left(\frac{1/S_t}{1/S_{t-1}}\right)$., it is observed from the normal Q-Q plots that he data is not normally distributed but it was closed to normal after the logarithm transformation.

Table 3: result of descriptive of the Returns of US Dollars against Nigeria Naira

	Raw	Log return
Min	0.005038	-0.56340
Max	0.015130	0.56340
Mean	0.007145	0.00016
Std.dev	0.000963	0.19137
Skew	-0.06282	0.19137
Kurtosis	1.305108	800.516
Q_1	0.006442	0.00000
Q_2	0.006897	0.00000
Q_3	0.007895	0.00000

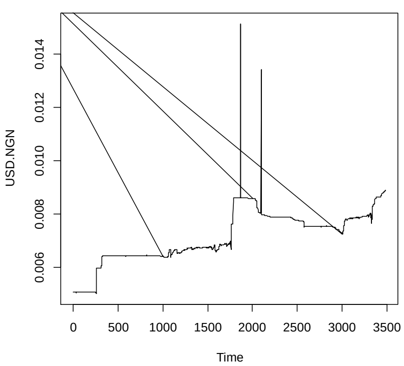

Fig 4: Time series plot of USD-NGN exchange rate

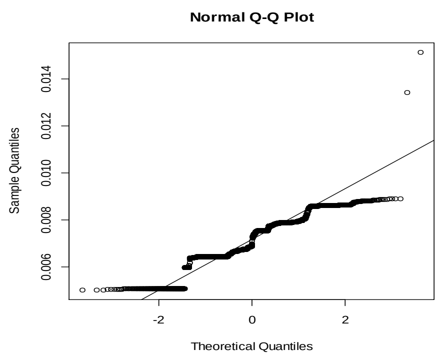

Fig 5: Q-Q normal plot of USD-NGN raw exchange rate

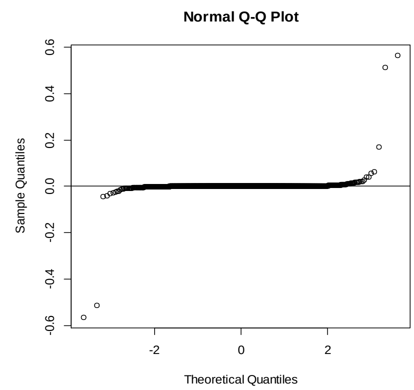

Fig 5: Q-Q normal plot of USD-NGN log transformed exchange rate

Table 4: Result beforelogarithm transformation

n	$E_{k=2}$	$E_{k=3}$	$E_{k=4}$	$E^*_{\alpha,k=2}$	$E^*_{\alpha,k=3}$	$E^*_{\alpha,k=4}$
10	0.9999907	0.9999864	0.9999822	1.00000	1.00000	1.00000
				1.00000	1.00000	1.00000
20	0.9999811	0.9999717	0.9999623	1	0.9999999	0.9999999
				1	0.9999999	0.9999999
50	0.9999974	0.9999609	0.9999479	0.9999997	0.9999996	0.9999995
				0.9999997	0.9999996	0.9999995
100	0.9999558	0.9999338	0.9999116	0.9999989	0.9999984	0.9999979
				0.9999989	0.9999984	0.9999979
200	0.9999424	0.9999135	0.9998847	0.9999986	0.9999949	0.9999932
				0.9999986	0.9999949	0.9999932
500	0.9998968	0.9998452	0.9997936	0.9999825	0.9999974	0.9999655
				0.9999825	0.9999974	0.9999655
1000	0.9998531	0.9997799	0.9997072	0.9999264	0.9998897	0.9998531
				0.9999264	0.9998897	0.9999907
2000	0.9996886	0.9995328	0.9993768	0.9996546	0.999482	0.9993101
			0.9993768	0.9996546	0.999482	0.9993101
3000	0.9986702	0.9980048	0.9973389	0.9987269	0.9980912	0.9974561
				0.9987269	0.9980912	0.9974561

Note: The critical values at 1% and 5% are recorded at first and second entry respectively in columns 5, 6 and 7 while Tietjen-Moore test statistics are computed in the first three columns.

Table 5: Result after logarithm transformation

n	$E_{k=2}$	$E_{k=3}$	$E_{k=4}$	$E^*_{\alpha,k=2}$	$E^*_{\alpha,k=3}$	$E^*_{\alpha,k=4}$
10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
			1.00000	1.00000	1.00000	1.00000
20	1.00000	1.00000	1.00000	1.00000	0.9999999	0.9999999
			1.00000	1.00000	0.9999999	0.9999999
50	1.00000	1.00000	0.9999999	0.9999997	0.9999996	0.9999995
				0.9999997	0.9999996	0.9999995
100	1.00000	1.00000	0.9999999	0.9999989	0.9999984	0.9999979
				0.9999989	0.9999984	0.9999979
200	1.00000	0.9999999	0.9999999	0.9999986	0.9999949	0.9999932
				0.9999986	0.9999949	0.9999932
500	1.00000	0.9999999	0.9999999	0.9999825	0.9999974	0.9999655
				0.9999825	0.9999974	0.9999655
1000	1.00000	0.9999999	0.9999999	0.9999264	0.9998897	0.9998531
				0.9999264	0.9998897	0.9999907
2000	1.00000	0.9999999	0.9999999	0.9996546	0.999482	0.9993101
				0.9996546	0.999482	0.9993101
3000	0.9999997	0.9999996	0.9999996	0.9987269	0.9980912	0.9974561
				0.9987269	0.9980912	0.9974561

Note: The critical values at 1% and 5% are recorded at first and second entry respectively in columns 5, 6 and 7 while Tietjen-Moore test statistics are computed in the first three columns.

5. DISCUSSION AND RECOMMENDATION

In this study algorithm from Tietjen-Moore test statistics was carried out after 10, 000 replications of Normal, Exponential and Gamma sample has been implemented in the R programming language, R core team (2016). The critical values were also gotten from the simulated sample at specified value of n. Almost all points falling on a straight line on normal Q-Qplots for Normal sample reveals how plot of a data set that is approximately normally distributed should look like, exponential and gamma sample deviates from this.

Plots of raw exchange rate deviates from normality as much as descriptive statistics shows in table 3 with skewness and kurtosis. However, normal Q-Q plots for the logarithm transformed data shows that the data is normally distributed.

Table 4 shows test statistics of Tietjen-Moore (E_k) at k =2,3 and 4 and the critical values are equally computed. It is observed that $E_k < E^*_{\alpha i}$ where i=2, 3 and 4 at all values of n. therefore is k-upper outliers at k =2,3, and 4 at all values of n observed.

Table 5 above shows test statistics of Tietjen-Moore (E_k) at k=2,3 and 4 and the critical values are equally computed. It is observed that $E_k > E^*_{\alpha i}$ at $n > 50$ and $E_k \le E^*_{\alpha i}$ at $n \le 20$ where i=2, 3 and 4. Therefore is no k-upper outliers at k=2,3, and 4 at all values of n observed when the data is logarithmic transformed.

Therefore, it is recommended that proper evaluation of nature of data be done so as not to have error in estimation or forecast.

REFERENCES

Acuna E., Rodriguez C. A (2004) ., ”Meta analysis study of outlier detection methods in classification,” Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez, Retrived from academic.uprm.edu/ eacuna/ paperout.pdf. In proceedings IPSI 2004, Venice.

Adeleke, I, Adesina, S. O &Oladeji T. F (2015). Estimating Nigerian Foreign Exchange Risk with Extreme Value Theory. Botswana Journal of Business Volume 8, No 1, pp. 61-80

Barnett, V., Lewis, T., (1994), Outliers in Statistical Data. Third edition: Chichester, John Wiley, 584 p.

Ben-Gal (2005). Outlier detection, In: Maimon O. and Rockach L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers," Kluwer Academic Publishers, 2005, ISBN 0-387-24435-2.

David H. A and Nagaraja H. N (2003), Order Statistics, Wiley, NewYork,

Dixon, W.J., (1950), Analysis of extreme values: Annals of Mathematical Statistics, 21(4), 488–506.

Grubbs, F. E. (1969), Procedures for detecting outlying observations in samples, Technometrics, 1(1), 1–21.

Grubbs, F. E., and G. Beck (1972), Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics, 14(4), 847–854.

Hawkins D. M (1972), Analysis of a slippage test for the chi squared distribution, South Afr. Stat. J. 6 (1972), pp. 11–17.

Jain R. B and Pingel L.A (1981), A procedure for estimating the number of outliers, Commun. Statist.Theor.Meth. 10 (1981), pp. 1029–1041.
JabbariNooghabi, M., H. JabbariNooghabi, and Nasiri. P (2010), Detecting outliers in gamma distribution." Communications in Statistics-Theory and Methods, 39, 698{706
Kale B. K (1976), Detection of outliers, Sankhya B 38, pp. 356–363
Kumar, N. and S. Lalhita (2012), \Testing for upper outliers in gamma sample" Communications in Statistics-Theory and Methods, 41, 820{828.
Likes. J (1966), Distribution of Dixon’s statistics in the case of an exponential population, Metrika 11, pp. 46–54.
Lalitha .S & Kumar. N (2012): Multiple outlier test for upper outliers in an exponential sample, Journal of Applied Statistics, 39:6, 1323-133
Liu. H, Shah. S, and Jiang (2014), On-line outlier detection and data cleaning, Comput. Chem. Eng. 28, pp. 1635–1647.
Lu C., Chen D., Kou Y. (2003), ”Algorithms for spatial outlier detection,” In Proceedings of the 3rd IEEE International Conference on Data-mining (ICDM’03), Melbourne, FL
Lucini M. M Alejandro Frery A. C (2015). Comments on \Detecting Outliers in Gamma Distribution" by M.JabbariNooghabi et al. (2010) Provost F. and Fawcett. T (1997) Analysis and visualization of classifier performance: Comparison under imprecise class and cost distribution. In Proceeding of Knowledge discovery and Data Mining pp. 43-48
R Core Team (2016), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rosner. B (1975), On the detection of many outliers, Technometrics 17 (1975), pp. 221–227.
N.L. Johnson N. L, Kotz S, and N. Balakrishnan N (1994), Continuous Univariate Distributions, Wiley, NewYork
Ruts I., Rousseeuw P. (1996), ”Computing Depth Contours of Bivariate Point Clouds,” In Computational Statistics and Data Analysis, 23,153-168.
Surendra P. V,*, Alfredo Q, and Lorena D-G (2006). Critical values for 33 discordancy test variants for outliers in normal samples up to sizes 1000, and applications in quality control in Earth Sciences. Revista Mexicana de CienciasGeológicas, v. 25, núm. 1, 2008, p. 82-96
Tietjen G. L and R.H (1972). Moore, Some Grubbs-type statistics for the detection of several outliers, Technometrics 14, pp. 583–597.
Woolley T W. Jr (2010). The Effect Of Swamping On Outlier Detection In Normal Samples. Joint Statistical Meetings - Biometrics Section-to include ENAR & WNAR pp. 3793-3798
Zerbet A and Nikulin (2003). M, A new statistic for detecting outliers in exponential case, Commun.Statist.Theor. Meth.32, pp. 573–583
Zerbet A and M. Nikulin. M (2003), A new statistic for detecting outliers in exponential case,Commun. Statist.Theor.Meth. 32, pp. 573–583.