Distribution of Nairobi Stock Exchange 20 Share Index Returns: 1998-2011

Cox Lwaka Tamba *clwaka@yahoo.com*
Faculty of Science/Department of Mathematics
Division of Statistics
Egerton University
P.o. Box 536-20115, Egerton, Kenya

Dickson Siele Cheruiyot *dicbelio09@gmail.com*
Faculty of Science/Department of Mathematics
Division of Statistics
Egerton University
P.o. Box 536-20115, Egerton, Kenya

Martin Wafula Nandelenga *nmartin200884@gmail.com*
Faculty of Arts and Social Sciences
Department of Economics
Egerton University
P.o. Box 536, Egerton, Kenya

Abstract

The assumption that returns of daily share index prices are normally distributed has long been disputed by the data. In this paper, the normality assumption has been tested using time series data of daily Nairobi Stock Exchange 20-Share Index (NSE 20-Share Index) for the period 1998-2011. It has been confirmed that the share price index returns does not follow the normal distribution. Other symmetrical distributions have been fit to the data i.e. logistic distribution and t-location scale distribution. With the aid of a programming language; Matlab we have computed the various Maximum Likelihood (ML) estimates from this distributions and tested how well they fit to the data. It has been established that the NSE 20 Share Index returns follows a t-location scale distribution (normal inverse gamma mixture). We recommend that since we have found that the normal inverse gamma mixture best fits the NSE 20 Share Index return, other normal mixtures can be investigated how well they fit this data.

Keywords: Returns, Leptokurtosis, Tail, 20 Share Index, t-location Scale.

1. INTRODUCTION

A stock market index is a measure of changes in the stocks markets and is usually considered to be a reasonably representative of the market as a whole. Indexes are usually tabulated on a daily basis and involve summarizing sample share price movements (NSE 20 share index) or all the share prices movements (NSE All Share Index, (NASI)). The NSE 20 Share index measures the average performance of 20 large cap stocks drawn from different industries. Stock prices and returns have been assumed to be normally distributed.

The first complete development of a theory of random walks in security prices is due to Bachelier (1900), whose original work first appeared around the turn of the century. Unfortunately his work did not receive much attention from economists. The Bachelier (1900) model assumed that under the central-limit theorem the daily, weekly, and monthly price changes will each have normal or Gaussian distributions. However it has been found out that most of the distributions of price changes are leptokurtic; that is, there are too many values near the mean and too many out in the extreme tails.

Mandelbrot (1962) asserts that, in the past, academic research has too readily neglected the implications of the leptokurtosis usually observed in empirical distributions of price changes. Mandelbrot (1963) studied the stable Paretian distribution and showed that it is the distribution of price changes provided the characteristic exponent is less than two.

Fama (1965) discussed first in more detail the theory underlying the random-walk model and then tested the model's empirical validity. The past behaviour of a security's price is rich in information concerning its future behaviour (Fama, 1965). The theory of random walk shows that the successive price changes are independent, identically distributed random variables. Most simply this implies that the series of price changes has no memory, that is, the past cannot be used to predict the future in any meaningful way. The probability distribution of the price changes during time period *t* is independent of the sequence of price changes during the previous time periods i.e,

$$\Pr\left(S_t = \mathrm{s}|\mathrm{S}_{t-1},\ \mathrm{s}_{t-1},\ \ldots\right) = \Pr\left(S_t = \mathrm{s}\right) \tag{1}$$

The actual tests were not performed on the daily prices themselves but on the first differences of their natural logarithms. The variable of interest was

$$u_{t+1} = \log_e S_{t+1} - \log_e S_t\,, \tag{2}$$

where S_{t+1} is the price of the security at the end of day *t+1*.

Fama (1965) demonstrated that first differences of stock prices seem to follow stable Paretian distributions with characteristic exponent is less than two.

According to Officer, (1972), the distribution of stock returns has some characteristics of a non-normal generating process i.e. the results indicate the distribution is "fat- tailed" relative to a normal distribution. However, characteristics were also observed which are inconsistent with a stable non-normal generating process. Evidence is presented illustrating a tendency for longitudinal sums of daily stock returns to become "thinner-tailed" for larger sums, but not to the extent that a normal distribution approximates the distribution. This confirms that the normal distribution is not a good fit for the distribution of stock and index prices.

Praetz (1972) presented both theoretical and empirical evidence about a probability distribution which describes the behaviour of share price changes. Osborne's Brownian motion theory of share price changes was modified to account for the changing variance of the share market. This produced a scaled t-distribution which is an excellent fit to series of share price indices. This distribution was the only known simple distribution to fit changes in share prices. It provided a far better fit to the data than the stable Paretian, compound process, and normal distributions (Praetz, 1972).

The Student and symmetric-stable distributions, as models for daily rates of return on common stocks, have been discussed and empirically evaluated (Blattberg, 1974). Both models were derived using the framework of subordinated stochastic processes. Some important theoretical and empirical implications of these models were also discussed. The descriptive validity of each model, relative to the other, was assessed by applying each model to actual daily rates of return. Interpretations of empirical results were guided by results from a Monte Carlo investigation of the properties of estimators and model-comparison methods. The major inference of this report was that, for daily rates of return, the Student model has greater descriptive validity than the symmetric-stable model.

Hung et al, (2007) studied the variation of Taiwan stock market using the statistical methods developed by econophysicists. The Taiwan market was found to have a fat tail as found in the markets of other countries, but it did not follow a power law as the others. The cumulative distribution of daily returns in Taiwan stock index could be fitted quite well using the log-normal distribution, and even better by a power law with an exponential cut-off. They believed that the distinct behaviour of Taiwan market was mainly due to the protective measures taken by the government.

Barndorff, (1977) introduced a family of continuous type distributions such that the logarithm of the probability (density) function is a hyperbola (or, in several dimensions, a hyperboloid). The focus was on the mass-size distribution of aeolian sand deposits however, this distribution has been widely used to model stock returns. Barndorff (1978) discussed the generalised hyperbolic distributions which includes the hyperbolic distributions and some distributions which induce distributions on hyperbolae or hyperboloids analogous to the von Mises-Fisher distributions on spheres. It is, among other things, shown that distributions of this kind are a mixture of normal distributions. Eberlin (1995) based on a data set consisting of the daily prices of the 30 DAX share over three years period, investigated the distributional form of compound returns. A class of hyperbolic distribution was fit to the empirical returns with high accuracy. Hyperbolic distributions fit empirical return adequately, (Kuchler et al. (1999) and Bibby (2003)).

The logistic distribution is a general stochastic measurement model; it has been used in measuring risk incurred in financial assets returns, (Osu, 2010). It has been observed that the initial stage of growth of the worth of a business enterprise is approximately exponential. At a time the growth slows down. Osu, 2010 asserts that this could be due to diversification (investing in more than one stock), since the returns on different stocks do not move exactly in the way all the time. Ultimately the growth of the firm is stable and it may not be affected by risk since diversification reduces risk.

Stock returns turn out to be quite sensitive to the degree to which distributions are thick tailed and asymmetric. Lack of encoding information about asymmetry and leptokurtosis is a well-known drawback of the Normal distribution. This has led to a search for alternative distributions. In this work, we test the normality assumption on the NSE 20 Share Price Index. We also fit the logistic and the t- location scale distribution to find the best fit to this data.

2. MODELS OF NSE 20 SHARE INDEX RETURNS

In this section we present the distributions used in this study. The normal distribution is fully discussed in Section 2.1. Section 2.2 presents the logistic distribution whereas the t- location scale is presented in Section 2.3. In all this sections, we review the construction, properties and estimation of these distributions. It is important to mention that from this point onwards, the sample herein referred to consists of the NSE 20 Share Index for the year 1998 to 2011. These indices are published daily in the Kenya local newspapers. The series analyzed, is the series of returns, where returns are defined as,

$$R_t = 100\left(\log\, S_{t+1} - \log\, S_t\right). \tag{3}$$

The behavior of the NSE 20 Share Index considered during this period is shown below in Figure 1. The series analyzed for each market is the series of returns, where returns are defined as, given in Equation (3).

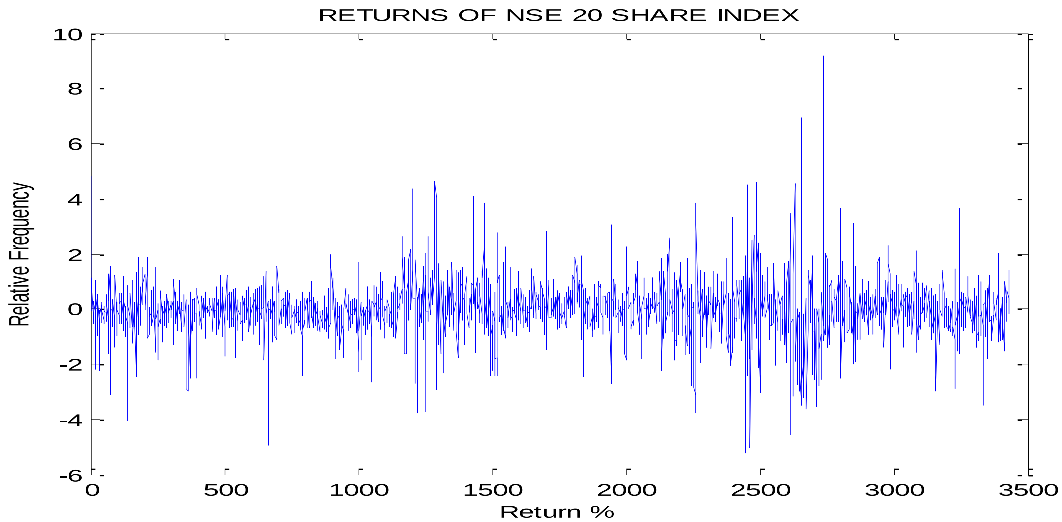

FIGURE 1: Percentage Returns, R_t of NSE 20 Share Index.

where R_t and S_t are the return and the index in day t, respectively. The histogram of this data is shown below;

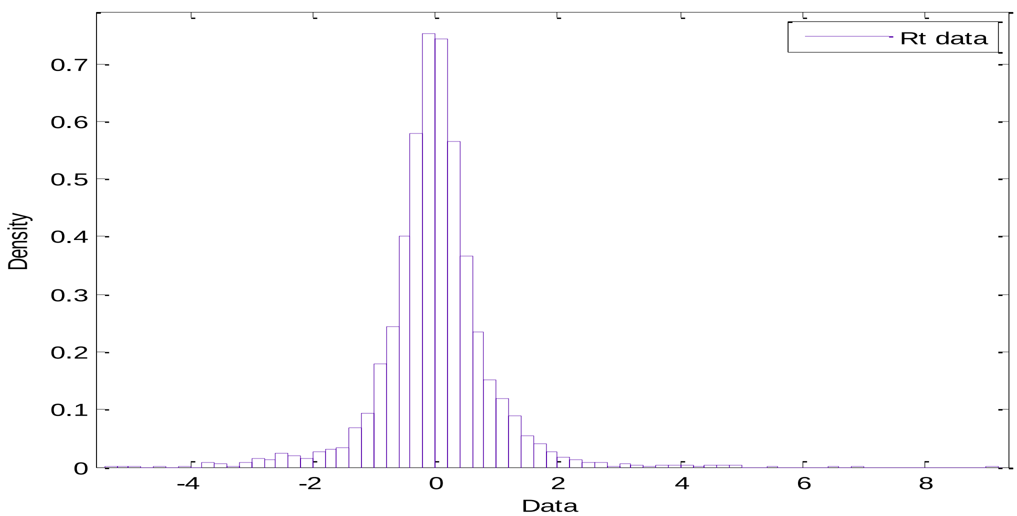

FIGURE 2: A Histogram of Returns Rt of NSE 20 Share Index.

Table 1 below summarizes some relevant information about the empirical distributions of stock returns under consideration. The statistics reported are the mean, standard deviation, minimum and maximum return during the sample period, coefficients of skewness and kurtosis.

Mean	Variance	Minimum Value	Maximum Value	Skewness	Kurtosis
7.9454×10^{-4}	0.7881	-5.2339	9.1782	0.5970	13.1608

TABLE 1: Sample Moments of the Distributions NSE 20 Share Index.

The third central moment is often called a measure of asymmetry or skewness in the distribution. From our descriptive results shown above, the distribution of this data is approximately symmetrical though it shows signs of being positively skewed. The coefficient of skewness is almost 0.5970. This is well depicted from the histogram in Figure 2 above.

Kurtosis measures a different type of departure from normality by indicating the extent of the peak (or the degree of flatness near its center) in a distribution. We see that this is the ratio of the fourth central moment divided by the square of the variance. If the distribution is normal, then this ratio is equal to 3. A ratio greater than 3 indicates more values in the neighborhood of the mean (is more peaked than the normal distribution). Our data has a fat tail or excess kurtosis as shown in the coefficient of kurtosis (i.e. kurtosis of 13.1608). The distribution of this data is leptokurtic.

2.1 Normality Assumption Test

A random variable is said to be normal distributed with mean, μ and variance, σ^2 (Mood, 2001) if its probability distribution is

$$f(r_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \left(\frac{\left(r_{t_i} - \mu\right)^2}{2\sigma^2} \right) \qquad for \quad -\infty \le r_{t_i} \le \infty, \mu \in \mathbb{R}, \sigma > 0. \qquad (4)$$

The maximum likelihood estimates of the normal distribution with their corresponding standard errors are shown below in Table 2 below;

Parameter	Estimate	Std error
μ	0.000794536	0.0151517
σ	0.887763	0.0107162

TABLE 2: ML Estimates of the Normal Distribution.

The distribution of financial returns over horizons shorter than a month is not well described by a normal distribution. In particular, the empirical return distributions, while unimodal and approximately symmetric, are typically found to exhibit considerable leptokurtosis. The typical shape of the return distribution, as compared to a fitted normal is presented in Figure 3 below.

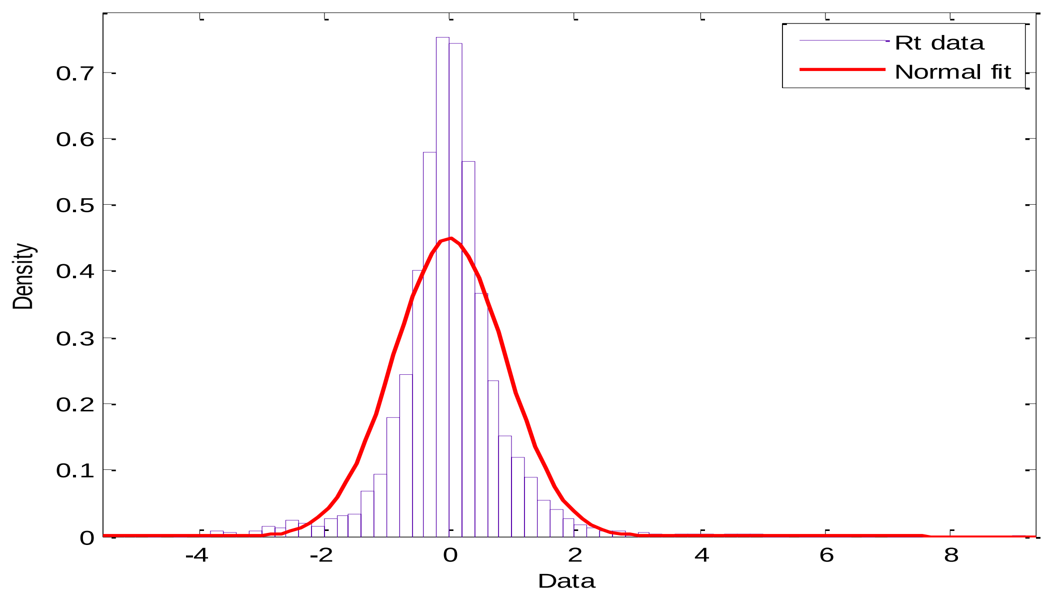

FIGURE 3: A normal Fit to the Returns R_t Data.

The coefficients of standardized skewness and kurtosis provide strong evidence about departures from normality, but more formal conclusions can be reached through the tests of normality reported below in Table 3.

	Ansari Bradley		Kolmogorov-Smirnov		Jarque-Bera	
Market	Statistic	P-value	Statistic	P-value	Statistic	P-value
NSE 20 Share Index	6516176	8.6874×10^{-052}	0.1181	4.0112×10^{-42}	$\mathbf{1.4972}\times10^{4}$	1.0×10^{-003}

TABLE 2: Goodness of Fit Tests for the Normal Distribution.

All these tests are done at the 5% significance level. The results in Table 3 above does not come as no surprise; virtually all studies that use daily data also reject the normality of stock returns. This is due to its failure to capture the fat tail of the returns data. Another reason why this model failed is due to the assumption that the variance σ^2 of price changes over unit time interval is a constant. This is clearly depicted when we plot a probability plot to compare the distribution of our data and the normal distribution as shown in Figure 4.

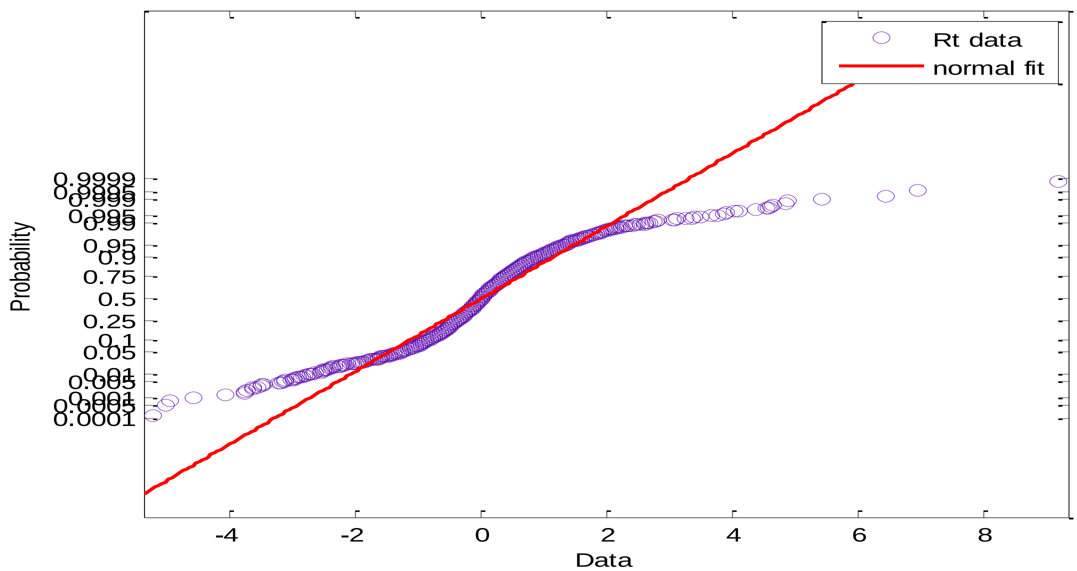

FIGURE 4: A Probability Plot of the Normal Fit.

The goodness of fit coupled with the probability plot give a clear evidence against the normal distribution as a fit of the Returns of NSE 20 Share Price Index. In order to test what specification describes the data better than the Normal distribution, we consider in the next part two alternative distributions that allow for the characteristics of the data discussed above; we then fit such distributions to the data in the following part.

2.2 Fitting a Logistic Distribution

The density function of the logistic distribution is given by

$$f(r_{t_i} \mid \mu, \sigma) = \frac{\exp\left(\frac{(r_{t_i} - \mu)}{\sigma}\right)}{\sigma\left[1 + \exp\left(\frac{(r_{t_i} - \mu)}{\sigma}\right)\right]^2} \quad , \; for \;\; -\infty \leq r_{t_i} \leq \infty \tag{5}$$

where μ (-∞ < μ < ∞) is a location parameter and σ (σ >0) is a dispersion (or scale) parameter. (Walck, 2007). This distribution, which is very similar to the normal in that it is symmetric but has thicker tails, and it has been first suggested as appropriate to model stock return. The maximum likelihood estimates of this distribution are obtained using the Expectation Maximization (EM) algorithm.

Parameter	Estimate	Std error
μ	-0.00461803	0.0121681
σ	0.423535	0.00620854

TABLE 4: ML Estimates for the Logistic Distribution.

The estimates in Table 4 above are used to fit the logistic distribution to the returns NSE 20 Share Index data. The shape of the return distribution, as compared to a fitted logistic distribution is presented in Figure 4 below.

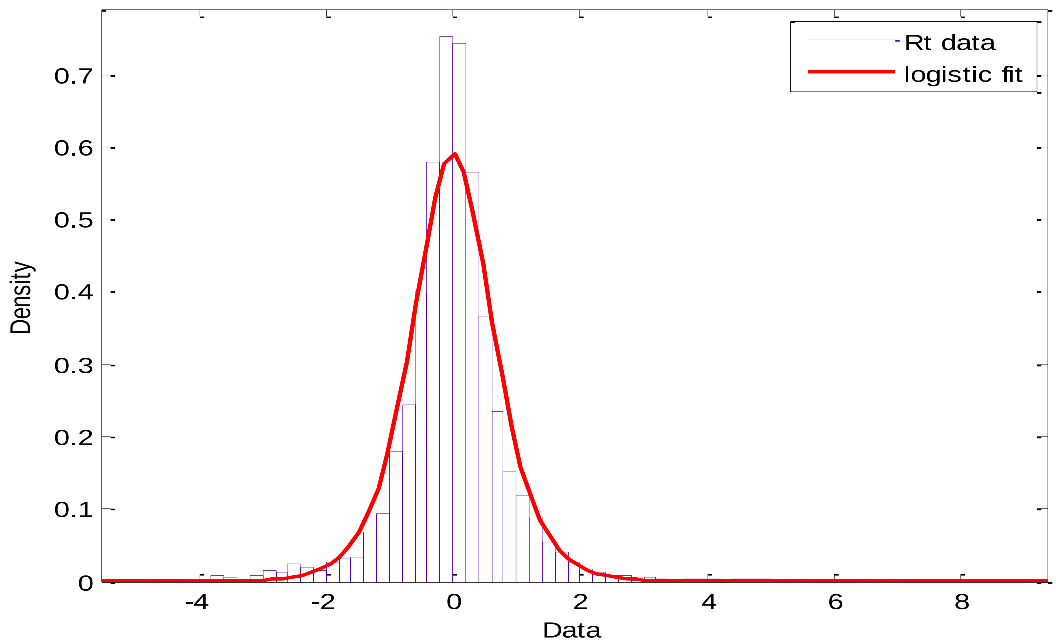

FIGURE 5: A logistic Distribution Fit to the Returns R_t Data.

Tests of goodness of fit results at 5% level of significance are presented in Table 5 below;

	Ansari Bradley		Kolmogorov-Smirnov	
Market	Statistic	P-value	Statistic	P-value
NSE 20 Share Index	6155059	2.2026×10^{-010}	0.0460	0.0013

TABLE 5: Goodness of Fit Tests for the Logistic Distribution.

The probability plot of the distribution of our data and the logistic distribution is represented in Figure 6 below;

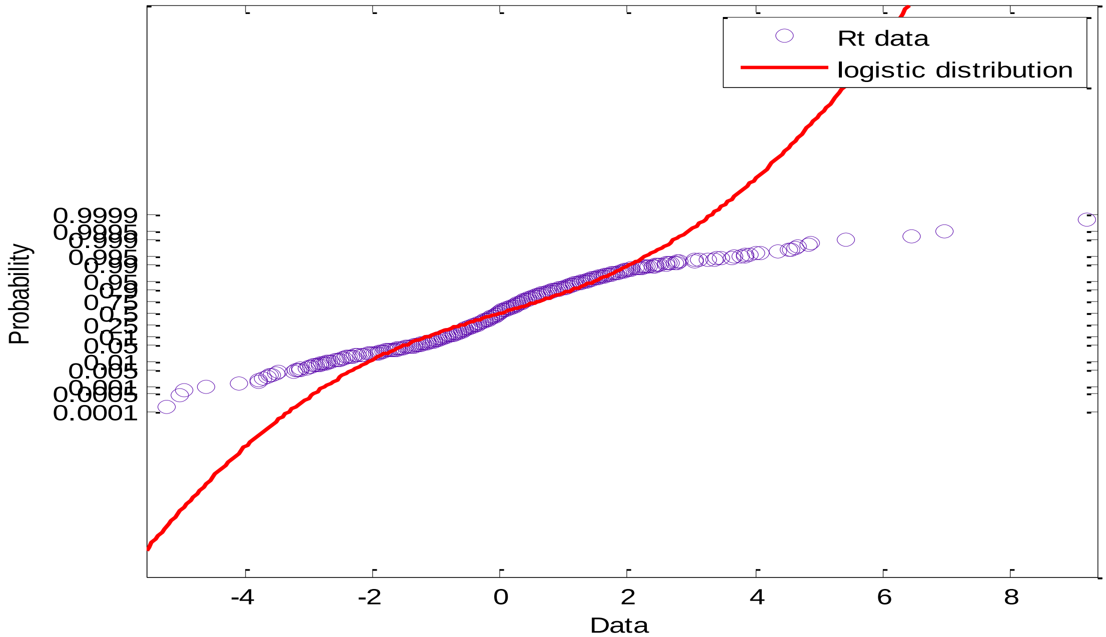

FIGURE 6: A Probability Plot of the Logistic Distribution Fit.

It is also clear that at any level of significance the logistic distribution fails to fit the returns of NSE 20 Share Price Index. The logistic distribution has a fatter tail than normal distribution. It failed to fit the returns of NSE 20 Share Index returns due to the assumption that the variance σ^2 of price changes over unit time interval is a constant.

2.3 Fitting a t-Location Scale Distribution

The t-location-scale distribution is useful for modeling data distributions with heavier tails (more prone to outliers) than the normal distribution. It is a normal inverse gamma mixture.

Let us assume that the returns of share price or an index R_t has a distribution whose variance, $\sigma^2\tau_i$ over unit time interval, is not a constant which is the case. Really, in practice this is so, because any share market often has long periods of relative activity, followed by long periods of relative inactivity. The information which affects prices does not come uniformly, but rather in bursts of activity.

Let R_t be normally distributed with the variance $\sigma^2\tau_i$ of returns of share price changing, i.e. it's a random variable with distribution function $g(\tau_i)$. A t-location scale is a mean variance mixture with a reciprocal gamma distribution as a mixing distribution. It is one of the standard non-normal distributions used in financial economics. Let R_t be distributed as follows

$$f(r_{t_i} \mid \mu, \sigma^2\tau_i) = \frac{1}{\sqrt{2\pi\sigma^2\tau_i}} \exp - \left(\frac{\left(r_{t_i} - \mu\right)^2}{2\sigma^2\tau_i} \right) \qquad for \quad -\infty \le r_{t_i} \le \infty. \tag{6}$$

with an inverse gamma distribution as the mixing distribution. Consider a special inverse gamma with shape parameter $\nu/2$ and scale parameter $\nu/2$ as shown below:

$$g\left(\tau_i\right)=\frac{\left(\nu/2\right)^{\nu/2}}{\Gamma\left(\nu/2\right)}\tau_i^{\left(-\nu/2-1\right)}e^{-\nu/2\tau_i} \tag{7}$$

and so the distribution,

$$f\left(r_{t_i},\mu,\sigma^2\right)=\int_0^\infty f\left(r_{t_i},\mu,\sigma^2\tau_i\right)g\left(\tau_i\right)d\tau_i \tag{8}$$

which simplifies to,

$$f\left(r_{t_i},\mu,\sigma^2\right)=\frac{\Gamma\left(\nu+1/2\right)}{\sigma}\frac{\left(\frac{\left(r_{t_i}-\mu\right)^2}{\nu\sigma^2}+1\right)^{-\frac{\nu+1}{2}}}{\sqrt{\nu\pi}}\frac{1}{\Gamma\left(\nu/2\right)}\quad,\quad for\ -\infty\leq r_{t_i}\leq\infty,\mu\in R,\sigma>0 \tag{9}$$

This is a t- location scale distribution as shown by (Barndorff, 1988). If a random variable R_t has a t- location with parameters μ,σ and ν then the random variable $\frac{r_t-\mu}{\sigma}$ has a student t-distribution with ν degrees of freedom. The mean, variance and kurtosis for this distribution are μ, $\frac{\nu\sigma^2}{\nu-2}$ and $\frac{3(\nu-2)}{\nu-4}$ respectively (Wenbo, 2006).

Using the Expectation Maximization algorithms of the t- location scale distribution, we obtain the maximum likelihood estimates of the t- location scale distribution.

Parameter	Estimate	Std. Error
μ	-0.00889354	0.0102819
σ	0.480765	0.0115711
ν	2.42046	0.124132

TABLE 6: ML Estimates for the t-location Scale Distribution.

Using these estimates we fit the t-location scale distribution to the NSE 20 Share Index data. The shape of the return distribution, as compared to a fitted t-location scale distribution is presented in Figure 7 below.

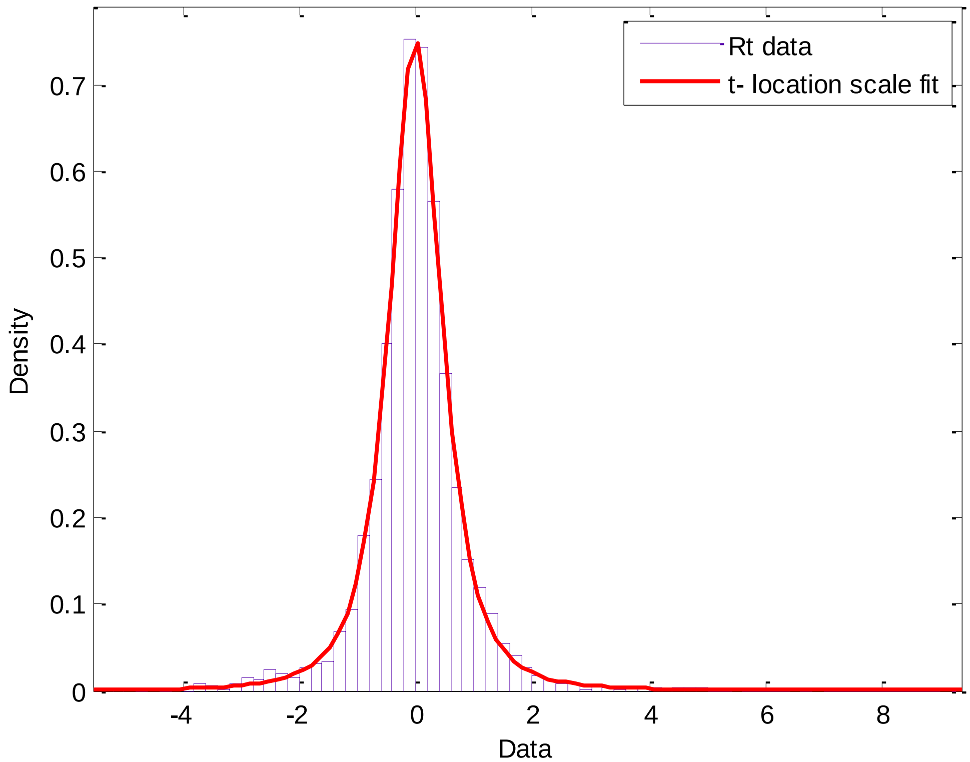

FIGURE 7: A t- location Scale Distribution fit to the Returns R_t Data.

Tests of goodness of fit results at 5% level of significance are presented in Table 7 below;

	Ansari Bradley		Kolmogorov-Smirnov	
Market	Statistic	P-value	Statistic	P-value
NSE 20 Share Index	5922916	0.4883	0.0192	0.5460

TABLE 7: Goodness of Fit Tests for the t-location Scale Distribution.

The probability plot of the return distribution and the t-location scale distribution is given Figure (8) below;

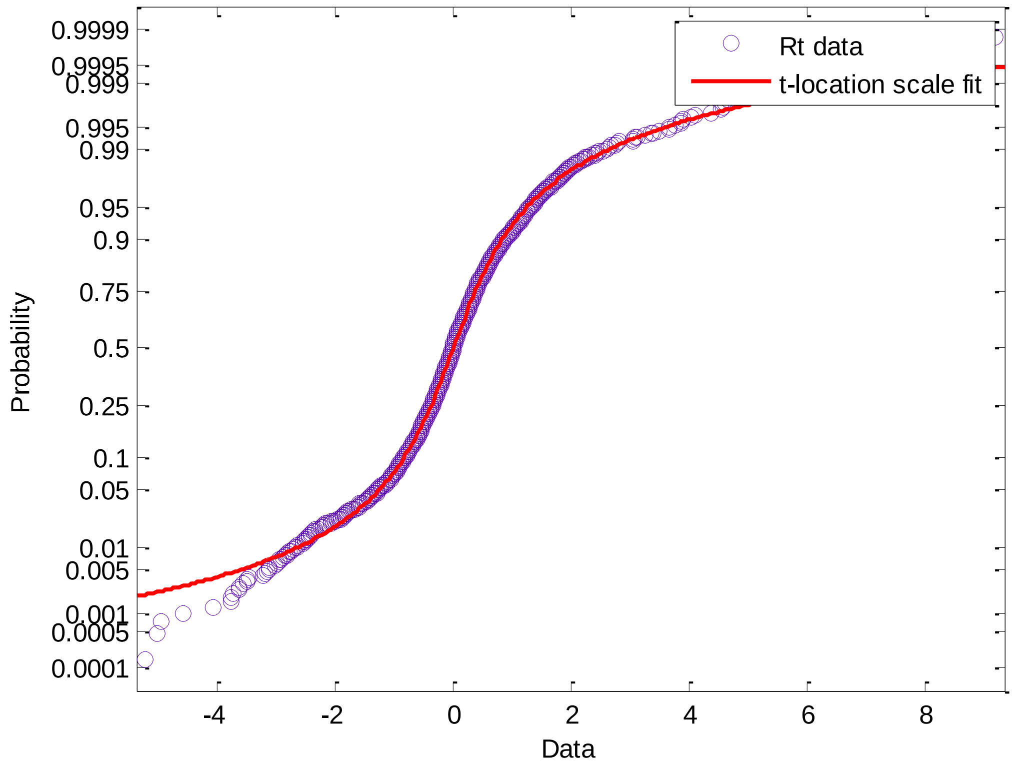

FIGURE 8: A Probability Plot of the t-location Scale Distribution Fit.

We have found strong support for the t-location scale distribution, which cannot be rejected at any reasonable significance level. This is because the t- location scale captures the fat tail exhibited in the NSE 20 Share Index returns. This also provides a clear evidence of the fact that the variance of price changes over unit time interval is a not constant. This is the case in practice because any share market often has long periods of relative activity, followed by long periods of relative inactivity. The information which affects prices does not come uniformly, but rather in bursts of activity. This is a formal evidence of the fact σ^2 varies significantly from year to year, as the degree of activity in the market also varies. Having established that the t-location scale (rather than the Normal) distribution properly describes daily NSE 20 Share Index, we conclude any predictions on the returns of NSE 20 Share Index should be based on the t-location scale distribution and not the normal distribution. Studies in financial economics can be based on the t-location scale distribution.

3. CONCLUSION

This study was interested in fitting an empirical distribution to the NSE 20 share index. In finding returns we used changes in logarithms prices instead of simple price changes. This is because;

i) The change in log price is the yield, with continuous compounding, from holding the security for that day.

ii) It has been shown that the variability of simple price changes for a given stock is an increasing function of the price level of the stock. Taking logarithms seems to neutralize most of this price level effect.

iii) For changes less than 15 percent, the change in log price is very close to the percentage price change, and for many purposes it is convenient to look at the data in terms of percentage price changes.

After thorough descriptive analysis it was clear that the NSE 20 share price index data is approximately symmetrical though it shows signs of being positively skewed. The coefficient of skewness is almost 0.5970. The data also exhibited a fat tail or excess kurtosis hence leptokurtic (i.e. kurtosis of 13.1608).

In an attempt to fit a normal distribution to the NSE 20 share price index data all tests done at the 5% significance level led to rejection of this distribution. This is due to the fact that the normal distribution fails to capture the fat tail of the returns data. Another reason why this model failed is due to the assumption that the variance σ^2 of price changes over unit time interval is a constant.

Though the logistic distribution has a fatter tail than normal distribution, it was also rejected at 5% level of significance. It failed to fit the returns of NSE 20 Share Index returns due to the assumption that the variance σ^2 of price changes over unit time interval is a constant.

The t-location scale distribution has a fatter tail than the normal and logistic distributions. In the construction of this distribution, the scale parameter (i.e. the variance) is not assumed constant. From our results, we found strong support for the t-location scale distribution, which could not be rejected at any reasonable significance level. This is because it captures the fat tail exhibited in the NSE 20 Share Index returns. This also provides a clear evidence of the fact that the variance of price changes over unit time interval is not a constant. This is a formal evidence of the fact σ^2 varies significantly from year to year, as the degree of activity in the market also varies. From these results, we conclude that any predictions on the returns of NSE 20 Share Index should be based on the t-location scale distribution and not the normal distribution. Studies in financial economics could be based on the t- location scale distribution. We further recommend that since we have found that the normal inverse gamma mixture best fits the NSE 20 Share Index return, other normal mixtures can be investigated how well they fit this data.

4. REFERENCES

[1] A. M. Mood, F.A. Graybill, and C. D. Boes. “An Introduction to the Theory of Statistics”. Tata McGraw Publishing Company Limited 2001, pp 107.

[2] B. Mandelbrot. "The Variation of Certain Speculative Prices," Journal of Business, 394-419 October 1963.

[3] B. Mandelbrot. “Paretian Distributions and Income Maximization”. The Quarterly Journal of Economics, Vol. 76, No. 1., pp. 57-85 Feb., 1962.

[4] C. Blattberg C., and N. J. Gonedes. “A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices”. Source: The Journal of Business, Vol. 47, No. 2, pp. 244-280. Apr., 1974.

[5] C. Hung, and S. Liaw.). “Statistical properties of Taiwan Stock Index”. Department of Physics, National Chung-Hsing University, 250 Guo-Kuang Road, Taichung, Taiwan Jan 9, 2007.

[6] C. Walck. “Handbook on Statistical Distributions for Experimentalists”.. University of Stockholm, Particle Physics Group. 2007, pp 84.

[7] F. Aparicio, and J. Estrada. “Empirical Distributions of Stock Returns: Scandinavian Securities Markets”. (1990-95).

[8] F. E. Fama. “The Behavior of Stock-Market Prices”. The Journal of Business, Vol. 38, No. 1., pp. 34-105 Jan., 1965.

[9] H. Wenbo, N. K. Alec. “The Skewed t-distribution for Portfolio Credit Risk”. Florida State University, 2006.

[10] L. Bachelier. “In the random character of stock market prices” (edited by P.H Cootner), Theorie de la speculation, Gauthiers, MA, (1964) Paris Cambridge; pp 17-18, 1900.

[11] M. Bibby. “Hyperbolic Processes in Finance”. Institute of Mathematics and Physics, The Royal Veterinary and Agricultural University Thorvaldsensvej 40 DK-1871 Frederiksberg C, DenmarkMichael Sørensen Department of Statistics and Operations Research Institute of Mathematical Sciences University of Copenhagen Universitetsparken 5 DK-2100 København Ø, Denmark, 2003.

[12] O. B. Osu. “Application of the logistic function to the Assessment of the Risk of Financial Assets Returns”. Journal of Modern Mathematics and Statistics, 4 (1) 7-10, 2010.

[13] O. Barndorff-Nielsen. “Exponentially Decreasing Distributions for the Logarithm of Particle Size”. Source: Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, Vol. 353, No. 1674, pp. 401-419, Mar. 25, 1977.

[14] O. Barndorff-Nielsen. “Hyperbolic Distributions and Distributions on Hyperbolae”. Source: Scandinavian Journal of Statistics, Vol. 5, No. 3, pp. 151-157 Published by: Wiley on behalf of Board of the Foundation of the Scandinavian Journal of Statistics,1978.

[15] O. Barndorff-Nielsen. Processes of normal inverse Gaussian type. Finance Stoch 2:41–68, 1988.

[16] P. D. Praetz. “The Distribution of Share Price Changes”. Source: The Journal of Business, Vol. 45, No. 1, pp. 49-55 Published by: The University of Chicago Press Jan., 1972.

[17] R. R. Officer. “The Distribution of Stock Returns”: R. R. Source: Journal of the American Statistical Association, Vol. 67, No. 340, pp. 807- 812 ,Dec., 1972.

[18] U. Keller. “Hyperbolic distributions in finance”- Ernst eberlin 1995.