Autoregressive Conditional Heteroscedasticity (ARCH) Modelling of the Nigerian Stock Indices using Seven Distributions for Innovation

Odum, Jane Ogechukwu[1]* and Okoli, Cecilia Nchedo[2],

[1]*Department of Mathematics, Nwafor Orizu College of Education, Nsugbe, Anambra State, Nigeria*
[2]*Department of statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria*

Abstract: **The characterization and quantification of stock market behaviour are important issues in financial risk management. This study analyzes the behaviour of Nigerian Stock Exchange (NSE) market returns over the period from 1985 to 2017. The autoregressive conditional heteroscedasticity model (ARCH) with seven unusual parametric distributions for innovations including the Gaussian distribution, the Skewed Gaussian distribution, the Students t distribution, the Generalized error distribution, the Skewed generalized error distribution, the Standardized normal inverse Gaussian distribution and the Skewed Students t distribution was considered in the study. The models were fitted to the data using the maximum likelihood method. The ARCH model with the Student's t as the distribution for the innovation gave the best fit. The assessment of the general predictive ability of the best-fitting model was carried out on the basis of the value-at-risk and the expected miscalculations using some loss functions.**

***Keywords*: Stock Return, Volatility, Maximum Likelihood, Parametric Distributions**

I. INTRODUCTION

Modeling financial time series is a complex problem. This complexity is not only due to the variety of the series in use (stocks, exchange rates, interest rates, etc.), to the importance of the frequency of the observation (second, minute, hour, day, etc) or to the availability of very large data sets. It is mainly due to the existence of statistical regularities (stylized facts) which are common to a large number of financial series and are difficult to reproduce artificially using stochastic models. Most of these stylized facts were put forward in a paper by Mandelbrot (1963). Since then, they have been documented, and completed, by many empirical studies. They can be observed more or less clearly depending on the nature of the series and its frequency. The properties that we now present are mainly concerned with daily stock prices.

Let p_t denote the price of an asset or stock at time t and let $\varepsilon_t = \log(p_t / p_{t-1})$ be the continuously compounded or log return (also simply called the return). The series (ε_t) is often close to the series of relative price variations $r_t = (p_t - p_{t1}) / p_{t1,}\text{since} = \varepsilon = \log(1 + r_t)$.The following properties have been amply commented upon in the financial literature; Non – stationarity of price series, Fat – tailed distribution, leverage effect, volatility clustering and absence of autocorrelation for the price variations.

Any satisfactory statistical model for daily returns must be able to capture the main stylized factsOf particular importance are the leptokurticity, the unpredictability of returns, and the existence of positive autocorrelations in the squared and absolute returns. Classical formulations (such as ARMA models) centered on the second-order structure are inappropriate. Indeed, the second-order structure of most financial time series is close to that of white noise. The fact that large absolute returns tend to be followed by large absolute returns (whatever the sign of the price variations) is hardly compatible with the assumption of constant conditional variance. This phenomenon is called conditional heteroscedasticity. One of the classical models introduced in the econometric literature to account for the very specific nature of financial series (price variations, log-returns, interest rates, stock indices, etc) is the Autoregressive Conditional Heteroskedasticity (ARCH)-type of model. This class of models were first proposed by Engle (1982) (ARCH). Both ARCH(p) and its generalization have received a wide range of applications compared to other classes of volatility models such as the historical method, exponentially weighted moving average (EWMA), etc. This is probably because of their ability to capture volatility clustering and leptokurtosis as well as track degrees of volatility variations over time. Also, ARCH models can be used to improve the modeling and prediction of other simple time series models like ARMA and ARIMA models (Tsay, 2011).

The motivation for this study was drawn from the fact that a lot of attention have been paid in the analysis of volatility associated with the Nigerian Stock Indices using ARCH model with Gaussian distribution, Student's t distribution and generalized error distribution but the skewed version of these distribution are rarely applied. From both practical and economic perspectives, the study based on the empirical evidence will help the policy make and researchers to make informed evidence based decisions at the most appropriate times as Nigerian Stock Markets evolves.Hence, the aim of this study is to characterize the all share price index of the Nigerian Stock Exchange using the Autoregressive Conditional Heteroskedasticity class of model with seven uncommon distribution for innovation, Gaussian distribution, Skewed Gaussian, Student t distribution, Skewed student t distribution, Generalized error, the skewed generalized error distribution and standardized normal inverse Gaussian distribution.

II. LITERATURE REVIEW

Many research works have been carried out, which in one way or the other relate to this research; hence there is need to review some past studies concerning modelling volatility based on both NSE and other stock indices across the world.

Moffat and Akpan (2018) in their study examined the influence of excess Kutorsis on the distributions of the innovations. They considered the presence of outliers in the data on daily closing prices of a share of Skye Bank, Sterling Bank, and Zenith Bank, starting from January 03, 2006, to November 24, 2006. The data consist of 2690 observations each obtained from the Nigerian Stock Exchange website. The result of their findings revealed that GARCH (1,1) Model under normal distribution, EGARCH (1,1) Model under a normal distribution and EGARCH (1,1) Model under student t distribution fitted adequately to the returns of Skye Bank, Sterling Bank and Zenith Bank respectively.

Okonkwo (2019) in their study examined the causal nexus between stock return volatility and selected macroeconomic variables in an emerging stock market from 1981 to 2017. The results of their findings showed that Johansen co-integration indicates the presence of a causal nexus between stock return volatility and selected macroeconomic variables in an emerging stock market in the long run. The Granger Causality Assessment Test revealed the index of industrial production and exchange rate as the statistically significant macroeconomic variables that influence stock return volatility to a high extent. The result on the significant effect of industrial productions and exchange rate lays credence to the existence of a positive and statistically significant relationship on stock return volatility.

Based on these findings, they recommended that the monetary authority should continually work towards the stabilization of the exchange rate of Naira against other currencies of the world as this significantly impact on stock return volatility.

Latha (2019) in his work studied the volatility pattern of thirteen emerging economics which are predominately oil-exporting countries using the time series consisting of monthly closing of price data from January 1st 2008 to 31st December 2018.

The emerging markets are considered as investment destinations due to the presence of risk premium which has made the stock markets of these countries mere volatile. They employed both symmetric and asymmetric models of generalized autoregressive heteroscedastic models. From their findings, it was revealed that there was evidence of volatility clustering and leptokurtic in all the countries considered. The result of their study also showed that GARCH (1, 1) and TGARCH (1,1) were found to be the most appropriate model that fits symmetric and asymmetric volatility for the thirteen countries.

Arum and Uche (2019) in their study investigated the volatility in equity prices of insurance stocks traded on the floor of the Nigerian Stock Exchange, using the time series data from 2011 to 2015 excluding weekend and public holidays. The result of their findings showed that GARCH (1,3) and GARCH (1, 1) were the best models that captured the volatility that exists in the insurance stocks through the information criteria of Akaike, Bayesian, Shibata and Hanna Quinn. Their findings suggested that potential investors should invest in insurance stock as they show calm tranquillity and the future is relatively stable.

III. MATERIALS AND METHODS

3.1 Method of Data collection

This study used secondary data which was obtained from the Central Bank of Nigeria Statistical Bulletin for various years. The data comprises of the Nigerian Stock Market Indices from 1985 to 2017.

3.2 Method of Data Analysis

3.2.1 ARCH Model

An ARCH model is composed of two components: the volatility component and the innovation component. The simplest and the most commonly used model for volatility is of order (1). The innovation is commonly assumed to come from the Gaussian distribution, the student’s distribution or some skewed extension of these distributions.

ARCH (p) is a popular financial time series model for weakly stationary financial data. It can be specified by

$$X_t = \sigma_t Z_t \qquad (1)$$

Where$\{X_t\}$ is the observed data, $\{\sigma_t\}$ is the volatility process specified by

$$\sigma_t^2 = \propto_0 + \propto_1 X_{t-1}^2 + \propto_2 X_{t-2}^2 + \cdots + \propto_p X_{t-p}^2 \qquad (2)$$

And Z_t is an innovation process.

Now, if ρ in equation (2) is equal to 1, then equations (1) & (2) reduce to ARCH(1) process with its evolution described by the following pair of equations.

$$X_t = \sigma_t Z_t \qquad (3)$$

$$\sigma_t^2 = \propto_0 + \propto_1 X_{t-1}^2 \qquad (4)$$

Where $\sigma_t > o$ and Z_t is an iid process with $E(Z_t)$ = 0 and var(Z_t) = 1. It is further assumed that (Z_t) is independent of the past of the process X_t.

The fourth equation makes sense only when its right-hand side is positive. To ensure this for all values of X_{t-1}, it is necessary to assume that the $\propto_0 > 0$ and $\propto_1 \geq 0$. Since we assume that the process X_t is stationary, its variance does not depend on t. Here, we will donate it by σ^2. Combining equations (3) and (4), we can write X_t more explicitly:

$$X_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} Z_t$$

3.2.1 ARCH (1) process is white noise

$E(X_t) = E(\sigma_t Z_t)$

$= \mathrm{E}\big(\mathrm{E}(\sigma_t z_t | F_{t-1})\big)$

$= \mathrm{E}\big(\sigma_t \mathrm{E}(z_t | \mathrm{F}_{t-1})\big)$ (because $\delta_t \in F_{t-1}$)

$= \mathrm{E}(\sigma_t \mathrm{E}\, z_t)$ (because Z_t is independent of F_{t-1})

$= 0$

By stationary, EX_t^2 is a constant. For $K \geq 1$we have

$\mathrm{E}(X_t X_{t-k}) = \mathrm{E}(\sigma_t Z_t \sigma_{t-k} Z_{t-k})$

$= \mathrm{E}\big(\mathrm{E}(\sigma_t Z_t \sigma_{t-k} Z_{t-k} | F_{t-1})\big)$

$= \mathrm{E}\big(\sigma_t \sigma_{t-k} Z_{t-k} \mathrm{E}(z_t | F_{t-1})\big)$

$= \mathrm{E}(\sigma_t \sigma_{t-k} Z_{t-k} \mathrm{E}\, z_t)$

$= 0.$

Hence $\{X_t\}$ is white noise.

3.2.2 Variance of the ARCH (1) process

From the equation of the ARCH (1) process we get

$\mathrm{E}(X_t^2) = \mathrm{E}(\sigma_t^2 Z_t^2)$

$= \mathrm{E}\big(\mathrm{E}(\sigma_t^2 z_t | \mathrm{F}_{t-1})\big)$

$= \mathrm{E}\big(\sigma_t^2 \mathrm{E}(Z_t^2 | \mathrm{F}_{t-1})\big)$ (because $\delta_t \in F_{t-1}$)

$= \mathrm{E}(\sigma_t^2 \mathrm{E}\, Z_t^2)$ (because Z_t is independent of F_{t-1})

$= \mathrm{E}(\sigma_t^2)$
(because$\mathrm{E}\,(Z_t^2) = 1)$

$= \propto_0 + \propto_1 \mathrm{E}\,(X_{t-1}^2)$(from the fourth equation of ARCH process)

From the last equation we can see that

$\mathrm{E}(X_t^2) = \propto_0 + \propto_1 \mathrm{E}(X_{t-1}^2)$

By the stationarity of X_t we also have $\mathrm{E}(X_t^2) = \mathrm{E}(X_{t-1}^2) = \sigma^2$

Hence

$\sigma^2 = \propto_0 + \propto_1 \sigma^2$. So $\sigma^2 = \frac{\propto_0}{1-\propto_1}$

Since $\sigma^2 > 0$ and $\propto_0 > 0$. We must have $\propto_1 < 1$.

3.2.3. Volatility of ARCH (1)

The conditional variance of X_t, given the information about X_s up to time $t-1$, is

$\mathrm{Var}(X_t | F_{t-1}) = \mathrm{E}(X_t^2 | \mathrm{F}_{t-1})$, since $\mathrm{E}(X_t | \mathrm{F}_{t-1}) = 0$

$= \mathrm{E}(\sigma_t^2 Z_t | \mathrm{F}_{t-1})$

$= \sigma_t^2 \mathrm{E}(Z_t^2 / F_{t-1})$

$= \sigma_t^2 \mathrm{E}(z_t^2)$

$= \sigma_t^2$

The volatility is the conditional standard deviation of X_t, given the information about X_s up to time $t-1$. The above calculation shows that volatility is equal to σ_t.

Note that sometimes σ_t^2 is also called volatility

3.2.4. Kurtosis of ARCH (1) process

$\mathrm{E}(X_t^4) = \mathrm{E}(\sigma_t^4 Z_t^4)$

$= \mathrm{E}\big(\mathrm{E}(\sigma_t^4 z_t^4 | \mathrm{F}_{t-1})\big)$

$= \mathrm{E}\big(\sigma_t^4 \mathrm{E}(z_t^4 | F_{t-1})\big)$ because $\delta_t \in F_{t-1}$

$= \mathrm{E}(\sigma_t^4 \mathrm{E}(Z_t^4))$ because Z_t is independent of F_{t-1}

$= \mu_4 \mathrm{E}(\sigma_t^4)$ (where $\mu_4 = \mathrm{E}(Z_t^4)$)

But

$\mathrm{E}(\sigma_t^4) = \mathrm{E}(\propto_0 + \propto_1 X_{t-1}^2)^2 = \propto_0^2 + 2 \propto_0 \propto_1 \mathrm{E}(X_{t-1}^2) + \propto_1^2 \mathrm{E}\,(X_{t-1}^4)$

$= \propto_0^2 + 2 \propto_0 \propto_1 \sigma^2 + \propto_1^2 \mathrm{E}(X_{t-1}^4)$ (by stationarity)

$= \propto_0^2 + 2 \propto_0 \propto_1 \sigma^2 + \propto_1^2 \mu_4 \mathrm{E}(\sigma_{t-1}^4)$

It is not difficult to see that this expression relates $\mathrm{E}(\sigma_t^4)$ to$\mathrm{E}(\sigma_{t-1}^4)$. In particular, it is clear to that $\mathrm{E}(\sigma_t^4)$ may be changed over time, even though $\mathrm{E}(\sigma_t^2)$ is constant. Now if we assume that our process $\{X_t\}$ is stationary up to fourth order. In that case$\mathrm{E}(X_t^4) = \mathrm{E}(X_{t-1}^4)$, which in turn implies that$\mathrm{E}(\sigma_t^4) = \mathrm{E}(\sigma_{t-1}^4)$.

Hence,

$\mathrm{E}(\sigma_t^4) = \propto_0^2 + 2 \propto_0 \propto_1 \sigma^2 + \propto_1^2 \mu_4 \mathrm{E}(\sigma_t^4)$

i.e. $(1 - \propto_1^2 \mu_4)\, \mathrm{E}(\sigma_t^4) = \propto_0^2 + 2 \propto_0 \propto_1 \sigma^2$.

Solving for $\mathrm{E}(\sigma_t^4)$ gives:

$\mathrm{E}(\sigma_t^4) = (\propto_0^2 + 2 \propto_0 \propto_1 \sigma^2)/(1 - \propto_1^2 \mu_4)$

Hence,

$$\mathrm{E}(X_t^4) = \mu_4 \mathrm{E}(\sigma_t^4) = \frac{\mu_4(\propto_0^2 + 2 \propto_0 \propto_1 \sigma^2)}{(1 - \propto_1^2 \mu_4)}$$

$= \frac{\mu_4\left(\propto_0^2 + 2\propto_0 \propto_1 \propto_0/(1-\propto_1)\right)}{(1-\propto_1^2 \mu_4)}$, since $\sigma^2 = {}^{\propto_0}/_{1-\propto_1}$

$= \frac{\mu_4 \propto_0^2 (1 + 2\propto_1 / 1 - \propto_1)}{(1-\propto_1^2 \mu_4)}$

$$= \frac{\mu_4 \propto_0^2 (1+\propto_1)}{(1-\propto_1^2 \mu_4)(1-\propto_1)} \tag{5}$$

The above equation (5) needs to be positive. So in addition to the requirements that

$0 \leq \propto_1 < 1$, we also have $(1 - \propto_1^2 \mu_4) > 0$ i.e. $\propto_1^2 < {}^1/_{\mu_4}$. If this inequality does not hold, then $\mathrm{E}(X_t^4)$ does not exist.

Consequently, to find the kurtosis of X_t, we combine the obtained results so far, that is

$$K_{xt} = \frac{E(X_t^4)}{\left(EX_t^2\right)^2}$$

$$= \frac{\mu_4 \alpha_0^2 (1-\alpha_1)(1-\alpha_1)^2}{(1-\alpha_1^2 \mu_4)(1-\alpha_1)\alpha_0^2}$$

$$= \frac{\mu_4 (1+\alpha_1)(1-\alpha_1)}{(1-\alpha_1^2 \mu_4)}$$

$$= \frac{\mu_4 (1-\alpha_1^2)}{(1-\alpha_1^2 \mu_4)}$$

Earlier, we noted that distributions with heavy tails have large kurtosis. The kurtosis of the normal distribution is equal to 3. The question now is for what values of $\propto$ is K_{x_t} larger than 3. It is easy to check that

$$\frac{\mu_4 (1-\alpha_1^2)}{(1-\alpha_1^2 \mu_4)} > 3, \text{ if and only if}$$

$$\alpha_1^2 = \frac{3-\mu_4}{2-\mu_4}$$

If $\mu_4 \geq 3$ this inequality holds for any $\propto$. In particular, if Z_t has standard normal distribution, then $\mu_4 = 3$ and the kurtosis of X_t is greater than 3. The practical consequences of this is that if $\mu_4 > 3$ the ARCH (1) process has heavy tails, which is a useful property for financial time series.

3.2.5 Distribution for the innovation process

For this work, we considered seven different distributions for Z_t: the Gaussian distribution due to de Moivre (1738) and Gaussian distribution due to Azzalini (1986); the student's t distribution due to Gosset (1908); the skewed student's t distribution due to Fernandez and Steel (1998); the generalized error distribution due to Stubbotin (1923); the skewed generalized error distribution due to Theodossiou (1998) and the standardized normal inverse Gaussian distribution due to Barndoff-Nielsen (1977).

Suppose, $\mathrm{X}, X_2 . . . . X_n$ are independent observations representing the daily log returns of the stock, prices from NSE. Now, if we assume these observation to follow ARCH (1) process with the distribution of Z_t following any one of the seven distribution for innovation process with the probability density function $f(x)$. For each distribution for Z_t, we give explicit expressions for $\mathrm{E}(Z_t), \mathrm{E}(Z_t^2), \mathrm{E}(Z_t^3), \mathrm{E}(Z_t^4)$, vaRp (Z_t) and $\mathrm{E}\mathcal{S}_p(Z_t)$.

3.2.6. Gaussian distribution

The probability density function for the Gaussian distribution is given by

$$f(Z) = \frac{1}{\sqrt{2\pi}} e - \frac{Z^2}{2}$$

If Z_t and iid Gaussian random variable with mean μ and variance then

$$\mathrm{E}(Z_t) = \mu$$

$$\mathrm{E}(Z_t^2) = \mu^2 + 1$$

$$\mathrm{E}(Z_t^3) = \mu^3 + 3\mu$$

$$\mathrm{E}(Z_t^4) = u^4 + 6u^2 + 3$$

$$V_a \mathrm{R}_p(Z_t) = \mu + \phi^{-1}(p)$$

$$\mathrm{E}(\mathrm{S}_p)(Z_t) = up + \phi(\phi^{-1}(p))$$

Where $\emptyset(\cdot)$ is the pdf of a standard Gaussian random variable and $\phi(\cdot)$ is the CDF of a standard Gaussian random variable. Gaussian distribution is due to de moivre (1738) and Gauss (1809).

3.2.7. Skewed Gaussian distribution

The probability density function for the location-scale skewed Gaussian is given by

$$\mathrm{f}(z_t, \mu, \sigma, \propto) = \frac{2}{\sigma} \emptyset \left(\frac{z_t - \mu}{\sigma}\right) \Phi \left(\propto \frac{z_t - \mu}{\sigma}\right)$$

where $\emptyset$ *and* Φ denote as usual, the pdf and the cumulative distribution function (CDF) of the standard normal distribution, respectively. If $\mu = 0$ and $\sigma = 1$, we obtain the standard skew-Gaussian distribution denoted by $\mathrm{SN}(\propto)$.

If Z_t are independent and identical skewed Gaussian random variables with location parameter μ and skewness parameter $\propto$ then

$$\mathrm{E}(Z_t) = \mu + \sqrt{\frac{2}{\pi}} \frac{\propto}{\sqrt{1+\propto^2}},$$

$$\mathrm{E}(Z_t^2) = 1 + \mu^2 + 2\mu \sqrt{\frac{2}{\pi}} \frac{\propto}{\sqrt{1+\propto^2}}$$

$$\mathrm{E}(Z_t^3) = \mu^3 + 3\mu^2 \frac{\propto}{\sqrt{1+\propto^2}} + 3\mu + \sqrt{\frac{2}{\pi}} \frac{\propto (3 + 2\propto^2)}{(1+\propto^2)^{3/2}}$$

$$E(Z_t^4) = \mu^4 + 6\mu^3 \frac{\propto}{\sqrt{1+\propto^2}} + 6\mu^2 + 4\mu\sqrt{\tfrac{2}{\pi}}\frac{\propto(3+2\propto^2)}{(1+\propto^2)\,{}^3\!/_2} + 3$$

$$E(S_p)(Z_t) = 2\int_{-\propto}^{VaR} x\,\phi(x-\mu)\Phi(\propto(x-\mu))dx$$

Where $T(h,a)$ is Owen's function defined in Owen (1980). The skewed Gaussian distribution is due to Azzalini (1986).

3.2.8. The Student t Distribution

The probability density function for the students t is given by

$$f(Z_t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x_t^2}{2}\right)^{-\frac{v+1}{2}}$$

Where $v$is the number of degrees of freedom and Γ is the gamma function.

If $Z_t$are independent and identical student's t random variables with location parameter μ and degrees of freedom v then

$$E(Z_t) = \mu$$

$$E(Z_t^2) = \mu^2 + \frac{v}{v-2}$$

$$E(Z_t^3) = \mu^3 + \frac{3uv}{v-2}$$

$$E(Z_t^4) = \mu^4 + \frac{6u^2v}{v-2} + \frac{3v^2}{(v-2)(v-4)}$$

$$V_aR_p(Z_t) = \mu + \sqrt{v}\,sign\left(P - {}^1\!/_2\right)\sqrt{\frac{1}{I_a^{-1}\left(\frac{v}{2} > {}^1\!/_2\right) - 1}}$$

$$E(S_p)(Z_t) = \mu p + \frac{\sqrt{v}\Gamma((v+1)/2)}{(1-v)\sqrt{\pi}\Gamma({}^v\!/_2)}\left[1+\frac{(VaR-\mu)2}{v}\right]^{\frac{1-v}{2}}$$

Where $a = 2P\ if\ P < {}^1\!/_2, a = 2(1-P)\ if\ P \geq {}^1\!/_2$, and $I_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt/B(a,b)$ is the incomplete beta function ratio and $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ is the beta function. The student's t distribution is due to Gosset (1908).

3.2.9. The Skewed Students t Distribution

The probability distribution density function for the skewed student's t is given by

$$F(Z_t) = \frac{2}{\Upsilon + \frac{1}{\Upsilon}}\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma({}^v\!/_2)(\pi v)^{1/2}}\left[1 + \frac{(Z_t-\mu)^2}{v}\left\{\frac{1}{\Upsilon^2}\mathrm{I}_{(o,\infty)}(Z_t-\mu) + \Upsilon^2\mathrm{I}_{(-\infty,0)}(Z_t-\mu)\right\}\right]^{-(v+1)/2}$$

Or

$$F(z_t)\frac{2\Upsilon}{(\Upsilon^2+1)}\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma({}^v\!/_2)\sqrt{\pi}\sqrt{v}}\left[1 + \frac{(z-\mu)^2}{v}\left(\frac{1}{\Upsilon^2}\mathrm{I}_{(z_t \geq u)} + \Upsilon^2\mathrm{I}(z_t < \mu)\right)\right] - \frac{v+1}{2}$$

Where $\mathrm{I}_{(\cdot)}$ is the indicator function. If $Z_t$are independent and identical skewed student's t random variables with location parameter μ, skewness parameter λ and degrees of freedom v then

$$E(Z_t) = \mu + \frac{\Gamma\left({}^{(v-1)}\!/_2\right)[\Upsilon^2 - \Upsilon^{-2}]}{\Gamma({}^v\!/_2)[\Upsilon + \Upsilon^{-1}]}$$

$$E(z_t^2) = \mu^2 + \frac{2\mu\Gamma\left({}^{(v-1)}\!/_2\right)[\Upsilon^2 - \Upsilon^{-2}]}{\Gamma({}^v\!/_2)[\Upsilon + \Upsilon^{-1}]} + \frac{\sqrt{\pi v}\,\Gamma\left({}^{(v-2)}\!/_2\right)[\Upsilon^3 - \Upsilon^{-3}]}{2\Gamma({}^v\!/_2)[\Upsilon + \Upsilon^{-1}]}$$

$$E(z_t^3) = \mu^3 + \frac{3\mu^2\Gamma\left({}^{(v-1)}\!/_2\right)[\Upsilon^2-\Upsilon^{-2}]}{\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]} + \frac{3\mu\sqrt{\pi v}\,\Gamma\left({}^{(v-2)}\!/_2\right)[\Upsilon^3-\Upsilon^{-3}]}{2\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]} + \frac{v\Gamma\left({}^{(v-3)}\!/_2\right)[\Upsilon^4-\Upsilon^{-4}]}{\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]}$$

$$E(z_t^4) = \mu^4 + \frac{4\mu^3\Gamma\left({}^{(v-1)}\!/_2\right)[\Upsilon^2-\Upsilon^{-2}]}{\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]} + \frac{6\mu^2\sqrt{\pi v}\,\Gamma\left({}^{(v-2)}\!/_2\right)[\Upsilon^3-\Upsilon^{-3}]}{2\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]}$$

$$+ \frac{6\mu v\Gamma\left({}^{(v-3)}\!/_2\right)[\Upsilon^4-\Upsilon^{-4}]}{\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]} + \frac{3v^{3/2}\sqrt{\pi}\,\Gamma\left({}^{(v-4)}\!/_2\right)[\Upsilon^5-\Upsilon^{-5}]}{4\Gamma({}^v\!/_2)[\Upsilon+\Upsilon^{-1}]}$$

$$VaR_p(Z_t) = \begin{cases} \mu + \sqrt{\gamma^2 v\left[\left\{1_{2\gamma p}^{-1}\left(\frac{v}{2},\frac{1}{2}\right)\right\}^{-1} - 1\right]}, if\ \ p \leq \frac{1}{(2\gamma)} \\ \mu + \sqrt{\gamma^2 v\left[\left\{1_{1+\gamma^{-2}-2\gamma^{-1}p}^{-1}\left(\frac{v}{2},\frac{1}{2}\right)\right\}^{-1} - 1\right]}, if\ \ p > \frac{1}{(2\gamma)} \end{cases}$$

$$E(S_p)(Z_t)=\begin{cases}\mu p+\dfrac{\sqrt{V}\Gamma\left(\dfrac{(V+1)}{2}\right)}{\gamma^2(1-V)\sqrt{\pi}\Gamma\left(\dfrac{V}{2}\right)}\left(1+\dfrac{\gamma^2 VaR^2}{v}\right)^{\frac{t-v}{2}}, if\ VaR\le\mu\\ \mu p+\dfrac{\sqrt{v}\Gamma\left(\dfrac{(v+1)}{2}\right)}{(1-v)\sqrt{\pi}\Gamma\left(\dfrac{v}{2}\right)}\left[\gamma^2\left(1+\dfrac{VaR^2}{\gamma^2 v}\right)^{\frac{1-2}{2}}-\gamma^2+\gamma^{-2}\right], if\ \ VaR>\mu\end{cases}$$

Where $I_x(a,b)=\int_0^x t^{a-1}(1-t)^{b-1}dt/B(a,b)$ is the incomplete beta function ratio and $B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}dt$ is the beta function. The skewed student's t distribution is due to Fernanderz and steel (1998).

3.2.9. Generalized Error Distribution

The probability density function for the generalized error distribution is give

$$F(Z_t)=\frac{a\exp\left[-\left(\frac{1}{2}\right)\left|Z/\lambda\right|^{\alpha}\right]}{\lambda 2^{(1+1/a)}\Gamma(1/a)}\quad ,-\infty<Z\ge\infty;0<a\le\infty$$

Where $\Gamma(.)$ is the gamma function and $\lambda\equiv\left[2^{(-2/a)}\Gamma\left(1/a\right)/\right.$ $^{\Gamma\ 3}a12$

If Z_t are independent and identical generalized error random variable with location parameter $\mu$and shape parameter a then

$E(Z_t)=\mu$,

$E(z_t^2)=\mu^2+\dfrac{a^2/_{a-1}\Gamma(3/a)}{\Gamma(1+1/a)}$,

$E(z_t^3)=\mu^3+\dfrac{3\mu a^{2}/_{a-1}\Gamma(3/a)}{\Gamma(1+1/a)}$,

$E(z_t^4)=\mu^4+\dfrac{6\mu^2 a^{2}/_{a-1}\Gamma(3/a)}{\Gamma(1+1/a)}+\dfrac{a^{4}/_{a-1}\Gamma(5/a)}{\Gamma(1+1/a)}$

$$V_aR_p(Z_t)=\begin{cases}\mu-a^{\frac{1}{a}}\left[Q^{-1}\left(\dfrac{1}{a},2p\right)\right]^{\frac{1}{a}}, if\ \ p\le\dfrac{1}{2}\\ \mu+a^{\frac{1}{a}}\left[Q^{-1}\left(\dfrac{1}{a},2(1-p)\right)\right]^{\frac{1}{a}}, if\ \ p>\dfrac{1}{2}\end{cases}$$

$$ES_p(Z_t)=\begin{cases}\mu p-\dfrac{a^{\frac{1}{a}}}{2\Gamma\left(\dfrac{1}{a}\right)}\Gamma\left(\dfrac{1}{a},\dfrac{(\mu-VAR)^a}{a}\right), if\ VAR\le\mu\\ \mu p-\dfrac{a^{\frac{1}{a}}}{2}+\dfrac{a^{\frac{1}{a}}}{2\Gamma\left(\dfrac{1}{a}\right)}\gamma\left(\dfrac{1}{a},\dfrac{(VAR-\mu)^a}{a}\right), if\ VAR>\mu\end{cases}$$

Where $Q(a,x)=\int_x^\infty t^{a-1}\exp(-t)dt/\Gamma(a)$ is the\regularized complementary incomplete gamma function.

$\gamma(a,x)=\int_0^x t^{a-1}\exp(-t)dt$ is the incomplete gamma function and $\Gamma(a,x)=\int_x^\infty t^{a-1}\exp(-t)dt$ is the complementary incomplete gamma function. The generalized error distribution is due to Subbotin (1923).

3.2.10. Skewed Generalized Error Distribution

The probability density function for the skewed generalized error distribution is given by

$$F(z_t)=\frac{K^{1-\frac{1}{k}}}{2}\Gamma\left(\frac{1}{k}\right)^{-1}\exp\left(-\frac{1}{k}\frac{|z-\mu+\delta|}{(1+\operatorname{sgn}(z-\mu+\delta)\lambda)^k}\right)$$

Where sign is the sgn function taking the value of -1 for μ<0 and 1 for μ>0 and $\Gamma(a)=\int_0^\infty x^{a-1}e^{-x}dx$ is the gamma function.

If Z_t are independent and identical skewed generalized error random variable with location parameter μ and shape parameter k than

$$E(Z_t)=\mu-\delta+\frac{C\theta^2}{k}\left[-(1-\lambda)^2+(1+\lambda)^2\right]\Gamma\left(\frac{2}{k}\right)$$

$$E(Z_t^2)=(\mu-\delta)^2+\frac{2C(\mu-\delta)\theta^2}{k}\left[-(1-\lambda)^2+(1+\lambda)^2\right]\Gamma\left(\frac{2}{k}\right)+\frac{C\theta^3}{k}\left[(1-\lambda)^3+(1+\lambda)^3\right]\Gamma\left(\frac{3}{k}\right)$$

,

$$E(Z_t^3)=(\mu-\delta)^3+\frac{3C(\mu-\delta)\theta^3}{k}\left[-(1-\lambda)^2+(1+\lambda)^2\right]\Gamma\left(\frac{2}{k}\right)+\frac{3C(\mu-\delta)\theta^3}{k}\left[(1-\lambda)^3+(1+\lambda)^3\right]\Gamma\left(\frac{3}{k}\right)$$

$$+\frac{C\theta^4}{k}\left[-(1-\lambda)^4+(1+\lambda)^4\right]\Gamma\left(\frac{4}{k}\right)$$

$$E(Z_t^4)=(\mu-\delta)^4+\frac{4C(\mu-\delta)^3\theta^2}{k}\left[-(1-\lambda)^2+(1+\lambda)^2\right]\Gamma\left(\frac{2}{k}\right)+\frac{6C(\mu-\delta)^2\theta^3}{k}\left[(1-\lambda)^3+(1+\lambda)^3\right]\Gamma\left(\frac{3}{k}\right)$$
$$+\frac{C(\mu-\delta)\theta^4}{k}\left[-(1-\lambda)^4+(1+\lambda)^4\right]\Gamma\left(\frac{3}{k}\right)$$
$$+\frac{C\theta^5}{k}\left[-(1-\lambda)^5+(1+\lambda)^5\right]\Gamma\left(\frac{5}{k}\right),$$

$$V_aR_p(Z_t)=\begin{cases}\mu-\delta-(1+\lambda)\theta\left[Q^{-1}\left(\frac{1}{k},\frac{2p}{1+\lambda}\right)\right]^{1/k}, if\ p\le\frac{1+\lambda}{2}\\ \mu-\delta+(1-\lambda)\theta\left[Q^{-1}\left(\frac{1}{k},\frac{2(1-p)}{1-\lambda}\right)\right], if\ p>\frac{1+\lambda}{2}\end{cases}$$

$$E(S_p)(Z_t)=\begin{cases}-\frac{C(1+\lambda)^2\theta^2}{k}\Gamma\left(\frac{2}{k},\frac{(\mu-V_aR-\delta)^2}{(1+\lambda)^k\theta^k}\right),\\ if V_aR\le\mu-\delta\\ -\frac{C(1+\lambda)^2\theta^2}{k}\Gamma\left(\frac{2}{k}\right)+\frac{C(1-\lambda)^2\theta^2}{k}\gamma\left(\frac{2}{k},\frac{(V_aR-\mu+\delta)^2}{(1-\lambda)^k\theta^k}\right),\ if\ V_aR>\mu-\delta\end{cases}$$

Where
$C=k/\left\{2\theta\Gamma\left(1/k\right)\right\}, \theta=\sqrt{\Gamma(1/k)/\Gamma(3/k)}/S(\lambda), \delta=2\lambda A/S(\lambda), S(\lambda)=\sqrt{1+3\lambda^2-4A^2\lambda^2}$
and

$A=\Gamma(2/k)/\sqrt{\Gamma(1/k)\Gamma(3/k)}$, The skewed generalized error distribution is due to Theodossiou (1998).

3.2.11. Standardized Normal Inverse Gaussian Distribution

The probability density function for the standardized normal inverse Gaussian distribution is given by

$$F(Z_t)=\frac{\alpha\exp\left(\sqrt{\alpha^2-\beta^2}\right)K_1\left(\alpha\sqrt{1+(z-\mu)^2}\right)\exp(\beta(z-\mu))}{\pi\sqrt{1+(x-\mu)^2}}$$

Where o< |β| <∝, $k_1(.)$ is the modified bessel function of the second kind of order one.

If z_t are independent and identical SNIG random variable the

$$E(Z_t)=\mu+\frac{\beta}{\gamma}$$

$$E(Z_t^2)=\left(\mu+\frac{\beta}{\gamma}\right)^2+\frac{\alpha^2}{\gamma^2}$$

$$E(Z_t^3)=\left(\mu+\frac{\beta}{\gamma}\right)^3+3\left(\mu+\frac{\beta}{\gamma}\right)\frac{\alpha^2}{\gamma^2}+\frac{3\alpha^2\beta}{\gamma^{7/2}}$$

$$E(Z_t^4)=\frac{3\alpha^4}{\gamma^5}\left(\gamma+1+\frac{4\beta^2}{\alpha}\right)+4\mu\left(\mu^2+\frac{3\alpha^2}{\gamma^2}\right)\left(\mu+\frac{\beta}{\gamma}\right)+4\mu\left(\mu+\frac{\beta}{\gamma}\right)^3-\frac{6\alpha^2\mu^2}{\gamma^2}-6\mu^2\left(\mu+\frac{\beta}{\gamma}\right)^2$$

$$E(S_p)(Z_t)=\frac{\alpha}{\pi}\int_{-\infty}^{VAR}\frac{K_1\left(\alpha\sqrt{1+(x-\mu)^2}\right)}{\sqrt{1+(x-\mu)^2}}\exp(\beta x+\gamma)dx$$

Where $\gamma=\sqrt{\alpha^2-\beta^2}, K1(.)$ is the modified bessel function of the second kind of order one and $VAR_p(Z_t)$ is the root of $\int_{-\infty}^{x}\frac{K_1\left(\alpha\sqrt{1+(y-\mu)^2}\right)}{\sqrt{1+(y-\mu)^2}}\exp(\beta y+\gamma)dy=P$

The normal inverse Gaussian distribution is due to Barndorff-Nielson (1977)

Fitting the ARCH (1) model, previously discussed, the method of maximum likelihood was employed that is if it is assumed that X_t, for $t=1,....,n$ are independent observations on X, then the parameters of ARCH (1) model are the values maximizing the likelihood

$$L(\theta)=\prod_{t=1}^{n}D(x_t/\delta_t;\theta)$$

or the log–likelihood which is sometimes more convenient, that is $LogL(\theta)=\sum_{t=1}^{n}LogD(x_t/\delta_t:\theta)$

Where D is the distribution for innovating as discussed earlier, $\theta=(\theta_1...\theta_k)$ is log likelihood function of the parameter vector.

Here and thereafter, $\hat{\theta}=\left(\hat{\phi}_1,\hat{\phi}_2,...,\hat{\phi}_k\right)$ will be used to denote the maximum likelihood estimate of θ. The standard errors of $\hat{\theta}$ were computed by approximating the covariance matrix of θ̂by the inverse of observed information matrix i.e

$$Cov\left(\hat{\theta}\right)\approx\begin{pmatrix}\frac{\partial^2\log L}{\partial\phi^2} & \frac{\partial^2\log L}{\partial\phi_1\partial\phi_2} & \cdots\frac{\partial\log L}{\partial\phi_1\partial\phi_k}\\ \frac{\partial^2\log L}{\partial\phi_2\partial\phi_1} & \frac{\partial^2\log L}{\partial\phi_2^2}\cdots & \frac{\partial^2\log L}{\partial\phi_2\partial\phi_k}\\ \vdots & \vdots & \vdots\\ \frac{\partial^2\log L}{\partial\phi_k\partial\phi_1} & \frac{\partial^2\log L}{\partial\phi_k\partial\phi_2}\cdots & \frac{\partial^2\log L}{\partial\phi_k^2}\end{pmatrix}^{-1}$$

$\theta = \hat{\theta}$

Discriminating among the fitted ARCH (1) model with respect to distributions for innovation was performed using different criteria. That is, the fit of the innovation distribution for the ARCH (1) model is compared in terms of the following criteria.

TheAkaike information criterion due to Akaike (1974) defined by

$$AIC = 2k - 2\log L \left(\hat{\theta}\right);$$

the Bayesian information criterion due to Schwarz (1978) defined by

$$BIC = k \log n - 2 \log L \left(\hat{\theta}\right);$$

The Hannan –Quinn criterion due to Hannan and Quinn (1979) defined by

$$HQC = -2\log L\left(\hat{\theta}\right) + 2K\ Log\ Log\ n$$

The smaller the values of these criterions the better the fit. Furthermore, various loss functions are employed to evaluate the selected model performances in terms of forecasting ability, and these include: the quasi-likelihood, defined by

$$QLike = n^{-1}\sum_{t-}^{n}\left(\log\left(S_t^2\right) + O_t^2\hat{S}_t^{-2}\right)$$

the R^2 Log, defined by

$$R^2 Log = n^{-1}\sum_{t=1}^{n}\left[\log\left(\delta_t^2\hat{S}_t^{-2}\right)\right]^2;$$

the root mean squared error defined by

$$RMSE = \sqrt{n^{-1}\sum_{t=1}^{n}\left(\delta_t^2 - \hat{S}^2\right)^2}$$

the mean squared error defined by

$$MSE = n^{-1}\sum_{t=1}^{n}\left(\hat{\delta}_t - \hat{S}_t\right)^2;$$

the mean absolute deviation defined by

$$MAD = n^{-1}\sum_{t=1}^{n}\left|\delta_t - \hat{S}_t\right|$$

It is important to note that of those functions could be used to evaluate model forecasting ability. However, there is no specific measure that is universally considered best. Hence, it is better to use as much as possible. This is because some of these measures are so sensitive to outlier's e.g MAD.

IV. DATA ANALYSIS AND RESULT

Time series analysis is performed on the share price index of the NSE using the Autoregressive conditional heteroskedasticity (ARCH) with seven different distributions for the innovations. The distributions for the innovations include: the Gaussian distribution, the Skewed Gaussian distribution, the Students t distribution, the Generalized error distribution, the Skewed generalized error distribution, the Standardized normal inverse Gaussian distribution, the

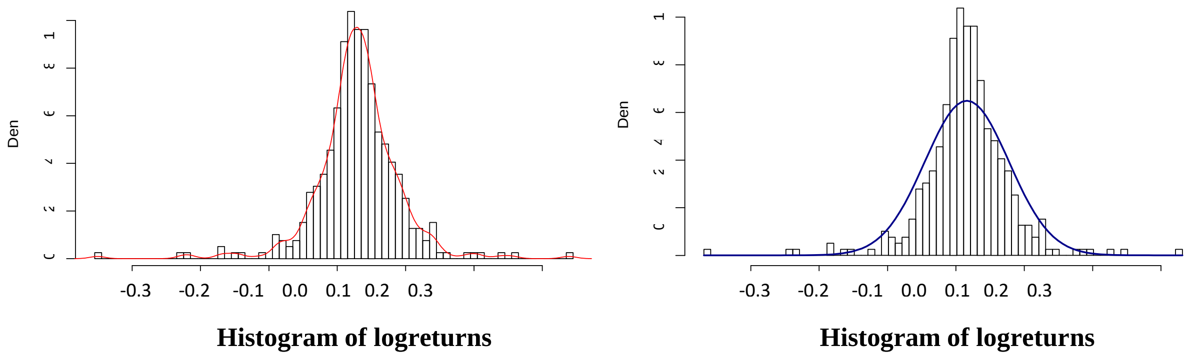
Skewed Students t distribution. The models are fitted on the data from the year 1985 to 2017, then the forecast accuracy will be done in terms of the two most widely used risk measures, the value at risk and expected shortfall. The flexibility and performanceof these distributionsfor innovation aredone using some selected criteria: the Akaike information criterion of Akaike (AIC), the Bayesian information criterion of Schwarz (BIC), the Hannan Quinn criterion (HQC).

Table 1: Descriptive statistics for monthly log returns for the period 1985-2017 for all share price index of the NSE.

	Statistics	NSE
Daily log returns	n	395
	Min	−0.365882
	Q1	−0.011696
	Median	0.015991
	Mean	0.014784
	Q3	0.042945
	Max	0.323515
	SD	0.061484
	CV	4.15882
	Skewness	−0.509779
	Kurtosis	6.833218
	IQR	0.054641
	Range	0.689397
	Variance	0.003780

4.1 Descriptive Analysis of the Data Set

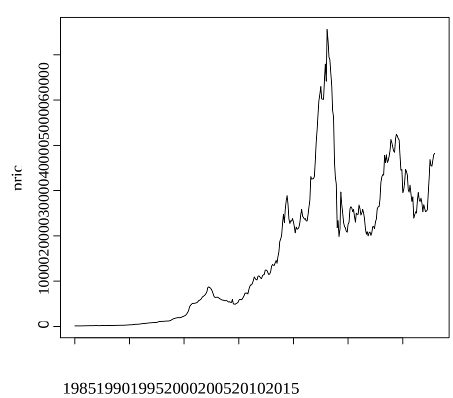
Figure1: Histogram of log returns for indices of the NSE with kernel density (left) and a normal (right) curve superimposed to the data.



Following common practice, the data were transformed by taking logarithms and then first-order differences. The histograms of the transformed data set are shown in Figure 1. Each histogram appears more or less symmetric about zero. Note that the histograms are the same but with different curves superimposed. The one on the left-hand side bears a kernel plot on top of the histogram. A kernel plot produces a smooth curved without assuming normality. Generally, in statistics, kernel density estimation is a non-parametric way to estimate the probability density function of a random variable. It is usually a much more effective way to view the distribution of a variable when compared to the normal distribution. Then, the one on the right-hand side has a normal curve over it. The following summary statistics for the daily log returns are computed and given in Table 1, the number of observations (n), the minimum (Min), first quartile (Q1), median, mean, third quartile(Q3), the maximum (Max), standard deviation, coefficient of variation (CV), skewness, kurtosis, interquartile range (IQR), range and variance. The minimum value for our data set is negative. The first quartile value for our data set is also negative. The median and mean values are positive but close to zero, which is evidence of the stock booming within the period of study. The maximum is a positive value. The coefficient of variation is relatively high. It is important to note that the coefficient of variation is one of the statistics that allows investors to determine how much volatility, or risk, is assumed in comparison to the amount of return expected from investments. Of course the lower the ratio of the standard deviation to mean return, the better risk-return trade-off. The data is negatively skewed and is less than zero.

The kurtosis value is significantly greater than three, the kurtosis value corresponding to the normal distribution. With both the skewness and kurtosis values, we note that our data is not normally distributed. This confirms that the data is heavy-tailed. Normality of stock price returns was tested using the Anderson-Darling test (Anderson and Darling, 1954), the Cramer-von Mises test, the Kolmogorov-Smirnov test, the Pearson chi-square test, the Jarque-Bera test (Jarque and Bera, 1980) and the data-driven smooth test. None of the tests for the data set on the stock price returns followed the normal distribution. This is not surprising as financial data are generally known to exhibit heavy tails, time-varying volatility, and long-range dependence with significant kurtosis and asymmetry.

Figure 2: Plots of the daily prices of the NSE shares



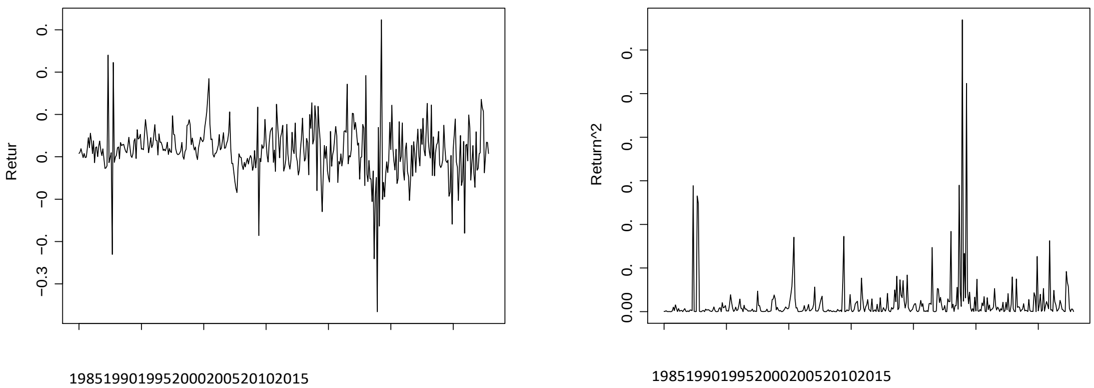
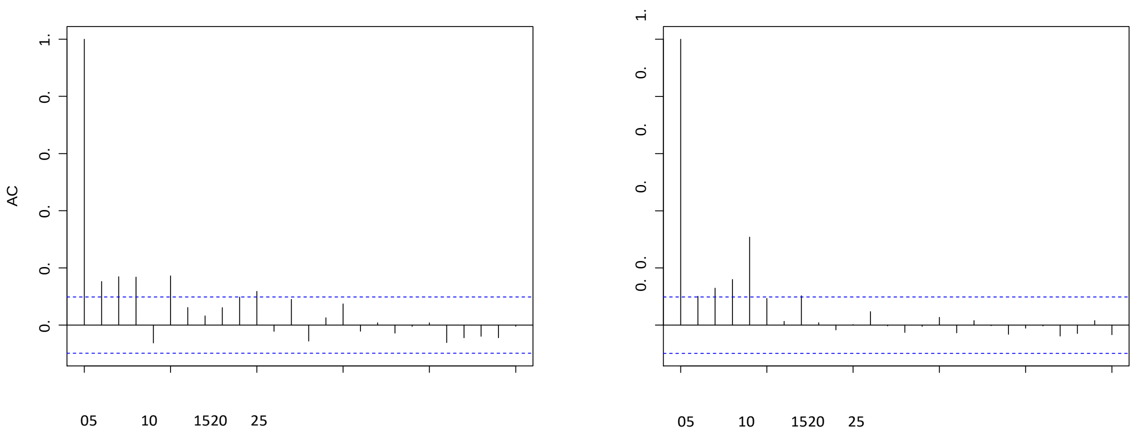


Figure 3: Plots of the daily log returns and squared log returns of the NSE shares.

Plot of the daily log returns of the NSE shares. *Plot of the daily log returns^2 of the NSE shares*

Figure 4: Autocorrelations of the NSE returns and Autocorrelations of the squared NSE returns.



Autocorrelations of the NSE returns **Autocorrelations of the squared returns**

Notably, the plot of daily prices of the NSE share in Figure 2 appears not to be stationary. That is the samples paths of the prices are generally close to a random walk without intercept. On the other hand, the plots of returns and squared returns are given in Figure3 and are generally compatible with the second-order stationarity assumption. For instance, it shows that the returns of the NSE index oscillate around zero, with no visible trend. The oscillations vary a great deal in magnitude, but are almost constant in average over long subperiods. The extreme volatility of prices, induced by the financial crisis of 1987 and 2008, are worth noting.

Figure 4 gives the autocorrelation function of the returns and squared returns. The ACF of the returns does not appear to be significantly non-zero anywhere (other than at lag 1 of course). In other words, there is weak or no serial correlation of the returns. On the other hand, the second chart which gives the ACF of squared returns does appear to be significantly non-zero at certain lags(e.g at lag 4). So this gives us hope that we can use these squared returns to predict something (volatility) by using them. Interestingly, if we plot the ACF for absolute returns, we would find something similar to the ACF of squared returns. This is because the absolute value and the squared both discard the sign to measure some sort of "deviation". And in general, the intuition behind looking at the square of the series when we are searching for the "ARCH" effect is unravelled. This is because we can now use them to predict "volatility" (i.e. the conditional variance) using ARCH models. And this is what ARCH models do.

The ARCH model in presented in section 3.2.1 were fitted to log returns from NSE share indices. We considered seven distributions for the innovation. So, the ARCH model with their respective innovations were fitted to log returns. The method of maximum likelihood was used. The log-likelihood values, the AIC values, the BIC values and the HQC values for the fitted models are given in Tables 2 to 8.

Table 2: Returns from NSE shares: Model 1 with innovation as normal.

Distribution for Zt:Normal	Let Xt be our time series. The fitted model is Xt=0.0182951+t t=σtZt Zt=N(0,1) t=0.0018205+0.8764411Xt−1
$\hat{\mu}$	0.0182951(0.0025300),p-value=4.78×10−13
$\hat{\omega}$	0.0018205(0.0002255)p-value=6.66×10−16
$\hat{\alpha}$	0.8764411(0.1782579)p-value=8.80×10−07
Log-likelihood&criteria	-566.3567,AIC=−2.852439,BIC=−2.822219,HQC=−2.840466
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value=1.20×10−14
Ljung-BoxTest	p-value=0.571607
Ljung-BoxTest R2	p-value=0.575411
LMArchTest	p-value=0.7249127

Table 3: Returns from NSE shares: Model 1 with innovation as Skewed normal.

Distribution for Zt: Skewed Normal	Let Xt be our time series. The fitted model is Xt=0.0172345 + t t=σtZt Zt= SN(0,1) t=0.0018042+0.809729Xt−1
$\hat{\mu}$	0.0172345(0.0026305),p-value= 5.69×10−11
$\hat{\omega}$	0.0018042(0.0002261)p-value=1.55×10−15
$\hat{\alpha}_1$	0.8709729(0.1783902)p-value=1.05×10−06
$\hat{\lambda}$	0.9276077(0.0459791) p-value=2.00 × 10−16
Log-likelihood & criteria	-567.5182, AIC = −2.853256, BIC = −2.812964,HQC = −2.837292
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value=1.04×10−14
Ljung-BoxTest	p-value=0.554252
Ljung-BoxTest R2	p-value=0.2931929
LMArchTest	p-value=0.472741

Table 4: Returns from NSE shares: Model 1 with innovation as Student's t.

Distribution for Zt: Student' t	Let Xt be our time series. The fitted model is Xt=0.0163458 + t t=σtZt Zt=t(0,1) t=0.0017359+1.0000000Xt−1
$\hat{\mu}$	0.01363458(0.0021833),p-value= 7.06×10−14
$\hat{\omega}$	0.0017359(0.0004204)p-value=3.64×10−05
$\hat{\alpha}_1$	1.0000000(0.3052518)p-value=0.00510
$\hat{\lambda}$	3.3137649(0.5322359) p-value=4.78 × 10−10
Log-likelihood&criteria	-615.6642, AIC = −3.097034, BIC = −3.056742,HQC = −3.081070
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value= 7.56×10−15
Ljung-BoxTest	p-value=0.684248
Ljung-BoxTest R2	p-value=0.29296983
LMArchTest	p-value=0.212483

Table 5: Returns from NSE shares: Model 1 with innovation as skewed Student's t.

Distribution for Zt: skewed Student't	Let Xt be our time series. The fitted model is Xt=0.018501 + t t=σtZt Zt= Sst(0,1) t=0.001745+1.0000000Xt−1
$\hat{\mu}$	0.0185011(0.0028756),p-value= 1.24 × 10−10
$\hat{\omega}$	0.0017450(0.0004258) p-value=4.16 × 10−05
$\hat{\alpha}_1$	1.0000000(0.3015904) p-value=0.000914
$\hat{\nu}$	3.2971615(0.5145715) p-value=1.48 × 10−10
$\hat{\lambda}$	1.0911472(0.0826959) p-value=2.00 × 10−16
Log-likelihood&criteria	- 615.6642, AIC = −3.095378, BIC = −3.045013,HQC = −3.075423
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value= 1.04×10−14
Ljung-BoxTest	p-value=0.1130496
Ljung-BoxTest R2	p-value=0.9839326
LMArchTest	p-value=0.53627

Table 6: Returns from NSE shares: Model 1 with innovation as GED**.**

Distribution for Zt: GED	LetXt beourtimeseries.Thefittedmodelis Xt=0.0203914 + t t=σtZt Zt=*GED*(0,1) t=0.0002088+ 0.5637341Xt−1
$\hat{\mu}$	0.0239(0.002353),p-value= 2.00 × 10−16
$\hat{\omega}$	0.000288(0.00009332) p-value=0.0253
$\hat{\alpha}_1$	0.5637(0.1276) p-value=9.91 × 10−06
$\hat{\lambda}$	1.265(0.1355) p-value=2.00 × 10−16
Log-likelihood&criteria	-577.6642, AIC = −2.095378, BIC = −2.045013, HQC = −2.075423

Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value= 1.95×10^{-06}
Ljung-BoxTest	p-value=0.687861
Ljung-BoxTest R2	p-value=0.5242079
LMArchTest	p-value=0.2899992

Table 7: Returns from NSE shares: Model 1 with innovation as skewed GED.

Distribution for Zt: skewed GED	Let Xt beourtimeseries.The fitted model is Xt=0.02448864 + t t=σtZt Zt=*SGED*(0,1) t=0.0001421604+ 0.6506868433Xt−1
$\hat{\mu}$	0.02449(0.002272),p-value= 2.0×10^{-16}
$\hat{\omega}$	0.0001422(0.00007223) p-value=0.049
$\hat{\alpha}_1$	0.6507(0.1373) p-value= 2.14×10^{-06}
$\hat{a}$	1.192(0.1236) p-value= 2.0×10^{-06}
$\hat{\lambda}$	1.266(0.0602) p-value= 2.00×10^{-16}
Log-likelihood&criteria	-- 531.2661, AIC = −2.165905, BIC = −2.105466, HQC = −2.141958
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value= 3.355×10^{-06}
Ljung-BoxTest	p-value=0.1529313
Ljung-BoxTest R2	p-value=0.2714512
LMArchTest	p-value=0.24475478

Table 8: Returns from NSE shares: Model 1 with innovation as standardized NIG.

Distribution for Zt: SNIG	LetXt beourtimeseries.Thefittedmodelis Xt=0.02436185 + t t=σtZt Zt=*SNIG*(0,1) t=0.00016738+ 0.68368297Xt−1
$\hat{\mu}$	0.02436(0.002206),p-value= 2.0×10^{-16}
$\hat{\omega}$	0.0001674(0.00008429) p-value=0.047063
$\hat{\alpha}_1$	0.6837(0.1629) p-value= 2.70×10^{-05}
$\hat{a}$	1.045(0.4070) p-value=0.010217
$\hat{\lambda}$	0.3516(0.09373) p-value=0.000176
Log-likelihood&criteria	- 530.1106, AIC = −2.160054, BIC = −2.099615, HQC = −2.136107
Jarque-BeraTest	p-value=0
Shapiro-WilkTest	p-value= 3.605723×10^{-06}
Ljung-BoxTest	p-value=0.1508468
Ljung-BoxTest R2	p-value=0.2800082
LMArchTest	p-value=0.238415

From the result obtained in Tables 2 - 8, we observe that the ARCH parameters are highly statistically significant (extremely small p-values) for all the distributions. Each parameter has a p-value of less than 0.05 and so is statistically significant at 5% significance level. Both, Jarque-Bera tests and Shapiro-Wilk tests strongly suggest that the residuals are non-Gaussian for all distributions. The Ljung-Box of the residuals supports the hypothesis that they are uncorrelated (or weak dependence). The Ljung-Box of the squared residuals supports the hypothesis that they are uncorrelated, as well. The LM Arch tests indicates that there is no ARCH effect in the residuals, that is each model(distribution) explains any ARCH effects present in the data. So comparing the performance of these distributions, it can be observed that the ARCH model with the Student's t as the distribution for the innovation gives the smallest values for the negative log-likelihood, the AIC, the BIC, and the HQC. For the best performing model, estimate of the shape parameter is 3.31376. It represents the degrees of freedom of the estimated standardized t-distribution. The t-distributions have heavier tails than the normal distribution, especially for small number of degrees of freedom. The estimate 3.31376 suggests that the tails are indeed heavier than that of the normal distribution. For example, moments of order higher than 3 do not exist for t-distributions with less than 4 degrees of freedom.

To access the forecasting ability of these models, we employed different loss functions and compared their observed values with their corresponding fitted values in terms of two most widely used risk measures, value at risk and expected short fall.

Table 9: Loss function as overall model evaluation criteria for predictive ability of the NSE share indices

Log returns of NSE share indices							
Loss function	Model	$VaR_{0.95,wl=50}$	$ES_{.95,wl=50}$	$VaR_{.95,wl=100}$	$ES_{.95,wl=100}$	$VaR_{.95,wl=150}$	$ES_{.95,wl=150}$
	ARCH NORM	1.441×10^{-1}	-	1.380×10^{-1}	-	$1.858\text{x}10^{-1}$	-
	ARCH SNORM						-
	ARCH STD	1.609×10^{-1}	-	1.509×10^{-1}	-	1.973×10^{-1}	-

	ARCH SSTD	3.692×10^{-3}	489.58	3.225×10^{-1}	0.0004	3.400×10^{-3}	239.93
	ARCH GED	1.586×10^{-1}	9363.288	1.500×10^{-1}	10783.53	1.619×10^{-1}	12239.01
	ARCH SGED ARCH SNIG	1.990×10^{-1}	-	1.579×10^{-1}	-	1.579×10^{-1}	-
QLIKE		1.476×10^{-1}	9224.826	1.395×10^{-1}	10602.8	1.886×10^{-1}	12037.74
		1.601×10^{-1}	-	1.447×10^{-1}	-	$1.941\times10-1$	-
	ARCH NORM	3.258×10^{-1}	-	3.230×10^{-1}	-	4.292×10^{-1}	-
	ARCH SNORM	3.694×10^{-1}	-	3.531×10^{-1}	-	4.582×10^{-1}	-
	ARCH STD	8.232×10^{-3}	392.93	6.475×10^{-4}	472.73	6.455×10^{-1}	216.06
	ARCH SSTD	3.391×10^{-1}	57009.21	3.228×10^{-1}	72251	4.237×10^{-1}	76337.82
	ARCH GED	4.389×10^{-1}	-	3.389×10^{-1}	-	3.700×10^{-1}	-
R^2LOG	ARCH SGED	3.108×10^{-1}	56073.26	2.992×10^{-1}	70876.93	3.943×10^{-1}	74655.99
	ARCH SNIG	3.388×10^{-1}	-	3.011×10^{-1}	-	3.857×10^{-1}	-
	ARCH NORM	8.551×10^{-3}	2.293×10^{-1}	8.862×10^{-3}	1.952×10^{-1}	1.057×10^{-2}	1.777×10^{-1}
	ARCH SNORM	8.885×10^{-3}	2.293×10^{-1}	9.174×10^{-3}	1.951×10^{-1}	1.086×10^{-2}	1.777×10^{-1}
	ARCH STD	1.064×10^{-5}	1.962×10^{-3}	9.749×10^{-5}	1.943×10^{-6}	9.706×10^{-4}	1.990×10^{-3}
	ARCH SSTD	8.171×10^{-3}	3.019×10^{-1}	8.277×10^{-3}	3.013×10^{-1}	9.954×10^{-3}	3.033×10^{-1}
	ARCH GED	8.329×10^{-3}	2.297×10^{-1}	7.595×10^{-3}	1.959×10^{-1}	8.727×10^{-3}	1.778×10^{-1}
RMSE	ARCH SGED	7.680×10^{-3}	2.938×10^{-1}	7.776×10^{-3}	1.895×10^{-1}	9.395×10^{-3}	1.898×10^{-1}
	ARCH SNIG	7.248×10^{-3}	2.278×10^{-1}	7.169×10^{-3}	1.987×10^{-1}	8.714×10^{-3}	1.764×10^{-1}
	ARCH NORM	5.654×10^{-3}	11.409	5.650×10^{-3}	9.555	6.508×10^{-3}	8.585
	ARCH SNORM	5.965×10^{-3}	11.400	5.913×10^{-3}	9.546	6.758×10^{-3}	8.589
	ARCH STD	7.987×10^{-5}	11.978	7.241×10^{-5}	9.926	7.506×10^{-5}	3.217
	ARCH SSTD	5.451×10^{-3}	18.489	5.400×10^{-3}	18.870	6.262×10^{-3}	18.653
	ARCH GED	5.874×10^{-3}	11.479	5.256×10^{-3}	9.625	5.572×10^{-3}	8.611
MAD	ARCH SGED	5.170×10^{-3}	17.949	5.069×10^{-3}	9.885	5.936×10^{-3}	9.985
	ARCH SNIG	5.171×10^{-3}	11.454	4.909×10^{-3}	9.555	5.774×10^{-3}	8.432
	ARCH NORM	7.312×10^{-5}	5.261×10^{-2}	7.855×10^{-5}	3.812×10^{-2}	1.118×10^{-4}	3.159×10^{-2}
	ARCH SNORM	7.895×10^{-5}	5.259×10^{-2}	8.417×10^{-5}	3.809×10^{-2}	1.181×10^{-4}	3.159×10^{-2}
	ARCH STD	1.132×10^{-6}	3.852×10^{-5}	9.505×10^{-5}	3.382×10^{-5}	9.422×10^{-6}	3.664×10^{-7}
	ARCH SSTD	6.677×10^{-5}	9.119×10^{-2}	6.851×10^{-5}	9.083×10^{-2}	9.908×10^{-5}	9.203×10^{-2}
	ARCH GED	6.938×10^{-5}	5.279×10^{-2}	5.768×10^{-5}	3.837×10^{-2}	7.971×10^{-5}	3.162×10^{-2}
MSE	ARCH SGED	5.899×10^{-5}	8.633×10^{-2}	6.047×10^{-5}	8.383×10^{-2}	8.826×10^{-5}	8.403×10^{-2}
	ARCH SNIG	5.253×10^{-5}	5.289×10^{-2}	5.140×10^{-5}	3.754×10^{-2}	7.558×10^{-5}	3.114×10^{-2}

Table 9 shows that the ARCH STD model provides the smallest values for all loss functions at p = 0.05 in different window periods w = 50; 100; 150. Therefore, we conclude that ARCH STD is the best model for characterizing NSE equity indices in terms of performance and predictive ability.

V. CONCLUSION

This study used the ARCH model with seven distributions for innovation as a tool to characterize NSE and provide an accurate prediction for the series.

The findings of the present study show that the ARCH model the Student's t-distribution yields lower values for AIC, BIC, and HQC. We, therefore, conclude that the ARCH model with Student's t-distribution outperforms other distributions in terms of performance and predictability. Hence, the study recommends carrying out future work with the artificial neural network and comparing the approach with the one used in the present study.

REFERENCES

[1] Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE transaction on Automatic Control*, 6(19): 716-723

[2] Anderson, T.W and Darling, D.A (1954).A Test of Goodness of Fit.*J Am Stat Assoc.*, 49: 765-69.

[3] Arum, K.C. and Uche, P.I. (2019). Volatility Modeling Using ARCH and GARCH MODELS (A Case Study of the Nigerian Stock Exchange) *International Journal of Mathematics Trends and Technology*, 65(4): 511-522.

[4] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones.*Statistica*, 46, 199-208.

[5] Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size.*Proceedings of the Royal Society of London: Series A, Mathematical and Physical Sciences*, 353, 401-409.

[6] De Moivre, A. (1738). *The Doctrine of Chances.*ISBN 0821821032.

[7] Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with estimates of variance of United Kingdom inflations. *Econometrica*, 50(4): 987-1008.

[8] Fernandez, C and Steel, M.F.J. (1998).On Bayesian modeling of fat tails and skewness.*Journal of the American Statistical Association*, 93, 359-371.

[9] Gauss, C. F. (1809). Theoriamotvscorporvmcoelestivm in sectionibvsconicisSolemamientivm (in Latin).

[10] Gosset, W. S. (1908). The probable error of a mean.*Biometrika*, 6, 1-25.

[11] Jarque C. M and Bera A. K (1980).Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals.*Economics Letters*, 6(3): 255-259.

[12] Latha, S. (2019).Volatility Estimation Using Symmetric and Asymmetric Models in Oil Exporting Emerging.*Asian Journal of Finance and Accounting*, 11(1): 41. DOI: 10.5296/ajfa.v11i1.14202

[13] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices.*Journal of Business*, 36: 394-419.

[14] Moffat, I.U. and Akpa, E. A. (2018).Modeling Heteroscedasticity of Discrete-Time Series in the Face of Excess Kurtosis.*Global Journal of Science Frontier Research*, 18(7): 21-32.

[15] Okonkwo, J.J. (2019). Volatility of Stock Return and Selected Macroeconomic Variables: International *Journal of Academic Research in Business and Social Science*, 9(6): 185-201. DOI: 10.6007/IJARBSS/v9-i6/5934

[16] Owen, D. (1980). A table of normal integrals.*Communications in Statistics – Simulation and Computation*, 9, 389-419.

[17] Subbotin, M. T. (1923). On the law of Frequency of errors.*MatematicheskiiSbornik*, 31, 296-301.

[18] Theodossiou, P. (1998). Financial data and the skewed generalized t distribution.*Management Science*, 44, 1650-1661.

[19] Tsay, R.S. (2011). Analysis of Financial Time Series. (3rd ed). New York: John Wiley & Sons Inc. 97-140.