Modelling Returns of Four Africa's Stock Market Using a Flexible Distribution

By

Charles Chinedu Nworu (charles.c.nworu@aims-senegal.org)
African Institute for Mathematical Sciences (AIMS), Senegal

Supervised by: Dr. Idika E. Okorie

University of Manchester, United Kingdom

February 2020

AN ESSAY PRESENTED TO AIMS SENEGAL IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MATHEMATICAL SCIENCES

DECLARATION

This work was carried out at AIMS Senegal in partial fulfilment of the requirements for a Master of Science Degree.

I hereby declare that except where due acknowledgement is made, this work has never been presented wholly or in part for the award of a degree at AIMS Senegal or any other University.

Student: Charles Chinedu Nworu (scan) student's signature

Supervisor: Idika E. Okorie (scan) supervisor's signature

Tutor: Firstname Middlename Surname (scan) tutor's signature

ACKNOWLEDGEMENTS

I want to use this opportunity to express my deep gratitude to the following people for helping and supporting me during this research project, writing process and my stay at AIMS Senegal.

Special thanks goes to my Supervisor, Dr. Idika E. Okorie for providing enthusiastic encouragement and useful critiques of this research work. I would like to say a very big thank you to Prof. Saralees Nadarajah who ignited my interest in probability during our course work.

To my tutor, Mouhamad Allaya for providing guidance and feedback. My sincere gratitude to my family for their continuous and unparalleled love, help and support.

I am also indebted to Prof. J.C. Nwabueze, Dr. C.K. Acha, Dr. F.C. Nwadire, Mr. John Chisimkwuo, Mr. George Uchechukwu for their morale support.

Finally, the friendship and support of my friends and relatives has been wonderful, I would specially want to thank Joy Agwu (Joy kaa!), Samuel Nworu, Mercy Ekekwem, Joana Ewuntomah from Ghana, Olyne Niyonwonyere from Burundi, Mansour Dia from Senegal, Ememmanuel Ndahimana from Rwanda, whao!, the list is unexhaustive. A big thank you for making my stay at AIMS Senegal a memorable one...

DEDICATION

This work is dedicated to God Almighty for seeing me through my Master's program.

Abstract

The aim of this project work is to find more flexible distributions that adequately describe stock returns of four emerging markets in Africa, namely Cote D'Ivoire, Egypt, Nigeria and South Africa. The sample period covers 5203, 5387, 4856, and 6041 trading days respectively. The obtained models were considered in estimating Value at Risk (VaR) at various confidence levels. Evaluation of VaR model accuracy was based on Kupiec likelihood ratio test.

Contents

1. Introduction

After the 1987 financial crisis that rocked the financial service firms, financial institutions and regulatory bodies adopted the metric, value-at-risk (VaR) as a standard tool in financial risk management. This is because VaR provides a single value that summarizes the overall market risk in individual stocks and for portfolios (Hull, 2012).

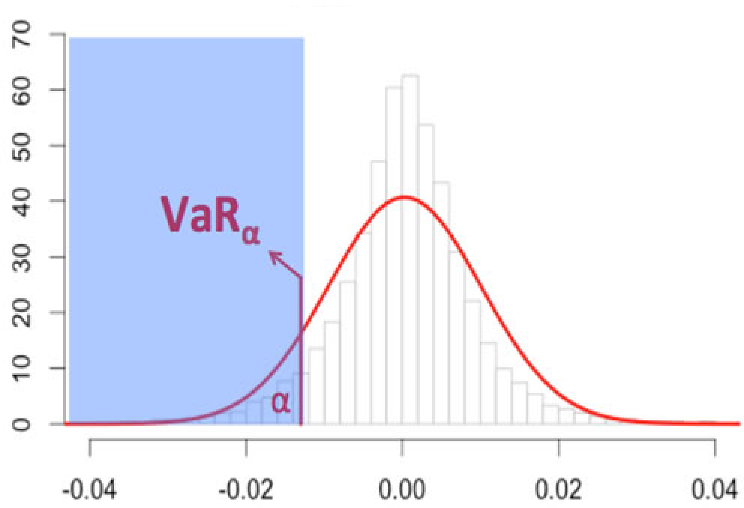

Figure 1.1: The Value at Risk of a hypothetical profit-and-loss probability density function at α level of significance.

For a given portfolio, time horizon, and probability p, the VaR calculates the maximum loss expected (or worst case scenario) on an investment during that time after we exclude all worst outcomes whose combined probability is at most p (Wikipedia, 2019). The correct estimation of VaR enables financial institutions to provide adequate capital requirements and to keep at pace the random behaviours associated with the stock market (Jadhav and Ramanathan, 2009). Jadhav and Ramanathan (2009) noted that if the underlying risk is under-estimated, this may lead to a suboptimal capital allocation thereby putting the financial stability of the institution in jeopardy, on the otherhand, over-estimation of risk may lead to unnecessary extra capital requirements. While the concept of VaR is easy to understand, the statistical estimation has become a topic of discourse by researchers. VaR analysis is done in three phases: First, we express the asset's profit and loss in terms of returns, model the returns statistically and then estimate the VaR of returns using the best fitted model. Jadhav and Ramanathan (2009) pointed out that the estimation of VaR involves two procedures: parametric and non-parametric.

The parametric VaR assumes that the financial returns are modelled using a statistical distribution, whereas the non-parametric VaR assumes that the financial returns are modelled from the empirical distribution (Mabitsela et al., 2015). The conventional parametric VaR assumes that asset returns are normally distributed but a lot of studies have shown that daily financial returns exhibit non-zero skewness and excess kutosis as noted by Thompson (2013) and hence, the VaR measure either underestimates or overestimates the true risk (Alexander, 2001). The distribution of stock returns are known to be characterised by heavy tails and higher peak (lepkurtosis). This implies that extreme values and values close to the mean are more probable than under the normal distribution. Therefore, the quality of VaR estimate hugely depend on how well a flexible statistical distribution captures the lepkurtic behaviour of the financial returns not just in the centre of the distribution but in the extreme tail as well (Dorić, 2016). The introduction of VaR has attracted a lot of scholarly contributions with the aim of finding a suitable distribution as an alternative to the Normal distribution.

Terzić et al. (2014) carried out an empirical study on the estimation of VaR using the Serbian daily stock return from January, 2011 to January, 2014. They estimated a 1-day Normal and Student t VaR for different significance levels. The lepkurtic behaviour evident in returns data were tested and found. The outcome of their work showed that for low significance levels, the normality assumption overestimate VaR, additionally, for higher significance levels, the normality assumption seriously underestimate VaR. They concluded that the Student t distribution produced better VaR estimates than the normal distribution.

Corlu and Corlu (2015) investigated the performance of the Generalized Lambda Distribution (GLD) to capture the skewed and high peak behaviour of exchange rate of nine currencies for the period 2006-2011. A comparison study was conducted to show the superiority of the GLD over the skewed t distribution, the unbounded Johnson family of distributions and the Normal Inverse Gaussian (NIG). They computed their various VaR and Expected Shortfall (ES) values for different significance levels. They concluded that in terms of risk management of exchange rate returns, the GLD provided an alternative fit.

Mabitsela et al. (2015) evaluated the South African equity markets in a Value-at-Risk (VaR) framework. VaR was estimated on four equity stocks listed on the Johannesburg Stock Exchange, including the FTSE/JSE TOP40 index and the S & P 500 index. The statistical distribution of the financial returns was modelled using the Normal Inverse Gaussian and was compared to the financial returns modelled using the Normal, Skew t-distribution and Student t-distribution. They estimated VaR values under the assumption that financial returns follow the Normal Inverse Gaussian, Normal, Skew t-distribution and Student t-distribution and their qualities were assessed by backtesting the VaR estimates under each distribution assumption. Their results showed that the performance of each model was dependent on a type of dataset.

Up until now, there has not been a universally accepted distribution for modelling financial time series. The performance of a distribution may be a function of the properties in a particular dataset, therefore, the goal of this work is to model stock price return data from four emerging markets, namely Cote D'Ivoire, Egypt, Nigeria and South Africa using five flexible statistical distributions, that is the Student t, Generalized Lambda, Normal Inverse Gaussian, Hyperbolic and Skewed t distributions with the aim of selecting the best fitted model. The best fitted model is used to predict future losses of these stock prices using the VaR metric. Finally, we backtest the estimates of the VaR in order to assess the usefulness of the best VaR model using the Kupiec's likelihood ratio test

2. Notations

In this chapter, we will discuss the basic concept of time series analysis, log returns, distributions of financial time series data and the concept of continuous random variables.

2.1 Time Series

The data used for this work is a time series data collected from DataStream, therefore it is important that we briefly discuss the concept of time series. A time series is a sequence of observations taken at equally spaced in discrete time, that is, we will have a single realization of X at each second, hour, day, month, year, etc. It is a realization of the stochastic process.

Definition 2.1.1: A time series process is a set of random variables $\{X_t, t \in T\}$, where T is a set of times at which the process is observed. Each random variable X_t is assumed to be distributed according to some univariate distribution function F_t.

2.1.1 Stationarity.

One of the conditions of the joint distribution F_t is the concept of stationarity. Consider the joint probability distribution of the collection of random variables

$$F(x_{t_1}, x_{t_2}, ..., x_{t_n}) = P(X_{t_1} \leq x_{t_1}, X_{t_2} \leq x_{t_2}, ..., X_{t_n} \leq x_{t_n}) \tag{2.1.1}$$

Definition 2.1.1.1 A time series is said to be **strictly stationary** if the joint distribution of $X(t_1), ..., X(t_n)$ is the same as the joint distribution of $X(t_1+\tau), ..., X(t_n+\tau)$ for all $t_1, ..., t_n, \tau$ (Chatfield, 2013). In otherwords the joint distribution function is not affected by shifting the time origin by an amount τ, which must therefore depend only on the intervals between $t_1, t_2, ..., t_n$.

If n=1, strict stationarity implies that the distribution of $X(t)$ is the same for all t, that is

$$\begin{aligned} \mu(t) &= \mu \\ \sigma^2(t) &= \sigma^2 \end{aligned} \tag{2.1.2}$$

are both constants which do not depend on the value of t, provided that $E[X_t] < \infty$ and $E[X_t^2] < \infty$. Furthermore, if n=2 the joint distribution of $X(t_1)$ and $X(t_2)$ depends only on the lag, $(t_2 - t_1)$. The autocovariance function $\gamma(t_1, t_2)$ depends only on $(t_2 - t_1)$ and may be written as $\gamma(\tau)$, where

$$\begin{aligned} \gamma(\tau) &= E\{[X(t) - \mu][X(t+\tau) - \mu]\} \\ &= Cov[X(t), X(t+\tau)] \end{aligned} \tag{2.1.3}$$

is the autocovariance coefficient at lag τ.

The autocorrelation function is given by:

$$\rho(\tau) = \frac{\gamma(t)}{\gamma(0)} \tag{2.1.4}$$

measures the correlation between X(t) and $X(t+\tau)$

Definition 2.1.1.2 A process is said to be n-order **weakly stationary** if all its joint moments up to order n exist and are time invariant (González-Rivera and Gonzalo, 2002). In otherwords, we say a process is called second-order stationary (or weakly stationary) if has constant mean and variance and its autovariance function depends only on the lag. For a covariance stationary process:

$$
\begin{aligned}
& E[X(t)] = \mu \\
& Var[X(t)] = \sigma^2 \\
& Cov[X(t), X(t+\tau)] = \gamma(t) \\
\text{and} \\
& \rho(\tau) = \frac{\gamma(t)}{\gamma(0)}
\end{aligned}
\tag{2.1.5}
$$

$$
\begin{aligned}
\gamma(\tau) &: \text{ autocovariance function} \\
& \gamma : \tau \rightarrow \mathrm{R} \\
\rho(\tau) &: \text{ autocorrelation function} \\
& \rho : \tau \rightarrow [-1, 1]
\end{aligned}
$$

In order to carry out a 'neat' statistical analyses with time series, weak stationarity is a necessary condition. If this condition fails, it is impossible to learn from it. Unfortunately, stock market data are often non-stationary. The way out is to study the log-returns, an approximation to the relative changes in the series.

2.1.2 Stationarity Test.

The augmented Dickey Fuller test is a common statistical test used to test whether a given time series is stationary or not. Given an observed time series $X_1, X_2, ..., X_n$, Dickey and Fuller considered three differential form autoregressive equations to detect the presence of a unit root (non-stationarity).

$$\Delta X_t = \gamma X_{t-1} + \sum_{j=1}^{p} (\delta_j \Delta X_{t-j}) + e_t \tag{2.1.6}$$

$$\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{j=1}^{p} (\delta_j \Delta X_{t-j}) + e_t \tag{2.1.7}$$

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{j=1}^{p} (\delta_j \Delta X_{t-j}) + e_t \tag{2.1.8}$$

where α is the constant called the drift, β is the coefficient on the time trend, p is the lag order of the autoregressive process, γ is the coefficient presenting process root, that is the focus of testing, e_t is the independent identically distributed residual term, and t is the time index.

The null and alternative hypotheses corresponding to these models are:

$$\begin{aligned} (h1) \quad & H_0 : X_t \text{ is random walk or } \ \gamma = 0 \\ & H_1 : X_t \text{ is stationary} \end{aligned} \tag{2.1.9}$$

$$\begin{aligned} (h2) \quad & H_0 : X_t \text{ is random walk around a drift or } \ \{\gamma = 0, \alpha \neq 0\} \\ & H_1 : X_t \text{ is level stationary process or } \ \{\gamma < 0, \alpha \neq 0\} \end{aligned} \tag{2.1.10}$$

$$\begin{aligned} (h2) \quad & H_0 : X_t \text{ is random walk around a trend or } \ \{\gamma = 0, \beta \neq 0\} \\ & H_1 : X_t \text{ is trend stationary process or } \ \{\gamma < 0, \beta \neq 0\} \end{aligned} \tag{2.1.11}$$

The ADF tests ensures that the null hypothesis is accepted unless there is a strong evidence against it to reject in favour of the alternative stationary process.

2.2 Simple Returns and Log Returns

Studies have shown that financial time series data are non-stationary in nature Iordanova and Data (2009). Using such data in financial models often produces spurious and unreliable results and thus leads to poor understanding and forecasting. The solution to this problem is to transform the time series data so that it becomes stationary.

Definition 2.2.1: Consider a financial market data (which consists of instruments such as stocks, shares or derivatives). Let P_t be the price of the security at time t.

The linear or simple return between times t and t - 1 is defined as (**?**):

$$R_t = \frac{P_t}{P_{t-1}} - 1 \tag{2.2.1}$$

The log return is defined as (Dettling, 2017):

$$r_t = 100 * \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2.2.2}$$

where P_t is the price at time t, r_t is the return between time t - 1 and t.

2.3 The Random Walk Model

One of the most important models in time series forecasting is the random walk model (Nau, 2014). This model is based on the notion that single-period log returns $r_{1,t}$ are independent and normally distributed.

For a multi-period returns expressed as follows:

$$r_{k,t} = r_{1,t} + ... + r_{1,t-k+1} \tag{2.3.1}$$

This implies that the k -period return at time t is the sum of all single-period returns back to time t-k+1 . If one (conveniently) further assumes that the log returns follow a Gaussian distribution, we have $r_{1,t} \sim N(\mu, \sigma^2)$. Because sums of independent normal random variables are themselves normal, it is obvious that $r_{k,t} \sim N(k\mu, k\sigma^2)$. If the time series is an i.i.d. sequence, then a random walk model is a potentially good candidate.

2.3.1 Test of Independence.

Many ways exist to test the independence assumption of the random walk hypothesis. One of them is the runs test. The runs test (Bradley, 1968) is a statistical procedure that examines whether a string of data is occuring randomly given a specific distribution. The runs test analyzes the occurrence of similar events that are separated by events that are different. A run is defined as a succession of similar events proceeded and followed by a different event. The hypothesis is given by:

H_0 : Independence (log returns are random or follow a random walk process)

H_1 : Dependence (log returns are not random or do not follow a random walk process)

The test statistic is given by:

$$Z = \frac{R - \bar{R}}{S_R} \tag{2.3.2}$$

where R is the observed number of runs, $\bar{R}$, is the expected number of runs, and S_R is the standard deviation of the number of runs.

The values of $\bar{R}$ and S_R are computed as follows:

$$\bar{R} = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{2.3.3}$$

$$S_R^2 = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \tag{2.3.4}$$

with n_1 and n_2 denoting the number of positive and negative values in the series. The runs test rejects the null hypothesis if $|Z| > Z_{1-\frac{\alpha}{2}}$.

2.4 Stylised Facts of Stock Market Returns

The stylised facts of the stock market returns are statistical properties expected to be present in any sufficiently long series of observed stock prices. For the purpose of this work, we will examine some of them that are very common to a wide set of financial returns. They include:

1. Heavy tails
2. Absence of autocorrelations
3. Volatility clustering.

4. Slow decay of autocorrelation in absolute returns

5. The standard deviation of returns completely dominates the mean of returns at short horizons such as daily.

6. Presence of some extreme spikes, that is, outliers that correspond to very big/small returns.

2.4.1 Non-normality of Returns.

1. Skewness and Kurtosis

One of the most predominant stylised facts of stock returns is its peaked and heavy tailed nature. Empirical results have shown that most daily stock return series tend to exhibit these characteristics (Pagan, 1996). When returns fall outside of a normal distribution, the distribution exhibits skewness or kurtosis.

The sample Kurtosis of returns can be defined as:

$$\hat{K_r} = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t - \bar{r})^4}{S_r^4} \tag{2.4.1}$$

equals three if the distribution is symmetric.

Due to the popularity of the Gaussian, it is common to compute the excess kurtosis, which is simply: $\hat{K}_{rexcess} = \hat{K_r} - 3$, For heavy tail distributions, if $\hat{K}_{rexcess} > 0$, it is called a *leptokurtic* distribution.

Another statistic relevant to the distribution of a stock returns is the skewness. Skewness measures the degree of asymmetry of a distribution around its mean. Positive skewness describes a return distribution where frequent small losses and a few extreme gains are common while negative skewness highlights frequent small gains and a few extreme losses.

The skewness point estimator is the following statistics

$$S\hat{K}_r = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t - \bar{r})^3}{S_r^3} \tag{2.4.2}$$

where $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and $S_r = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2}$

For $S\hat{K}_r > 0$, it indicates positive skewness, i.e. distributions that have a heavy tail on the right hand side. Conversely, $S\hat{K}_r < 0$ indicates a left-skewed distribution.

2. Q-Q plot

The Q-Q plot or quantile-quantile plot is a simple graphical method of testing the goodness of fit of observed returns to the Normal distribution. The Q-Q plot reveals how each stock deviate from the straight line at the tails of the distribution. They are easy to implement, but they are less formal and hence reliance on them can lead to poor conlusions Mabitsela (2015). The Q-Q plot is defined by:

$$\{\left(r_{(i)}, \Phi^{-1}\left(\frac{n-i+1}{n+1}\right)\right), i = 1, 2, ..., n\} \tag{2.4.3}$$

where $r_{(i)}$ denotes the ith order statistics and

Φ^{-1} is the inverse of the cummulative standard gaussian distribution function.

3. The Jarque-Bera (JB) Test of Normality

The JB test is one of the formal techniques used to test that the log returns follows a normal distribution (Mabitsela et al., 2015). In the formal framework of hypothesis testing the null hypothesis H_0 is that the log returns are normally distributed verses the alternative hypothesis H_1 which states that the log returns are not normally distributed.

The JB test is based on the third and fourth sample moments. The test combines the sample skewness and kurtosis to that of the Normal distribution and it is mostly used when the samples is very large ($n > 2000$). The test is defined as:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \sim \chi_2^2 \tag{2.4.4}$$

where S and K are sample skewness and kurtosis respectively. If our data is drawn from a normally distributed population then the value of JB increases and the null hypothesis will be rejected.

2.4.2 Absence of autocorrelations.

Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals (Smith, 2019). An autocorrelation of $+1$ represents a perfect positive correlation (an increase seen in one time series leads to a proportionate increase in the other time series). An autocorrelation of negative 1, on the other hand, represents perfect negative correlation (an increase seen in one time series results in a proportionate decrease in the other time series).

Daily returns have very little autocorrelation . We therefore write

$$Corr(r_{t+1}, r_{t+1-r}) \approx 0, \text{for } \tau = 1, 2, 3, ... \tag{2.4.5}$$

Technically speaking this means that the returns are impossible to predict from their own past. The Autocorrelation (ACF) plot is a visual way to show serial correlation in data that changes over time.

2.4.3 Volatility clustering.

Variance measured for example by the squared returns displays positive correlation with its own past.

Mathematically, this is defined as

$$Corr(r_{t+1}^2, r_{t+1-\tau}^2) > 0, \text{for small } \tau \tag{2.4.6}$$

Technically, this could be thought of, as the volatility of the past being used to predict the standard deviation of the future values

2.5 Continuous Random Variables

If the set of all possible values of X is an interval or union of two or more nonoverlapping intervals, then X is called a continuous random variable.

Definition 2.5.1. Any real valued function f(x) that meets the following conditions is called a probability density function (pdf):

$$f(x) \geq 0 \quad \text{for all x, and} \quad \int_{-\infty}^{\infty} f(x)dx = 1 \tag{2.5.1}$$

Definition 2.5.2. The cummulative density function (cdf) of a random variable X is defined by:

$$F(x) = P(X \leq x) \quad \text{for all x} \tag{2.5.2}$$

For a continuous random variable X with the probability density function f(x), the cdf is given by:

$$P(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad \text{for all x} \tag{2.5.3}$$

Definition 2.5.3: Let X be a random variable with the cdf F(x). For a given $0 < p < 1$, the inverse of the distribution function is defined by:

$$F^{-1} = inf\{x : P(X \leq x) = p\} \tag{2.5.4}$$

Definition 2.5.4: The moments about the origin are obtained by finding the expected value of the random variable with the power k, k = 1, 2,.... That is,

$$\mu_k' = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx \tag{2.5.5}$$

is called the kth moment about the origin or the kth raw moment of X.

Definition 2.5.5: The kth moment about the mean or the kth central moment of a random variable X is defined by

$$\mu_k = E(X - \mu)^k, \quad \text{k = 1, 2,...} \tag{2.5.6}$$

Definition 2.5.6 (Law of Large Numbers, LLN) If the distribution of the iid sample $X_1, ..., X_n$ is such that X_1 has finite expectation, that is, $|\mathbb{E}X_1| < \infty$, then the sample average,

$$\bar{X}_n = \frac{X_1 + ... + X_n}{n} \to \mathbb{E}X_1 \tag{2.5.7}$$

converges to its expectation in probability, which means that for any arbitrary small $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - EX_1 > \varepsilon) \to 0 \ \text{ as } \ n \to \infty \tag{2.5.8}$$

Definition 2.5.7 (Central Limit Theorem, CLT) If the distribution of the iid sample $X_1, ..., X_n$ is such that X_1 has finite expectation and variance, that is, $|\mathbb{E}X_1| < \infty$ and $Var(X) < \infty$, then

$$\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

converges in distribution to normal distribution with zero mean and variance σ^2, which means that for any interval [a, b],

$$\mathbb{P}(\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \ \in [a,b]) \to \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} dx \tag{2.5.9}$$

3. Financial Models, Parameter Estimation, Goodness of Fit Tests

In this chapter, we are going to examine some continuous models that can handle financial time series data, their method of parameter estimation and the approaches for evaluating two or more distributions.

3.1 Financial Models

We will examine five parametric distributions for modeling financial data. The best fitted distribution enables us to describe the behaviour of the underlying financial data.

Let X denote a continuous random variable representing the daily log returns of the stock market data. Let $f(x)$ denote the probability density function of X and $F(x)$ the cummulative distribution function of X.

3.1.1 Normal distribution.

The p.d.f is given by:

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \begin{aligned} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{aligned} \tag{3.1.1}$$

where μ is the mean or average value of the daily log returns

σ is the volatility.

Normal distributions are continuous and have tails that are asymptotic, which means that they approach but never touch the x-axis. Generally, if a variable has a higher variance (that is, if a wider spread of values is possible), then the curve will be broader and shorter. However, if the variance is small (where most values occur very close to the mean), the curve will be narrow and tall in the middle. The behaviour of a normal distribution with zero mean for different volatility values is demonstrated in Fig. 3.1

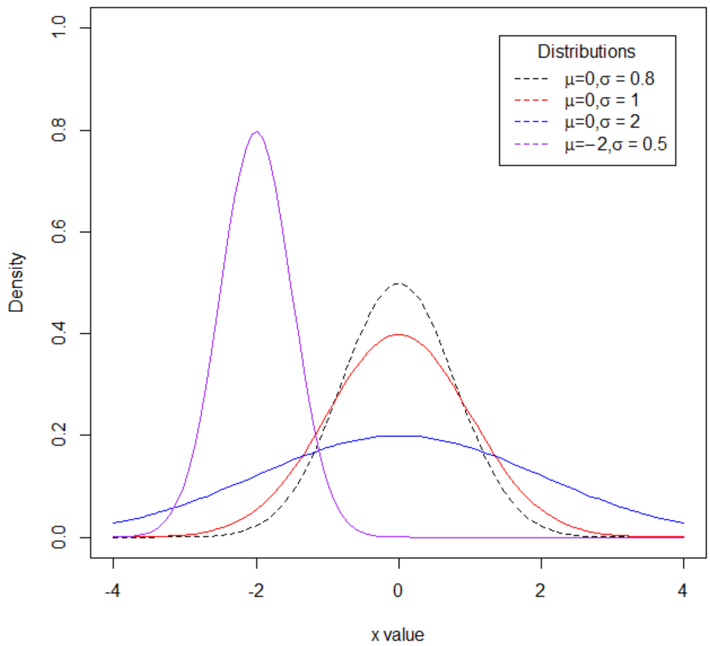

Figure 3.1: The density function of a normal distribution with varying location parameters and scaling parameters.

3.1.2 The Student's t-distribution.

The t-distribution due to Blattberg and Gonedes (1974) is an alternative distribution to model financial returns, is characterized by the shape-defining parameter known as the degree of freedom k¿0. The pdf is given by:

$$f_k(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k\pi)\ \Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \qquad x \in (-\infty, \infty) \tag{3.1.2}$$

where k is the degree of freedom and Γ is the gamma function. The t-distribution is similar to the Normal distribution, it is symmetrical about the mean and exhibits heavier tails. It has mean zero when $k > 1$,otherwise the mean is undefined.

The variance is given by

$$\frac{k}{k-2} \quad \text{for} \quad k > 2, \quad \infty \quad \text{for} \quad 0 < k \leq 2 \tag{3.1.3}$$

otherwise undefined.

The skewness is zero (0) for $k > 0$, otherwise undefined. The excess kurtosis is given by

$$\frac{6}{k-4} \quad \text{for} \quad k > 4, \quad \infty \quad \text{for} \quad 2 < k \leq 4 \tag{3.1.4}$$

otherwise undefined.

The degree of freedom k controls the fat tails of the distribution. The smaller the value of k the fatter the tails of the distribution.

The t-distribution could be very useful for financial analysis because it can adapt to the tail behavior of a financial data, though it's not a very flexible model because of the absence of a location and/or scale parameter.

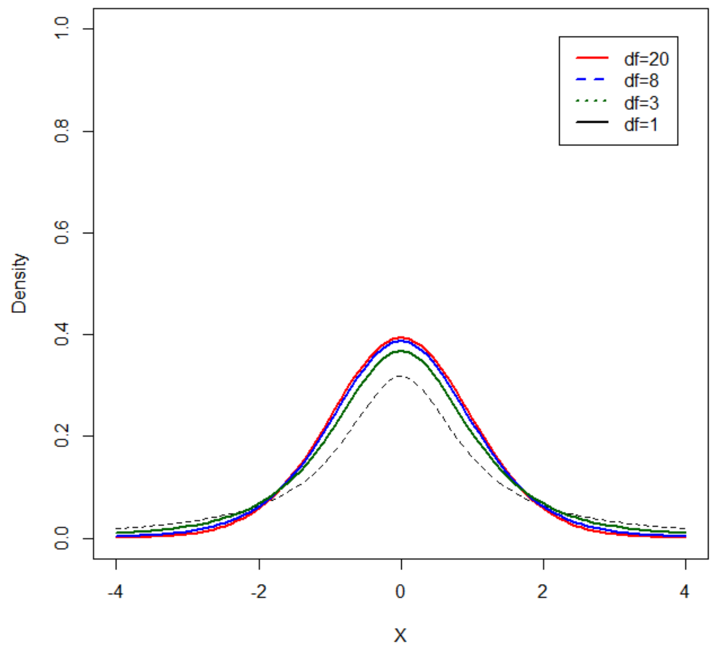

Figure 3.2: The density function of a t-distribution with degrees of freedom 1,3,8,20 respectively.

3.1.3 Generalized Lambda Distribution (GLD).

The GLD is an extension of the Tukey lambda distribution (TLD) (Hastings et al., 1947). Ramberg and Schmeiser (1974) generalized the TLD to four parameters called the RS parameterization:

$$Q_{RS}(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \tag{3.1.5}$$

where $Q_{RS} = F_{RS}^{-1}$ is the quantile function for the probabilities u; $0 \leq u \leq 1$; λ_1 is the location parameter, λ_2 is the scale parameter, λ_3 and λ_4 are related to the skewness and kurtosis respectively which are the shape parameters.

However of the limitations of (3.1.5) is that it does not specify a proper probability density function (pdf) for all combinations of λ_3 and λ_4. To overcome this limitation, Freimer et al. (1988) proposed a different parameterization for the GLD denoted by Freimer-Mudholkar-Kollia-

Lin Generalized Lambda Distribution (FMKL GLD), which is given by:

$$Q_{FMKL}(u) = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{u^{\lambda_3}-1}{\lambda_3} - \frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right) \tag{3.1.6}$$

As in the previous parameterization, λ_1 and λ_2 are the location and scale parameters respectively, and λ_3 and λ_4 are the tail index parameters; the only constraint is that λ_2 must be positive. One of the advantages of the GLD in financial modeling is that the shape range of the GLD family is so large that it can accommodate almost any financial time series (Chalabi et al., 2012).

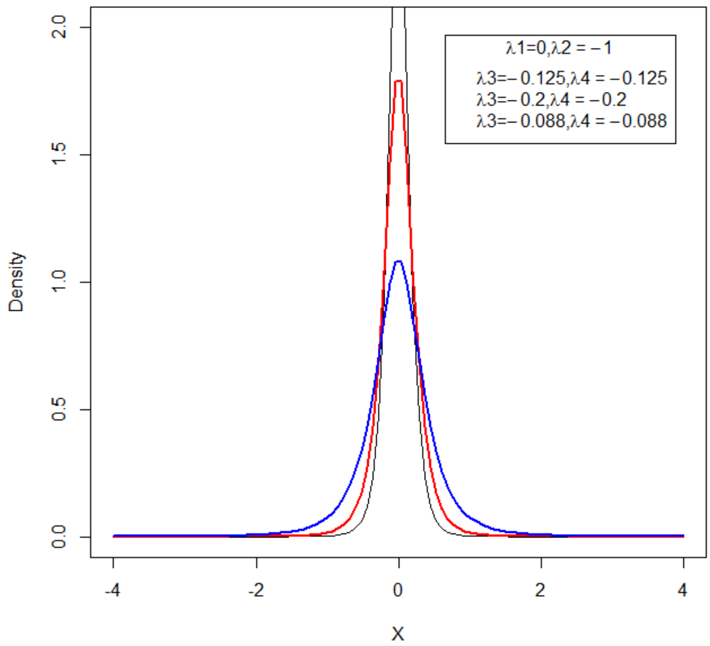

Figure 3.3: The behaviour of the density function of the GLD with fixed location and scale parameters and varying shape parameters

The moments of the FMKL parameterizations of the GLD is given as follows:

$$v_1 = \frac{1}{\lambda_3+1} - \frac{1}{\lambda_4+1} \tag{3.1.7}$$

$$v_2 = \frac{1}{\lambda_3^2(3\lambda_3+1)} + \frac{1}{\lambda_4^2(3\lambda_4+1)} - \frac{2}{\lambda_3\lambda_4}B(\lambda_3+1,\lambda_4+1) \tag{3.1.8}$$

$$\begin{aligned} v_3 = \frac{1}{\lambda_3^3(3\lambda_3+1)} - \frac{1}{\lambda_4^3(3\lambda_4+1)} - \frac{3}{\lambda_3^2\lambda_4}B(2\lambda_3+1,\lambda_4+1) \\ + \frac{3}{\lambda_3\lambda_4^2}B(\lambda_3+1,\lambda_4+1) \end{aligned} \tag{3.1.9}$$

$$\begin{aligned} v_4 = \frac{1}{\lambda_3^4(4\lambda_3+1)} + \frac{1}{\lambda_4^4(4\lambda_4+1)} + \frac{6}{\lambda_3^2\lambda_4^2}B(2\lambda_3+1,\lambda_4+1) \\ - \frac{4}{\lambda_3^3\lambda_4}B(3\lambda_3+1,\lambda_4+1) - \frac{4}{\lambda_3\lambda_4^3}B(\lambda_3+1,3\lambda_4+1) \end{aligned} \tag{3.1.10}$$

3.1.4 Normal Inverse Gaussian (NIG) distribution.

The univariate NIG due to Barndorff-Nielsen and Halgreen (1977) is a subclass of the generalized hyperbolic distribution. Its probability density function is given by:

$$f(x,\mu,\alpha,\beta,\delta) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}e^{\delta\gamma+\beta(x-\mu)}, \quad x\in(-\infty;+\infty) \tag{3.1.11}$$

where $K_1(x)$ denotes a modified Bessel function of the third kind indexed by 1 and is given by

$$\frac{1}{2}\int_0^\infty exp(-\frac{1}{2}x(\tau+\tau^{-1}))d\tau \tag{3.1.12}$$

and

$$\gamma = \sqrt{\alpha^2-\beta^2} \tag{3.1.13}$$

The parameters must satisfy

$$0 \le |\beta| \le \alpha$$
$$\mu \in \mathbb{R}$$
$$0 < \alpha \text{ and } 0 < \delta$$ (Prause et al., 1999)

μ is the location parameter, α, the tail heaviness, β is the asymmetry parameter, and δ is the scale parameter.

Definition 2.6.1 The moment generating function of NIG distribution is given by:

$$\begin{aligned} M_x(t) &= \int_{-\infty}^{\infty} e^{tx}.f(x;\mu,\alpha,\beta,\delta)dx \\ &= exp\left[t\mu + \delta\left(\sqrt{\alpha^2+\beta^2} - \sqrt{\alpha^2-(\beta+t)^2}\right)\right] \end{aligned} \tag{3.1.14}$$

The mean, variance, skewness and the kurtosis of random variable x can be obtained by successively differentiating the moment generating function.

Definition 2.6.2: The mean, variance, skewness and kurtosis of the random variable x distributed according to (3.1.11) are given by the following equations

$$E[X] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} \tag{3.1.15}$$

$$Var[X] = \frac{\delta\alpha^2}{(\alpha^2-\beta^2)^{3/2}} \tag{3.1.16}$$

$$Skew[X] = 3\alpha^{-1/4}\frac{\beta/\alpha}{(1-(\beta/\alpha)^2)^{1/2}} \tag{3.1.17}$$

$$Kurt[X]_{excess} = 3\,\frac{1+4(\beta^2/\alpha^2)}{\delta\sqrt{(\alpha^2+\beta^2)}} \tag{3.1.18}$$

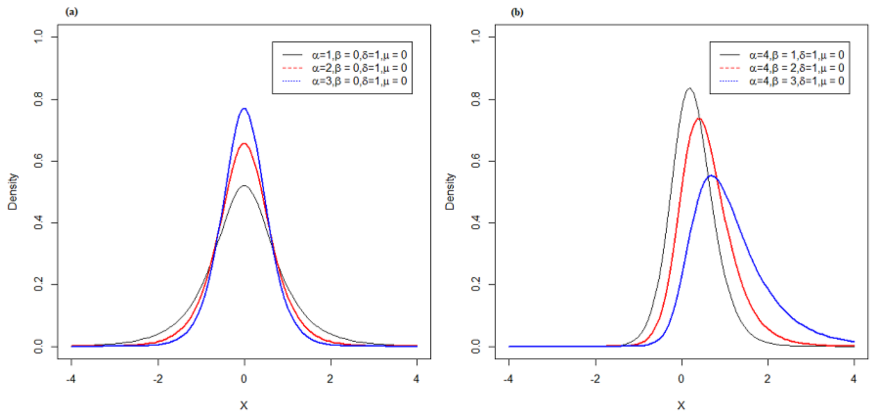

Figure 3.4: The density function of an NIG distribution with (a) fixed location, asymmetry, and scale parameters with different values of tail heaviness (b) its behaviour with different values of asymmetry parameter.

The class of NIG distributions is a flexible system of distributions that includes fat tailed and skewed distributions, and the Normal distribution $\mathcal{N}(\mu, \sigma^2)$ arises as a special case by setting $\beta = 0$, $\delta = \sigma^2\alpha$, and letting $\alpha \to \infty$. Its flexibility allows to fit financial data well at any frequency.

3.1.5 Hyperbolic distribution.

The hyperbolic distributions is a subclass of the generalized hyperbolic distributions. A random variable X has the hyperbolic distribution if its pdf is defined as (Barndorff-Nielsen and Halgreen, 1977):

$$\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)} exp\{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\} \tag{3.1.19}$$

where K_1 denotes the Bessel function of the third kind with index 1, $\alpha > 0$ and $0 \leq |\beta| < \alpha$, α and β are the shape parameters with α representing the gradient and β, the skewness, $\delta > 0$ is the scale parameter, $\mu \in \mathbb{R}$ is the location parameter.

The mean and variance of the hyperbolic distribution is given by:

$$E[X] = \mu + \frac{\delta\beta K_2(\delta\gamma)}{\gamma K_1(\delta\gamma)} \tag{3.1.20}$$

and

$$Var[X] = \frac{\delta K_2(\delta\gamma)}{\gamma K_1(\delta\gamma)} + \frac{\beta^2}{\gamma^2}\left(\frac{K_3(\delta\gamma)}{K_1(\delta\gamma)} - \frac{K_2^2(\delta\gamma)}{K_1^2(\delta\gamma)}\right) \tag{3.1.21}$$

respectively.

where $\gamma = \sqrt{\alpha^2-\beta^2}$

The skewness is given by:

$$Var[X]^{-3/2}\left[\frac{\beta^3\delta^6}{\zeta^3}\left(\frac{K_4(\zeta)}{K_1(\zeta)}-\frac{3K_3(\zeta)K_2(\zeta)}{K_1^2(\zeta)}+\frac{2K_2^3(\zeta)}{K_1^3(\zeta)}\right)\right.$$
$$\left.+\frac{3\beta\delta^4}{\zeta^2}\left(\frac{K_3(\zeta)}{K_1(\zeta)}-\frac{K_2^2(\zeta)}{K_1^2(\zeta)}\right)\right] \quad (3.1.22)$$

The kurtosis is given by:

$$-3+Var[X]^{-2}\left[\frac{\delta^8\beta^4}{\zeta^4}\left(\frac{K_5(\zeta)}{K_1(\zeta)}-\frac{4K_4(\zeta)K_2(\zeta)}{K_1^2(\zeta)}+\frac{6K_3(\zeta)K_2^2(\zeta)}{K_1^3(\zeta)}\right.\right.$$
$$\left.-\frac{3K_2^4(\zeta)}{K_1^4(\zeta)}\right)+\frac{\delta^6\beta^2}{\zeta^3}\left(\frac{6K_4(\zeta)}{K_1(\zeta)}-\frac{12K_3(\zeta)K_2(\zeta)}{K_1^2(\zeta)}\right. \quad (3.1.23)$$
$$\left.\left.+\frac{6K_2^3(\zeta)}{K_1^3(\zeta)}\right)+\frac{3\delta^4K_3(\zeta)}{\zeta^2K_1(\zeta)}\right]$$

where $\zeta=\delta\sqrt{\alpha^2-\beta^2}$

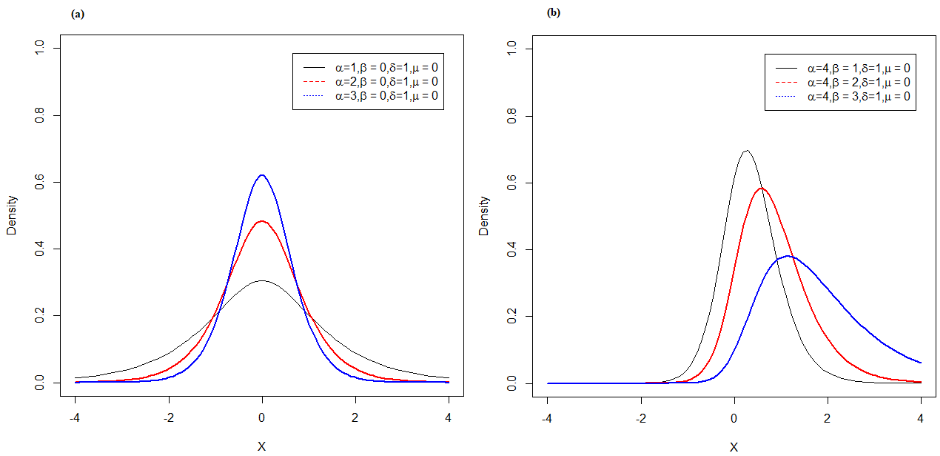

Figure 3.5: The behaviour of the probability density function of an hyperbolic distribution when (a) the gradient is varied with the location, scale and skewness parameters fixed (b) the skewness changes with other parameters constant

As usual, we expect the hyperbolic distribution to be an alternative for heavy tailed data such as the stock log-returns.

3.1.6 Student skewed t-distribution.

The skewed student t distribution proposed in this work is the parameterization done by Azzalini and Capitanio (2003). A random variable X from the skewed student t distribution has the density of the form:

$$f(x;\xi,\omega,\alpha,\nu)=\frac{1}{\omega}t_v\left(\frac{x-\xi}{\omega}\right)2T_{v+1}\left(\alpha\left(\frac{x-\xi}{\omega}\right)\sqrt{\frac{v+1}{\left(\frac{x-\xi}{\omega}\right)^2+v}}\right) \quad (3.1.24)$$

where t_v is the density of standard Student t-distribution with v degrees of freedom and T_{v+1} is the distribution function of the standard Student t-distribution with $v+1$ degrees of freedom, ξ is the location parameter, ω is the scale parameter, α is the slant, and ν is the degree of freedom

The mean, variance, skewness and kurtosis are defined as (Aas and Hobæk Haff, 2005)

$$E[X] = \xi + \frac{\alpha\omega^2}{v-2} \tag{3.1.25}$$

$$\text{Var}[X] = \frac{2\alpha^2\omega^4}{(v-2)^2(v-4)} + \frac{\omega^2}{v-2} \tag{3.1.26}$$

The variance is only finite when $v > 4$

$$S[X] = \frac{2(v-4)^{1/2}\alpha\omega}{[2\alpha^2\omega^2 + (v-2)(v-4)]^{3/2}}\left[3(v-2) + \frac{8\alpha^2\omega^2}{v-6}\right] \tag{3.1.27}$$

$$\begin{aligned} K[X] = \frac{6}{[2\alpha^2\omega^2 + (v-2)(v-4)]^2}\Bigg[&(v-2)^2(v-4) + \frac{16\alpha^2\omega^2(v-2)(v-4)}{v-6} \\ &+ \frac{8\alpha^4\omega^4(5v-22)}{(v-6)(v-8)}\Bigg] \end{aligned} \tag{3.1.28}$$

respectively. The skewness and kurtosis do not exist when $v \leq 6$, and $v \leq 8$ respectively.

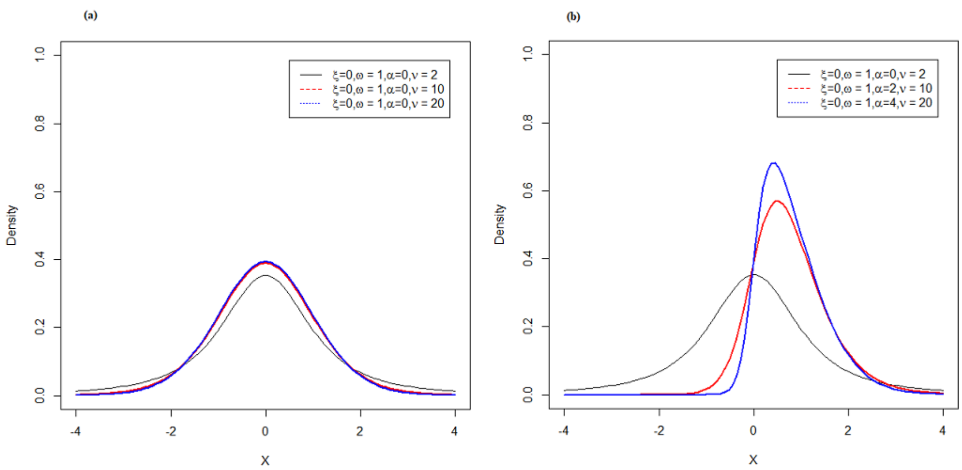

Figure 3.6: The performance of the probability density function of an skewed-t distribution when (a) the degree of freedom is increased while keeping other parameters constant (b) the slant and degree of freedom are increased simultaneously

3.2 Method of Parameter Estimation

Let $(x_1, ..., x_n)$ be a realization of the distribution $\mathbb{P}_{\theta_0}$ and consider a parametric statistical model $\{\mathbb{P}_{\theta_0},\ \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$ and $d < \infty$. We present the approach for the construction of an estimator of θ_0 based on the data $(x_1, ..., x_n)$ known as the **maximum likelihood estimator** (Fisher, 1922).

Definition 3.2.1: Let $X_1, ..., X_n$ be iid with pdf $f(x; \theta)$. The joint density function is given by:

$$f(x_1, x_2, ..., x_n|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times ... \times f(x_n|\theta) \tag{3.2.1}$$

The likelihood function is given by:

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i|\theta) \tag{3.2.2}$$

The log likelihood function is given by:

$$\sum_{i=1}^{n} \log f(X_i|\theta) \tag{3.2.3}$$

Therefore, the maximum likelihood estimator $\hat{\theta}$ is the value that maximizes $\mathcal{L}_n(\theta)$

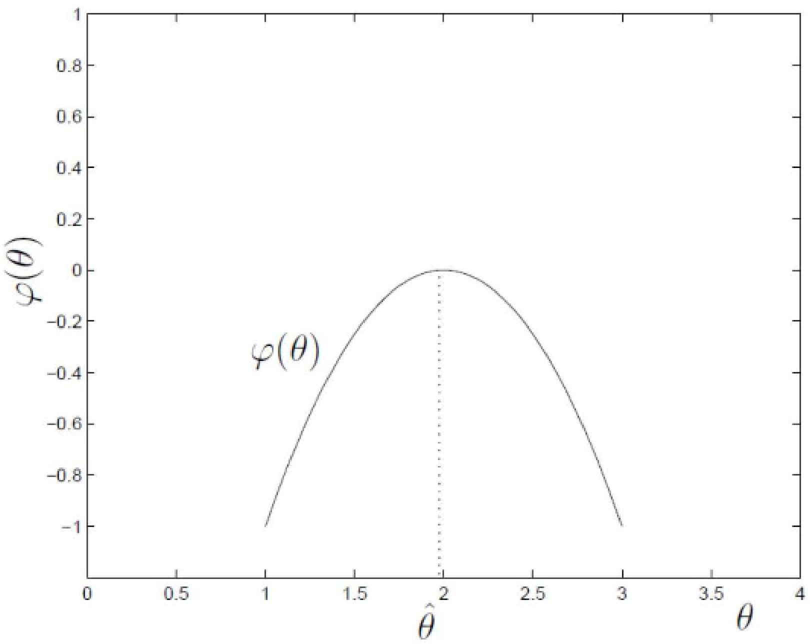

Figure 3.7: Maximum likelihood estimate

The maximum likelihood estimation has the following nice properties (Max, nd)

1. Consistency
2. Asymptoticly Normal
3. Asymptotic optimality

4. Invariance property

Definition 3.2.2: Let $X_1, ..., X_n$ be a sequence of observations. Let $\hat{\theta}_n$ be the estimator using $X_1, ..., X_n$. We say that $\hat{\theta}_n$ is consistent if $\hat{\theta}_n \xrightarrow{\mathsf{P}} \theta$, that is,

$$\mathbb{P}\left(|\hat{\theta}_n - \theta| > \varepsilon\right) \to 0, \quad \text{as n} \to \infty \tag{3.2.4}$$

In practice, this means as the number of sample size increases, the distributions of the estimators become more and more concentrated near the true value of the parameter being estimated.

Proof:

1. $\hat{\theta}$ is the maximizer of $\mathcal{L}_n(\theta)$ (by definition 3.2.1)

2. θ_0 is the maximizer of $\mathcal{L}(\theta)$ We will use these equations for our proof The scaled log-likelihood function given by

$$\mathcal{L}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i|\theta) \tag{3.2.5}$$

$$\mathcal{L}(X|\theta) = \log f(X|\theta) \tag{3.2.6}$$

The expectation of the likelihood function is given by

$$\mathcal{L}(\theta) = \int (\log f(x|\theta)) f(x|\theta_0) dx) = \mathbb{E}_{\theta_0}\mathcal{L}(X|\theta) \tag{3.2.7}$$

For any θ, we have $\mathcal{L}(\theta) \leq \mathcal{L}(\theta_0)$

$$\mathcal{L}(\theta) - \mathcal{L}(\theta_0) = \mathbb{E}_{\theta_0}\left(\log f(X|\theta) - \log f(X|\theta_0)\right) \tag{3.2.8}$$

simplifying the term of the right hand side, we have

$$= \mathbb{E}_{\theta_0} \log \frac{f(X|\theta)}{f(X|\theta_0)} \tag{3.2.9}$$

Since,

$$\log t \leq t - 1 \tag{3.2.10}$$

Applying equation (3.2.10) to equation (3.2.9), we have:

$$\mathbb{E}_{\theta_0} \log \frac{f(X|\theta)}{f(X|\theta_0)} \leq \mathbb{E}_{\theta_0}\left(\frac{f(X|\theta)}{f(X|\theta_0)} - 1\right) \tag{3.2.11}$$

Using the expectation formula in equation (3.2.7), we have

$$= \int \left(\frac{f(x|\theta)}{f(x|\theta_0)} - 1\right) f(x|\theta_0) dx \tag{3.2.12}$$

By linearity, we have:

$$= \int f(x|\theta) dx - \int f(x|\theta_0) dx \tag{3.2.13}$$

Using the theorem in equation (2.5.1), we have:

$$= 1 - 1 = 0 \tag{3.2.14}$$

3. $\forall\theta$, $\mathcal{L}_n(\theta) \to \mathcal{L}(\theta)$ in probability by law of large numbers

   A sufficient condition to have equation (3.2.4) is that Chan (2015):

$$\mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right] \to 0, \quad \text{as n} \to \infty \tag{3.2.15}$$

   According to Chebyshev's inequality, we have

$$\mathbb{P}\left(|\hat{\theta}_n - \theta| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right]}{\varepsilon^2} \tag{3.2.16}$$

   Since $\mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right] \to 0$, we have

$$0 \leq \mathbb{P}\left(|\hat{\theta}_n - \theta| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right]}{\varepsilon^2} \to 0 \tag{3.2.17}$$

   Therefore, $\mathbb{P}\left(|\hat{\theta}_n - \theta| > \varepsilon\right) \to 0$, as n $\to \infty$

Based on proofs 1,2 and 3, we can conclude that $\hat{\theta}$ is a consistent estimator.

Definition 3.2.3: Let $\{X_1, ..., X_n\}$ be a sequence of iid observations where $X_k \overset{iid}{\sim} f_\theta(x)$, $\hat{\sigma}$ is the MLE of θ, then $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right)$

We will use the following notations

$$l'(X|\theta) = \frac{\partial \log f(x;\theta)}{\partial\theta} = \frac{f'(x|\theta)}{f(x|\theta)} \tag{3.2.18}$$

Fisher Information is defined as:

$$I(\theta_0) = \mathbb{E}(l'(X|\theta_0))^2 = \mathbb{E}\left(\frac{\partial \log f(x;\theta)}{\partial\theta}|_{\theta=\theta_0}\right)^2 \tag{3.2.19}$$

and

$$I(\theta) = Var\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right) \tag{3.2.20}$$

$$\mathbb{E}_{\theta_0} l''(X|\theta_0) = \mathbb{E}_{\theta_0} \frac{\partial^2}{\partial\theta^2} \log f(X|\theta_0) = -I(\theta_0) \tag{3.2.21}$$

Since $f(x|\theta)$ is a valid pdf:

$$\int_{-\infty}^{\infty} f(x|\theta)dx = 1 \tag{3.2.22}$$

Taking the derivative of equation (3.2.22), we have,

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial\theta} dx \tag{3.2.23}$$

Equation (3.2.23) is equivalent to:

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)/\partial\theta}{f(x;\theta)} f(x;\theta)dx \tag{3.2.24}$$

The first term in the integral sign is equivalent to equation (3.2.18), we rewrite (3.2.24) as

$$0 = \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)dx \tag{3.2.25}$$

Writing (3.2.25) as an expectation, we have,

$$\mathbb{E}\left[\frac{\partial \log f(x;\theta)}{\partial\theta}\right] = 0 \tag{3.2.26}$$

Differentiating (3.2.25), we have,

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)dx \; + \; \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)dx \tag{3.2.27}$$

Expressing equation (3.2.27) is expectation form, we have,

$$\text{Second term:} \;\; I(\theta_0) = \mathbb{E}(l'(X|\theta_0))^2 = \mathbb{E}\left(\frac{\partial \log f(x;\theta)}{\partial\theta}|_{\theta=\theta_0}\right)^2 \tag{3.2.28}$$

and

$$\text{First term:} \;\; I(\theta_0) = -\mathbb{E}(l''(X|\theta_0))^2 = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)|_{\theta=\theta_0}\right) \tag{3.2.29}$$

Since MLE $\hat{\theta}$ is maximizer of $\mathcal{L}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\log f(X_i|\theta)$, we have $\mathcal{L}_n'(\hat{\theta}) = 0$.

By mean value theory;

$$0 = \mathcal{L}_n'(\hat{\theta}) \tag{3.2.30}$$

$$= \mathcal{L}_n'(\theta_0) + \mathcal{L}_n''(\hat{\theta}_1)(\hat{\theta} - \theta_0) \quad \hat{\theta}_1 \in \left[\hat{\theta}, \theta_0\right] \tag{3.2.31}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\frac{\sqrt{n}\mathcal{L}_n'(\theta_0)}{\mathcal{L}_n''(\hat{\theta}_1)} \tag{3.2.32}$$

Let us consider the numerator in equation (3.2.32)

$$\begin{aligned} -\sqrt{n}l_n'(\theta_0) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} l'(X_i|\theta_0) - 0\right) \\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} l'(X_i|\theta_0) - \mathbb{E}_{\theta_0}l'(X_i|\theta_0)\right) \end{aligned} \tag{3.2.33}$$

$$\to \mathcal{N}(0, Var_{\theta_0}(l'(X_i|\theta_0))) \quad \text{(convergence in distribution by central limit theory)}$$

Next let us consider the denominator in equation (3.2.32) Since,

$$\mathcal{L}''_n(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} l''(X_i|\theta) \to \mathbb{E}_{\theta_0} l''(X_i|\theta)$$

Convergence in probability by law of large numbers

$$\mathcal{L}''_n(\hat{\theta}_1) \to \mathbb{E}_{\theta_0} l''(X_i|\theta_0) = -I(\theta_0) \tag{3.2.34}$$

Combine equation (3.2.33) and equation (3.2.34), we get:

$$\frac{-\sqrt{n}\mathcal{L}'_n(\theta_0)}{\mathcal{L}''_n(\hat{\theta}_1)} \xrightarrow{d} \mathcal{N}\left(0, \frac{Var_{\theta_0}(l'(X_i|\theta_0))}{(I(\theta_0))^2}\right) \tag{3.2.35}$$

$$\theta - \theta_0 \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_n(\theta_0)}\right)$$

Therefore, $\hat{\theta}$ is asymptoticly normal.

Definition 3.2.4: Let X is a statistic with mean E(X) $= \mathbb{E}[u(X_1, ..., X_n)] = k(\theta_0)$ then we have,

$$\text{var(X)} \geq \frac{[k'(\theta_0)]^2}{nI(\theta_0)} \tag{3.2.36}$$

Statistic X is called efficient estimator of θ if and only if the variance of X attains the Rao-Cramer lower bound.

When X is an unbiased estimator of θ, then the Rao-Cramer inequality becomes

$$\text{var(X)} \geq \frac{1}{nI(\theta_0} \tag{3.2.37}$$

$$\theta - \theta_0 \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_n(\theta_0)}\right)$$

where $I_n(\theta_0) = -E\left[\frac{\partial^2 \ln p(x;\theta_0)}{\partial \theta_0^2}\right]$ = Fisher Information.

Technically speaking, when n converges to infinity, MLE is an unbiased estimator with smallest variance.

Theorem 3.2.1 **(Invariance property of MLE)** Let $\tau =$ g(θ) be a function of θ. Let $\hat{\theta}_n$ be the MLE of θ. Then $\hat{\tau}_n = g(\hat{\theta}_n)$ is the MLE of τ.

Proof

$$\begin{aligned} L^*_x(\tau) &= \prod f(x, \tau) = \prod f(x|g^{-1}(\tau)) \\ &= L_x(g^{-1}(\tau)) = L_x(\theta) \end{aligned} \tag{3.2.38}$$

Thus, the maximum of $L^*_x(\tau)$ is attained at $\tau = g(\hat{\theta})$

3.3 Parameter Estimation of Financial Models

3.3.1 Normal distribution.

Let $X_1, ..., X_n$ be independent and identically distributed $\mathcal{N}(\mu, \sigma^2)$ random variables, and let x_i be the value X_i takes (Orloff and Bloom) The density for each X_i is

$$f_{X_i}(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.3.1}$$

Since the X_i are independent their joint pdf is the product of the individual pdf's

$$f(x_1, ..., x_n|\mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\ e^{-\sum\limits_{i=1}^{n}\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.3.2}$$

For the fixed data, $x_1, ..., x_n$, the likelihood and log likelihood are:

$$\ln(f(x_1, ..., x_2|\mu, \sigma)) = -n\ln(\sqrt{2\pi}) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x-\mu)^2}{2\sigma^2} \tag{3.3.3}$$

Since $\ln(f(x_1, ..., x_2|\mu, \sigma))$ is a function of the two variables μ, σ we use partial derivatives to find the MLE. The easy value to find is $\hat{\mu}$:

$$\frac{\partial f(x_1, ..., x_n|\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} = 0 \tag{3.3.4}$$

This implies that:

$$\sum_{i=1}^{n} x_i = n\mu \tag{3.3.5}$$

Therefore,

$$\hat{\mu} = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \bar{x} \tag{3.3.6}$$

To find $\hat{\sigma}$, we differentiate equation (3.3.3) with respect to σ and solve for σ:

$$\frac{\partial f(x_1, ..., x_n|\mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3} = 0 \tag{3.3.7}$$

Therefore,

$$\hat{\sigma}^2 = \frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{n} \tag{3.3.8}$$

We already know that $\mu = \bar{x}$, so we use that as the value of μ in the equation (3.3.8). Therefore, the maximum likelihood estimates are as follows:

$$\begin{aligned} \hat{\mu} &= \bar{x} \quad \text{(the mean of the log-returns)} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \quad \text{(the volatility of the log-returns)} \end{aligned} \tag{3.3.9}$$

3.3.2 Student t-distribution.

The standard Student's t random variable can be written as a normal distribution whose variance is equal to the reciprocal of a Gamma random variable. Let X be a continuous random variable, we say that X has a Student's t distribution with mean μ, scale σ^2 and k degrees of freedom if and only if its probability density function is:

Student$(x|\mu, \lambda, k)$

$$\equiv \int_0^\infty \text{Normal}(x|\mu, (\lambda\eta)^{-1})\ Gamma(\eta|\ {}^k/_2, {}^k/_2)\ d\eta \tag{3.3.10}$$

where $\lambda = \sigma^2$ for simplicity

$$= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma({}^k/_2)} \left(\frac{\lambda}{\pi k}\right)^{1/2} \left(1 + \frac{\lambda(x-\mu)^2}{k}\right)^{-\frac{k+1}{2}} \tag{3.3.11}$$

We want to find the maximum likelihood estimate for a set of parameters Θ given a set of observed data X by maximizing

$$P(X|\Theta)$$

We assume that it is hard to solve this problem directly but that it is relatively easy to evaluate

$$P(X, Z|\Theta)$$

where Z is a set of latent variables such that

$$P(X|\Theta) = \int_Z P(X, Z|\Theta) \tag{3.3.12}$$

To find the MLE for t, we use a stable approach called the Expectation Maximization (EM) approach since the closed form of the distribution does not exist (Scheffler, 2008).

$$P(x_i|\Theta) = \text{Student}(x_i|\mu, \lambda, v) \tag{3.3.13}$$

By viewing this as an infinite mixture of Normal distributions,

$$P(x_i|\Theta) = \int_{\eta_i} Normal(x_i|\mu, (\lambda\eta_i)^{-1})\ Gamma(\eta_i|{}^k/_2, {}^k/_2) \tag{3.3.14}$$

Let X = $\{x_i\}$, Z = $\{\eta_i\}$, $\Theta = \{\mu, \lambda, v\}$

Therefore the complete likelihood function is:

$$P(X, Z|\Theta) = \prod_{i=1}^{N} Normal\left(x_i|\mu, (\lambda\eta_i)^{-1}\right)\ Gamma(\eta_i|{}^k/_2, {}^k/_2) \tag{3.3.15}$$

The loglikelihood function is given by:

$$\log P(X, Z|\Theta) = \sum_{i=1}^{N} \log Normal\left(x_i|\mu, (\lambda\eta_i)^{-1}\right) + \log Gamma(\eta_i|\ {}^k/_2, {}^k/_2) \tag{3.3.16}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} - {}^{1}\!/_{2}\log(2\pi) + {}^{1}\!/_{2}\log\lambda + {}^{1}\!/_{2}\log\eta_i - \frac{\lambda\eta_i}{2}(x_i-\mu)^2 \\
&\qquad - \log\Gamma({}^{k}\!/_{2}) + \frac{k}{2}\log({}^{k}\!/_{2}) + \left(\frac{k}{2}-1\right)\log\eta_i - \frac{k}{2}\eta_i
\end{aligned} \tag{3.3.17}
$$

Next the posterior latent distribution is given by:

$$
\begin{aligned}
P(Z|X,\Theta) &\propto P(X,Z|\Theta) \\
&= \prod_{i=1}^{N} Normal\left(x_i|\mu,(\lambda\eta_i)^{-1}\right)\; Gamma(\eta_i|{}^{k}\!/_{2},{}^{k}\!/_{2}) \\
&\propto \prod_{i=1}^{N}\; Gamma(\eta_i|a_i,b_i)
\end{aligned} \tag{3.3.18}
$$

Since the Gamma distribution is the conjugate prior to a Normal distribution with unknown precision, we find the parameters a_i and b_i by combining the factors from the Normal and Gamma distributions.

$$
\begin{aligned}
P(X,Z|\Theta) &= \prod_{i=1}^{N} Normal(x_i|\mu,(\lambda\eta_i)^{-1})\; Gamma(\eta_i|{}^{k}\!/_{2},{}^{k}\!/_{2}) \\
&\propto \prod_{i=1}^{N}\left[\eta_i^{\frac{v-1}{2}} exp\left(-n_i({}^{v}\!/_{2} + {}^{\lambda}\!/_{2}(x_i-\mu)^2)\right)\right]
\end{aligned} \tag{3.3.19}
$$

All factors independent of η_i are taken up in the proportionality.

$$
\propto \prod_{i=1}^{N} Gamma\left(\eta_i|\frac{v+1}{2},\frac{v}{2}+\frac{\lambda}{2}(x_i-\mu)^2\right) \tag{3.3.20}
$$

and hence

$$
\begin{aligned}
a_i &= \frac{v+1}{2} \\
b_i &= \frac{v}{2} + \frac{\lambda}{2}(x_i-\mu)^2
\end{aligned} \tag{3.3.21}
$$

Next, we calculate the expectations of η_i and $\log\eta_i$ under the posterior latent distribution

$$
\begin{aligned}
E[\eta_i] &= \int_Z \eta_i \prod_{j=1}^{N} Gamma(\eta_j|a_j,b_j) \\
&= \int_Z \eta_i\; Gamma(\eta_i|a_i,b_i) \\
&= {}^{a_i}\!/_{b_i} \\
&= \frac{v_0+1}{v_0+\lambda_0(x_i-\mu_0)^2}
\end{aligned} \tag{3.3.22}
$$

$$
\begin{aligned}
E[\log \eta_i] &= \int_Z \log \eta_i \prod_{j=1}^{N} Gamma(\eta_j | a_j, b_j) \\
&= \int_Z \log \eta_i \; Gamma(\eta_i | a_i, b_i) \\
&= \psi(a_i) - \log b_i \\
&= \psi\left(\frac{v+1}{2}\right) - \log \frac{v}{2} + \frac{\lambda}{2}(x_i - \mu)^2
\end{aligned}
\tag{3.3.23}
$$

See Appendix for the definition of the digamma function, $\psi(.)$

Next is to find a function to optimize the posterior latent distribution

$$
Q(\Theta, \Theta_0) = \int_Z P(Z|X, |\Theta_0) \log P(X, Z|\Theta) \tag{3.3.24}
$$

$$
\begin{aligned}
= -\frac{N}{2}\log 2\pi + \frac{N}{2}\log\lambda + \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}[\log\eta_i] - \frac{\lambda}{2}\sum_{i=1}^{N}(x_i-\mu)^2\mathbb{E}[\eta_i] \\
-N\log\Gamma(v/2) - \frac{Nv}{2}\log v/2 + \left(\frac{v}{2}-1\right)\sum_{i=1}^{N}\mathbb{E}[\log\eta_i] - \frac{v}{2}\sum_{i=1}^{N}\mathbb{E}[\eta_i]
\end{aligned}
\tag{3.3.25}
$$

The last step is the maximization step

$$
\begin{aligned}
\frac{\partial Q}{\partial \mu} = 0 \quad &\Longrightarrow \lambda \sum_{i=1}^{N}(x_i - \mu)\mathbb{E}[\eta_i] = 0 \\
&\Longrightarrow \quad \mu = \frac{\sum_{i=1}^{N} x_i \mathbb{E}[\eta_i]}{\sum_{i=1}^{N} \mathbb{E}[\eta_i]}
\end{aligned}
\tag{3.3.26}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \lambda} = 0 \quad &\Longrightarrow \frac{N}{2\lambda} - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\mathbb{E}[\eta_i] = 0 \\
&\Longrightarrow \quad \lambda = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\mathbb{E}[\eta_i]\right)^{-1}
\end{aligned}
\tag{3.3.27}
$$

Note that we require the updated μ value to find λ.

$$
\begin{aligned}
\frac{\partial Q}{\partial v} = 0 \quad \Longrightarrow -\frac{N}{2}\psi(\frac{v}{2}) + \frac{N}{2}\log\frac{v}{2} + \frac{N}{2} + \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}[\log\eta_i] \\
-\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}[\eta_i] = 0
\end{aligned}
\tag{3.3.28}
$$

$$\implies \psi(\frac{v}{2}) - \log\frac{v}{2} = 1 + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\log\eta_i] - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\eta_i] \tag{3.3.29}$$

Therefore the unbiased estimators of μ, λ and η are:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i\ x_i}{\sum_{i=1}^{n} w_i} \tag{3.3.30}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} w_i\ (x_i - \hat{\mu})^2}{n} \tag{3.3.31}$$

$$\hat{w}_i^2 = \frac{(v+1)\sigma^2}{v\sigma^2 + (x_i-\mu)^2} \tag{3.3.32}$$

3.3.3 Generalized Lambda Distribution (GLD).

Dedduwakumara et al. (2019) noted that obtaining the maximum likelihood estimates of the GLD is computationally difficult. It involves two steps. First, initial estimated parameter values are obtained using the method of moments. The secondly step involves using those initial values as a starting point to seek the values that maximize the numerical log likelihood. However, Su (2007) derived a log-likelihood that can be used to estimate the parameters of the GLD numerically (See Cecchinato (2010) for more details)

3.3.4 Normal Inverse Gaussian (NIG) distribution.

The likelihood function is given by:

$$\mathcal{L}(x_i,\theta) = \ln\left(\prod_{i=1}^{n} f(x_i,\alpha,\beta,\mu,\delta)\right) \tag{3.3.33}$$

$$= \ln\left(\prod_{i=1}^{n}\frac{\delta\alpha}{\pi} exp(\delta\sqrt{\alpha^2+\beta^2} + \beta(x_i-\mu))\ K_1\frac{(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}{\sqrt{\delta^2+(x_i-\mu)^2}}\right) \tag{3.3.34}$$

Simplifying the expression we have:

$$\begin{aligned} &= \sum_{i=1}^{n}\ln\left(\frac{\delta\alpha}{\pi}\right) + \ln\left(\prod_{i=1}^{n} exp(\delta\sqrt{\alpha^2+\beta^2} + \beta(x_i-\mu))\right) \\ &\quad + \ln\left(\prod_{i=1}^{n} K_1\left(\alpha\sqrt{\delta^2-(x_i-\mu)^2}\right)\right) - \frac{1}{2}\ln\left(\prod_{i=1}^{n}\left(\delta^2+(x_i-\mu)^2\right)\right) \end{aligned} \tag{3.3.35}$$

$$\begin{aligned} &= n\ln(\delta\alpha) - n\ln(\pi) + n\left(\delta\sqrt{\alpha^2-\beta^2} - \beta\mu\right) + \beta\sum_{i=1}^{n} x_i \\ &\quad + \sum_{i=1}^{n}\ln\left(K_1\left(\alpha\sqrt{\delta^2+(x_i-\mu)^2}\right)\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\delta^2+(x_i-\mu)^2\right) \end{aligned} \tag{3.3.36}$$

To obtain the maximum likelihood parameter estimates, we differentiate the log-likelihood function in (3.3.36)

$$\begin{aligned} \text{For } \beta: \quad \frac{\partial}{\partial \beta} L &= n\delta \left(\frac{1}{2\sqrt{\alpha^2 - \beta^2}} \right) (-2\beta) - n\mu + \sum_{i=1}^{n} x_i \\ &= -\frac{n\delta\beta}{\sqrt{\alpha^2 - \beta^2}} - n\mu + \sum_{i=1}^{n} x_i \end{aligned} \tag{3.3.37}$$

Setting $\frac{\partial}{\partial \beta} L = 0$ and solving for β we obtain the following likelihood estimate for β:

$$\hat{\beta} = \frac{\alpha \left(n\mu - \sum_{i=1}^{n} x_i \right)}{\sqrt{n^2\delta^2 + \left(n\mu - \sum_{i=1}^{n} x_i \right)^2}} \tag{3.3.38}$$

Apply the properties of the modified Bessel function to obtain the derivative of the loglikehood function with respect to the remaining parameters.

$$\frac{\partial}{\partial \mu} L = -n\beta + \sum_{i=1}^{n} \frac{x_i - \mu}{\sqrt{\delta^2 + (x_i - \mu)^2}} \left[\frac{2}{\sqrt{\delta^2 + (x_i - \mu)^2}} + \frac{\alpha K_0 \left(\alpha \sqrt{\delta^2 + (x_i - \mu)^2} \right)}{K_1 \left(\alpha \sqrt{\delta^2 + (x_i - \mu)^2} \right)} \right] \tag{3.3.39}$$

Setting $\frac{\partial}{\partial \mu} L = 0$ and solving for μ we obtain the estimate of μ as:

$$\hat{\mu} = -\frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} + \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.3.40}$$

The derivatives of the loglihood function in (3.3.36) with respect to δ and α are given by (Prause, 1999).

$$\frac{\partial}{\partial \delta} L = \frac{n}{\delta} + n\sqrt{\alpha^2 - \beta^2} - 2\sum_{i=1}^{n} \frac{\delta}{\delta^2 + (x_i - \mu)^2} - \sum_{i=1}^{n} \frac{\alpha\delta}{\sqrt{\delta^2 + (x_i - \mu)^2}} \tag{3.3.41}$$

and

$$\frac{\partial}{\partial \alpha} L = \frac{n\delta\alpha}{\sqrt{\alpha^2 - \beta^2}} - \sum_{i=1}^{n} \sqrt{\delta^2 + (x_i - \mu)^2} \frac{K_0 \left(\alpha \sqrt{\delta^2 + (x_i - \mu)^2} \right)}{K_1 \left(\alpha \sqrt{\delta^2 + (x_i - \mu)^2} \right)} \tag{3.3.42}$$

According to Prause (1999), it is preferable to maximize the $\ln(\alpha)$ and $\ln(\delta)$ to maintain the positivity condition of α and β.

3.3.5 Hyperbolic distribution.

The parameters of the hyperbolic distribution is obtained using the parameters of the generalized hyperbolic distribution and setting λ to one. The likelihood is given by:

$$\mathcal{L}(x_i;\theta) = \ln\left(\prod_{i=1}^{n} f(x_i;\alpha,\beta,\delta,\mu)\right) \tag{3.3.43}$$

$$= \ln\left(\prod_{i=1}^{n} \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2-\beta^2})} \exp(-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu))\right) \tag{3.3.44}$$

$$\begin{aligned} &= \frac{1}{2}\sum_{i=1}^{n}\left(\ln(\sqrt{(\alpha^2-\beta^2)}-(\alpha\delta))\right) - \sum_{i=1}^{n}\left(\ln K_1(\delta\sqrt{\alpha^2-\beta^2})\right) \\ &\quad + \ln\left(\prod_{i=1}^{n} exp(-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu))\right) \end{aligned} \tag{3.3.45}$$

We differentiate the log-likehood function above,

$$\frac{\partial}{\partial\alpha}L = n\frac{\delta\alpha}{\sqrt{\alpha^2-\beta^2}}R_1(\delta\sqrt{\alpha^2-\beta^2}) - \sum_{i=1}^{n}\sqrt{\sigma^2-(x_i-\mu)^2}\ .R_{1/2} \tag{3.3.46}$$

$$\frac{\partial}{\partial\beta}L = n\left[-\frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}R_1(\delta\sqrt{\alpha^2-\beta^2}) - \mu\right] + \sum_{i=1}^{n} x_i \tag{3.3.47}$$

$$\begin{aligned} \frac{\partial}{\partial\delta}L =& n\left[-\frac{2}{\delta} + \sqrt{\alpha^2-\beta^2}\ R_1(\delta\sqrt{\alpha^2-\beta^2})\right] \\ & \sum_{i=1}^{n}\left[\frac{\delta}{\delta^2+(x_i-\mu)^2} - \frac{\alpha\delta R_1(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}{\sqrt{\delta^2+(x_i-\mu)^2}}\right] \end{aligned} \tag{3.3.48}$$

$$\begin{aligned} \frac{\partial}{\partial\mu}L = -n\beta + \sum_{i=1}^{n}\frac{x_i-\mu}{\sqrt{\delta^2+(x_i-\mu)^2}}&\left[\frac{1}{\sqrt{\delta^2+(x_i-\mu)^2}} - \right. \\ &\left. \alpha R_{1/2}(\alpha\sqrt{\delta^2+(x_i-\mu)^2})\right] \end{aligned} \tag{3.3.49}$$

Therefore, we obtain direct solutions of the likelihood equations for β and μ.

From $\frac{\partial}{\partial\beta}L = 0$ and $\frac{\partial}{\partial\mu}L = 0$, we have

$$\hat{\mu} = -\frac{\delta\beta}{\sqrt{\alpha^2-\beta}}\ R_1\ (\delta\sqrt{\alpha^2-\beta^2}) + \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.3.50}$$

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\mu}{\sqrt{\delta^2+(x_i-\mu)^2}}\left[\frac{1}{\sqrt{\delta^2+(x_i-\mu)^2}} - \alpha\ R_{1/2}(\alpha\sqrt{\delta^2+(x_i-\mu)^2})\right] \tag{3.3.51}$$

Prause (1999) suggested that the solutions of the likelihood equations for the other parameters can be obtained by maximizing the log-likehood equation with respect to (α,δ,β)

3.3.6 Skewed t-distribution.

Let $x \sim StN(\xi, \omega, \alpha, \nu)$ denote that x follow a skewed t- distribution with location paramater ξ, scale parameter ω, skewness parameter α, and degrees of freedom parameter ν. Then the density of x given by:

$$f(z|\xi, \omega, \alpha, \nu) = 2\ t_v(x|\xi, \sqrt{\omega})\ \Phi\left(\frac{x-\xi}{\omega}\alpha\right) \tag{3.3.52}$$

where $t_v(x|\xi, \omega)$ denotes the t density with v degrees of freedom, location parameter ξ and scale parameter ω. Φ denotes the Gaussian cumulative distribution function.

There is no explicit solutions using the likelihood equations of the skewed t distribution, therefore Gómez et al. (2007) gave the following hierarchical decomposition using the EM algorithm:

$$\tau \sim \Gamma\left({}^{v}\!/_{2}, {}^{v}\!/_{2}\right) \tag{3.3.53}$$

$$\gamma|\tau \sim TN\left(0, \frac{\tau+\alpha^2}{\tau}; (0,\infty)\right) \tag{3.3.54}$$

$$z|\tau,\gamma \sim N\left(\xi + \frac{\sqrt{\omega}\ \alpha}{\tau+\alpha^2}\gamma, \frac{\omega}{\tau+\alpha^2}\right) \tag{3.3.55}$$

where $TN(\xi, \omega; (a, b))$ denotes the truncated normal distribution within the interval (a, b). Taking τ and γ as latent variables, the maximum likelihood estimators of the complete log likelihood are given below:

$$\hat{\xi} = \frac{\sum_{i=1}^{n}[\tau_i x_i] + \alpha^2 \sum_{i=1}^{n} x_i - \alpha\sqrt{\omega}\sum_{i=1}^{n}\gamma_i}{n\alpha_i^2 + \sum_{i=1}^{n}\tau_i} \tag{3.3.56}$$

$$\hat{\omega} = \frac{1}{n}\sum_{i=1}^{n}\tau_i(x_i-\xi)^2 \tag{3.3.57}$$

$$\hat{\delta} = \frac{\sum_{i=1}^{n}(x_i-\xi)\gamma_i}{\sum_{i=1}^{n}(x_i-\xi)^2} \tag{3.3.58}$$

$$\hat{\alpha} = \hat{\delta}\sqrt{\hat{\omega}} \tag{3.3.59}$$

There is no closed form estimator ν, hence we numerically maximize:

$${}^{\nu}\!/_{2}\log({}^{\nu}\!/_{2}) - \log\Gamma({}^{\nu}\!/_{2}) + \frac{\nu-1}{2}\log\tau_i - \frac{\nu}{2}\ \tau_i \tag{3.3.60}$$

The conditional expectations of the latent variables are given by:

$$E[\gamma_i|x_i] = (x_i-\xi)\frac{\alpha}{\sqrt{\omega}} + \frac{\phi\left((x_i-\xi)^{\alpha}\!/\!_{\sqrt{\omega}}\right)}{\Phi\left((x_i-\xi)^{\alpha}\!/\!_{\sqrt{\omega}}\right)} \tag{3.3.61}$$

$$E[\tau_i|x_i] = \frac{\nu+1}{\nu + \frac{(x_i-\xi)^2}{\omega}} \tag{3.3.62}$$

3.4 Goodness of Fit Tests

The goodness of fit tests is used to measure the closeness of the theoretical distribution and empirical distribution. In this case, we select the model with the best goodness of fit value.

3.4.1 Log-likelihood.

The log-likelihood (LL) is an overall measure of goodness-of-fit, with higher values of LL implying a more likely distribution candidate to model the data.

3.4.2 Kullback-Leibler (KL) divergence.

The relative entropy, also known as the Kullback-Leibler divergence, between two probability distributions on a random variable is a measure of the distance between them (Rao, 2010). Formally, given two probability distributions p(x) and q(x) over a continuous random variable X, the relative entropy given by K$\mathcal{L}(p||q)$ is defined as follows (Popkes, 2010):

$$K\mathcal{L}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{3.4.1}$$

where p and q denote the probability density functions of P and Q.

Definition 3.4.1.1 Given a function f with x being a continuous variable, the expected value of f(x) is defined by:

$$\mathbb{E}[f(x)] = \int_{-\infty}^{\infty} f(x)p(x)dx \tag{3.4.2}$$

where p(x) is the probability density function of the variable.

Definition 3.4.1.2 The Kullback-Leibler (KL) divergence has the following properties:

1. $K\mathcal{L}(p||q) \neq K\mathcal{L}(q||p)$. This means that the K-L divergence is non-symmetric
2. $K\mathcal{L}(p||q) \geq 0$: This means that the KL divergence is always non-negative.

   Proof: We can prove this using the Jensen's inequality.

   Jensen's inequality states that, if a function $f(x)$ is convex, then

   $$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x]) \tag{3.4.3}$$

   Let us use the expected value from equation (3.4.2)

   $$\begin{aligned} K\mathcal{L}(p||q) = \int_{-\infty}^{\infty} p(x) log \left(\frac{p(x)}{q(x)}\right) dx &= \mathbb{E}_{x\sim p(x)} \left[\log \left(\frac{p(x)}{q(x)}\right)\right] \\ &= -\mathbb{E}_{x\sim p(x)} \left[\log \left(\frac{q(x)}{p(x)}\right)\right] \end{aligned} \tag{3.4.4}$$

Because $-\log(x)$ is a convex function, we apply the Jensen's inequality:

$$
\begin{aligned}
-\mathbb{E}_{x\sim p(x)}\left[\log\left(\frac{q(x)}{p(x)}\right)\right] &\geq -\log\left[\mathbb{E}_{x\sim p(x)}\left(\frac{q(x)}{p(x)}\right)\right] \\
&= -\log\left(\int_{-\infty}^{\infty} p(x)\frac{q(x)}{p(x)}dx\right) \\
&= -\log\left(\int_{-\infty}^{\infty} q(x)dx\right)
\end{aligned}
\tag{3.4.5}
$$

Since it is a proper pdf, we have:

$$
\begin{aligned}
&= -\log(1) \\
&= 0
\end{aligned}
$$

3. $K\mathcal{L}(p||p) = 0$

   Let $p(x)$ and $q(x)$ be two probability density functions such that their KL divergence is well defined. If the two probability density function coincide *almost surely*, that is, if

$$\int_A p(x)dx = \int_A q(x)dx$$

   for all measurable sets $A \subseteq R_x$, then $K\mathcal{L}(p||q) = 0$.

$$\frac{q(x)}{p(x)} = \frac{\int_A q(x)dx}{\int_A p(x)dx} = 1 \quad \text{(Cover and Thomas, 2012)}$$

$$
\begin{aligned}
K\mathcal{L}(p||q) &= -\mathbb{E}_{x\sim p(x)}\left[\log\left(\frac{q(x)}{p(x)}\right)\right] \\
&= -\int_{-\infty}^{\infty} p(x).\log\frac{q(x)}{p(x)}\;dx \\
&= -\int_{-\infty}^{\infty} p(x).\log(1)\;dx \\
&= -\int_{-\infty}^{\infty} p(x)\;dx.(0) = 0
\end{aligned}
\tag{3.4.6}
$$

3.4.3 Akaike Information Criteria (AIC).

The AIC (Akaike, 1973) is one of the tools used in model selection. The AIC model selection is motivated by the KL divergence principle. In this context, let us denote the KL-divergence as:

$$K\mathcal{L}(f||g) = I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)dx \tag{3.4.7}$$

$I(f,g)$ is the "information" lost when g is used to approximate f.

$$I(f,g) = \int f(x)\log(f(x))dx - \int f(x)\log(g(x|\theta))dx \tag{3.4.8}$$

Each of the two terms on the right hand side of equation (3.4.8) is a statistical expectation with respect to f. Thus we have,

$$I(f,g) = \mathbb{E}_f[log(f(x))] - \mathbb{E}_f[log(g(x|\theta))] \tag{3.4.9}$$

The first expectation $\mathbb{E}_f[log(f(x))]$ is a constant across models, thus,

$$\begin{aligned} I(f,g) &= constant - \mathbb{E}_f[\log(g(x|\theta))] \\ \text{or} \quad I(f,g) - constant &= -\mathbb{E}_f[\log(g(x|\theta))] \end{aligned} \tag{3.4.10}$$

where $I(f,g) - constant$ is a relative, directed distance between f and g; and $\mathbb{E}_f[\log(g(x|\theta))]$ is the quantity of interest. To find an estimate of $\mathbb{E}_f[log(g(x|\theta))]$, Akaike found its expectation as $\mathbb{E}_f\mathbb{E}_f[log(g(x|\theta))]$

An asympototically unbiased estimator of the relative, expected K-L information is given by:

$$\log(\mathcal{L}(\hat{\theta}|data)) - k \tag{3.4.11}$$

where k is the number of estimable parameters in the model, g.

Akaike (1973) thus defined "an information criterion" AIC by multiplying equation (3.4.11) by -2

$$AIC = -2\log(\mathcal{L}(\hat{\theta}|data)) + 2k \tag{3.4.12}$$

To use AIC for model selection, we simply choose the model with the smallest AIC over the set of models considered (Hastie et al., 2009)

3.4.4 Corrected Akaike Information Criteria (AIC$_c$).

Akaike derived an asymptotically unbiased estimator of K-L information; however AIC may perform poorly if there are too many parameters in relation to the size of the sample. A small sample bias adjustment which led to a criterion that is called AIC$_c$ (Hurvich and Tsai, 1989).

$$AIC_c = -2\log(\mathcal{L}(\hat{\theta})) + 2k\left(\frac{n}{n-k-1}\right) \tag{3.4.13}$$

where the penalty term is multiplied by the correction factor $\frac{n}{n-k-1}$

This can be rewrritten equivalently as

$$AIC_c = -2\log(\mathcal{L}(\hat{\theta})) + 2k + \frac{2k(k+1)}{n-k-1} \tag{3.4.14}$$

or equivalently,

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{3.4.15}$$

where n is the sample size, the value of AIC and AIC$_c$ converges as the sample size becomes large.

3.4.5 Bayesian Information Criteria.

The Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. The BIC resolves this problem by introducing a stronger penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC and AIC$_c$.

The theory behind the BIC is that it combines the KL-divergence and the Bayes theorem to obtain an unbiased estimator of the quantity of interest in equation (3.4.10). it uses the maximum likelihood estimate of the parameter, rather than averaging over the posterior distribution (Barum, 2018)

Let $p(x|q_k)$ be denoted as the "model evidence" or the "marginal likelihood" given by:

$$p(x|q_k) = \int p(x, \theta_k|q_k)d\theta_k = \int p(x|\theta_k, q_k)p(\theta_k|q_k)d\theta_k \tag{3.4.16}$$

where $p(x|\theta_k, q_k)$ is the likelihood of the data and $p(\theta_k|q_k)$ is the prior distribution of the parameter vector under the model q_k.

Note that $p(x|\theta) = p(x|\theta, q)$ and $p(\theta) = p(\theta|q)$ we drop to reference to q for clarity in what follows.

In this context, the equivalent equation to the one in (3.4.10) becomes:

$$g(\theta) = \log[p(x|\theta)p(\theta)] \tag{3.4.17}$$

Note that $p(x|\theta)p(\theta) \quad \alpha \quad p(\theta|x)$, the posterior distribution of θ The marginal likelihood is the normalization constant of the posterior and, thus, can be written as

$$p(x) = \int e^{g(\theta)}d\theta \tag{3.4.18}$$

To approximate the marginal likelihood in (3.4.18), we use the Laplace approximation.

Our estimator to approximate the log marginal likelihood is therefore:

$$\log p(x|\hat{\theta}) - \frac{k}{2}\log n \tag{3.4.19}$$

For "historical reasons", BIC is defined as -2 times the approximated log marginal likelihood:

$$BIC = -2\log p(x|\hat{\theta}) + k\log n \tag{3.4.20}$$

3.5 Value at Risk Estimates and Backtesting

3.5.1 Value at Risk.

After carrying out a robust statistical tests of model fits based mainly on the centre of the distribution, a further analysis is required at the extreme tails of the distribution. One of them is the analysis of the Value-at-Risk (VaR) estimates. VaR is a frequently used measure of potential risk for losses in financial markets (Duffie and Pan, 1997). It is used by financial institutions to calculate the maximum loss over a given time horizon. Hence, its calculations concentrate on the tails of the distribution. For an accurate VaR estimation, we expect the underlying theoretical model for the financial returns to be a good representation of the data at the extreme points (Huang et al., 2014).

The definition of VaR calls for a confidence level $\alpha \in (0, 1)$. Then, the VaR of portolio return at a confidence level α is defined as the smallest number x_0 such that the probability that the loss X exceeds x_0 is not longer than (1 - α). That is, in general,

$$\begin{aligned} \text{VaR}_\alpha(X) &= \inf\{x_0 : P(X > x_0) \leq 1 - \alpha\} \\ &= \inf\{x_0 : F_X(x_0) \geq \alpha\} \\ &= F_X^{-1}(\alpha) \end{aligned} \tag{3.5.1}$$

where $F_X(.)$ is the cummulative distribution function of X, F_X^{-1} is the inverse function of F_X provided it exists. In probabilitic terms, VaR is the α - quantile of the loss distribution.

In terms of long and short financial market positions, given a confidence level; of p $\in$ (0,1), with time index of t and t+ α, we want to find the change stock of the $\Delta V(\alpha)$ in the financial position over the time period α. Let $F_\alpha(x)$ be the cummulative distribution function (CDF) of $\Delta V(\alpha)$. If we consider a long position in a given time α, with probability p, and the financial position $\Delta V(\alpha) \leq 0$, then we can define the VaR as (Wenquing, 2016):

$$p = \mathbb{P}[\Delta V(\alpha) \leq \text{VaR}] = F_\alpha(\text{VaR}) \tag{3.5.2}$$

Wenquing (2016) defines VaR for short term position $\Delta V(\alpha) \geq 0$ as:

$$\begin{aligned} p = \mathbb{P}[\Delta V(\alpha) \geq \text{VaR}] &= 1 - \mathbb{P}[\Delta V(\alpha) \leq \text{VaR}] \\ &= F_\alpha(\text{VaR}) \end{aligned} \tag{3.5.3}$$

3.5.2 Backtesting.

Backtesting is the process of gauging the accuracy of a value-at-risk measure's predictions when applied to a particular portfolio over time. The most widely used method is the Kupiec likelihood ratio (LR) test Kupiec (1995). The Kupiec LR test utilizes the fact that a good model should have its proportion of violations of VaR estimates close to the corresponding tail probability. The method consists of calculating x^α the number of times the observed returns fall below (for long positions) or above (for short positions) the VaR estimate at level α; that is $r_t < \text{VaR}_\alpha$ or $r_t >$ VaR_α, and compare the corresponding failure rates to α. The null hypothesis is that the expected proportion of violations is equal to α Under this null hypothesis, the Kupiec LR statistic, given by (Huang et al., 2014):

$$2\ln\left(\left(\frac{x^\alpha}{N}\right)^{x^\alpha}\left(1-\frac{x^\alpha}{N}\right)^{N-x^\alpha}\right) - 2\ln\left(\alpha^{x^\alpha}(1-\alpha)^{N-x^\alpha}\right) \tag{3.5.4}$$

is asymptotically distributed to a Chi-square distribution with one degree of freedom. Larger p-values and smaller Kupiec LR statistic correspond to better fits. See (Corlu and Corlu, 2015)

4. Empirical Results

4.1 Data

The data used in the research work are the stock market indices of four African countries, namely: Cote D'Ivoire, Egypt, Nigeria and South Africa respectively and they are obtained from the Datastream (database for financial data). The sample period covers 5203, 5387, 4856 and 6041 trading days for each of the country's stock markets respectively with their time-frames shown under each plot in Figure 4.1.

1. Time Series Plots

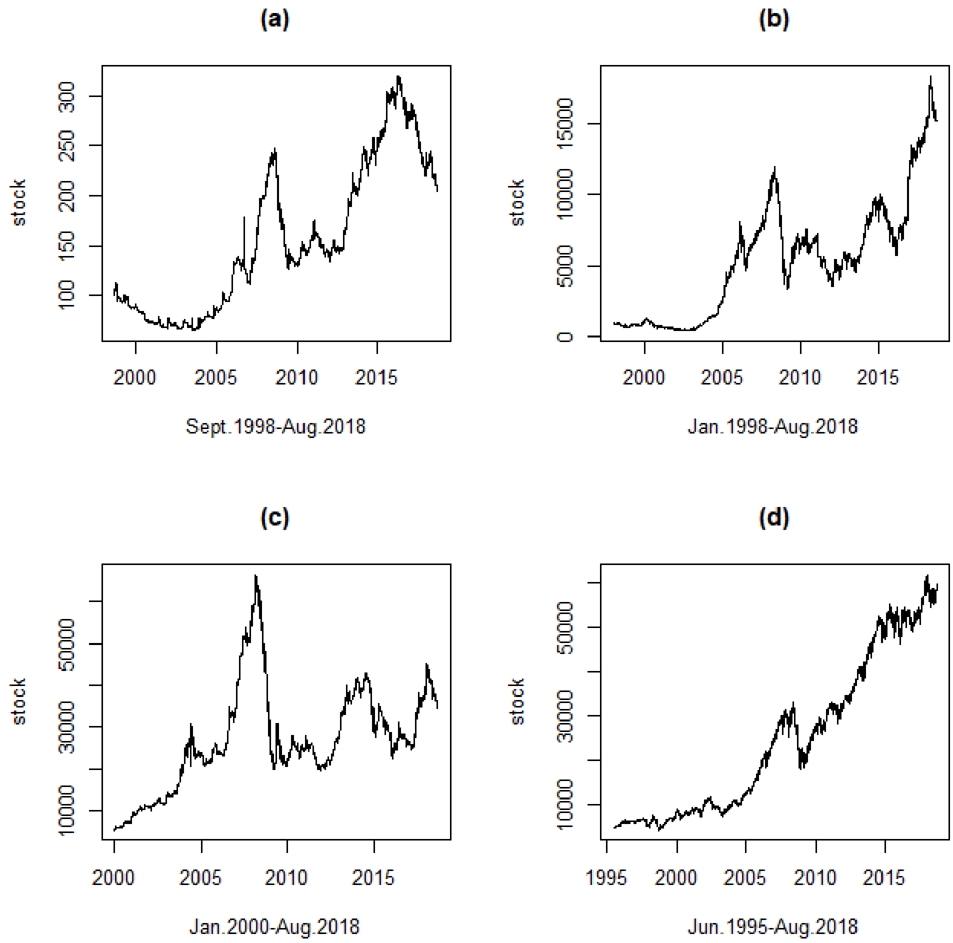

Figure 4.1: Time Series plots of the (a) Ivorian Stock Market (b) Egyptian Stock Market (c) Nigerian Stock Market (d) South African Stock Market

Figure 4.1 shows the stock market series of Cote D'Ivoire, Egypt, Nigeria and South Africa respectively. The stock market data for each country shows a steady increase in stock at the beginning with slight fluctuations and appear non-stationary. There is also an indication of large boom bust periods associated to the run-up to the global economic crisis of 2008. Furthermore, there is a strong common upward trend behaviour of the four stock series.

4.1.1 Summary Statistics of the Stock Returns.

The returns r_t are expressed in percentages, that is, $r_t = 100 * (\log P_t - \log P_{t-1})$.

Table 4.1: Descriptive Statistics of the log-returns of the four stock markets

Log-returns	**n**	**Min**	**Max**	**Mean**	**Variance**	**Stdev**	**Skewness**	**Excess Kurtosis**
Cote d'Ivoire	5203	-15.6315	15.6963	0.01377	0.8233	0.9074	-0.1020	76.5066
Egypt	5387	-17.9916	18.3692	0.05056	2.6471	1.6270	-0.3328	9.6875
Nigeria	4856	-39.5684	38.1842	0.03869	1.6298	1.2767	-0.6298	354.7905
S/ Africa	6041	-12.6900	7.4230	0.04143	1.3839	1.1764	-0.4532	6.5435

Table 4.1 provides descriptive statistics for the return series in consideration. We observe that the mean of all the stock returns are positive, indicating that the overall price stock was slightly increasing. In addition, we observe the variance of the series completely dominates the mean. Furthermore, the estimates of the skewness are all negative. This highlights frequent small gains and a few extreme losses of the stock prices. The positive values of the excess kurtosis indicates the leptokurtic behavior of the returns. This implies that the series have a distribution with tails that are significantly fatter than those of the normal distribution.

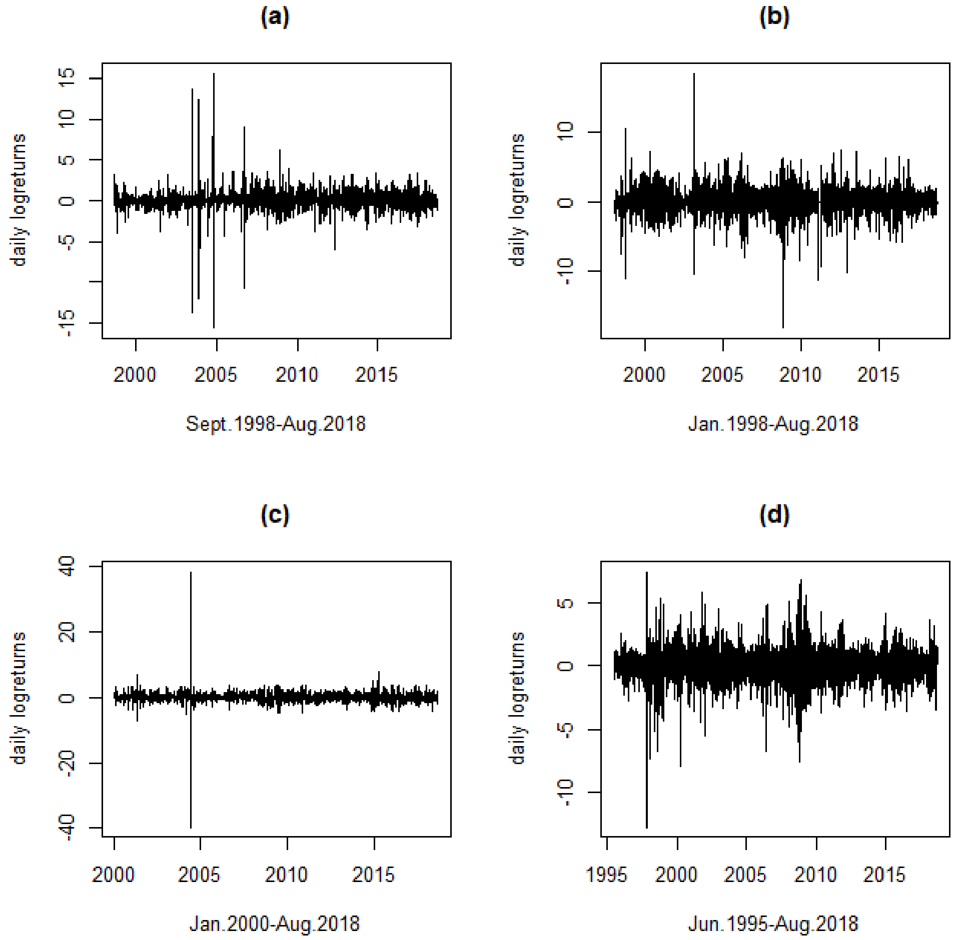

Figure 4.2: Time Series plots of the daily log returns of (a) Ivorian Stock (b) Egyptian Stock (c) Nigerian Stock series and the (d) South African Stock series respectively

Figure 4.2 shows the time series plots of the returns for each stock market. The patterns of these plots tend to fluctuate about a mean close to 0. There is a presence of extreme spikes that correspond to very small or big returns at some points, mostly caused by unforeseen events.

4.1.2 Stationarity Tests.

Table 4.2: Results for ADF tests for each Stock Returns

Log-returns	Test Statistic	p-value
Cote D'Ivoire	-15.866	0.01
Egypt	-15.948	0.01
Nigeria	-15.929	0.01
South Africa	-17.890	0.01

Stationarity of the return series are tested using the Augmented Dickey-Fuller (ADF) test. The

ADF test is set to lag 0 using the Schwartz Information Criterion (SIC). Results reported in Table 4.2 indicate that the null hypothesis of unit root is rejected at $\alpha = 0.05$. Therefore, the return series for all the stock returns can be considered to be stationary. Hence, we can statistically learn from these series.

4.1.3 Independence Test.

Table 4.3: Results for Runs test of Independence for each Stock Returns

Log-returns	Test Statistic	p-value
Cote D'Ivoire	-3.4788	0.0005036
Egypt	-10.2490	2.2e-16
Nigeria	-15.8820	2.2e-16
South Africa	-3.6803	0.000233

The results for the independence test is shown in Table 4.3. The null hypothesis of independence is rejected at values of α. This shows that the series show dependence behaviour and hence do not follow a random walk model.

4.1.4 Normality Tests.

Table 4.4: Results for Jarque test of Normality for each Stock Returns

Log-returns	Jarque-Bera Statistic	p-value
Cote D'Ivoire	127000	2.2e-16
Egypt	21185	2.2e-16
Nigeria	25490000	2.2e-16
South Africa	10995	2.2e-16

Table 4.4 provides the results of Jarque-Bera test of normality. The Jarque-Bera (JB) test statistic and p-values indicates that the returns are non-normal for all the return series. A more larger JB statistic and a smaller p-value is an indication that the series are far from normality. Hence, based on our results, we can say that the returns are heavy tailed. This is also evident by visualising the normal q-q plots in Figure 4.3.

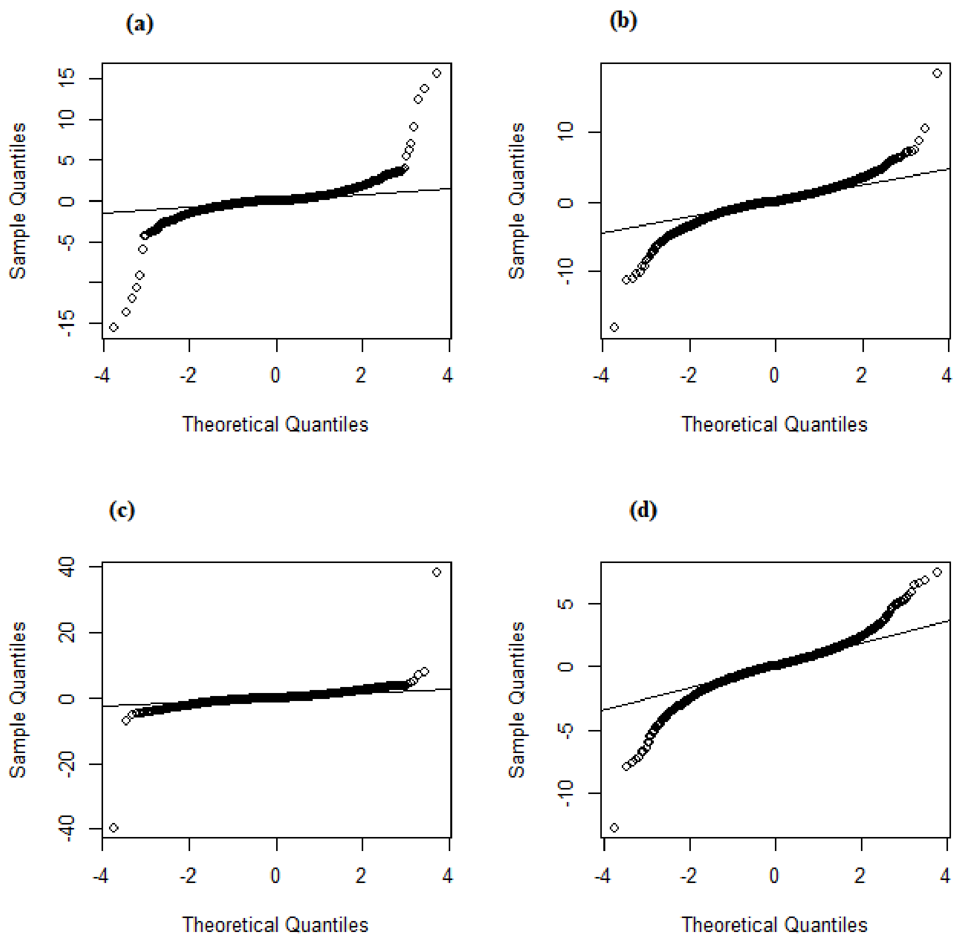

Figure 4.3: Normal Q-Q Plot of (a) Ivorian Stock series (b) Egyptian Stock series (c) Nigerian Stock series (d) South African Stock series Returns

4.1.5 Autocorrelation of Returns.

Visualizing the autocorrelation function (ACF) plots of the log returns in Figure 4.4 addresses the issue of non-existence of autocorrelation. At lag zero (0), the correlation coeffiecient is one. Secondly, the dependency in the conditional variance of the process can be captured by showing the ACF of the squared log returns in Figure 4.5. We observe that the autocorrelation coefficients are positive. In particular, whenever volatility clusters do exist, the squared log returns will show autocorrelation.

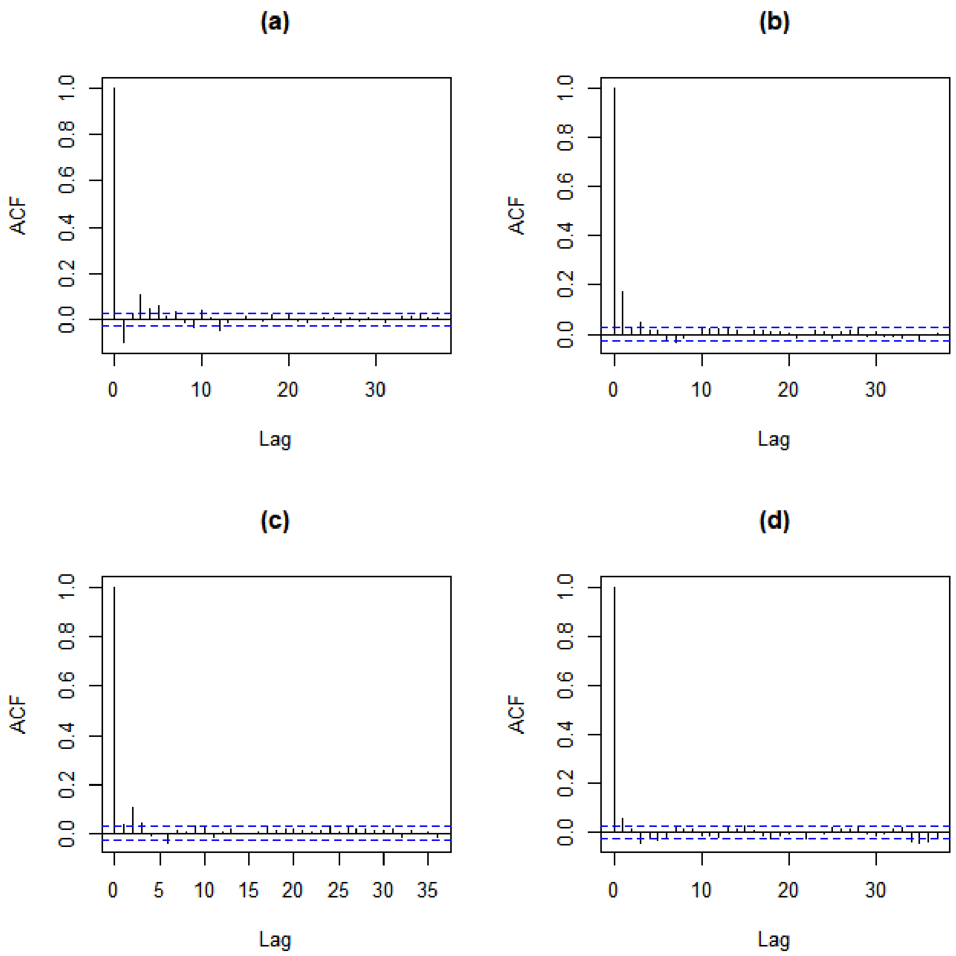

Figure 4.4: ACF plots of the returns for (a) Cote d'Ivoire (b) Egypt (c) Nigeria and (d) South Africa

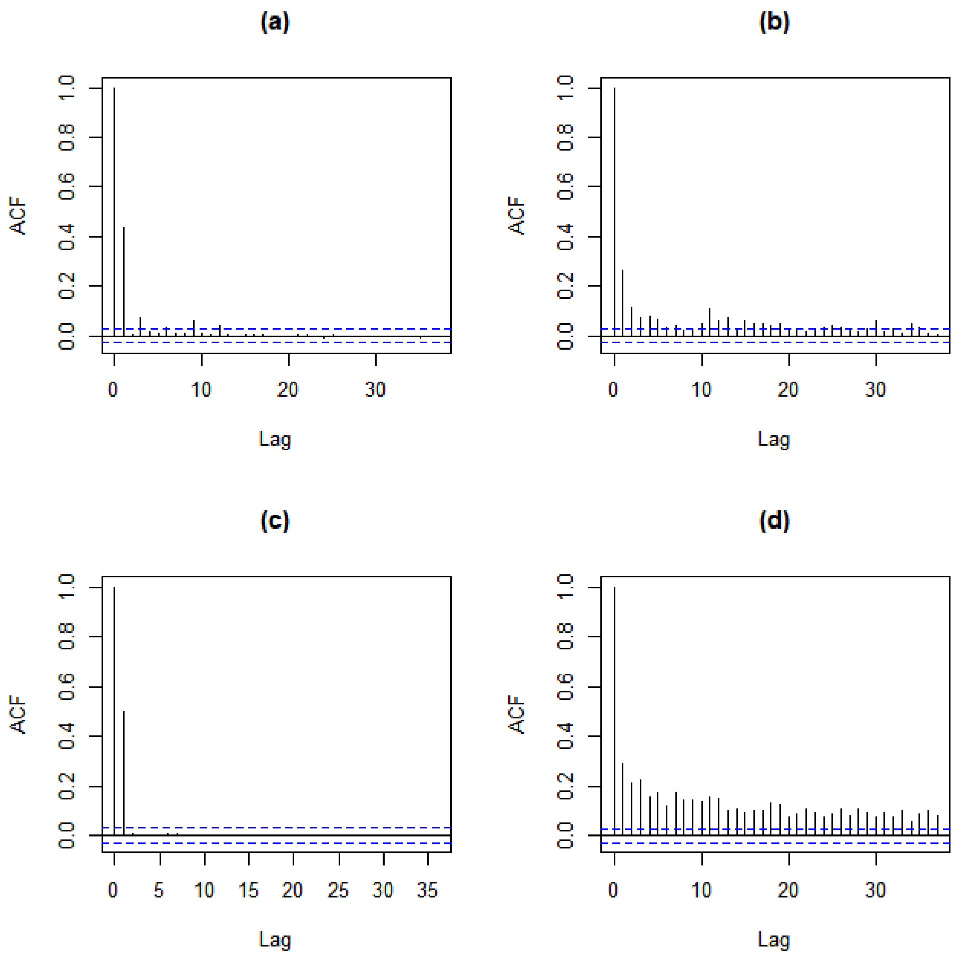

Figure 4.5: Squared ACF plots of the returns for (a) Cote d'Ivoire (b) Egypt (c) Nigeria and (d) South Africa

4.2 Parameter Estimates and Goodness of Fit Tests

We first visualize the performance of the six distributions on each stock return using the histogram and the Q-Q plot. Figures 4.6, 4.7, 4.9, 4.11, 4.13, 4.15 shows the histogram of the stock return series with their respective empirical distributions. It is observed that the NIG, Hyperbolic, GLD, Skewed t and Student t distribution fits the empirical histogram better than that of the Normal distribution. In addition, these five distributions are able the capture the skewness and the peak of the returns compared to the Normal distribution while the NIG, GLD and the hyperbolic distributions shows higher peak. Though none of the distributions captured the high peak of the Egyptian and South African stock return series.

Figures 4.8, 4.10, 4.12, 4.14, and 4.16 presents the Q-Q plots for different daily stock returns.

The Q-Q plot gives an insight of how well the tails of the data can adequately be modelled. We see that the empirical returns of South Africa fit into the GLD, the hyperbolic distribution properly captures the tails of the Nigerian daily stock returns while the empirical returns of Egypt and South Africa could be well modeled by the NIG.

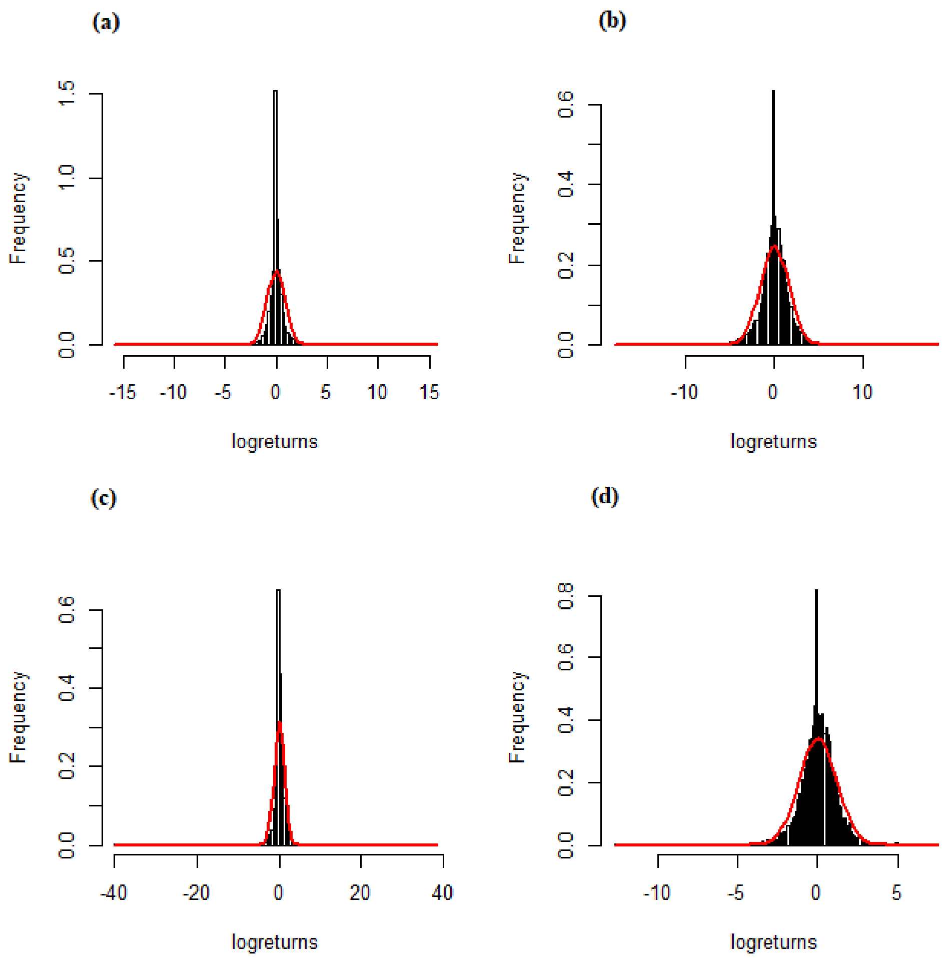

Figure 4.6: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Normal Distribution

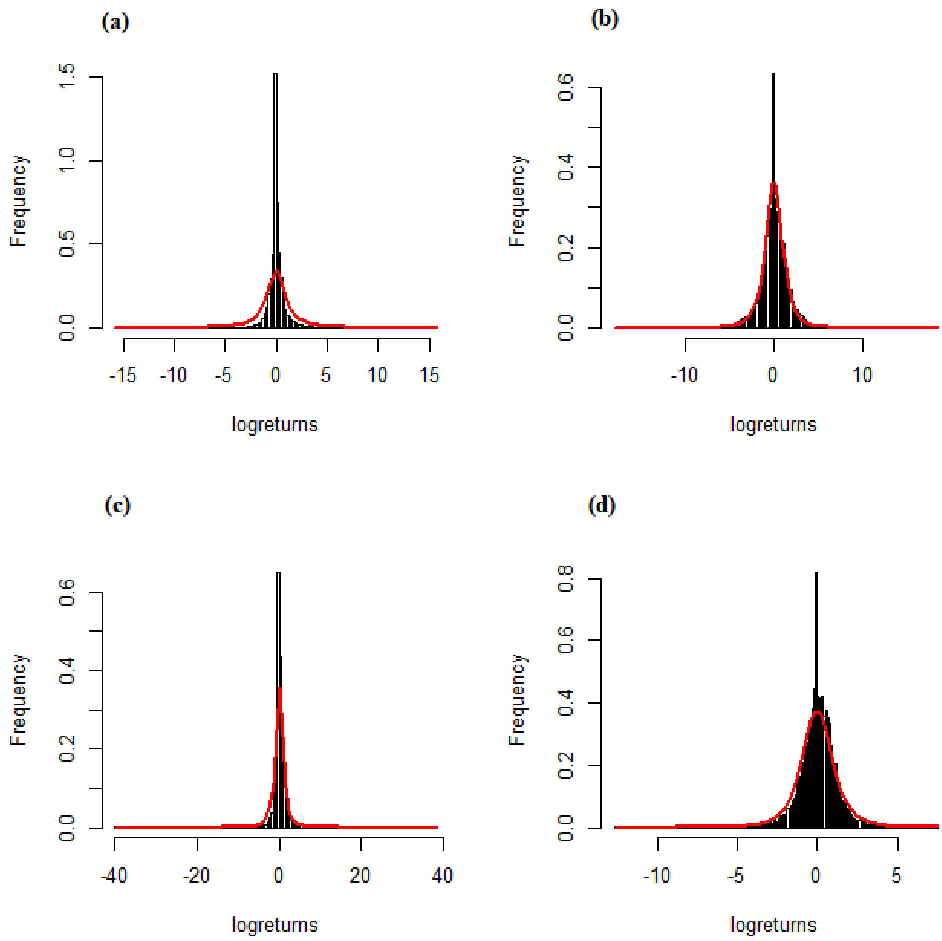

Figure 4.7: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa with the Student-t distribution superimposed.

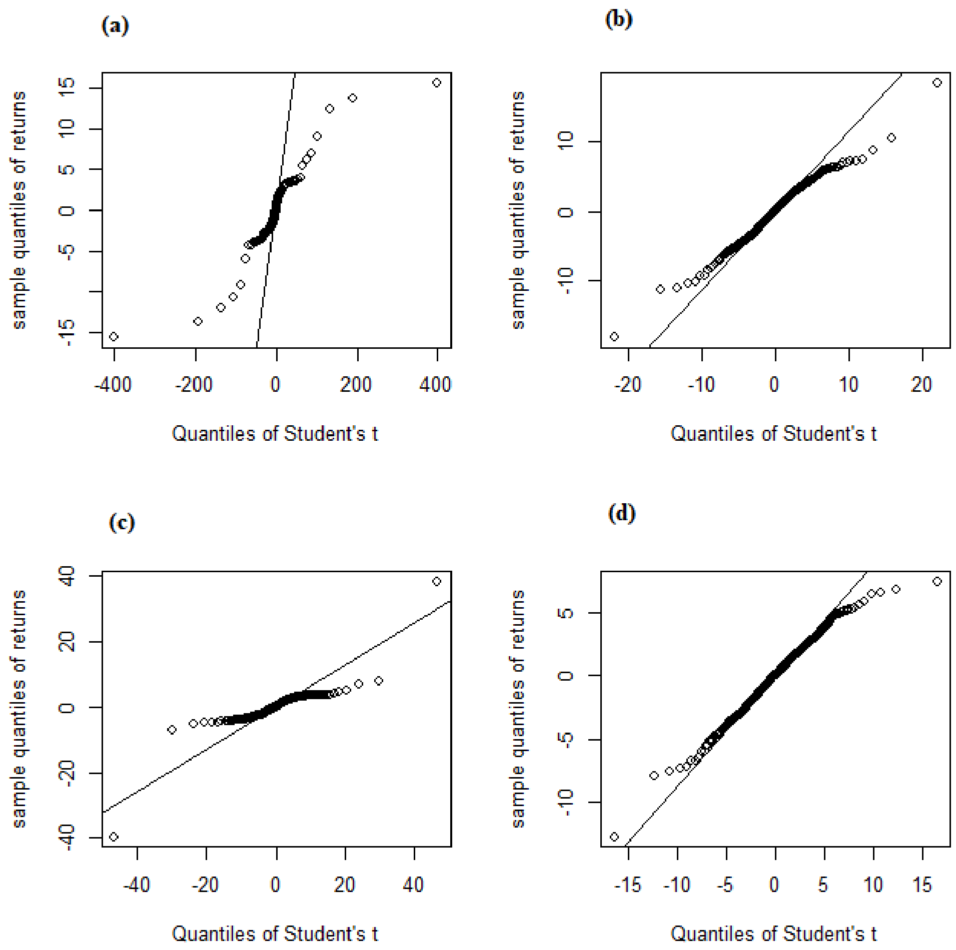

Figure 4.8: Q-Q Plots of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Student-t Distribution

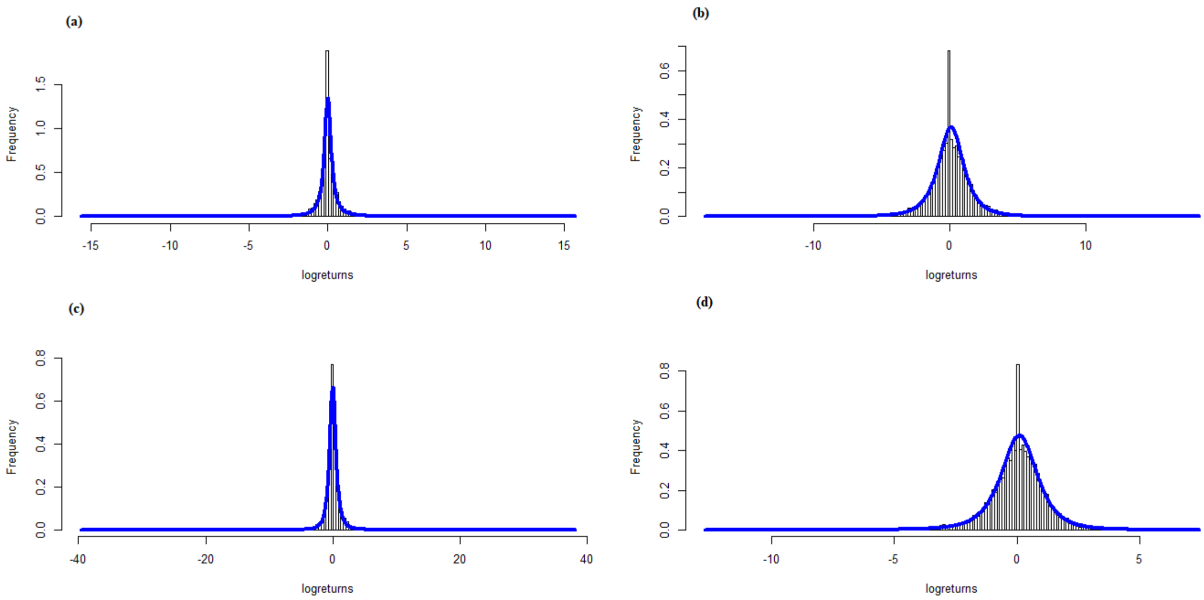

Figure 4.9: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa with the Generalized Lambda curve overlay.

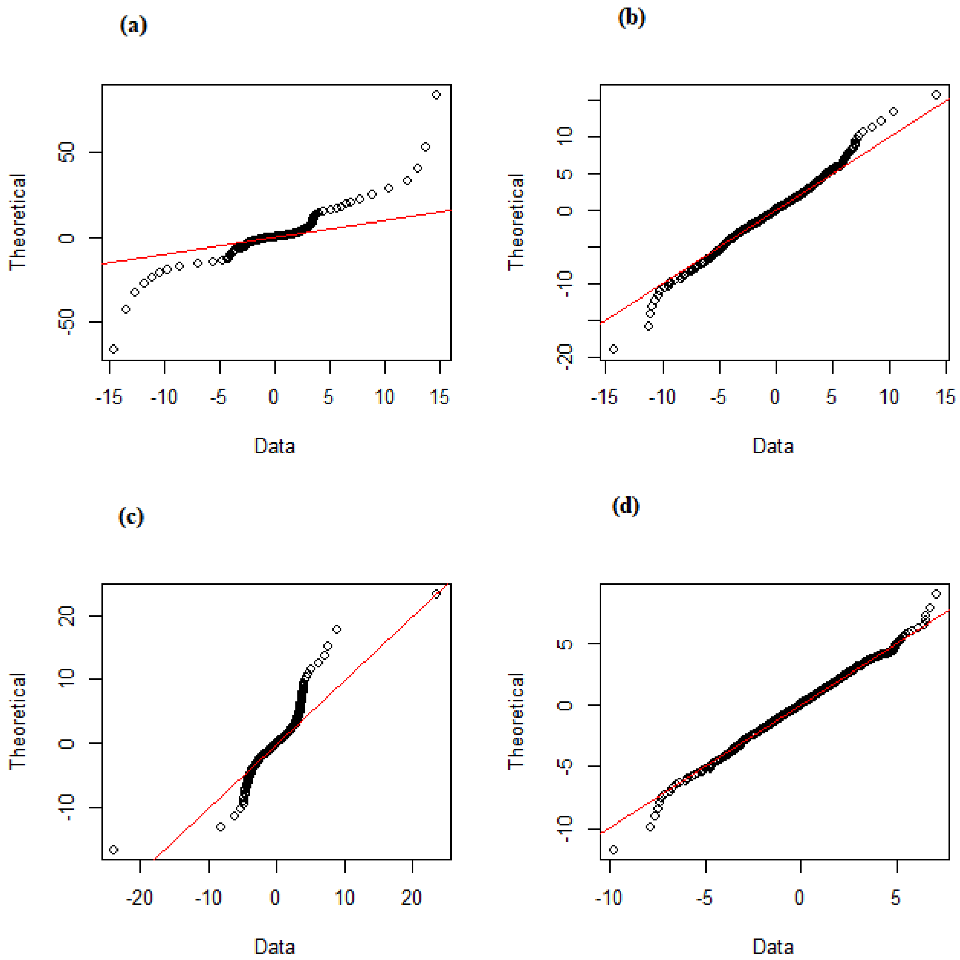

Figure 4.10: Q-Q Plots of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Generalized Lambda distribution

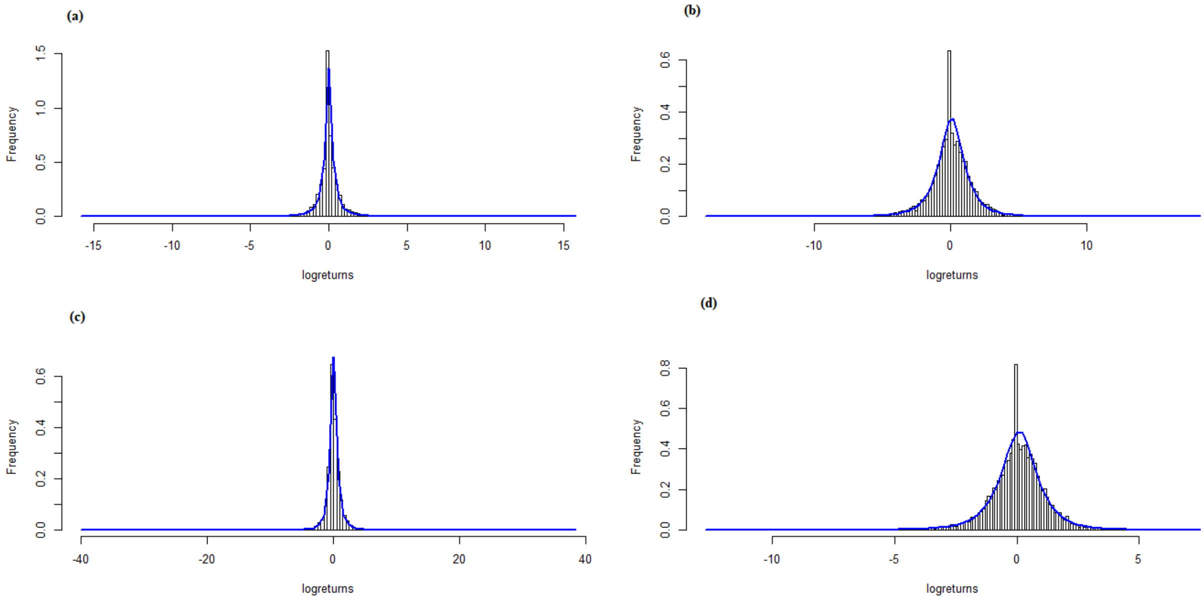

Figure 4.11: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa with the Normal Inverse Gaussian curve overlay.

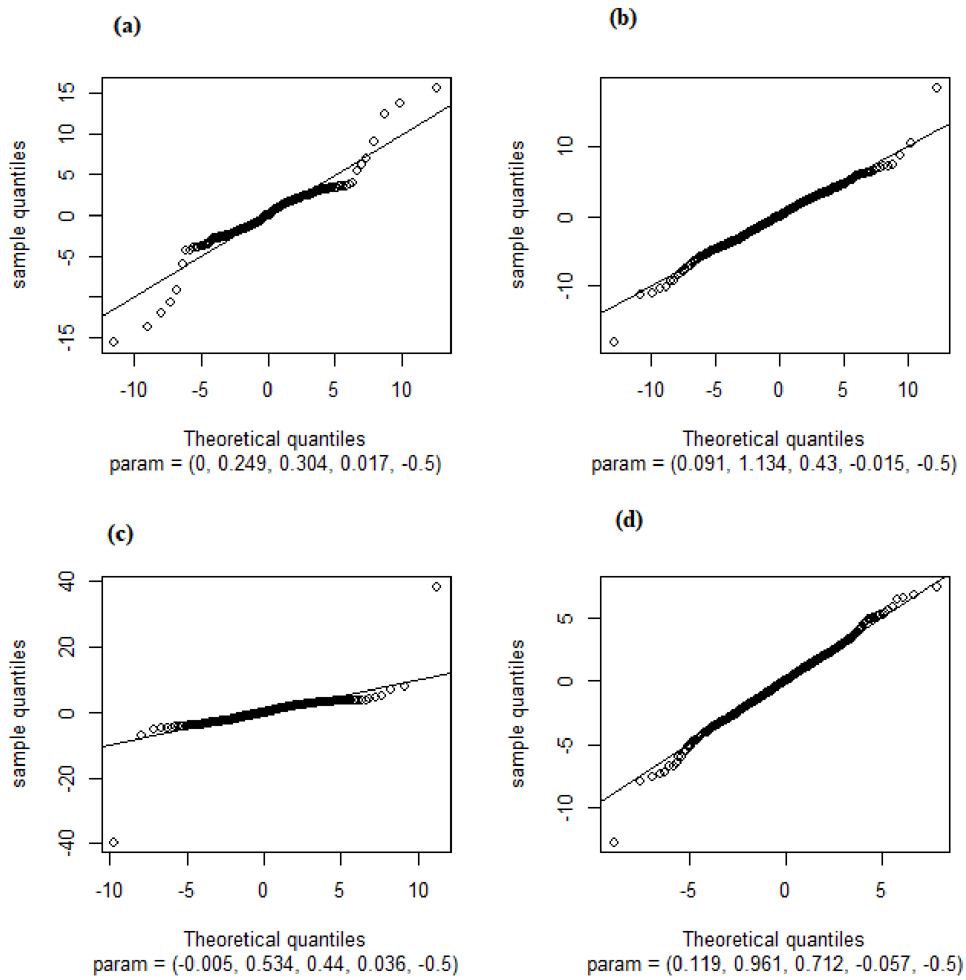

Figure 4.12: Q-Q Plots of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Normal Inverse Gaussian distribution

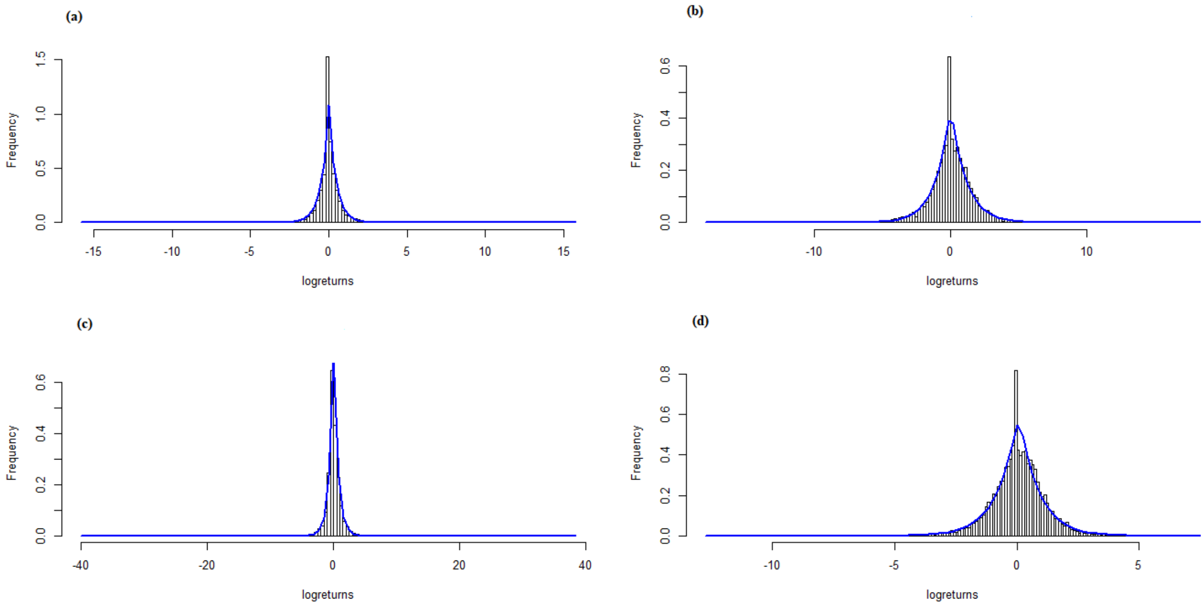

Figure 4.13: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa with the hyperbolic curve overlay.

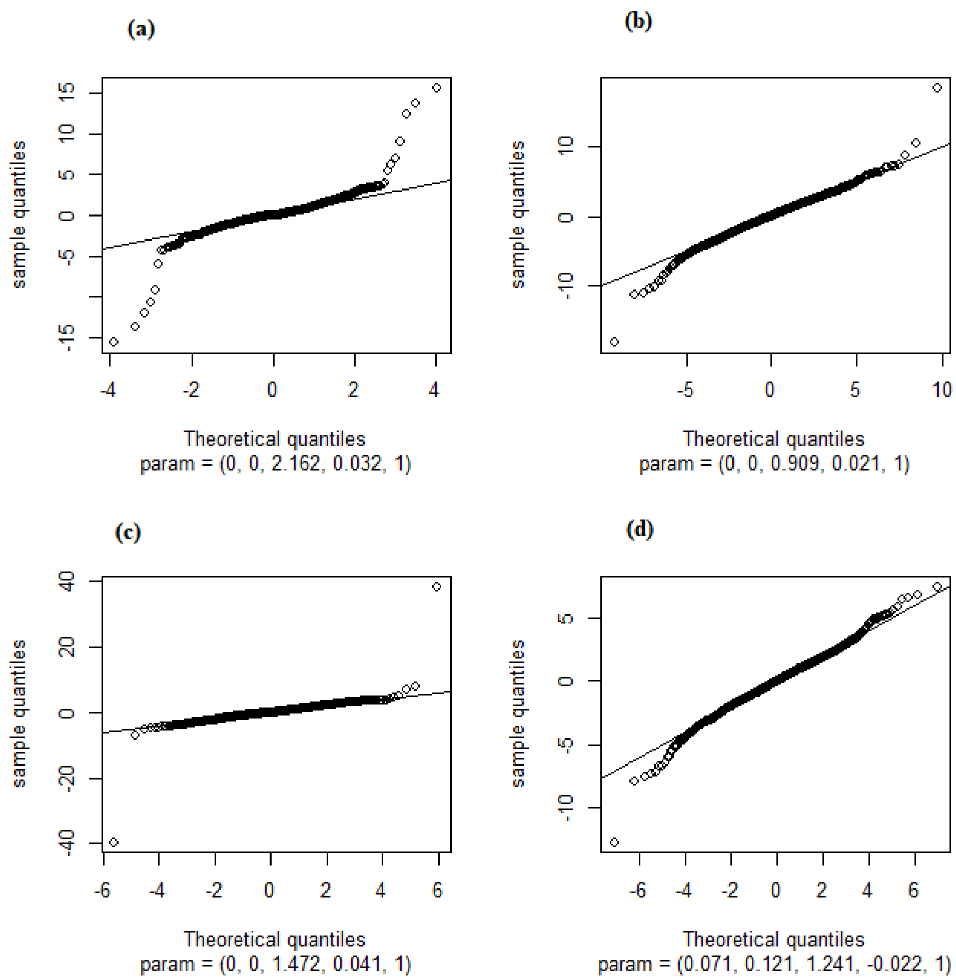

Figure 4.14: Q-Q Plots of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Hyperbolic distribution

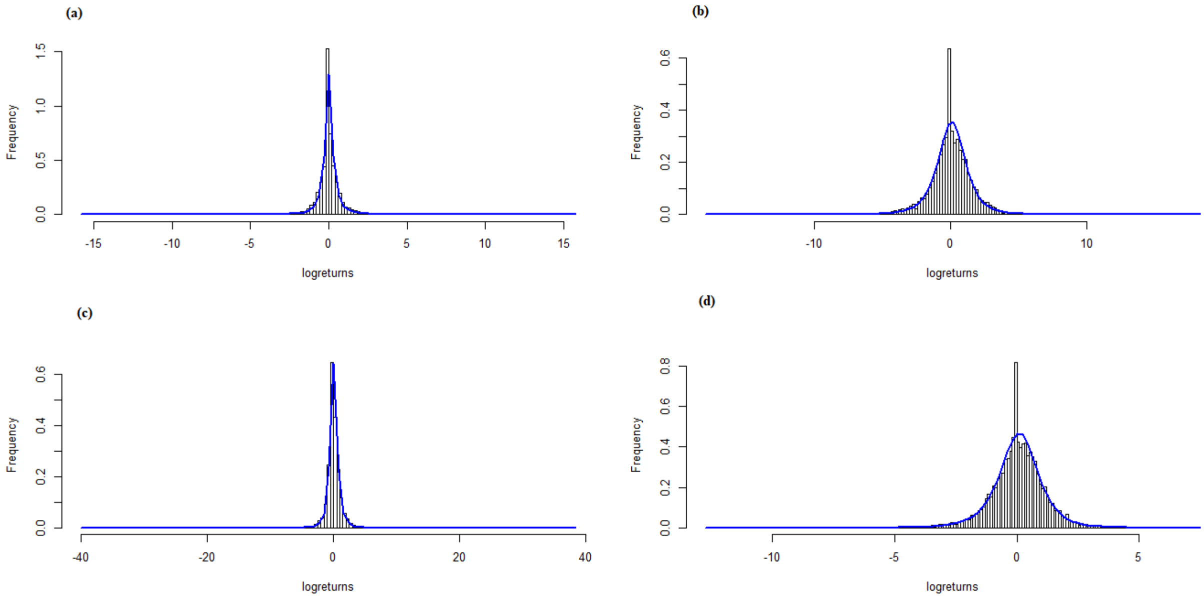

Figure 4.15: Histograms of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa with the Skewed-t curve overlay.

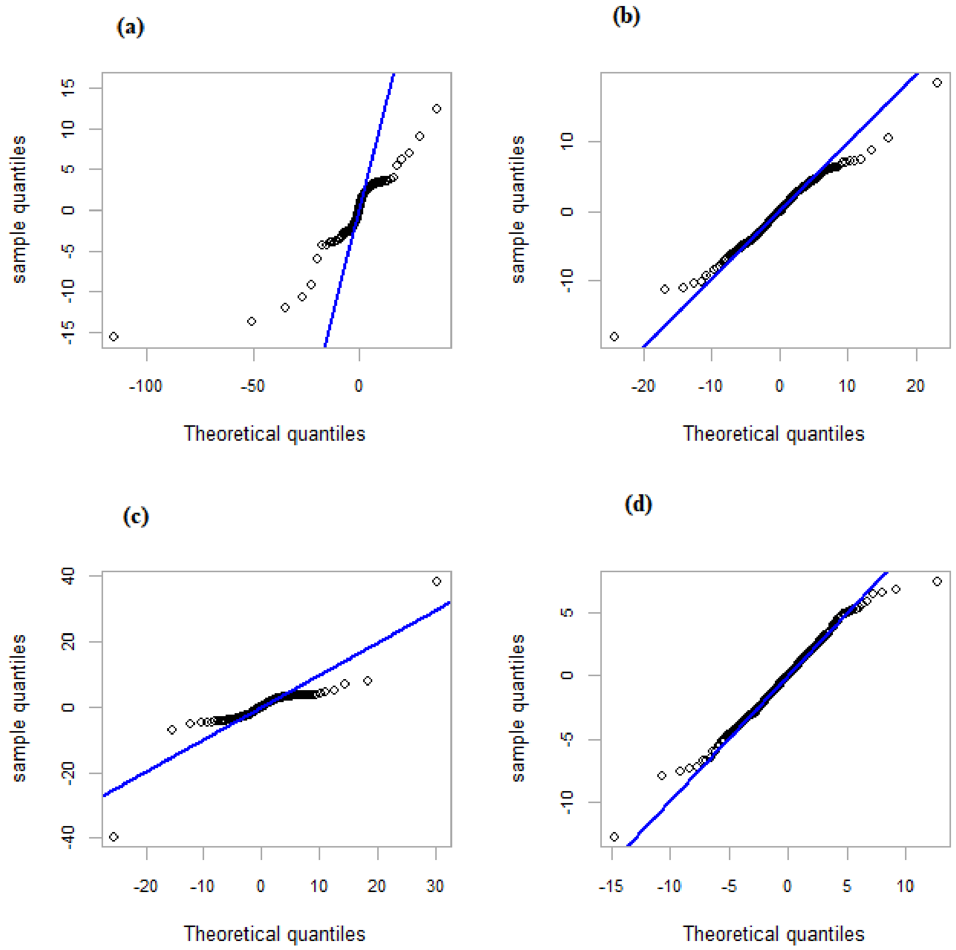

Figure 4.16: Q-Q Plots of the stock returns of (a) Cote d'Ivoire (b) Egypt (c) Nigeria (d) South Africa using the Skewed-t distribution

Table 4.5 displays the fitted distributions with their corresponding parameter estimates and standard errors of the log returns for Cote d'Ivoire. We have also included the goodness of fit criteria for the different fitted distributions. Our result shows that the normal inverse gaussian (NIG) distribution has the lowest values for AIC, BIC, and AICc and highest log-likelihood values. The generalized lambda distribution (GLD) gives the second lowest values for AIC, BIC, and AICc. The normal distribution producing the highest values for AIC, BIC, and AICc and least highest log-likelihood values. The NIG distribution may be deemed the best fit for the Ivorian stock returns.

In Table 4.6, the hyperbolic distribution gives the lowest values for AIC, BIC, and AICc and the highest log-likelihood value. This is followed by the NIG distribution which gives the second lowest values for AIC, BIC, and AICc while the skewed t distribution gives the highest values for –ln L, AIC, BIC, and AICc. Therefore, the hyperbolic distribution outperforms other models for the Egyptian stock returns. The hyperbolic distribution and the GLD favours the Nigerian and South African stock returns respectively based on the lowest AIC, BIC and AICc values and the highest loglikehood values as shown in Tables 4.7 and 4.8 respectively

Table 4.5: Fitted distributions, parameter estimates, standard errors, and goodness-of-fit criteria of the log returns of Cote d'Ivoire

					$L(\hat{\theta})$	AIC	BIC	AIC_c
Normal	$\hat{\mu}$: 0.013773; [0.012578]	$\hat{\sigma}$: 0.907277; [0.008894]			-6876.444	13756.89	13770	13756.89
Student t	$\hat{\mu}$: 0.001723; [0.004610]	$\hat{\sigma}$: 0.260075; [0.008894]	$\hat{\nu}$: 1.339881 [0.049596]		-4650.697	9307.394	9327.065	9307.399
GLD	$\hat{\lambda}_1$: 0.001708; [0.004318]	$\hat{\lambda}_2$: 8.516726; [0.357879]	$\hat{\lambda}_3$: -0.639761; [0.031516]	$\hat{\lambda}_4$: -0.670651 [0.032007]	-4609.333	9226.667	9252.895	9226.674
NIG	$\hat{\mu}$: -0.000201; [0.004595]	$\hat{\delta}$: 0.249079; [0.006550]	$\hat{\alpha}$: 0.304330; [0.02729]	$\hat{\beta}$:0.016885 [0.01691]	**-4577.854**	**9163.708**	**9189.936**	**9163.716**
Hyperbolic	$\hat{\mu}$: 9.409e-09; [NaN]	$\hat{\delta}$: 1.406e-07; [NaN]	$\hat{\alpha}$: 2.162e+00; [0.02995]	$\hat{\beta}$:3.241e-02 [0.02121]	-4799.162	9606.324	9632.5519	9606.3317
Skewed t	$\hat{\xi}$: -0.005924; [0.00876]	$\hat{\omega}$: 0.2602; [0.0080]	$\hat{\alpha}$: 0.0377; [0.0370]	$\hat{\nu}$: 1.3401 [0.0500]	-4650.17	9308.34	9334.568	9308.348

Table 4.6: Fitted distributions, parameter estimates, standard errors, and goodness-of-fit criteria of the log returns of Egypt

					$L(\hat{\theta})$	AIC	BIC	AIC_c
Normal	$\hat{\mu}$: 0.050556; [0.022165]	$\hat{\sigma}$: 1.626844 [0.015673]			-10265.36	20534.72	20547.91	20534.73
Student t	$\hat{\mu}$: 0.074543; [0.016965]	$\hat{\sigma}$: 1.022299; [0.019869]	$\hat{\nu}$: 2.966546 [0.147648]		-9694.969	19395.94	19415.71	19395.94
GLD	$\hat{\lambda}_1$: 0.076854; [0.016989]	$\hat{\lambda}_2$: 1.708591; [0.044468]	$\hat{\lambda}_3$: -0.233038; [0.019204]	$\hat{\lambda}_4$: -0.201760 [0.019691]	-9682.371	19372.74	19399.11	19372.75
NIG	$\hat{\mu}$: 0.09072; [0.02423]	$\hat{\delta}$: 1.13393; [0.03859]	$\hat{\alpha}$: 0.43015; [0.02478]	$\hat{\beta}$: -0.01521 [0.01230]	-9675.1	19358.2	19384.57	19358.21
Hyperbolic	$\hat{\mu}$: 1.017e-07; [NaN]	$\hat{\delta}$: 5.885e-07; [NaN]	$\hat{\alpha}$: 9.089e-01; [0.01239]	$\hat{\beta}$: 2.087e-02 [0.008758]	**-9638.344**	**19284.688**	**19311.055**	**19284.695**
Skewed t	$\hat{\xi}$: 0.13087; [0.05786]	$\hat{\omega}$: 1.02502; [0.020]	$\hat{\alpha}$: -0.06313; [0.062]	$\hat{\nu}$:: 2.97963 [0.149]	-9694.448	19396.89	19423.26	19396.90

Table 4.7: Fitted distributions, parameter estimates, standard errors, and goodness-of-fit criteria of the log returns of Nigeria

					$L(\hat{\theta})$	AIC	BIC	AIC_c
Normal	$\hat{\mu}$: 0.038687; [0.018318]	$\hat{\sigma}$: 1.276521 [0.012953]			-8075.902	16155.80	16168.78	16155.81
Student t	$\hat{\mu}$: 0.013319; [0.009919]	$\hat{\sigma}$: 0.551649; [0.012861]	$\hat{\nu}$: 2.198364 [0.100409]		-6383.363	12772.73	12792.19	12772.73
GLD	$\hat{\lambda}_1$: 0.010152; [0.009651]	$\hat{\lambda}_2$: 3.399923; [0.108035]	$\hat{\lambda}_3$: -0.320303; [0.023151]	$\hat{\lambda}_4$: -0.370253 [0.024823]	-6361.009	12730.02	12755.97	12730.03
NIG	$\hat{\mu}$: -0.004978; [0.011136]	$\hat{\delta}$: 0.533903; [0.016431]	$\hat{\alpha}$: 0.440320; [0.029727]	$\hat{\beta}$: 0.035537 [0.016125]	-6352.459	12712.92	12738.87	12712.93
Hyperbolic	$\hat{\mu}$: -1.328e-07; [0.0001757]	$\hat{\delta}$: 9.470e-07; [0.001219]	$\hat{\alpha}$: 1.472e+00; [0.02114]	$\hat{\beta}$: 4.145e-02 [0.01497]	**-6348.897**	**12705.794**	**12730.640**	**12705.802**
Skewed t	$\hat{\xi}$: -0.05792; [0.02624]	$\hat{\omega}$: 0.5547; [0.013]	$\hat{\alpha}$: 0.1519; [0.052]	$\hat{\nu}$:: 2.1994 [0.100]	-6379.094	12766.19	12792.14	12766.20

Table 4.8: Fitted distributions, parameter estimates, standard errors, and goodness-of-fit criteria of the log returns of South Africa

					$L(\hat{\theta})$	AIC	BIC	AIC_c
Normal	$\hat{\mu}$: 0.041429; [0.015134]	$\hat{\sigma}$: 1.176274 [0.010701]			-9552.575	19109.15	19122.56	19109.15
Student t	$\hat{\mu}$: 0.067861; [0.012307]	$\hat{\sigma}$: 0.796639; [0.013558]	$\hat{\nu}$:: 3.481227 [0.174362]		-9032.693	18071.39	18091.51	18071.39
GLD	$\hat{\lambda}_1$: 0.075618; [0.012448]	$\hat{\lambda}_2$:: 2.136536; [0.048065]	$\hat{\lambda}_3$: -0.191522; [0.017108]	$\hat{\lambda}_4$: -0.141141 [0.016371]	**-9023.989**	**18055.98**	**18082.80**	**18055.99**
NIG	$\hat{\mu}$: 0.11870; [0.02069]	$\hat{\delta}$: 0.96055; [0.03289]	$\hat{\alpha}$: 0.71183; [0.03749]	$\hat{\beta}$: -0.05699 [0.01911]	-9024.656	18057.31	18084.14	18057.32
Hyperbolic	$\hat{\mu}$: 0.07109; [NaN]	$\hat{\delta}$: 0.12108; [NaN]	$\hat{\alpha}$: 1.24104; [NaN]	$\hat{\beta}$: -0.02228 [NaN]	-9035.836	18079.672	18106.4973	18079.6786
Skewed t	$\hat{\xi}$: 0.20402; [0.04724]	$\hat{\omega}$: 0.80610; [0.0156]	$\hat{\alpha}$: -0.1970; [0.067036]	$\hat{\nu}$: 3.5008 [0.175885]	-9028.331	18064.66	18091.49	18064.67

4.3 Risk Estimation

In this section, we present the comparison of Value-at-Risk (VaR) estimates under the Normal, Student t, GLD, NIG, Hyperbolic and Skew t distribution to the empirical distribution of the stocks and indices. We further verify the accuracy of the VaR models using the backtesting technique.

4.3.1 Value-at-Risk Estimates.

Table 4.9: VaR Values of the Stock Returns Based on Empirical Distribution

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Cote d'Ivoire	-0.703	-1.110	-1.736	-2.323	-2.740	-5.624
Egypt	-1.666	-2.583	-3.653	-4.486	-5.529	-9.317
Nigeria	-0.975	-1.486	-2.360	-2.882	-3.590	-4.658
South Africa	-1.246	-1.803	-2.666	-3.233	-4.019	-6.689
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Cote d'Ivoire	0.712	1.202	1.848	2.428	3.097	6.075
Egypt	1.804	2.572	3.433	4.208	5.111	7.008
Nigeria	1.169	1.736	2.437	2.994	3.402	4.002
South Africa	1.298	1.791	2.470	3.057	3.798	5.408

Table 4.9 presents the VaR estimates using the empirical distribution at $\alpha = 0.1$, 0.05, 0.02, 0.01, 0.005 and 0.001. The empirical VaR values are based on quantile estimates of returns. Also, Table 4.10 - 4.13 display the VaR estimates for the four different daily stock returns using the theoretical distributions

For the Gaussian distribution and other heavy tail distributions, we predict the 1-day VaR at levels 10%, 5%, 2%, 1%, 0.5%, and 0.1%. The overall VaR estimates over the whole period of our data are compared to the empirical VaR values in Table 4.9 . It can be seen, as well-known in the literature, that the Gaussian distribution tends to underestimate the VaR values. Whereas, the existence of heavy tails in the NIG, GLD and the Hyperbolic distributions makes them better candidates for risk management.

Table 4.14 - 4.17 presents the number of violations of VaR for the different models, at different levels.

Table 4.10: VaR Values of the Ivorian Stock Returns Based on Theoretical Distributions

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	-1.149	-1.479	-1.850	-2.097	-2.323	-2.790
Student t	-2.372	-4.182	-8.449	-14.233	-23.913	-79.546
GLD	-0.603	-1.056	-2.054	-3.307	-5.257	-15.055
NIG	-0.599	-1.033	-1.863	-2.691	-3.681	-6.490
Hyperbolic	-0.727	-1.042	-1.460	-1.776	-2.092	-2.825
Skewed t	-0.602	-1.059	-2.135	-3.593	-6.032	-20.047
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	1.176	1.506	1.877	2.124	2.351	2.817
Student t	2.372	4.182	8.449	14.233	23.913	79.546
GLD	0.634	1.126	2.238	3.607	5.941	17.823
NIG	0.626	1.090	1.990	2.899	3.991	7.108
Hyperbolic	0.763	1.088	1.519	1.844	2.169	2.925
Skewed t	2.372	4.182	8.449	14.233	23.913	79.546

Table 4.11: VaR Values of the Egyptian Stock Returns Based on Theoretical Distributions

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	-2.034	-2.625	-3.291	-3.734	-4.140	-4.977
Student t	-1.643	-2.364	-3.506	-4.582	-5.907	-10.386
GLD	-1.644	-2.430	-3.649	-4.751	-6.042	-9.973
NIG	-1.660	-2.477	-3.691	-4.703	-5.784	-8.509
Hyperbolic	-1.706	-2.452	-3.437	-4.183	-4.923	-6.659
Skewed t	-1.626	-2.381	-3.576	-4.703	-6.090	-10.774
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	2.135	2.726	2.392	3.835	4.241	5.078
Student t	1.643	2.364	3.506	4.582	5.907	10.386
GLD	1.730	2.455	3.551	4.516	5.622	8.865
NIG	1.744	2.516	3.659	4.607	5.619	8.165
Hyperbolic	1.838	2.618	3.650	4.431	5.211	7.024
Skewed t	1.734	2.453	3.585	4.648	5.954	10.357

Table 4.12: VaR Values of the Nigerian Stock Returns Based on Theoretical Distributions

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	-1.597	-2.061	-2.583	-2.931	-3.249	-3.906
Student t	-1.813	-2.748	-4.415	-6.171	-8.544	-17.923
GLD	-0.960	-1.453	-2.280	-3.083	-4.082	-7.464
NIG	-0.969	-1.504	-2.368	-3.129	-3.971	-6.172
Hyperbolic	-1.045	-1.503	-2.108	-2.566	-3.024	-4.087
Skewed t	-0.941	-1.418	-2.261	-3.144	-4.337	-9.041
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	1.675	2.138	2.660	3.008	3.327	3.983
Student t	1.813	2.748	4.415	6.171	8.544	17.923
GLD	1.048	1.609	2.591	3.583	4.864	9.467
NIG	1.063	1.667	2.657	3.538	4.517	7.088
Hyperbolic	1.144	1.629	2.270	2.754	3.239	4.364
Skewed t	1.060	1.615	2.611	3.664	5.089	10.725

Table 4.13: VaR Values of the South African Stock Returns Based on Theoretical Distributions

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	-1.466	-1.893	-2.374	-2.695	-2.988	-3.594
Student t	-1.578	-2.226	-3.201	-4.075	-5.108	-8.370
GLD	-1.229	-1.794	-2.641	-3.379	-4.220	-6.656
NIG	-1.250	-1.835	-2.677	-3.361	-4.083	-5.872
Hyperbolic	-1.284	-1.854	-2.606	-3.175	-3.744	-5.065
Skewed t	-1.229	-1.777	-2.608	-3.356	-4.242	-7.041
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	1.549	1.976	2.457	2.778	3.071	3.676
Student t	1.578	2.226	3.201	4.075	5.108	8.370
GLD	1.299	1.797	2.510	3.107	3.762	5.550
NIG	1.302	1.819	2.551	3.143	3.764	5.298
Hyperbolic	1.350	1.899	2.625	3.174	3.723	4.998
Skewed t	1.288	1.770	2.487	3.125	3.875	6.230

Table 4.14: Number of Violations for the Ivorian Stock VaR Estimates

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	241	138	87	62	53	25
Student t	611	272	58	1	5	0
GLD	631	287	65	18	6	1
NIG	635	293	87	29	14	5
Hyperbolic	498	291	145	98	63	24
Skewed t	633	283	62	16	5	0
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	266	167	100	78	61	35
Student t	633	287	73	9	6	0
GLD	613	277	71	11	6	0
NIG	630	287	86	33	8	4
Hyperbolic	480	287	166	106	73	33
Skewed t	60	7	4	1	0	0

Table 4.15: Number of Violations for the Egypt Stock VaR Estimates

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	405	263	153	102	75	36
Student t	568	313	127	51	21	3
GLD	549	291	110	44	20	5
NIG	540	285	105	49	22	7
Hyperbolic	518	288	138	73	36	14
Skewed t	562	299	119	49	20	3
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	394	230	314	75	53	29
Student t	572	284	87	36	13	1
GLD	581	295	95	41	23	2
NIG	578	282	87	37	23	3
Hyperbolic	518	256	87	44	26	6
Skewed t	580	296	94	36	16	2

Table 4.16: Number of Violations for the Nigerian Stock VaR Estimates

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	214	140	75	47	35	16
Student t	474	238	92	32	4	1
GLD	499	255	108	40	12	1
NIG	489	238	97	36	14	2
Hyperbolic	439	238	135	75	44	12
Skewed t	516	267	112	36	7	1
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	260	152	74	49	33	5
Student t	592	306	96	24	4	1
GLD	560	286	80	14	4	1
NIG	554	264	75	17	4	2
Hyperbolic	500	277	118	70	33	5
Skewed t	555	280	78	13	3	1

Table 4.17: Number of Violations for the South African Stock VaR Estimates

Long position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	456	269	156	118	90	43
Student t	649	339	148	65	32	7
GLD	621	308	126	54	25	7
NIG	601	289	118	55	29	10
Hyperbolic	575	283	133	65	38	15
Skewed t	621	312	133	56	25	5
Short position						
α	10%	5%	2%	1%	0.5%	0.1%
Normal	430	234	123	88	60	33
Student t	584	279	104	47	25	2
GLD	603	299	114	59	32	6
NIG	602	289	110	58	31	7
Hyperbolic	566	259	103	57	32	11
Skewed t	613	312	119	59	30	4

4.3.2 Backtesting.

Table 4.18 gives the Kupiec likelihood ratio (LR) test with their corresponding p-values for the six fitted distributions for the Ivorian VaR models. We observe the following from the table. For the long position, the normal and the NIG distributions are statistically significant at $\alpha = 0.02$, while the normal and NIG are statistically significant. The normal and the NIG are not rejected at 1% and 0.5% VaR levels. Also at $\alpha = 0.01$, the student t, GLD, NIG and skewed t distributions are statistically significant. For the short position, all the distributions are rejected at $\alpha = 0.02, 0.01,$ and 0.005 VaR levels, meanwhile at 0.1% level of significance, the normal and the hyperbolic distributions are rejected. On average, the NIG seems to be the most effective VaR model.

The Kupiec LR test with their corresponding p-values for the Egyptian returns are displayed in Table 4.19. For both long and short positions, the normal distribution is not statistical significant at all levels of α. For the long position, the hyperbolic distribution is not significant at 2% VaR level while at 1%, 0.5%, 0.1%, the student t distribution is not statistically significant for risk estimation of short position. But it is important to note that the GLD and the NIG distributions were consistent for the long position, while for the short position, the hyperbolic distribution showed superiority with large p-values for all the α level of significance except at 2% VaR level.

Table 4.20 shows the Kupiec LR test for the Nigerian stock returns with their corresponding p-values. At 2% level of significance, all the fitted distributions are statistically significant except for the normal and hyperbolic distributions. In addition, we observe that at $\alpha = 0.01$ all the fitted distributions are also significant except the hyperbolic distribution. At 0.5% level of significance, the normal, GLD and the NIG are not rejected for significance. At 0.1% level of significance, all are statistically significant except the normal distribution. This is for the long position. With respect to the short position, at 2% level of significance, all the fitted distributions are statistically significant except the normal distribution, at $\alpha = 0.01$, only the normal distribution is statistically significant, meanwhile, the normal and hyperbolic distributions are statistically significant. The hyperbolic distribution is statistically significant at $\alpha = 0.001$ with the largest p-value.

Table 4.21 displays the Kupiec LR test for the South African stock returns with their corresponding p-values. The normal distribution is not statistically significant at all levels while the student t is not statistically significant at 2% level of significance for the long position. On the other hand, the normal distribution is not statistically significant at $\alpha = 0.01, 0.005, 0.001$.In general, the GLD appear to be a more better VaR model since they have a larger p-value when compared with other fitted distributions.

Table 4.18: Kupiec's test of VaR Models for Long and Short Positions for Cote d'Ivoire stock returns at $\alpha = 2\%, 1\%, 0.5\%, 0.1\%$

Long Position								
Fitted distribution	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	3.0207	8.221e-02	1.8182	1.775e-02	21.6025	**3.354e-06**	38.9635	**4.318e-10**
Student t	24.7300	**6.595e-07**	94.6603	**2.260e-22**	25.6228	**4.150e-07**	10.4112	1.253e-03
GLD	17.2430	**3.289e-05**	30.0722	**4.163e-08**	22.5043	**2.096e-06**	5.1109	2.378e-02
NIG	3.0207	8.221e-02	12.2604	**4.626e-04**	6.7088	9.594e-03	0.0080	9.286e-01
Hyperbolic	14.6618	**1.286e-04**	32.5683	**1.151e-08**	37.7370	**8.095e-10**	35.8573	**2.123e-09**
Skewed t	20.2547	**6.779e-06**	34.5760	**4.099e-09**	25.6228	**4.151e-07**	10.4112	1.253e-03
Short Position								
	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	134.2161	**4.900e-31**	11.3537	**7.530e-04**	34.2354	**4.883e-09**	74.0050	**7.791e-18**
Student t	195.7616	**1.757e-44**	54.8357	**1.310e-13**	22.5043	**2.097e-06**	10.4112	1.253e-03
GLD	201.0495	**1.233e-45**	48.1996	**3.850e-12**	22.5043	**2.097e-06**	10.4112	1.253e-03
NIG	163.9763	**1.531e-37**	8.0796	**4.477e-03**	17.2249	**3.321e-05**	0.3028	5.822e-01
Hyperbolic	Inf	**0.0000**	43.4906	**4.260e-11**	57.0985	**4.145e-14**	66.4749	**3.544e-16**
Skewed t	491.9510	**5.362e-109**	94.6603	**2.260e-22**	52.1605	**5.114e-13**	10.4112	1.253-03

Table 4.19: Kupiec's test of VaR Models for Long and Short Positions for Egypt stock returns at $\alpha = 2\%, 1\%, 0.5\%, 0.1\%$

Long Position								
Fitted distribution	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	17.1885	**3.385e-05**	34.4091	**4.466-09**	57.9115	**2.742e-14**	75.7147	**3.278e-18**
Student t	3.3248	6.824e-02	0.1572	6.917e-01	1.4226	2.330e-01	1.2628	2.611e-01
GLD	0.0480	8.265e-01	1.9485	1.628e-01	1.9712	1.603e-01	0.0285	8.659e-01
NIG	0.0717	7.889e-01	0.4586	4.983e-01	0.9696	3.248e-01	0.4414	5.065e-01
Hyperbolic	7.9728	**4.748e-03**	6.1760	1.295e-02	2.7720	9.593e-02	9.5297	2.022e-03
Skewed t	1.1618	2.811e-01	0.4586	4.983e-01	1.9712	1.603e-01	1.2628	2.611e-01
Short Position								
	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	NaN	**NaN**	7.4610	**6.300e-03**	19.7447	**8.851e-06**	50.5096	**1.186e-12**
Student t	4.3579	3.684e-02	6.7799	**9.219e-03**	8.9658	**2.750e-03**	5.4096	2.003e-02
GLD	1.6004	2.059e-01	3.3849	6.580e-02	0.6080	4.355e-01	2.8128	9.352e-02
NIG	4.3579	3.684e-02	5.9948	1.435e-02	0.6080	4.355e-01	1.2628	2.611e-01
Hyperbolic	4.3579	3.684e-02	1.9485	1.628e-01	0.0330	8.558e-01	0.0673	7.953e-01
Skewed t	1.8676	1.717e-01	6.7799	9.219e-03	5.2255	2.226e-02	2.8128	9.352e-02

Table 4.20: Kupiec's test of VaR Models for Long and Short Positions for Nigeria stock returns at $\alpha = 2\%, 1\%, 0.5\%, 0.1\%$

Long Position								
Fitted distribution	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	5.5738	**1.823e-02**	0.0512	8.210e-01	4.1825	4.084e-02	15.8936	**6.700e-05**
Student t	0.2803	5.965e-01	6.4849	1.088e-02	26.2181	**3.049e-07**	4.5546	3.283e-02
GLD	1.2006	2.732e-01	1.6216	2.029e-01	7.6773	5.592e-03	4.5546	3.283e-02
NIG	0.0002	9.902e-01	3.6045	5.762e-02	5.1652	2.304e-02	2.1654	1.412e-01
Hyperbolic	13.4607	**2.436e-04**	12.4689	**4.138e-04**	12.9598	**3.182e-04**	7.4351	6.396e-03
Skewed t	2.2181	1.364e-01	3.6045	5.762e-02	17.2093	**3.348e-05**	4.5546	3.283e-02
Short Position								
	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	6.1136	**1.341e-02**	0.0040	9.495e-01	2.8282	9.263e-02	0.0042	9.481e-01
Student t	0.0132	9.084e-01	15.4174	**8.620e-05**	26.2181	**3.049e-07**	4.5546	3.280e-02
GLD	3.2742	7.037e-02	34.5431	**4.169e-09**	26.2181	**3.049e-07**	4.5546	3.280e-02
NIG	5.5738	1.823e-02	27.6408	**1.461e-07**	26.2181	**3.049e-07**	2.1654	1.411e-01
Hyperbolic	4.2897	3.834e-02	8.4131	**3.725e-03**	2.8282	9.263e-02	0.0042	9.481e-01
Skewed t	4.1155	4.249e-02	37.1183	**1.112e-09**	30.1073	**4.088e-08**	4.5546	3.283e-02

Table 4.21: Kupiec's test of VaR Models for Long and Short Positions for South Africa stock returns at $\alpha = 2\%, 1\%, 0.5\%, 0.1\%$

Long Position								
Fitted distribution	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	9.5824	**1.960e-03**	43.3854	**4.495e-11**	77.5312	**1.306e-18**	95.0950	**1.815e-22**
Student t	5.8265	**1.579e-02**	0.3438	5.577e-01	0.1052	7.457e-01	0.1449	7.034e-01
GLD	0.2235	6.364e-01	0.7124	3.986e-01	0.9579	3.277e-01	0.1449	7.034e-01
NIG	0.0677	7.947e-01	0.5045	4.775e-01	0.0490	8.249e-01	2.1649	1.412e-01
Hyperbolic	1.2137	2.706e-01	0.3438	5.577e-01	1.8681	1.717e-01	9.3797	2.194e-03
Skewed t	1.2137	2.706e-01	0.3333	5.637e-01	0.9579	3.277e-01	0.1909	6.622e-01
Short Position								
	2%		1%		0.5%		0.1%	
	LR	p value	LR	p value	LR	p value	LR	p value
Normal	0.0399	8.417e-01	11.1555	**8.378e-04**	22.9184	**1.690e-06**	22.9184	**2.289e-14**
Student t	2.5063	1.134e-01	3.2554	7.119e-02	0.9579	3.277e-01	0.9579	5.563e-02
GLD	0.4003	5.269e-01	0.0335	8.548e-01	0.1052	7.457e-01	0.1052	9.867e-01
NIG	1.0190	3.127e-01	0.0984	7.534e-01	0.0208	8.852e-01	0.0208	7.034e-01
Hyperbolic	2.8216	9.300e-02	0.1982	6.562e-01	0.1052	7.457e-01	0.1052	7.050e-02
Skewed t	0.0281	8.668e-01	0.0335	8.548e-01	0.0014	9.701e-01	0.0014	3.758e-01

5. Conclusion

In the work, we modelled selected stock returns data from four Africa's emerging markets, Cote d'Ivoire, Egypt, Nigeria and South Africa using the Normal distribution, Student t distribution, GLD, NIG, Hyperbolic and Skew t distribution. First, we analyzed the data in order to get an overview of the stylised facts associated with stoc market returns. These distributions are able to capture some stylized facts such as, skewness and lepkurtosis which is fundamental when fitted to financial returns data. Also, we visualised the behaviour of the models on the empirical distribution using the histogram and Q-Q plots. Furthermore, we estimated the parameters of the models and evaluated their performances using the log-likelihood, AICc and BIC values. Our results showed that the NIG favoured the Ivorian stock returns, the hyperbolic distribution is a good fit for the Egyptian and Nigerian stock returns while the South African stock return is more appropriate to be modelled by the GLD. Finally, we calculated VaR under the Normal, Student t, GLD, NIG, Hyperbolic and Skew t distribution assumptions. The results obtained showed that the VaR calculated under the five distributions outperformed those under the Normal distribution. The Kupiec LR test further showed that the NIG, GLD and hyperbolic distributions dominated other distributions in terms of the VaR estimates at different level of significance. The tail behavior of the GLD, NIG and hyperbolic distributions implies that these models may be successfully employed to improve financial derivatives pricing models and the estimation of market risk using VaR methodology. In order to quantify the market risk one has to take into consideration both the distribution of individual returns in the portfolio and the dependency between the assets. Consequently, further research will focus on modeling extreme events using Extreme value theory (EVT) and dependency structure using the copula functions using these datasets

References

Maximum likelihood estimator. https://people.eecs.ku.edu/~jhuan/EECS940_S12/slides/MLE.pdf, nd.

Kjersti Aas and Ingrid Hobæk Haff. Nig and skew student's t: Two special cases of the generalised hyperbolic distribution. 2005.

H Akaike. Information theory as an extension of the maximum likelihood principle. á in: Petrov, bn and csaki, f. In *Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 276Á281*, 1973.

Carol Alexander. *Market models: A guide to financial data analysis*. John Wiley & Sons, 2001.

Rao Anup. Cse533: Information theory in computer science. https://homes.cs.washington.edu/~anuprao/pubs/CSE533Autumn2010/lecture3.pdf, 2010.

Josef Arlt and Markéta Arltová. Financial time series and their features. 2001.

Adelchi Azzalini and Antonella Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389, 2003.

Ole Barndorff-Nielsen and Christian Halgreen. Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Probability Theory and Related Fields*, 38(4):309–311, 1977.

Ole Eiler Barndorff-Nielsen. Normal\\inverse gaussian processes and the modelling of stock returns. 1994.

Park Barum. Aic and bic. https://barumpark.com/blog/2018/aic-and-bic/, 2018.

Robert C Blattberg and Nicholas J Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *The journal of business*, 47(2):244–280, 1974.

James V Bradley. Distribution-free statistical tests. Technical report, 1968.

Nedda Cecchinato. *Forecasting time-varying value-at-risk*. PhD thesis, Queensland University of Technology, 2010.

Yohan Chalabi, David J Scott, and Diethelm Wuertz. Flexible distribution modeling with the generalized lambda distribution. 2012.

Stanley H. Chan. Properties of maximum likelihood estimation (mle). unpublished, 2015.

Christopher Chatfield. *The analysis of time series: theory and practice*. Springer, 2013.

Canan G Corlu and Alper Corlu. Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance*, 15(11):1851–1864, 2015.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Dilanka S Dedduwakumara, Luke A Prendergast, and Robert G Staudte. An efficient estimator of the parameters of the generalized lambda distribution. *arXiv preprint arXiv:1907.06336*, 2019.

Marcel Dettling. Statistical analysis of financial data. *Swiss Federal Institute of Technology Zurich*, 2017.

Dragan Dorić. Return distribution and value at risk estimation for belex15. *Yugoslav Journal of Operations Research*, 21(1), 2016.

Darrell Duffie and Jun Pan. An overview of value at risk. *Journal of derivatives*, 4(3):7–49, 1997.

Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.

Marshall Freimer, Georgia Kollia, Govind S Mudholkar, and C Thomas Lin. A study of the generalized tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10): 3547–3567, 1988.

Héctor W Gómez, Osvaldo Venegas, and Heleno Bolfarine. Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics: The official journal of the International Environmetrics Society*, 18(4):395–407, 2007.

Gloria González-Rivera and Jesús Gonzalo. Introduction to time series. www.eco.uc3m.es/jgonzalo/teaching/timeseriesMA/Introduction, 2002.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

Cecil Hastings, Frederick Mosteller, John W Tukey, Charles P Winsor, et al. Low moments for small samples: a comparative study of order statistics. *The Annals of Mathematical Statistics*, 18(3):413–426, 1947.

Chun-Kai Huang, Knowledge Chinhamu, Chun-Sung Huang, and Jahvaid Hammujuddy. Generalized hyperbolic distributions and value-at-risk estimation for the south african mining index. *International Business & Economics Research Journal (IBER)*, 13(2):319–328, 2014.

John Hull. *Risk management and financial institutions,+ Web Site*, volume 733. John Wiley & Sons, 2012.

Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

Tzveta Iordanova and Cooking Raw Data. Introduction to stationary and non-stationary processes. *Retrieved September*, 19:2013, 2009.

Deepak Jadhav and TV Ramanathan. Parametric and non-parametric estimation of value-at-risk. *The Journal of Risk Model Validation*, 3(1):51, 2009.

Roger Kaufmann and Pierre Patie. Strategic long-term financial risks the one-dimensional case. *RiskLab Report, ETH Zurich*, 2003.

Paul Kupiec. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2), 1995.

Lesedi Mabitsela. *Evaluation of the South African equity markets in a value-at-risk framework*. PhD thesis, University of Pretoria, 2015.

Lesedi Mabitsela, Eben Maré, and Rodwell Kufakunesu. Quantification of var: A note on var valuation in the south african equity market. *Journal of Risk and Financial Management*, 8(1): 103–126, 2015.

Robert Nau. Notes on the random walk model. *Fuqua School of business, Duke University. 19p*, 2014.

Jeremy Orloff and Jonathan Bloom. Gallery of continuous random variables class 5, 18.05, spring 2014.

Adrian Pagan. The econometrics of financial markets. *Journal of empirical finance*, 3(1):15–102, 1996.

Anna-Lena Popkes. Kullback-leibler divergence. alpoples.com/files/kl-divergence.pdf, 2010.

K Prause. How to use nig laws to measure market risk. *fDM Preprint*, 65, 1999.

Karsten Prause et al. *The generalized hyperbolic model: Estimation, financial derivatives and risk measures*. PhD thesis, Verlag nicht ermittelbar, 1999.

R-bloggers. Predictability of stock returns: Using runs.test().

John S Ramberg and Bruce W Schmeiser. An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82, 1974.

Anup Rao. Cse533: Information theory in computer science. https://homes.cs.washington.edu/~anuprao/pubs/CSE533Autumn2010/lecture3.pdf, 2010.

Carl Scheffler. A derivation of the em updates for finding the maximum likelihood parameter estimates of the student'st distribution. *Technical note. URL www. inference. phy. cam. ac. uk/cs482/publications/scheffler2008derivation. pdf*, 2008.

Tom Smith. Autocorrelation. investopedia.com/terms/a/autocorrelation.asp, 2019.

Steve Su. Numerical maximum log likelihood estimation for generalized lambda distributions. *Computational Statistics & Data Analysis*, 51(8):3983–3998, 2007.

Ivica Terzić et al. Empirical estimation and comparison of normal and student t linear var on the belgrade stock exchange. *Sinteza 2014-Impact of the Internet on Business Activities in Serbia and Worldwide*, pages 298–302, 2014.

Scott Thompson. The stylised facts of stock price movements. *The New Zealand Review of Economics and Finance*, 1, 2013.

Hu Wenquing. Value at risk (var) and its calculations: an overview. http://web.mst.edu/~huwen/teaching_VaR_Weiqian_Li.pdf, 2016.

Wikipedia. Value at risk. https://en.wikipedia.org/wiki/Value_at_risk, 2019.