Modeling financial data using the multivariate generalized hyperbolic distribution and copula

L. E. KEMDA

December, 2015

Modeling financial data using the multivariate generalized hyperbolic distribution and copula

by

L. E. KEMDA

Dissertation submitted to the University of KwaZulu-Natal in fulfilment of the requirements for the Masters degree in Statistics.

Degree Assessment Board

Dissertation supervisor Mr. K. CHINHAMU
Mr. C. -K. HUANG

UNIVERSITY OF KWAZULU-NATAL

WESTVILLE CAMPUS, DURBAN, SOUTH AFRICA

Declaration - Plagiarism

I, L. E. KEMDA, declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowlegded as being sourced from other persons.
4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then
   (a) their words have been re-written but the general information attributed to them has been referenced, or
   (b) where their exact words have been used, then their writing has been placed in italics and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

L. E. KEMDA

Disclaimer

This document describes work undertaken as part of a masters programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

L. E. KEMDA

As the candidate's main supervisor, I have approved this dissertation for submission.

Mr. K. CHINHAMU

Abstract

Financial data usually possess some characteristics, such as volatility clustering, asymmetry, heavy and semi-heavy tails thus, making it difficult, if not impossible, to use Normal distribution to model them. Statistical analyzis shows that the Generalized hyperbolic distribution is appropriate for capturing these characteristics. This research shows that the USD/ZAR, All shares, Gold mining as well as the the S&P 500 returns are best modeled with the Skew t, generalized hyperbolic, hyperbolic, generalized hyperbolic distributions respectively based on AIC and Value-at-Risk (VAR) backtesting. Further multivariate analyzis of these returns based on the kernel smoothing goodness of fit shows that; the multivariate affine normal inverse gaussian (MANIG) distribution provides the best fit for the affine models. Likewise, the multivariate normal inverse gaussian (MNIG) distribution based on AIC provides the best model for the four returns. Finally, the positive tail dependencies exhibited between the All shares and Gold mining returns as well as All shares and S&P 500 returns is best modeled with the Gumbel and Clayton copulas respectively. While the negative dependencies between the USD/ZAR returns and other returns is modeled with the Frank copula.

Keywords

Financial returns, univariate distribution, multivariate distribution, generalized hyperbolic distribution, affine transformations, Kolmogorov-Smirnov, kernel smoothing, tails, Value-at-Risk, Archimedean copula.

Acknowledgements

I thank and praise the almighty God for giving me the power and patience to do this work right up to the end. Secondly, I thank my supervisors Mr. Knowledge Chinhamu and Mr. Chun-Kai Huang for their guidance to make sure I am on track to conclude this work.

I thank my parents Mr. Ofimboudem Signing Leon and Mrs. Nguetsa Celine, including my siblings for making it possible for me to be here at UKZN today and for their advice.

Finally, the financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF

Contents

List of Figures

List of Tables

Background

The assumption that financial data is normally distributed was long challenged by analysts like Mandelbrot (1963) and Fama (1965). Using data consisting of daily prices for each of thirty stocks of the Dow-Jones Industrial Average, Fama showed that the empirical distribution of this data was more peaked in the center and had larger tails than the normal distribution (Fama, 1965). He suggested the use of more stable and heavy tailed distributions. Different econometric models over the years were also widely suggested by researchers but they all proved to be inadequate to fit financial returns. Thus, distributions that adequately represented these tails were proposed.

Barndorff-Nielsen (1977) introduced the family of continuous type distributions in which the logarithm of the probability density function is a hyperbola, called the generalized hyperbolic distribution (GHD). This was purely motivated by the research carried out by Bagnold to model the grain size distribution of wind blown sand (Bagnold, 1954). These distributions proved to fit financial returns more adequately compared to other distributions like the normal and student t-distributions. Over the years, these distributions were widely introduced in finance to study data consisting of daily prices of the 30 DAX over a period of three years (Eberlein & Keller, 1995). They showed that the GHD presented the best fit to model data with high frequency. Similar research was also carried out by Bibby & Sorensen (1997), as well as Prause (1999). He (Prause) described algorithms to model financial returns of NYSE Industrial Index and Bayer share. He found that the GHD presented a better fit to the data compared to the normal distribution.

In this present work, we do not focus on the univariate case. We rather seek to investigate the fit of multivariate generalized hyperbolic distributions (MGHDs). Within this context, we introduce an alternative class of multivariate distributions which consist of affine linearly transformed random vectors with independent and generalized hyperbolic marginals. Blæsild & Jensen (1981) showed that these distributions are closed under margining, conditioning as well as affine transformations. Thus, these affine transformed random variables/vectors of generalized hyperbolic marginals will be considered here.

These analysis are carried out in such a way that later on, we incorporate these into a copula model in order to assess the dependencies that exist between the variables. The notion of copula is one that analysis multivariate distributions in order to describe the dependence that exist between multivariate variables without studying the individual marginal distributions of the variables. The idea of copula came into existence when Fréchet (1951) posed the following problem; given a random vector $(X_1, X_2, \ldots, X_n)$, with marginal distributions given by $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$ respectively, what can we say about the distribution of

$$(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)),$$

denoted $\Gamma(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$, called the **Fréchet class of the F_i**. This set is defined as

$$\begin{aligned} T &\in \Gamma(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \\ &\Leftrightarrow T(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty) = F_I(x_i). \end{aligned}$$

It was not until 1959 that a significant result was obtained when Sklar introduced the concept of copula. Since then this concept has been extensively used within the area of multivariate analysis. Even though this concept has been used in many fields of multivariate analysis, the first applications in finance were brought about by Frees & Valdez (1998) among others. They showed how data could be fitted to copulas but also described their usefulness in pricing insurance. A lot of research in the field of finance was also carried out by Embrechts *et al.* (1999) in his paper "Correlation: pitfalls and alternatives.", with coauthors McNeil and Straumann.

Even though this concept applies to multivariate variables in general, most research using copula make use of the bivariate case, and so in this project we will also explain the interaction of bivariate variables.

0.1 Review of literature

As mentioned earlier, investigations carried out using the MGHDs showed that these distributions are closed under marginalizing, conditioning as well as affine transformations (Blæsild & Jensen, 1981). Over the years, these distributions have been significantly incorporated into the field of engineering and finance for modeling. Protassov (2003) used a series of five foreign exchange returns to fit into the GHD for fixed parameter $\lambda = -1/2$ (Normal Inverse Gaussian Distribution). This research was aimed at fitting this particular distribution to returns of dimensionality greater than three. This is because such algorithms had always proved to be unsuccessful in the past. Hence, a simple EM algorithm based on maximum likelihood was used for a fixed GHD parameter λ. It should be noted that the version/parametrization used to derive the GHD here were those generated by linear combination of mixing generalized inverse Gaussian distributions.

However, Schmidt *et al.* (2006), introduced a particular class of these distributions called the multivariate affine generalized hyperbolic distribution (MAGHD) which proved to be easier to implement especially when dependence of extreme events was considered. These affine distributions also proved to have better estimation properties than the MGHDs themselves. Hence, these properties together with others such as for some parameter restrictions, the marginal of the MAGHDs become independent (which is not the case with the MGHDs), all contribute to the fact that any research involving the MGHDs prefer to use the MAGHDs. Their research also provided a backbone for further research that is carried out using the MGHDs. Moreover, they provided clear distinctions about the power and weaknesses of these two families of distributions. They went further by describing algorithms which can be used for estimating these distributions with simulated data.

Practical application of these models was carried out by Fajardo *et al.* (2005) in order to estimate the MAGHDs using some financial indices comprising: BVSP, CAC, DAX, FTSE, NIKK, S&P 500. Within this context, univariate and bivariate, as well as the six-dimensional distributions, were estimated and, using the Kolmogorov-Smirnov distances, the goodness of fit was assessed for the Normal Inverse Gaussian (NIG), Generalized Hyperbolic (GHYP) and Hyperbolic (HYP) classes. However, they were more focused on how to model multivariate distributions in the presence of correlation. This was the primary motivation for using the MAGHDs (in particular the Multivariate Affine Normal Inverse Gaussian (MANIG) and Multivariate Affine Hyperbolic (MAHYP)). This research showed the accuracy of the MAGHD in fitting real financial data. However, this application of MAGHD to finance did not end there as they went further to use a similar approach in order to show how to price multivariate derivatives when the underlying assets follow a MAGHD. As an application, the Sao Paulo Stock Exchange as well as the Brazilian Real/US Dollar was considered (Fajardo & Farias, 2010).

Thus, most of the research mentioned above illustrate that the GHDs seem adequate to model financial data as a whole. However, none of them explained the dependence that might exist between the variables, and more especially, they do not focus on the South African data context.

Recently, Konlack & Wilcox (2014) carried out a research using daily log returns of seven of the most liquid mining stock indices from the Johannesburg Stock Exchange in order to calibrate the MGHD as well as it subclasses. But also, assessing the stability of the estimated parameters of the MGHD. They found that the marginal parameter estimates were not stable for the subclasses of the GHD, except for the

Variance Gamma (VG) distribution. An interesting aspect of this research was the calibration of the MGHD, in which principal component analysis was first employed in order to reduce the dimension of the set of variables/stocks (due to correlation as they all came from the same family of stock: mining). The reduced set was then fitted with MGHDs. In this way, MGHDs were utilized to describe the dependencies (Konlack & Wilcox, 2014).

One of the most fundamental concepts behind risk management is the identification and quantification of dependency between the variables involved. While the Pearson's correlation is adequate for linear correlation, it is very inaccurate when other forms of correlation is investigated. The Kendall's tau as well as the Spearman's rho (which represent rank correlation measures) play a very important role in such cases. According to Embrechts *et al.* (1999), the Kendall's tau is very significant especially within copulas, due to their flexibility and ability to model dependency. Though there is a lot of theory about different copula models and calibrations, there exists very little empirical analyzis.

In our present work, we will use an alternative to the MGHD. We introduce the MAGHD due to their flexibility, as they allow the marginals to have a choice of parameter estimation. Rather than inheriting the parameters of the multivariate family in general. We will also extend the MAGHD and incorporate them into the copula model which is more appropriate in describing the relationship that exist among variables. Our analysis will not only involve indices from the same financial section (as mining in the previous case), but will incorporate indices like exchange rate, All shares index as well as S&P 500. At this point in time, no research has focused on using copulas to model dependency between variables of a MAGH model, in particular within the South African context.

0.2 Statement of the problem

Modeling multivariate data is usually not an easy task especially when such data is characterized by heavy tails and volatility clustering, just to name a few. Even when this is done, a suitable goodness of fit test has to be used to determine the accuracy of the fit.

MGHD and MAGHD have been proposed in the literature of multivariate financial data analysis. However, very few applications of these distributions exist and do not really apply to South African financial indices. Moreover, the Kolmogorov-Smirnov as a goodness of fit test has no practical application for dimension bigger than two.

Thus, considering four financial indices (three from the JSE and the S&P 500 index), how do we fit these distributions to MGHD and MAGHD? How do we then assess the goodness of these distributions? Finally, how do these indices vary with respect to one another?

0.3 Aims and Objectives of the Study

Our main aim will be to fit financial indices from the Johannesburg Stock Exchange using MGHD and MAGHD as well as their subclasses. We seek to find out which member of these distributions provide the best possible fit by using the kernel smoothing technique as a goodness of fit test. Finally, using Archimedean copulas, we investigate pairwise dependencies between the tails (upper and lower) of these indices.

0.4 Significance of the research

This research provides an alternative class of multivariate distributions which adequately fit multivariate indices within the South African context but as well as some foreign indices. In view of risk management, we describe the relationship between pairwise indices. Hence, enabling us to understand how the change in one index can trigger a corresponding change in the other index. This enables practitioners to control some variables in order to obtain particular responses. Moreover, the estimated models can also serve for forecasting and predicting for future, enhancing economic management.

0.5 Research layout

Having outlined the core reasons behind the use of MGHDs above, the rest of the thesis is structured such that in Chapter 1, we introduce this general class of asymmetric distributions. Firstly, we describe the univariate GHD as well as the parameters that define this distribution. Then, we generalize this univariate distribution in order to define the MGHD. Lastly in this chapter, we introduce the MAGHD as a particular case of the MGHD with affine transformed variables. Moreover, we discuss the parameter estimation procedure for these affine linearly transformed distributions.

In Chapter 2, we introduce copula models that will be utilized later to analyze dependency between variables. However, emphasis is placed on the bivariate Archimedean copulas. Hence, preliminary properties of copulas are discussed here including generating functions of these bivariate copulas.

In Chapter 3, we describe the methodology that will be used; the different statistical tests, including stationarity, independence, goodness of fit as well as tail dependency parameters for Archimedean copulas. We also look at the different copula parameters estimation procedures as well as statistical model selection criterion. Finally in Chapter 4, we apply all the methodology described in the previous chapters to the four returns under consideration. We begin with univariate distributions, then move to MAGHDs. Then, we look at MGHDs fits without affine transformed variables. We carry out appropriate goodness of fit tests for these models. In addition, we fit the bivariate returns to copulas and extract the tail dependencies.

Finally, in Chapter 5, we summarize all the analysis carried out in the conclusion as well as useful recommendations. We finish by providing other useful information related to this research in the appendix.

Chapter 1

The Multivariate Generalized Hyperbolic Distributions (MGHDs)

Semi-heavy and heavy tails, non-Gaussianity, lack of autocorrelation and volatility clustering are some features that characterize financial returns. Hence, the Normal distribution proved inadequate to model such data (Prause, 1999). Thus, over the years, the GHD has extensively been used as an alternative distribution for modeling. In this chapter, we introduce the MGHDs as a family consisting of elliptical distributions. Their ability to model dependencies between variables (especially extreme events), make then ideal for modeling. We introduce the univariate GHDs and some of their properties as preliminaries to their higher dimensional counterparts.

1.1 Univariate Generalized Hyperbolic Distribution (GHD)

Most of the theory related to univariate GHDs is referenced from Prause (1999). The ***GHD*** is a five parameter continuous distribution. A random variable, X, follows a generalized hyperbolic distribution denoted

$$X \sim gh(x; \lambda, \alpha, \beta, \delta, \mu),$$

where μ is a location parameter, δ the scale parameter, α determines the shape, β determines the skewness, and λ influences the kurtosis and characterizes the classification of the GHDs (Necula, 2009).

A random variable following this distribution has probability density function given by (*following Prause (1999)*)

$$\begin{aligned} gh(x;\lambda,\alpha,\beta,\delta,\mu) &= a_\lambda(\delta^2+(x-\mu)^2)^{(\lambda-1/2)/2} \\ &\quad \times K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})\exp\left(\beta(x-\mu)\right), \end{aligned}$$

where $a_\lambda = a(\lambda;\alpha,\beta,\delta) = \frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}, \quad x,\mu\in\mathbb{R}$, and K_λ is the modified Bessel function of the third kind.

It should also be noted that the domain of the parameters are as follows;

$$\begin{array}{ccccccc} \delta & \geq & 0, & |\beta|<\alpha, & \text{if } \lambda & > & 0 \\ \delta & > & 0, & |\beta|<\alpha, & \text{if } \lambda & = & 0 \\ \delta & > & 0, & |\beta|\leq\alpha, & \text{if } \lambda & < & 0 \end{array}$$

The mean and variance of this distribution is given by (Barndorff-Nielsen & Stelzer, 2005)

$$\begin{aligned} E(X) &= \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)}; \\ Var(X) &= \delta^2\left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)} + \frac{\beta^2}{\alpha^2-\beta^2}\left[\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)}\right)^2\right]\right), \end{aligned}$$

where $\zeta = \delta\sqrt{\alpha^2-\beta^2}$. The GHD consists of the following main classes; the Hyperbolic distribution, the Normal-Inverse Gaussian distribution, the Skew Student t-distribution, the Variance-Gamma distribution.

1.1.1 The Normal Inverse-Gaussian (NIG) Distribution

This is a sub-class of the GHD obtained when the parameter $\lambda = -\frac{1}{2}$. A random variable X is said to follow a normal Inverse Gaussian distribution denoted

$$X \sim nig(x;\alpha,\beta,\delta,\mu),$$

if its probability density function is given by

$$\begin{aligned} nig(x;\alpha,\beta,\delta,\mu) &= \frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\delta^2}+\beta(x-\mu)\right) \\ &\times \frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\sqrt{\delta^2+(x-\mu)^2}}, \end{aligned}$$

with $x, \mu \in \mathbb{R}$ and $\delta > 0, \quad |\beta| \leq \alpha$. Also (Aas & Haff, 2005),

$$\begin{aligned} E(X) &= \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}; \\ Var(X) &= \frac{\delta\alpha^2}{(\alpha^2-\beta^2)^{3/2}}. \end{aligned}$$

Similarly, the skewness and kurtosis are given by (Aas & Haff, 2005)

$$\begin{aligned} S &= 3\frac{\beta}{\alpha(\alpha^2-\beta^2)^{1/4}\delta^{1/2}}, \\ K &= 3\frac{\left(1+4(\frac{\beta}{\alpha})^2\right)}{\delta(\alpha^2-\beta^2)^{1/2}}. \end{aligned}$$

It should also be noted that this distribution has semi-heavy tails which decay according to the equation (Aas & Haff, 2006)

$$f(x) \sim const|x|^{-3/2}\exp(-\alpha|x|+\beta x), \quad \text{for} \quad x \to \pm\infty. \tag{1.1}$$

1.1.2 The Hyperbolic (HYP) Distribution

This is a sub-class of the GHD obtained when the parameter $\lambda = 1$ (Barndorff-Nielsen, 1997). Thus, a random variable X is said to follow a hyperbolic distribution denoted

$$X \sim hyp(x;\alpha,\beta,\delta,\mu),$$

if its probability density function is given by (Eberlein & Keller, 1995)

$$hyp(x;\alpha,\beta,\delta,\mu) = \frac{\sqrt{\alpha^2-\beta^2}}{2\delta\alpha K_1(\delta\sqrt{\alpha^2-\beta^2})}\exp\left(-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\right), \tag{1.2}$$

with $x, \mu \in \mathbb{R}$.

1.1.3 The variance-gamma (VG) distribution

We also have a subclass of the GHD family which occurs when the parameter $\delta \to 0$; called the variance-gamma distribution. A random variable X follows a variance-gamma distribution, denoted

$$X \sim VG(\lambda, \alpha, \beta, \mu),$$

if its probability density function is given by (Madan & Seneta, 1990)

$$\begin{aligned} VG(\lambda, \alpha, \beta, \mu) = & \frac{\gamma^{2\lambda}}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-1/2}} |x-\mu|^{\lambda-1/2} \\ & \times K_{\lambda-1/2}(\alpha|x-\mu|) \exp(\beta(x-\mu)) \qquad x \in \mathbb{R}, \end{aligned}$$

where $\Gamma(\lambda)$ is the Gamma function and $\gamma^2 = \alpha^2 - \beta^2$. The parameter domain is also given by; $\lambda > 0$ and $\alpha > |\beta|$. The mean and variance of this distribution are given by

$$\begin{aligned} E(X) &= \mu + 2\frac{\beta\lambda}{\gamma^2}, \\ Var(X) &= \frac{2\lambda}{\gamma^2}\left(1 + 2\left(\frac{\beta}{\gamma}\right)^2\right) \end{aligned}$$

1.1.4 The generalized hyperbolic skew student t (skew t) distribution

This definition follows from Aas & Haff (2006). As a subclass of the GHD, we also have the skew student t distribution with the parameter $\lambda = -\frac{\nu}{2}$ and $\alpha \to |\beta|$. A random variable X is said to follow a skew student t distribution if its probability density function is given by

$$\begin{aligned} f(x) &= \frac{2^{\frac{1-\nu}{2}} \delta^\nu |\beta^{\frac{\nu+1}{2}}| K_{\frac{\nu+1}{2}}(\sqrt{\beta^2(\delta^2 + (x-\mu)^2)})}{\gamma(\frac{\nu}{2})\sqrt{\pi}(\sqrt{\delta^2 + (x-\mu)^2})^{\frac{\nu+1}{2}}} \exp(\beta(x-\mu)), \qquad \beta \neq 0, \\ f(x) &= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\delta\Gamma(\frac{\nu}{2})}\left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{-\frac{(\nu+1)}{2}} \qquad \beta = 0, \end{aligned}$$

where we have used the fact that

$$K_\nu(x) = \sqrt{\frac{\pi}{2x}} \exp(-x), \quad x \to \pm\infty.$$

These definitions are obtained using the properties of the Bessel functions documented in Appendix 5.2 brought about by Abramowitz & stegun (1972). The mean and variance of this distribution are given by (Aas & Haff, 2005)

$$E(X) = \mu + \frac{\beta\delta^2}{\nu - 2},$$
$$Var(X) = 2\frac{\beta^2\delta^4}{(\nu-2)(\nu-4)} + \frac{\delta^2}{\nu - 2}$$

This distribution is the only subclass of the GHD which has one polynomial and one exponential tail, thus enabling them to handle heavy tails data well. However, they do not handle skewness adequately.

We also note that the tails of this distribution decay according to the equation (Aas & Haff, 2006)

$$f(x) \sim const|x|^{-\nu/2-1}\exp(-|\beta||x| + \beta x).$$

The skewness and kurtosis of this distribution are given by

$$S = 2\frac{(\nu-4)^1/2\beta\delta}{(2\beta^2\delta^2 + (\nu-2)(\nu-4))^{3/2}}\left[3(\nu-2) + \frac{8\beta^2\delta^2}{\nu-6}\right],$$
$$K = \frac{6}{(2\beta^2\delta^2 + (\nu-2)(\nu-4))^2}\Big[(\nu-2)^2(\nu-4)$$
$$+\frac{16\beta^2\delta^2(\nu-2)(\nu-4)}{\nu-6} + \frac{8\beta^4\delta^4(5\nu-22)}{(\nu-6)(\nu-8)}\Big]$$

1.1.5 Parametrizations of the GHDs

It should also be noted that different parametrizations of the GHD and its subclasses are also available with

$$\xi = (1 + \delta\sqrt{\alpha^2 - \beta^2})^{-1/2},$$
$$\chi = \frac{\xi\beta}{\alpha};$$
$$\zeta = \delta\sqrt{\alpha^2 - \beta^2};$$
$$\rho = \frac{\beta}{\alpha}.$$

These parametrizations are very important in our analysis as they help us to determine the behavior or nature of our data. For instance, the parameter ξ determines the nature of the heaviness of the tails; the closer it is to 1, the heavier the tails are. Also, for

$$\begin{aligned} \chi < 0, &\quad \text{the left tail is heavier than the right tail,} \\ \chi = 0, &\quad \text{the distribution is symmetric,} \\ \chi > 0, &\quad \text{the right tail is heavier than the left tail.} \end{aligned}$$

Hence, we can now introduce the MGHD based on two definitions that are often used in the literature. We also look at the disadvantages of each, so as to motivate our choice.

1.2 The MGHDs

In this section, we introduce the MGHD. However in this case, we introduce this distribution following the definition of Generalized Inverse Gaussian (GIG) distribution by mean-variance mixture McNeil *et al.* (2005). This is very important as the linear transformations obtained from the GIG preserve the same properties as those of the GHDs. Thus, we start by introducing the GIG distribution and Normal Mean-Variance Mixture distributions which essentially are the building blocks of the MGHDs.

Definition 1.1. The Generalized Inverse Gaussian (GIG) Distribution. According to McNeil *et al.* (2005), a random variable $\mathbf{X}$ is said to follow a GIG distribution, denoted

$$X \sim GIG(x; \lambda, \chi, \psi)$$

if its probability density function is given by (Barndorff-Nielsen *et al.* 1992)

$$G(x; \lambda, \chi, \psi) = \frac{(\psi\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} x^\lambda \exp\left(-\frac{1}{2}(\psi x^{-1}) + \psi x)\right) \tag{1.3}$$

where K_λ is the modified Bessel function with index λ given by

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty u^{\lambda-1} e^{-\frac{x}{2}(u+u^{-1})}, \tag{1.4}$$

and the parameter domain for the GIG distribution is given by

$$\begin{aligned}
\chi > 0, \psi \geq 0 \quad &\text{if } \lambda < 0, \\
\chi > 0, \psi > 0 \quad &\text{if } \lambda = 0, \\
\chi \geq 0, \psi > 0 \quad &\text{if } \lambda > 0,.
\end{aligned}$$

It should be noted that different parameterizations of the GIG distribution also exist, and for this distribution,

$$E(X) = \frac{\sqrt{\chi/\psi} K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}, \tag{1.5}$$

where the parameters follow the same domain as above.

Definition 1.2. The Normal Mean-Variance Mixture Distribution. Let $X \in \mathbb{R}^d$, be a random vector. We say $\mathbf{X}$ follows a normal mean-variance mixture distribution if

$$\mathbf{X} = \mu + W\gamma + \sqrt{W} AZ, \tag{1.6}$$

where

- $\mathbf{Z} \sim \mathbb{N}_k(\mathbf{0}, \mathbf{I}_k)$,
- $W \geq 0$ is a positive, scalar valued random variable with $Cov(W, \mathbf{Z}) = \mathbf{0}$,
- $\mathbf{A} \in \mathbb{R}^{d*k}$ is a matrix,
- Finally, μ, $\gamma \in \mathbb{R}^d$ are some parameter vectors.

This definition is referenced from Hu (2005).
For this distribution,

$$\begin{aligned}
E(\mathbf{X}) &= \mu + E(\mathbf{W})\gamma \\
COV(\mathbf{X}) &= E(W)\Sigma + Var(W)\gamma\gamma',
\end{aligned}$$

where $\Sigma = \mathbf{AA}'$. In the formulation above, W is interpreted as the shock that changes the volatility and mean of the normal distribution (Hu & Kercheval, 2007). Also, the following can be drawn from the normal mean-variance mixture distribution;

$$\mathbf{X}|W = N_d(\mu + W\gamma, W\Sigma). \tag{1.7}$$

Definition 1.3. The Multivariate Generalized Hyperbolic Distribution (MGHD). Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector and suppose $W \sim GIG(\lambda, \chi, \psi)$ from Definition 1.2 above. Then we that say X follows the MGHD.

However, this definition of MGHD does not pass on some of the important properties possessed by the general multidimensional family to the subclasses. For instance, the multivariate hyperbolic distribution (MHYP) does not have generalized hyperbolic marginal. As such, we introduce an alternative definition to the MGHD family which is rather straight forward and does not involve other distributions like the variance-mean mixtures described above.

Definition 1.4. Let $\mathbf{X}$=$(X_1, X_2, \ldots, X_n)$ be a random vector. We say $\mathbf{X}$ follows the MGHD with location vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and scaling matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{n*n}$, denoted $\mathbf{X} \sim MGHD_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\omega})$, where $\boldsymbol{\omega} = (\lambda, \alpha, \beta)$, if it admits the representation (Blæsild & Jensen 1981),

$$\mathbf{X} = \mathbf{A}'\mathbf{Y} + \boldsymbol{\mu},$$

where

- $\mathbf{A}' \in \mathbb{R}^{n*n}$ is a lower triangular matrix with $\mathbf{A}'\mathbf{A}$=$\boldsymbol{\Sigma}$ positive definite,
- $\mathbf{Y} \in \mathbb{R}^n$ has density given by

$$g_Y(y) = \frac{\alpha^{n/2}(1-\beta'\beta)^{\lambda/2} K_{\lambda-n/2}(\alpha\sqrt{1+y'y})}{(2\pi)^{n/2} K_\lambda(\alpha\sqrt{1-\beta'\beta})(1+y'y)^{n/4-\lambda/2}} e^{\alpha\beta' y} \tag{1.8}$$

One important characteristic of this definition is that when $\lambda = \frac{n+1}{2}$, we obtain the multivariate hyperbolic (MHYP) distribution, and in particular $\lambda = 1 (n = 1)$ brings us back to the univariate HYP distribution discussed earlier. Also, for $\lambda = -\frac{1}{2}$, we obtain the multivariate NIG (MNIG) distribution. Thus, an important aspect to note about the parametrization above is that it is invariant under affine transformations. Indeed, the following theorem explains this property (Hu & Kercheval, 2007). We will state without proving.

Theorem 1.1. *According to McNeil* et al. (2005)*, suppose* $\boldsymbol{X} \sim GH_n(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ *and* $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{b}$, $\boldsymbol{B} \in \mathbb{R}^{t*n}$ *and* $\boldsymbol{b} \in \mathbb{R}^t$. *Then* $\boldsymbol{Y} \sim GH_n(\lambda, \chi, \psi, B\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}', \boldsymbol{B}\boldsymbol{\gamma})$.

Just like the univariate subclasses of the GHDs, the MGHD consists of the subclasses; multivariate generalized hyperbolic (MGHYP), multivariate hyperbolic (MHYP), multivariate normal inverse gaussian (MNIG), multivariate variance gamma (MVG) as well as the multivariate skew t (MST).

1.3 Multivariate Affine Generalized Hyperbolic Distribution (MAGHD)

Definition 1.5. According to Schmidt *et al.* (2006), suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. We say $\mathbf{X}$ follows the MAGHD with location vector $\boldsymbol{\nu} \in \mathbb{R}^n$ and scaling matrix $\mathbf{\Sigma} \in \mathbb{R}^{n*n}$, denoted $\mathbf{X} \sim MAGHD_n(\boldsymbol{\mu},\mathbf{\Sigma}, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$=($\boldsymbol{\omega_1}$, $\boldsymbol{\omega_2}$, ..., $\boldsymbol{\omega_n}$) and $\boldsymbol{\omega_i}$=$(\lambda_i, \alpha_i, \beta_i)$, if it admits the representation

$$\mathbf{X} = \mathbf{A}'\mathbf{Y} + \boldsymbol{\mu},$$

where $Y = (Y_1, Y_2, \ldots, Y_n)$ and each $Y_i \sim MGHD_1(0, 1, \omega_i)$; each Y_i is univariate generalized hyperbolic distributed, and more importantly are **mutually independent**.

Thus, the multivariate subclasses will consist of; multivariate affine generalized hyperbolic (MAGHYP), multivariate affine hyperbolic (MAHYP), multivariate affine normal inverse gaussian (MANIG), multivariate affine variance gamma (MAVG) as well as the multivariate affine skew t (MAST).

In most research carried out using MGHD, the multivariate affine version is mostly used as it is more flexible. This allows each ω_i to be calculated for each marginal distribution. Hence, providing more accurate fits as each marginal can now be fitted independently with a GHD. This is not the case with the usual MGHD, as the parameters are predefined for the general family. Thus, the individual variables inherit these parameters. If the scaling matrix $\mathbf{\Sigma}$=$\boldsymbol{I}$ (the identity matrix) from the definition above, then the marginals are independent, and hence can be modeled with the MAGHD. However, the MGHD cannot be used if the marginal are independent.

Another important point to note about these two families is that within the field of risk assessment, MAGHD are used to model tail dependence while their MGHD counterparts do not because they are always tail dependent (Schmidt *et al.* 2006).

If $\mathbf{X} \sim MGHD_n(\mu, \Sigma, \omega)$, then the mean and covariance matrix of X are given by

$$\begin{aligned} E(\mathbf{X}) &= \mu + \alpha R_{\lambda,1}\left(\sqrt{\alpha^2(1-\boldsymbol{\beta}'\boldsymbol{\beta})}\right)A'\boldsymbol{\beta}, \quad \text{and} \\ Cov(\mathbf{X}) &= R_{\lambda,1}\left(\sqrt{\alpha^2(1-\boldsymbol{\beta}'\boldsymbol{\beta})}\right)\Sigma + \left[R_{\lambda,2}\left(\sqrt{\alpha^2(1-\boldsymbol{\beta}'\boldsymbol{\beta})}\right)\right. \\ &\quad \left. -R^2_{\lambda,1}\left(\sqrt{\alpha^2(1-\boldsymbol{\beta}'\boldsymbol{\beta})}\right)\right]\frac{A'\boldsymbol{\beta}\boldsymbol{\beta}'A}{1-\boldsymbol{\beta}'\boldsymbol{\beta}}, \end{aligned}$$

where $R_{\lambda,i}(x) = \frac{K_{\lambda+i}(x)}{x^i K_\lambda(x)}$. In particular, $E(X) = 0$ and $Cov(X) = \frac{K_2(\alpha)}{\alpha K_1(\alpha)} \times \mathbf{\Sigma}$ if $\boldsymbol{\beta} = (0, 0, 0, \ldots, 0)'$ and $\lambda = 1$.

Similarly, the following are obtained for the MAGHD;

$$\begin{aligned} E(\mathbf{X}) &= \boldsymbol{\mu} + \mathbf{A}'E(\mathbf{Y}), \\ Cov(\mathbf{X}) &= \mathbf{A}'\mathbf{C}\mathbf{A}, \end{aligned}$$

where $E(\mathbf{Y})$ is a vector of expectation of independent $Y \sim MGHD_1(0, 1, \omega_i)$, given by

$$E(Y_i) = R_{\lambda_i,1}\left(\sqrt{\alpha_i^2(1-\beta_i^2)}\right)\alpha_i\beta_i,$$
$$C = diag(c_{11}, c_{22}, c_{33}, \ldots, c_{nn})$$

with

$$c_{ii} = R_{\lambda_i,1}\left(\sqrt{\alpha_i^2(1-\beta_i^2)}\right)\Sigma$$
$$+\left[R_{\lambda_i,2}\left(\sqrt{\alpha_i^2(1-\beta_i^2)}\right) - R^2_{\lambda_i,1}\left(\sqrt{\alpha_i^2(1-\beta_i^2)}\right)\right]\frac{\beta_i^2}{1-\beta_i^2}.$$

Furthermore, if $c_{ii} = c$, $\forall i = 1, 2, 3, \ldots, n$, then $Cov(X) = c\Sigma$, (Schmidt *et al.* 2006).

1.4 Parameter Estimation of MAGHD

As mentioned earlier, the key features of the MAGHD is the freedom in which each marginals can freely choose their parameters, which is not the case with the MGHD. A two step procedure can be used to estimate the parameters of these multivariate distributions (Schmidt *et al.* 2006). To this end, we state the following proposition.

Proposition 1.1. *(Schmidt* et al. 2006*). Suppose $\boldsymbol{X} \sim MAGHD_n(\boldsymbol{\mu},\boldsymbol{\Sigma}, \boldsymbol{\omega})$, then $\boldsymbol{Z}$=$\boldsymbol{BX}$ is a set of n independently $GH(\omega, \delta, \mu)$, where $\boldsymbol{B}$ is the inverse Cholesky factorization of the $\boldsymbol{X}$ covariance matrix, S, where $S^{-1} = B'B$.*

Proof. (Fajardo & Farias, 2009)

Indeed, suppose $\mathbf{X} \sim MAGH_n(\boldsymbol{\omega}, \boldsymbol{\Sigma}, \boldsymbol{\psi})$. Then, $\mathbf{X} = \mathbf{A}'\mathbf{Y} + \boldsymbol{\psi}$, where A is upper triangular and $\boldsymbol{\Sigma}$ is positive definite, with $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$, and $Y_i \sim GH(\omega_i, 1, 0)$ are mutually independent, where $\mathbf{A}'$ is the transpose of $\mathbf{A}$.

Suppose S is the covariance matrix of X, then by Cholesky decomposition, $\mathbf{S} = \mathbf{P}'\mathbf{P}$.

Hence, $\mathbf{S}^{-1} = \mathbf{P}^{-1}(\mathbf{P}')^{-1} = \mathbf{P}^{-1}(\mathbf{P}^{-1})'$. Thus, if we define $\mathbf{B} = (\mathbf{P}^{-1})'$, then $\mathbf{S}^{-1} = \mathbf{B}'\mathbf{B}$ and hence $\mathbf{Z} = \mathbf{BX} = \mathbf{B}(\mathbf{A}'\mathbf{Y} + \boldsymbol{\xi}) = \mathbf{BA}'\mathbf{Y}$+$\mathbf{B}\boldsymbol{\xi}$.

But we know that $\mathbf{A}'\mathbf{A}$ is positive definite, thus $\mathbf{BA}'(\mathbf{BA}')' = \mathbf{BA}'\mathbf{AB}'$ is also positive definite and hence, $\mathbf{Z}$=$\mathbf{BA}'\mathbf{Y}$+$\mathbf{B}\boldsymbol{\xi}$. Thus, $\mathbf{Z}$ is a multivariate affine generalized hyperbolic distributed vector and hence a vector of independently distributed $GH(\omega, \delta, \mu)$ (theorem 1.1). □

With this proposition in mind, a two step procedure is used for parameter estimation of the multivariate distributions. The steps involved are (Schmidt *et al.* 2006);

- In the first step, we compute/estimate the covariance matrix, $\mathbf{S}$, of $\mathbf{X}$ vector (random data sample). We then carry out the transformation described in the proposition above in order to construct the vector $\mathbf{Z}$=$\mathbf{BX}$. This consists of independent generalized hyperbolic distributed marginals, where the matrix $\mathbf{B}$ is obtained by Cholesky decomposition of the covariance matrix, $\mathbf{S}$. That is $\mathbf{S}^{-1} = \mathbf{B}'\mathbf{B}$.

- In the second step, we estimate the univariate marginal distributions in order to obtain the parameter estimates; $(\alpha_i, \beta_i, \lambda_i)$. Hence using these marginal parameter estimates, we can estimates the multivariate estimates. Indeed, since $\mathbf{Z}$=$\mathbf{BX} \quad \rightarrow \mathbf{X}$=$\mathbf{B}^{-1}\mathbf{Z}$. But from above, $\mathbf{X} \sim MAGH(\boldsymbol{\omega}, \boldsymbol{\Sigma}, \boldsymbol{\xi}) \implies \mathbf{X}$=$\mathbf{A}'\mathbf{Y} + \boldsymbol{\xi}$ and $\mathbf{Z} \sim MAGH(\boldsymbol{\omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) \implies \mathbf{Z}$=$\mathbf{CY} + \boldsymbol{\mu}$.

  Thus

  $$\begin{aligned} \mathbf{X}=\mathbf{B}^{-1}\mathbf{Z} \quad &\implies \mathbf{A}'\mathbf{Y} + \boldsymbol{\xi} = \mathbf{B}'\mathbf{CY} + \boldsymbol{\mu}), \\ &\iff \mathbf{A}' = \mathbf{B}^{-1}\mathbf{C} \qquad \boldsymbol{\xi} = \mathbf{B}^{-1}\boldsymbol{\mu}, \end{aligned}$$

  where $\boldsymbol{\mu}$ is the location vector for Z and $\boldsymbol{\xi}$ is the location vector for $\mathbf{X}$. The scaling matrix, Σ is obtained from the expression $\boldsymbol{\Sigma} = \mathbf{B}'^{-1}\mathbf{D}\mathbf{B}^{-1}$, where $\mathbf{D}$ is a diagonal matrix with diagonal elements consisting of the scaling parameters of each of the independent Z_i marginals. Thus in this way we obtain the estimates of the multivariate model.

It should be noted that this procedure is quite computationally easy as it requires the estimation of n univariate distributions consisting of only five parameters each.

Chapter 2

Copulas

Statistical analysis usually involves describing the nature in which a variable behaves over time in a given random experiment. However, such single variables usually do not vary individually, but depend on other variables as well. It becomes necessary to not only analyze such single variables separately, but also how one variable affects or interact with the other variables. As such, we can draw some pattern or dependencies between these variables. Multivariate analysis involves the study or analysis of two or more variables simultaneously; how these variables vary together, their dependencies on each other. The concept of copula is one that helps to determine the dependence that exists between variables in multivariate analysis. We begin with a few definitions that will enable us to understand the concept behind copulas.

The following definitions are from Tsay (2005).

Definition 2.1. The uniform distribution. Consider X, a random variable. We say X follows the uniform distribution on $[a, b]$, denoted $X \sim U(a, b)$ if its probability density function is given by

$$U(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b; \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

For this distribution, $E(X) = \frac{b+a}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$.

Definition 2.2. (Tsay, 2005)**The copula**. A copula is a multivariate Cumulative Distribution Function (CDF) of univariate $U(0, 1)$ marginal distributions.

Thus, if $X = (X_1, X_2, \ldots, X_n)$ is a random vector such that each $F(X_i) \sim U(0, 1)$, with CDF $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$ respectively then the CDF of $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, denoted

$$C(F_1(x_1), F_2(x_2), \ldots F_n(x_n)) = F(x_1, x_2, \ldots, x_n)$$

is a **copula** if the following conditions hold;

- $C : [0,1]^n \in \mathbb{R}^d \rightarrow [0,1]$
- Boundary condition;

$$\begin{aligned} &C(x_1, x_2, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_n) = 0 \\ &C(1, 1, \dots 1, x_i, 1, \dots, 1) = x_i, \quad \forall i = 1, 2, 3, \dots, n, \quad x_i \in [0,1] \end{aligned} \tag{2.2}$$

- $\sum_{i_1=1}^{2} \cdots \sum_{i_n} = 1^2 (-1)^{i_1+i_2+\cdots+i_n} C(x_{1_{i_1}}, \dots, x_{n_{i_n}}) \geq 0$

Hence, given a copula $C(x_1, x_2, x_3, \dots, x_n)$, our main aim is to explain the relationship that exists between the individual random variables with realizations $x_1, x_2, x_3, \dots, x_n$, through the copula (Tsay, 2005).

Unlike the n-dimensional copula with very few construction schemes like in the Archimedean and elliptical copulas, the 2-dimensional copula has widely and extensively been discussed and used in the past. In this study, we will also consider the 2-dimensional copula.

Definition 2.3. The Bivariate Copula. Let X and Y be two univariate uniformly distributed random variables on $[0,1]$, with CDF $F_X(x)$ and $F_Y(y)$ respectively. Let also $C : [0,1] \times [0,1] \rightarrow [0,1]$. Then $C(F_X(x), F_Y(y)) = F(x,y)$ is a copula if the following conditions hold;

- $C(x,0) = C(0,y) = 0, \quad C(x,1) = x$, and $C(1,y) = y$,
- For any $[x_1, x_2] * [y_1, y_2] \in U(0,1)^2$, such that $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1) \geq 0$$

Thus, this copula contains information about the dependency that exists between the two random variables X and Y.

It is also relevant to note that most of the definitions and concepts associated with copulas here are obtained from Ruppert (2011) and Tsay (2005)

Theorem 2.1. ***Sklar Theorem in n-Dimensions.*** *According to Ruppert (2011), let F be an n-dimensional distribution function with marginals $F_1, F_2, \dots F_n$. Then there exists a copula C, such that*

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

If the margins $F_1, F_2, \dots F_n$ are continuous, then the copula C is unique.

This theorem is very important in multivariate analysis as it provides the basis for the construction of copula.

However, our focus on the bivariate case allows us to give the 2 dimensional Sklar theorem.

Theorem 2.2. ***The bivariate Sklar Theorem.*** *Let X and Y be two random variables with distribution margins F_X and F_Y respectively, and joint distribution $F(x, y)$. Then there exists a copula C such that*

$$F(x, y) = C(F_X(x), F_Y(y)), \tag{2.3}$$

and if the marginal F_X and F_Y are continuous, then the copula is unique (Tsay, 2005).

We now outline some characteristic features of copula models. Some of these features may also describe some basic copula models that are usually included in the theory of copula models. Some of these properties are given by Nelsen (2003), we just list a few.

- **Strictly increasing:** If $X = (X_1, X_2, \ldots, X_n)$, and $Y = (Y_1, Y_2, \ldots, Y_n)$ are two random vectors such that f_i is a strictly increasing function such that $Y_i = f_i(X_i)$. Then X and Y have the same copula.

- **Symmetry:** In this case we distinguish two kinds of symmetry;

  ***Plain symmetry*;** $C(u, v) = C(v, u)$, $\forall u, v \in [0, 1]$

  ***Radial symmetry*;** $C(u, v) = u + v - 1 + C(1 - u, 1 - v)$, $\forall u, v \in [0, 1]$

- Fréchet Bound Copula: Let C be a copula function. We define the Fréchet lower and upper bound Copulas respectively as $C_L(u, v) = max(u + v - 1, 0)$ and $C_U(u, v) = min(u, v)$, $u, v \in [0, 1]$. Note that these are the bivariate cases of the Fréchet Bound Copula. Furthermore, for any bivariate copula C, we have

  $$C_L(u, v) \leq C(u, v) \leq C_U(u, v), \qquad u, v \in [0, 1]$$

- **Comonotonicity:** Suppose we have a random vector given by $Y = (X, X, X, \ldots, X)$, such that $X \sim U(0, 1)$; that is Y is the vector which contains n identical copies of X. Then the n-dimensional comonotonicity copula is the copula model of the cumulative distribution function (cdf) of Y, given by

  $$C(x_1, x_2, \ldots, x_n) = min(x_1, x_2, \ldots, x_n),$$

  that is the Fréchet upper bound Copula.

- **Counter-monotonicity:** Suppose now that $Y = (X, 1 - X)$, and $X \sim U(0, 1)$.

Then the counter-monotonicity copula is the cdf of Y, defined by

$$C(u, v) = max(u + v - 1, 0).$$

That is, the bivariate Fréchet lower bound Copula.

- **Independence Copula:** Let $X = (X_1, X_2, \ldots, X_n)$ be a random vector of n independent variables such that each $X_i \sim U(0, 1)$, $\forall i$. Then we define the n-dimensional copula as

$$C(x_1, x_2, \ldots, x_n) = \Pi_{i=1}^{n} x_i$$

As mentioned earlier, our main use of copula in this project is to model the dependency that exist between our variables of interest. The reasons for using copula to model dependency is due to factors which may be described as follows;

- the correlation described by such measures is ideal mostly in cases when the variables are linearly related. However, the correlation between financial variables may be non-linear. Hence, copula within the context of non linear correlation may be ideal.
- Secondly, financial variables are characterized by semi-heavy and heavy tails and copulas are known to be adequate when such distributions are of interest.
- Finally, different estimation procedures are possible with copulas ranging from non-parametric to semi-parametric and finally parametric estimations. Hence, estimation using copulas usually involves a two stage procedure which in general is quite easy and fast.

Correlation as a measure of dependency even though is commonly used, has some limitations such as the fact that a correlation of 0 between two variables does not necessarily mean the variables are uncorrelated (or independent). Similarly, a correlation coefficient of 1 between a pair of variables does not necessarily means the variables are perfectly correlated. Secondly, linear transformations of correlated variables may not result in variables which are correlated in the same way as the original variables. Furthermore, the correlation between a pair of variables is only defined in cases where the variables under consideration have finite variances and covariances. However, some variables have infinite variation; like that exhibited by extremely upper and lower tailed variables.

2.1 Implicit and explicit copulas

In this section, we consider some classical families of copula models that are often used by practitioners in the field of statistics, as they form the bases for construction of more complex copula models. Here, we classify them according to two sub groups; implicit and explicit copulas.

2.1.1 Implicit Copulas

As their name might suggest, these copula models have no particular closed forms. They are obtained from some known distribution functions. Amongst such copula models, we find the **Gaussian** and **student** t copulas.

Suppose ϕ is a function representing the univariate standard normal distribution. Then the density of the bivariate Gaussian copula is given by the integral equation (Aas, 2004);

$$C_\rho(u,v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dxdy, \tag{2.4}$$

where ϕ^{-1} is the inverse of the standard univariate Gaussian distribution and ρ is the copula parameter.

On the other hand, suppose t_ν represents the univariate student t distribution with ν degrees of freedom. Then, we define the bivariate Student t copula by the integral equation

$$C_{\rho,\nu}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2} dxdy; \tag{2.5}$$

where, ρ and ν represent the copula parameters and t^{-1} represents the inverse of the standard univariate t distribution with ν degrees of freedom.

It should also be noted that the student t copula allows for joint fat tailed distributions even if the individual marginal do not allow for such tails (Aas, 2004).

2.1.2 Explicit copulas

Unlike the implicit copulas which do not assume a particular closed form, the explicit copulas are characterized by having a particular closed form. In fact, they are completely specified by a generator function. Thus, unlike the implicit copulas, they are not obtained from multivariate distribution functions. Amongst these class of copula includes the well known Archimedean copulas.

Definition 2.4. Let ϕ be a continuous, strictly decreasing convex function such that

$$\phi : [0, 1] \to [0, \infty],$$

with $\phi(0) = \infty$ and $\phi(1) = 0$. The archimedean copula with **strick generator** is the copula defined by (Ruppert, 2011)

$$C(u_1, u_2, u_3, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \phi(u_3) + \cdots + \phi(u_n)),$$

where ϕ is defined as above is a strick generator.

In particular, the bivariate Archimedean copula is given by

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)).$$

It should be noted that such generator functions are not uniquely defined. Feller (1971) suggests that the inverse of Laplace transform of cumulative distribution functions also correspond to a very important class of generator functions for Archimedean copulas. There are many Archimedean copulas in the theory of copulas but we will only focus a few of them that are of direct interest to us. They include; Clayton copula, Gumbel copula and Frank copula.

Clayton Copula

The Clayton copula is an asymmetric copula model with generator given by

$$\phi(t) = \frac{(t^{-\theta} - 1)^{1/\theta}}{\theta}, \qquad \theta > 0.$$

Hence, the copula model is given by (Cherubini *et al.* 2004)

$$C(u_1, u_2, u_3, \ldots, u_n) = \left(u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} + \cdots + u_n^{-\theta} - n + 1\right)^{-1/\theta}. \tag{2.6}$$

This representation is very straight forward using the above generator of Clayton copula and the general copula model for Archimedean copulas. In particular, the bivariate Clayton copula is given by

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

One important point to note about this copula model is that when the parameter $\theta = 0$, we obtain the independence copula (Ruppert, 2011). While on the other hand, as $\theta \to \infty$, we obtain dependency between the variables. Hence, θ is a parameter that influences the dependency. This copula model is very often used in statistical analysis in order to asses variables with lower tail correlation, as they are well known for estimating such correlation. It should also be noted that this copula model (Clayton) is defined only when the generator is strict.

Gumbel copula

This is another asymmetric copula under the category of Archimedean copulas, with generator function given by;

$$\phi(t) = \left(-\log(t)\right)^{\theta}, \qquad \theta \geq 1.$$

In this case, the copula model is given by (Aas, 2004)

$$C(u_1, u_2, \ldots, u_n) = \exp\left(-\left((-\log u_1)^{\theta} + (-\log u_2)^{\theta} + \cdots + (-\log u_n)^{\theta}\right)^{1/\theta}\right). \tag{2.7}$$

In particular, the bivariate Gumbel copula is defined by

$$C(u, v) = \exp\left(-\left((-\log u)^{\theta} + (-\log v)^{\theta}\right)^{1/\theta}\right).$$

Just like the Clayton copula, the parameter θ in the Gumbel copula influences the dependence. Indeed, as $\theta = 0$, we obtain the independence copula, but when $\theta \to \infty$, we obtain the comonotonicity copula. Unlike the Clayton copula which are used to estimate lower tail dependency, the Gumbel copula on the other hand are used to estimate upper tail dependencies.

Frank copula

This is another member of the Archimedean family of copulas with generator function given by (Ruppert, 2011)

$$\phi(t) = -\log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right), \qquad -\infty < \theta < \infty.$$

Using this generator and the genera representation for Archimedean copulas, we obtain the following Frank copula model

$$C(u_1, u_2, \ldots, u_n) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)\ldots e^{-\theta u_n} - 1}{(e^{-\theta} - 1)^{n-1}}\right). \tag{2.8}$$

In particular, for the bivariate case ($n = 2$), we obtain

$$C(u, v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right).$$

An important point to note is that, amongst the Archimedean copulas listed above, this is the only one whose generator parameter can take any real number without any restrictions. On the other hand, this Frank copula has no lower or upper tail dependence like the other members. Hence, it cannot be used to model data with strong tail dependence. There are many more copulas model that could be discussed in the literature, however our analysis will only focus on these few copulas.

These implicit and explicit copulas are all referenced from Aas (2004). Details about copulas and tail dependence may be found on the next chapter.

Chapter 3

Methodology

3.1 Test For Stationarity

As mentioned earlier, financial returns like exchange rate are characterized by intermittency (or variability), meaning that at any given point in time, financial returns present a very high degree of variability. A series $\{r_1, r_2, \ldots, r_t\}$ is said to be strictly stationary if the joint density of $\{r_1, r_2, \ldots, r_t\}$ is invariant under time shift. However, this condition is often difficult to be proven and thus a simpler version is often used; namely weakly stationary. A series is *weakly stationary* if both its mean and covariance functions are time invariant. This is a very important feature of a time series as most analysis are carried out on the basis of stationarity. However, most financial time series literature assume that financial returns are weakly stationary and this assumption is checked empirically in cases where we have sufficient historical data (Tsay, 2005). This means that we need to transform our data to make it stationary (such as log, square root transformations, etc). Some techniques used to check whether a given series is stationary includes what follow in the sequel.

3.1.1 Autocorrelation Function (ACF)

This is a rather graphical method of analyzing the nature of stationarity in our data, as it gives an indication of stationarity beforehand. Firstly, we recall that if $r_1, r_2, \ldots, r_n$ is a series, then the autocorrelation function at lag k is defined by

$$\rho_t = \frac{\gamma_k}{\gamma_0}, \tag{3.1}$$

where γ_k is the covariance at lag k and γ_0 is the variance, and $-1 \leq \rho_k \leq 1$. However in general, we only deal with a sample of the observation, we talk about sample autocorrelation function and in this case, we use

$$\hat{\rho}_t = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \tag{3.2}$$

where (Tsay, 2005)

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{n}(r_t - \bar{r})(r_{t+k} - \bar{r})}{n},$$
$$\hat{\gamma}_0 = \frac{\sum_{t=1}^{t=n}(r_t - \bar{r})^2}{n},$$

with n being the sample size. A graph of $\hat{\gamma}_k$ against k is called the *sample correlogram.* Thus, analysis of the likelihood of stationarity by this approach rely on the nature of the correlogram;

- If the autocorrelations starts at a very high value and decay slowly but not to zero, the series is likely non-stationary.
- if the autocorrelations hover around zero, then the series is likely to be stationary.

Once we are aware of the likelihood of the nature of stationarity, we may then proceed to verify this. Two formal tests are often used in financial time series are the Augmented Dickey Fuller (ADF) test and the Phillips-Perron (PP) test.

3.1.2 Augmented Dickey Fuller (ADF) Test

Consider the simple model

$$r_t = \rho r_{t-1} + \mu_t, \tag{3.3}$$

μ_t is a white noise process. Suppose we have data which follows this model. Then we know that if $\rho = 1$, then the model is a pure random walk and thus is non-stationary. However, if $\rho < 1$, then the model is stationary. The Dickey-Fuller test assumes that the error terms, (u_t), in the regression model above are uncorrelated and thus runs the regression (Maddala & Kim, 1998)

$$\Delta r_t = \delta \Delta r_{t-1} + \epsilon_t, \tag{3.4}$$

where $\delta = \rho - 1$, $\Delta r_t = r_t - r_{t-1}$ and $\Delta r_{t-1} = r_{t-1} - r_{t-2}$. Under the null hypothesis that $\delta = 0$ (series is non-stationary) against $\delta < 0$ (series is stationary), the estimated $\hat{\delta}$ follows the τ statistic (which is in fact the Dickey-Fuller (DF) distribution). It should be noted that the values of this distribution are calculated using Monte Carlo Simulations.
However, the Augmented Dickey Fuller (ADF) test is more general as it does not

assume that the error terms are uncorrelated. It has a similar implementation as the DF test but rather estimates the model (Maddala & Kim, 1998)

$$\Delta r_t = \alpha_1 + \alpha_2 t + \delta r_{t-1} + \sum_{i=1}^{m} \beta_i \Delta r_{t-i} + \epsilon_t, \tag{3.5}$$

where $\Delta r_{t-i} = r_{t-i} - r_{t-i-1}$ (the addition of α_1 or $\alpha_1 + \alpha_2 t$ is just in case r_t is a process with drift or with drift around stochastic trends respectively). The value of m is obtained by minimizing the AIC or BIC or using the sequential procedure suggested by Campbell & Perron (1991). With this procedure, we start with a model with a large p_{max} value and then estimate the model, until the last included value of p_{max} (at this stage our $max = m$) is statistically significant at the 10% level of significance. In this case, $p = p_{max}$. Hence, the number of lagged difference values of r_t is determined empirically but basically, we add as many as possible in order to make the error terms serially uncorrelated. It should be noted that in our statistical package selects/calculates this value automatically. Under the hypotheses

$$\begin{aligned} H_0 &: \delta = 0, \quad \text{the data is non stationary,} \\ H_1 &: \delta < 0, \quad \text{the data is stationary,} \end{aligned}$$

the test statistic is given by

$$t_\delta = \frac{\hat{\delta}}{se(\hat{\delta})}. \tag{3.6}$$

This statistic follows the same DF distribution as the Dickey Fuller test above, and not the usual t distribution. The values of this distribution are obtained by Monte Carlo Simulations; however, the statistical package *EViews* provides the correct critical values for this test (Mohadeva & Robinson, 2004). Hence, we will use it for this test. We reject the null hypothesis of non stationarity if t_δ is bigger than the critical value at the chosen level of significance.

3.1.3 Phillips-Perron (PP) Test

This is the most widely used test for stationary as an alternative to the ADF test. It does not specify any assumption of serial correlation or about the error terms. It is a non parametric method of analyzing whether a series is stationary or not. It also does not include lagged values of the series, hence, is applicable to a wide variety of problems. Suppose we have the model

$$\Delta r_t = \delta \Delta r_{t-1} + u_t. \tag{3.7}$$

If there are any serial correlations or heteroscedasticity in the error terms u_t, then this test will correct it. Hence, both unit root tests are very similar, but differ in the way they both deal with serial correlation and heteroscedasticity. In general, the pp test is viewed as the DF test that has been made robust to serial correlation by using Newey-West (1987) heteroscedasticity and autocorrelation consistent covariance matrix estimator. Just as in the ADF, the PP test runs the auxiliary regression (Maddala & Kim, 1998)

$$\Delta r_t = \alpha_1 + \alpha_2 t + \delta r_{t-1} + u_t. \tag{3.8}$$

Similar to the ADF test, the hypotheses are

$$\begin{aligned} &H_0 : \delta = 0, \quad \text{the data is non stationary,} \\ &H_1 : \delta < 0, \quad \text{the data is stationary,} \end{aligned}$$

and the test statistics are given by (Maddala & Kim, 1998)

$$\begin{aligned} Z_t &= \left(\frac{\hat{\sigma}^2}{\hat{S}^2}\right)^{1/2} \times t_{\delta=0} - \frac{1}{2}\left(\frac{\hat{S}^2 - \hat{\sigma}^2}{\hat{S}^2}\right)\left(\frac{T.SE(\hat{\delta})}{\hat{\sigma}^2}\right) \\ Z_\delta &= T\hat{\delta} - \frac{T^2.SE(\hat{\delta})}{2\hat{\sigma}^2}(\hat{S}^2 - \hat{\sigma}^2), \end{aligned}$$

where $\hat{\sigma}^2$ and $\hat{S}^2$ are respectively consistent estimators of σ^2 and S^2 given by (Maddala & Kim, 1998)

$$\begin{aligned} \sigma^2 &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2). \\ S^2 &= \lim_{T\to\infty} \sum_{t=1}^{T} E(T^{-1} S_T^2), \end{aligned}$$

and $S_T = \sum_{t=1}^{T} u_t$. It is also worth noting that this test statistic has the same asymptotic distribution as the ADF test statistic. This test directly incorporates any serial correlation or heteroscedasticity that may be found in the error terms by calculating modified statistics above; for $t_{\delta=0}$ and $T\hat{\delta}$, given above.

3.2 Measure of dependence

Analysis of multivariate data usually involves not only fitting such data to distributions. It also involves analyzing the dependency that exist between these variables; explaining if and how two or more variables are related. This is very important as this might enable us to control a variable in case we know how it is related to other variables. Scatterplot matrices and correlation matrices are often used in this context to explain the relationship between variables.

3.2.1 Scatterplot matrices

This matrix is used to determine if there exists a pairwise relationship between a group of variables. In case there exists a relationship, it tells us the nature of the relationship and the presence of outliers.

Thus, given n variables, $X_1, X_2, \ldots, X_n$, a *scatterplot matrix* is a matrix in which each row and column represents a plot characterizing the relationship between two variables; row i contains the variable X_i and column j, the X_j variable. Usually, the main diagonal contains the name of the variable in that row/column or is left blank. This matrix is very important especially in case where the relation between two variables is not exactly linear.

3.2.2 Covariance/Correlation matrices

This is a way of analyzing in cases where the variables are pairwise linearly related. This technique is very accurate especially in cases where the underlying distribution of our variables is elliptical or multivariate normally distributed. Thus, given n variables, $X = (X_1, X_2, \ldots, X_n)'$, we have $E(X) = (E(X_1), E(X_2), \ldots, E(X_n))'$. Hence,

$$Cov(X) = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \ldots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \ldots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \ldots & Var(X_n), \end{pmatrix} \tag{3.9}$$

where $Cov(X_i, X_j) = E\left((X_i - E(X_i))(X_j - E(X_j))\right), \forall i = 1, 2, 3, \ldots, n$ and $Var(X_i) = Cov(X_i, X_i)$. In a similar manner, letting $\rho_{X_i,X_j} = \frac{Cov(X_i,X_j)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}}$; called the correlation coefficient and $Corr(X_i, X_i) = 1$, we obtain the correlation matrix as

$$Corr(X) = \begin{pmatrix} 1 & \rho_{X_1,X_2} & \cdots & \rho_{X_1,X_n} \\ \rho_{X_2,X_1} & 1 & \cdots & \rho_{X_2,X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{X_n,X_1} & \rho_{X_n,X_2} & \cdots & 1, \end{pmatrix} \tag{3.10}$$

Thus, dependency between a pair of variables (X, Y) can be explained by the correlation between X and Y. The closer the value of this correlation is to 1, the higher the variables are assumed to be related either positively or negatively depending on the sign of the correlation coefficient. The closer this value is to 0, the lower these variables are linearly related (also positively or negatively, depending on the sign of the coefficient). Thus, we have that $|\rho| \leq 1$.

Also, if we let $\psi = diag(\sigma_{X_1}, \sigma_{X_2}, \ldots, \sigma_{X_n})$, where $\sigma_{X_i} = \sqrt{Var(X_i)}$, that is, the standard deviation of X_i, then we have the following relationship between the correlation and covariance matrices;

$$Corr(X) = \psi^{-1} COV(X) \psi^{-1},$$

or equivalently

$$Cov(X) = \psi Corr(X) \psi.$$

One important use of correlation within the context of multivariate analysis is that lack of correlation means independence of the variables especially in the case where the variables have joint multivariate normal or elliptical distributions. Thus, one should be careful when using correlation as a measure of dependency between variables.

3.2.3 Kendall's tau

A significant aspect of this project is to assess the dependency that exist between variables using copulas. The Kendall's correlation is a very important way to assess the dependency that exist between two variables, especially within the field of copula. This is a non parameter measure of dependency between two variables. Unlike linear correlation described above, which is not constant or preserved within the copula, the Kendall's Tau (or correlation) is preserved under copula. This means that any pair of correlated variables under the same copula will have the same value of tau of that copula. Thus, the Kendall's tau simple measures the different between the probability of concordance and the probability of discordance. According to Nelsen (1992), Kendall's tau measures the average likelihood ratio dependence and is given by

$$\tau_k = \frac{2(a-b)}{n(n-1)}, \tag{3.11}$$

where a is the of concordant pairs, b is the number of discordant pairs and n is the sample size. Also, this Kendall's tau is equivalent to the expression below which in this case incorporates a copula model. Thus, for two random vectors, $X = (X_1, X_2)'$ and $\bar{X} = (\bar{X}_1, \bar{X}_2)'$, where $\bar{X}$ is the independent copy of X, with both vectors having a common continuous distribution, F, and copula, C, the Kendall's Tau is defined by (Aas, 2004)

$$\begin{aligned}\tau &= \Pr((X_1 - \bar{X_1})(X_2 - \bar{X_2}) > 0) - \Pr((X_1 - \bar{X_1})(X_2 - \bar{X_2}) < 0) \\ &= 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1 \\ &= 4E(C(u,v)) - 1.\end{aligned}$$

In particular, for the independent copulas, we know that $C(u,v) = uv$. Hence,

$$\begin{aligned}E(C(u,v)) &= \int_{[0,1]^2} C(u,v)dC(u,v) \\ &= \int_{[0,1]^2} uvdC(u,v) \\ &= \frac{1}{2}.\end{aligned}$$

Thus, $\tau = 1$. The following theorem holds for the MAGHD and MGHD (Schmidt *et al.* 2006).

Theorem 3.1. *(Schmidt* et al. 2006*) Let $\rho \in (-1,1)$, be the correlation coefficient between X_1 and X_2.*

1. *If $X \in MGH_2(\mu, \Sigma, \omega)$, with $\beta = 0$, then*

$$\tau = \frac{2}{\pi}\arcsin(\rho).$$

2. *If $X \in MAGH_2(\mu, \Sigma, \omega)$, where $X = A'Y + \mu$, then if $\rho \neq 0$, then*

$$\tau = \frac{4}{|c|}\int_{\mathbb{R}^2} f_{Y_1}(x_1)\left(\frac{x_2 - x_1}{c}\right) \times \int_{-\infty}^{x_1} F_{Y_2}\left(\frac{x_2 - z}{c}\right) f_{Y_1}(z)dzd(x_1, x_2) - 1$$

where $c = sgn(\rho)\sqrt{1/\rho^2 - 1}$ and in particular $\tau = 0$ when $\rho = 0$.

3.2.4 Tail dependency of Archimedean copulas

In financial risk analysis, we often are not just concerned about correlation of variables but more specifically, we are concerned with extreme tail behavior of the variables. That is, we try to explain the risk that exist between variables so as to correct for any loss that occur as a result of the loss in a particular variable. As we have noted earlier, we estimate the tail dependency between extreme events of variables through the use of Archimedean copulas. However, we restrict our analysis to only the three Archimedean copulas discussed in section 2.1.2.

Copulas are very often used in practice in order to model tail dependency between variables, as these copulas are related to some measures of dependency like the Kendall tau. Indeed, the copula generator function, ϕ, is related to the Kendall tau (in the bivariate case) by (Genest & Mackay, 1986),

$$\tau = 1 + 4\int_0^1 \frac{\phi(v)}{\phi'(v)}\,dv. \tag{3.12}$$

There are other measures of dependencies that one might look at like the Spearman correlation. However, this is similar to the Kendall's tau. The relationship between this Spearman's rho, rho_τ, and Kendall's tau as well as their respective copula models can be written as (Cherubini *et al.* 2004)

$$\rho_k = 12\int_0^1\int_0^1 C(u,v)dudv - 3,$$
$$\rho_\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1.$$

Furthermore, Joe (1997) showed that the relationship between Archimedean copulas and tail dependency could be established in a theorem which can be summarized as; If ϕ is a strict generator defined as in 3.12 and if $0 \neq \phi'(0) < \infty$, then the copula given by

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v))$$

has no tail dependency. However, if it does have an upper tail dependency, then $\frac{1}{\phi'(0)} = +\infty$. In this case, the values of the lower and upper tail dependencies are given by (Cherubini *et al.* 2004)

$$\lambda_L = 2\lim_{t\to+\infty}\frac{\phi'(t)}{\phi'(2t)} \qquad \text{and} \qquad \lambda_U = 2 - 2\lim_{t\to 0^+}\frac{\phi'(t)}{\phi'(2t)} \tag{3.13}$$

Thus, the table below summarizes the coefficient of tail dependency for the three members of the Archimedean family discussed above.

Table 3.1 *Tail dependency coefficients of Archimedean copula.*

Copula	Generator	Kendall Tau	Tail dependency
Clayton	$\frac{t^{-\theta}-1}{\theta}, \quad \theta > 0$	$\tau_k = \frac{\theta}{\theta+2}$	$\lambda_L = 2^{-1/\theta}$ $\lambda_U = 0$
Gumbel	$(-\log t)^{-\theta}, \quad \theta > 1$	$\tau_k = 1 - \theta^{-1}$	$\lambda_L = 0$ $\lambda_U = 2 - 2^{1/\theta}$
Frank	$-\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right), \quad -\infty < \theta < \infty$	$\tau_k = 1 - \frac{4(D_1(\theta)-1)}{\theta}$	$\lambda_L = 0$ $\lambda_U = 0$

where $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}\,dt$.

It is clear from the table that the Clayton copula has no estimate for the upper tail dependency; it cannot account for positive dependency, while the Gumbel copula has no lower tail dependency estimate and hence cannot account for negatively dependency. The Frank copulas neither accounts for upper nor lower tail dependencies.

3.3 Goodness of fit

We now look at some statistical tests that are often used to assess the fit of a given postulated distribution. They are used to determine how good a model or distribution fits a given set of data. This is very important since the implementation of statistical analysis depend on this test. Over the years, many goodness of fit tests involving univariate distributions have been proposed (Kolmogorov-Smirnov, Anderson and Darling, Chi square, just to name a few) and have proven adequate. However, not many tests have been proposed for their multivariate counterparts. Nevertheless, multivariate tests such as the Chi-square test and the multivariate Kolmogorov-Smirnov tests have been proposed too within the contest of multivariate distributions. In this project, we will describe the Anderson and Darling as well as the one dimensional and two dimensional Kolmogorov-Smirnov goodness of fit tests. But later on, we will also consider the multivariate Kolmogorov-Smirnov goodness of fit test for copula as well as the kernel smoothing test.

3.3.1 The Kolmogorov-Smirnov (KS) goodness of fit test

Given a random sample $X_1, X_2, X_3, \ldots, X_n$ from a population with unknown distribution, this sample is compared with a distribution function $H(x)$ to find out if it is reasonable to say that $H(x)$ is the distribution function of our sample. The way in which this is done is by use of a function, $S(t)$, called the *empirical distribution function.* The empirical distribution function, $S(x)$, is a function of a random sample $X_1, X_2, X_3, \ldots, X_n$ which gives the fraction of the X_i values that are less than or equal to x for all x. That is

$$S(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x},$$

where $I_{X_i \le x}$ is the *indicator function* defined by;

$$I_{X_i \le x} = \begin{cases} 1, & \text{if } X_i \le x \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

It should be noted that this function is very useful in estimating $S(x)$.

The K-S test simply compares the empirical distribution function, $S(x)$, with a hypothesized distribution function, $H(x)$, and see if these distributions agree. This is done by calculating the largest vertical distance between the two distribution functions. The steps involved in the test are summarized below;

$$\begin{aligned} H_0 &: S(x) = H(x), \quad \text{for all x;} \\ H_1 &: S(x) \neq H(x), \quad \text{for at least one x,} \end{aligned}$$

and our test statistic is given by;

$$T = \sup_x |S(x) - H(x)|. \tag{3.15}$$

The critical region, $C_r = [1-\alpha, \infty[$, where α is the level of significance.

However, we will also consider the fit of two dimensional KS test. This is considered in the sequel.

3.3.2 Two dimensional KS goodness of fit test

In this subsection, we discuss the two dimensional KS test following Peacock's test (Fasano & Franceschini, 1987). From the one dimensional test above, the test statistic T simply calculates the maximum distance between the cumulative distribution and the sample distribution. However, T turns out to be proportional to $1/\sqrt{n}$ (Fasano & Franceschini, 1987), and hence, one often is interested in the distribution of $K_n = T/\sqrt{n}$, where n is the sample size. The multivariate (two dimensional) approach proposed by Peacock (1983) stipulates that one uses the maximum absolute distance between the observed and predicted normalized cumulative distribution in all the four possible ways to cumulate our data in the directions of the coordinate axes; namely $(x < X_i, y < Y_j)$, $(x < X_i, y > Y_j)$, $(x > X_i, y < Y_j)$, $(x > X_i, y > Y_j)$, $i, j = 1, 2, 3, \ldots, n$. Thus in this case, we consider all the $4n^2$ quadrants of the plane. However, it is clear that if the sample size is large, the implementation of this procedure is quite expensive (both in computer time and memory). For this reason, instead of considering all the $4n^2$ point (x_i, y_j) in the plane to cumulate data points, we instead use n points where detection is found. It should be noted that this procedure does not compromise the efficiency/accuracy of the test. The maximum absolute distance between the quantiles which have been normalized to 1 is calculated along all the four quadrants. To the best of my knowledge, no research has been done on multivariate distributions with the implementation of the K-S test with dimension bigger than two. This is practically infeasible as will require a lot of resources. In addition, even the two dimensional case has not been implemented in R so far. Thus, the need to look for alternative goodness of fit tests for multi dimensional data.

3.3.3 Anderson and Darling (A-D) Test

This test is a modified version of the K-S test as it overcomes some of the shortcomings of the K-S test. It is a more sensitive test and puts more emphasis on the tails. Just like the K-S test, the A-D test compares a random sample to see if this sample comes from a population with known distribution.

Unlike the K-S test which only makes use of the level of significance for calculating the critical region, the A-D test makes us of the specific distribution.

The steps involved in the A-D test are as follows;

$$H_0 : F(x) = H(x),$$
$$H_1 : F(x) \neq H(x),$$

where $F(x)$ is the distribution function of the random sample $X_1 < X_2 < X_3 < \ldots < X_n$ (which is not known) and $H(x)$ is our hypothesized distribution function. It should be noted that the random sample is first arranged in ascending order. The test statistic is given by

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(X_i) + \ln\left(1 - F(X_{n-i+1})\right)\right]. \qquad (3.16)$$

Hence, since this test is distribution specific, the critical region depends on the hypothesized distribution function.

3.3.4 Kernel Smoothing goodness of fit test

Due to the inability or practical infeasibility of the K-S test to be implemented in higher dimension, an alternative test that can be used is based on the kernel density of the returns. Unlike the K-S test that deals with a particular distribution, this test is carried out with the assumption that no parametric distribution that describes the multi-dimensional data is known. This test is an extension of the histogram used in order to determine the shape of a particular multivariate data. Unlike the histogram that largely depends on the equal sub-intervals in which the whole data interval is divided as well as the end points of these intervals, the kernel density depends on the bandwidth.

For a given series $(X_1, \ldots, X_n)$, a multivariate fixed width kernel density estimate with kernel function, K, and fixed kernel width (bandwidth), h, gives the estimated density $\hat{f}_h(X)$ for a multivariate data $X \in \mathbb{R}^d$ based on (Hwang et al., 1994)

$$\hat{f}_h(X) = \frac{1}{nh^d} \sum_{r=1}^{n} K(\frac{X - X_i}{h}), \tag{3.17}$$

where K satisfies the conditions

$$K(X) \geq 0 \qquad \text{and} \int_{\mathbb{R}^d} K(X)dY = 1. \tag{3.18}$$

In most instance and also in *rStudio*, K is chosen as the Gaussian kernel defined by

$$K(X) = (2\pi)^{-p/2} \exp(-\frac{1}{2} X^T X). \tag{3.19}$$

According to Hwang *et al.* 1994, the kernel is symmetrical, with its value smoothly decaying away from the kernel center. However, this kernel density estimator is largely dependent on what bandwidth is used in the estimation. Incorrect values used may result in either undersmoothed or oversmoothed curves in the estimation. Hence, the value of the parameter that produces the most optimal density curves is used. The value of this parameter that produces the most optimal curve is called the *mean integrated squared error (MISE)* defined by

$$MISE(h) = E\left(\int \hat{f}_h(X) - f(X)\right)^2 dX. \tag{3.20}$$

However, this estimator has no closed form and as a result, the asymptotic version is used, which will not go into the details here (see Chacón & Duong (2010)).

This test is implemented up to dimension six, unlike the K-S test. It simply tests two data sets and evaluates if both data come from the same distribution. The R implementation of this test, called the *kernel density based global two sample comparison test* is straight forward. Suppose X and Y are two random vectors with density functions $f_1(x)$ and $f_2(x)$ respectively. Our aim is to find out if $f_1(x) = f_2(x)$. Under the hypotheses

$$H_0 : f_1(x) = f_2(x), \quad \text{for all } x;$$
$$H_1 : f_1(x) \neq f_2(x), \quad \text{for at least one } x;$$

the test statistic is given by (Duong *et al.* 2012);

$$T = \psi_{0,1} - \psi_{0,12} - \psi_{0,21} + \psi_{0,2}, \tag{3.21}$$

where $\psi_{0,uv} = \int f_u(x) f_v(x)\,dx$. This test statistic has a distribution which is asymptotically normal distributed. In this case, the level of discrepancy is the integral squared error given by

$$\int (f_1(x) - f_2(x))dx$$

The steps used in this project will consist of the following;

- The data is fitted with the multivariate generalized hyperbolic distribution and the parameters estimated.
- A random sample is then generated from the multivariate generalized hyperbolic distribution with the parameters estimated above.
- Then, using the kernel smoothing, this randomly generated sample is compared with our original data in order to assess if both data sets come from the same distribution

3.4 Copula parameter estimation

The analysis of financial returns using copulas basically starts with fitting the returns to a copula model and estimating the parameters of the copula model. In this section, we discuss two techniques that are employed for parameter estimation of parametric copulas such as Archimedean copulas. Various estimation techniques so far have been discussed by various research on copulas (Bouyé *et al.* 2000). In this current work, we will only discuss two techniques; full maximum likelihood and inference from the margins.

3.4.1 Full Maximum Likelihood (FML)

This method is a direct method of estimating parameters of a model. It simply involves calculating the likelihood function of the copula model and then using mathematical differentiation, it maximizes this likelihood function. The advantage of this method is that it fits data to a parametric copula model irrespective of whether the marginal distributions are parametric or not. Suppose we have a random vector $X_1, \ldots, X_n$, such that each $X_i = X_{i,1}, \ldots, X_{i,d}$, and marginal distribution functions $F_1, \ldots, F_n$ and copula C. We further assume that our parameter vector $\theta = (\theta_1, \ldots, \theta_N)$. Then the likelihood function is given by

$$l(\theta) = \sum_{i=1}^{n} l_i(\theta).$$

However, since we know by Sklar theorem that

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)),$$

for a series of T observations, we obtain the likelihood function of the copula model as

$$l(\theta) = \sum_{i=1}^{T} \ln c(F_1(x_{1,i}), \ldots, F_k(x_{k,i}), \ldots, F_n(x_{n,i})) + \sum_{i=1}^{T} \sum_{k=1}^{n} \ln f_k(x_{k,i}),$$

where c is the copula density function. Our aim is to maximize this function, $l(\theta)$. Using differentiation, we obtain the estimated vector

$$\hat{\theta}_t = \arg\max_{\theta} \sum_{i=1}^{n} \ln f_t(x_{t,i}; \theta_t).$$

However, the higher the dimension of the parameter vector to be estimated, the more difficult it is to estimate these parameters. Thus, this method can prove computationally difficult. We can reduce this likelihood function to the bivariate case as follows; let (X_1, X_2) be a random vector with marginal distributions F_1 and F_2 respectively and copula model C. Then by Sklar theorem, we know that $F(x_1, x_2) = C\left(F_1(x_1), F_2(x_2)\right)$, and

$$f(x_1, x_2) = c\left(F_1(x_1), F_2(x_2)\right) f_1(x_1) f_2(x_2),$$

where

$$c\left(F_1(x_1), F_2(x_2)\right) = \frac{\partial^2 C\left(F_1(x_1), F_2(x_2)\right)}{\partial F_1(x_1) \partial F_2(x_2)}.$$

Hence, the bivariate likelihood equations will be given by

$$l(\theta) = \sum_{i=1}^{T} \ln c\left(F_1(x_{1,i}), F_2(x_{2,i})\right) + \sum_{i=1}^{T} \sum_{k=1}^{2} \ln f_k(x_{k,i}).$$

Similarly as above, the parameters are estimated. It is also showed that these estimates are asymptotically normally distributed and consistent estimators (Bouyé *et al.* 2000).

In particular,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to N(\mathbf{0}, J^{-1}(\theta_0)),$$

where θ_0 is the actual parameter vector and $J(\theta_0)$ is the Fisher Information Matrix. Even though this approach is computationally intense, it however estimates the marginal and copula parameters jointly, unlike the other techniques like inference of the margins.

3.4.2 Inference from the margins (IFM)

This approach is very similar to the full maximum technique discussed above. They only differ in that parameters belonging to the marginal distribution and the copula model are estimated separately in a two step procedure. Thus, the parameters to be estimated for the whole model are split into marginal distributions parameters, which are estimated by maximum likelihood. These estimated marginal parameters are used to estimate the common copula parameters. Hence, if $\theta = (\theta_1, \ldots, \theta_N, \alpha)$, where $\theta_1, \ldots, \theta_N$ are the marginal parameters and α is the common copula parameter (which can be the copula generator parameter vector), then, in order to estimate this model, we begin by estimating the parameters originating from the marginal distribution by maximizing the likelihood function. In this case, we obtain

$$\begin{aligned} l(\theta) = & \sum_{i=1}^{T} \ln c\left(F_1(x_{1,i};\theta_1), \ldots, F_k(x_{k,i};\theta_k), \ldots, F_n(x_{n,i};\theta_n);\alpha\right), \\ & + \sum_{i=1}^{T}\sum_{k=1}^{N} \ln f_k(x_{k,i};\theta), \end{aligned}$$

and $\hat{\theta}_n = \arg\max \sum_{i=1}^{T} \ln f_n(x_{n,i};\theta_n)$. In the second step, we then use the estimated marginal parameter to estimate the common copula parameter by

$$\hat{\alpha} = \arg\max \sum_{i=1}^{T} \ln c\left(F_1(x_{1,i};\hat{\theta}_1), \ldots, F_k(x_{k,i};\hat{\theta}_k), \ldots, F_n(x_{n,i};\hat{\theta}_n);\alpha\right).$$

As this point, one may ask which technique then provides the best estimates for the copula model. Well depending on some regularity conditions (Joe & James, 1996), the estimators obtained by IFM are more efficient and also asymptotically and normally distributed, compared to those obtained by FML.

3.5 Model selection criteria

So far, we have discussed the concept of copula and how we fit data into a copula model. However, selecting the best possible model (the model that accurately fits the data, with the smallest number of parameters as possible) out of many models as we know is a very important aspect in statistical analyzis. Thus, selecting the best copula model describing the dependence between variables suggests that many copula models have to be fitted. Then, we use well known established selection criteria to select the best model. In this section, we present two of these information criteria; the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). It should be noted that there are many more criteria, and modifications of these. But in our context, we will only focus on the AIC and BIC.

3.5.1 Akaike's Information Criterion (AIC)

This is a criterion that is widely used in statistical analysis (as it measures the quality of a model for a given data set) to select the best possible model from a set of models. This criterion suggests that the best possible model is the one with the smallest AIC value, where AIC is given by

$$AIC = -2\ln(L) + 2k, \tag{3.22}$$

where k is the number of parameters in the model and L is the likelihood of the particular model. The value of AIC is often printed out by statistical packages.

3.5.2 Bayesian Information Criterion (BIC)

This is another approach to select the best model from a set of models. Its implementation is similar to that of the AIC, where the best model is that with the the smallest BIC value, with BIC given by

$$BIC = -2\ln(L) + k\ln(n), \tag{3.23}$$

where L is the likelihood of the fitted model, k is the number of parameters in the model and n is the sample size. The values of BIC are also provided by statistical packages. It should also be noted that some key differences exist between both criteria such the fact that the BIC tends to favor smaller models, as for $n \geq 8$, $k\ln(n) > 2k$. Hence, this means that BIC tends to choose models which are more parsimonious (fits the data more accurately with as few parameters as possible) than those selected by AIC (Cavanaugh, 2012).

3.5.3 Value-at-Risk (VaR) and Backtesting

Value-at-Risk (VaR) is the threshold value such that the probability of the market loss on a portfolio, over a given time horizon, exceeds this value is equal to the given probability level. It is widely used as a risk measure and utilized for assessments of extreme behavior in financial returns (Jorion, 2006). More importantly, it can be used to measure a distribution's level of adequacy for tail fits, i.e., VaR backtesting. It is relevant to note that that financial institutions are more prone to failure due to the shortage of capital resulting from underestimation of VaR. Furthermore, under the Basel framework, there is a negative profit impact due to the misestimation of VaR in either direction (Beling *et al.* 2010). Hence, an adequate model for assessing the risk of a return series, should neither underestimate VaR, nor should it overestimate VaR. In the analysis of maximum loss for a portfolio, Kupiec likelihood ratio test (Kupiec, 1995) is the most commonly used backtesting procedure. The test relies on unconditional coverage, meaning that it verifies if the reported VaR estimate is violated significantly more or less number of times compared to the level of significance, α. In this case, if the ratio of number of violations is not significantly different from the level of significance, then the overall adequacy of the model is verified. Thus, under the null hypothesis that the ratio of expected number of violations is α, the test statistic for the Kupiec test is given by

$$2\left[\ln\left(\left(\frac{r^{\alpha}}{N}\right)^{r^{\alpha}}\left(1-\frac{r^{\alpha}}{N}\right)^{N-r^{\alpha}}\right)-\ln(-\alpha^{r^{\alpha}}(1-\alpha)^{N-r^{\alpha}})\right], \tag{3.24}$$

where N is the sample size and r^{α} is the number of times the returns deflect below (for long position) or above (for short position) the estimated VaR value, at α level of significance. This test statistic asymptotically follows a chi-square distribution with one degree of freedom

Chapter 4

Application

Having discussed the statistical techniques and theory on multivariate generalized hyperbolic distributions and copulas, we may now consider the applications of these techniques to our data which consist of a series of three daily indices from the Johannesburg Stock Exchange (JSE) and the S&P 500 return. However before we get into the details of these techniques, a brief discussion of the data is done. The data consist of daily USD/ZAR exchange rate, South African All shares Index (ALSI), South African Gold Mining Index and finally the America All shares Index (S&P 500). All data considered here range from the period January 03, 2000 to August 30, 2014.

4.1 Data description

The daily USD/ZAR exchange rate represents the daily South African ZAR per US Dollar. The United States and South Africa are very much linked, especially when we deal with imports and exports; ranging from organic food materials such as fruits to inorganic compounds such as gold and diamonds. Thus, this index represents the rate of currency exchange between these two countries.

The South African gold mining has been the main driving force behind South African's economy in the past. However, according to Statistics South Africa, monthly gold production has decreased over the years due to temporal factors. The general trend has decreased so drastically from an index of 359.0 in 1980 to about 48.4 in 2015. Thus, producing about 87% less than in 1980, and contributing only 1.7% to the Gross Domestic Product in 2013. It is presently valued at about $4 billion USD in South Africa and accounting for about 50% of the world's gold reserves.

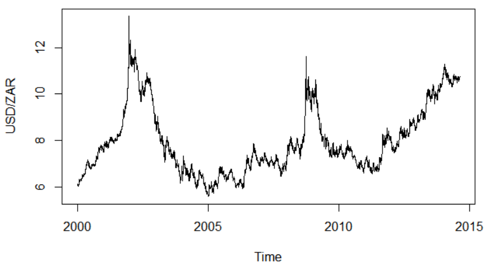
The South African All shares Index (ALSI) is a major index of the JSE measures/tracks the general performance of all the major companies within that establishment. These include domestic financial indices ranging from banks, to insurance companies, mining companies, etc. As a leading index in the South African economy, this index provides investors with the overall performance of the capital as well as market sectors of the South African economy.

The Standard and Poor's (S&P) 500 is an American stock market index consisting of 500 large companies (ranging from industrials, utilities, health care, information technology, just to name a few) having common stock listed in the New York Stock Exchange (NYSE). Having a market capital valued at $19.5 trillion USD, this stock is is highly regarded and followed as it is considered the best representation of the U.S. economy. Consisting of both growth and value stocks, is it regarded as the benchmark of stocks in the United States.

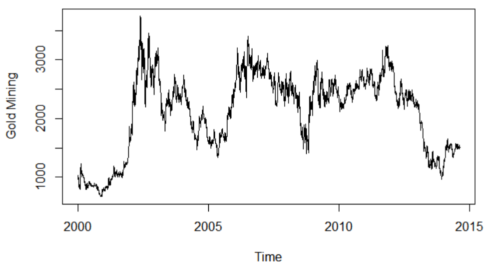
4.2 Empirical results

4.2.1 Time series plots

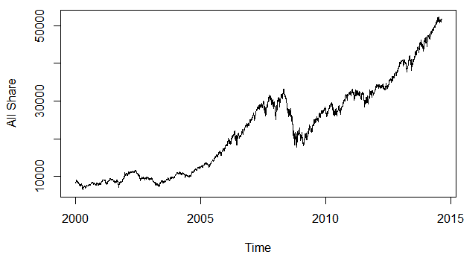
We begin our analyzis by considering the time series plots of the four indices mentioned above. It should be noted that in order to make our data smooth, a few data points were deleted from the Gold mining, All shares as well as the S&P 500 indices so that the data can tally with those of the daily USD/ZAR exchange rate. Deletion was opted as a statistical procedure to deal with missing values as these points were assumed to be missing completely at random and our sample was large (3803), hence resulting in unbiased data. Apart from the deletion, no other form of data manipulation or smoothing was carried out. Hence, the data presented here is exactly the same as that from the JSE. The time series plots of the four indices are given below.



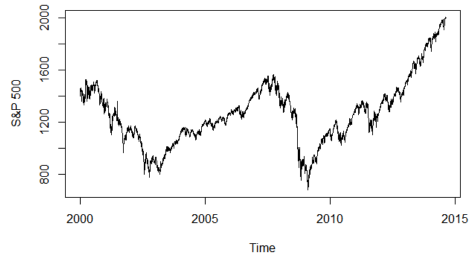
(a) USD/ZAR Exchange Rate.



(b) Gold Mining Index.



(c) All shares Index.



(d) S&P 500 Index.

Figure 4.1: *Time series plots of (a) Daily USD/ZAR Exchange Rate, (b) Daily Gold Mining Index, (c) All shares Index, and (d) S&P* 500 *Index for the Period* 03/01/2000 *to* 03/09/2014

Starting with the time series plot of the daily USD/ZAR exchange rate, we observe a rapid increase from about $R6$ in 2000 to about $R13$ in 2002. However from there onwards, we observe a rapid fall below $R6$ around 2005. Thus repeated increases are observed over time intervals. This highly suggests that the series is not stationary.

Similarly to the exchange rate, the Gold mining index also shows a systematic increase up to about the year 2002. However, the series starts decreasing thereafter. Contrary to the other plots however, we observe a periodic pattern which is repeated after about three and a half years. This also is an indication of non-stationarity.

The All shares index on the other hand shows an overall increase through out this period, characterized by the general upward trend as observed from the plot (c).

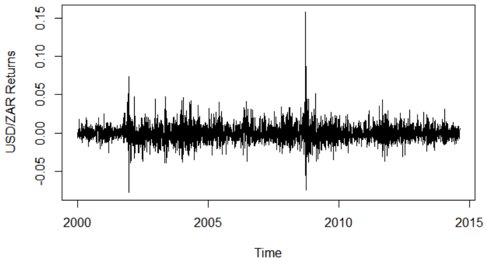
Finally, the S&P 500 plot contrary to the other plots shows an overall decrease from the year 2000 up to the year 2002. It increases for the next six years and then starts decreasing again. However, the series shows an overall increase for the past four years, reaching and overall high value of about 2000 in 2014. Thus, these observations culminates to the point that these plots are all non stationary. On the other hand though, we will be analyzing the log returns of these indices, which are stationary.

4.2.2 Descriptive statistics of log returns

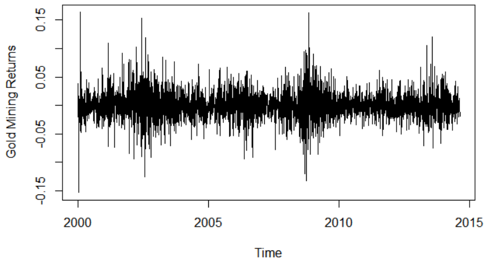
Suppose $x_{tt=1,\ldots,n}$ is a series which represents the price of an asset at time t over a period of time. Then the log returns or simple returns of this series is defined by

$$r_t = \ln(x_t) - \ln(x_{t-1}). \tag{4.1}$$

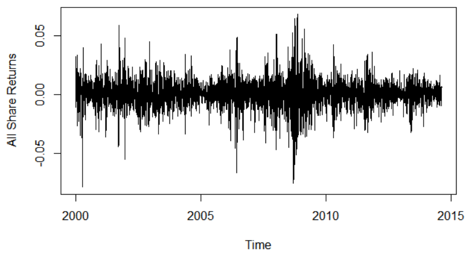
Thus, in our case, X_t will simply represent the daily closing values of the individual indices. Hence, the equation above will simply represent the one day returns. The returns plots are given below.



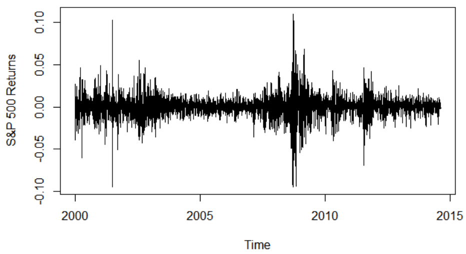
(a) USD/ZAR Exchange Returns.



(b) Gold Mining Returns.



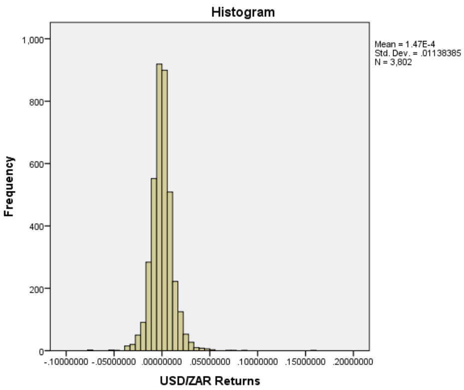
(c) All shares Returns.



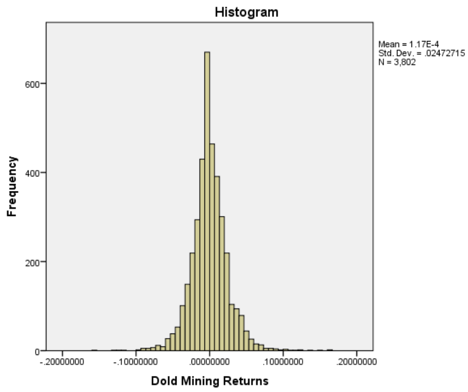
(d) S&P 500 Returns.

Figure 4.2: *Time series plots of (a) Daily USD/ZAR Exchange Returns, (b) Daily Gold Mining Returns, (c) All Share Returns, and (d) S&P* 500 *Returns for the Period* 03/01/2000 *to* 03/09/2014

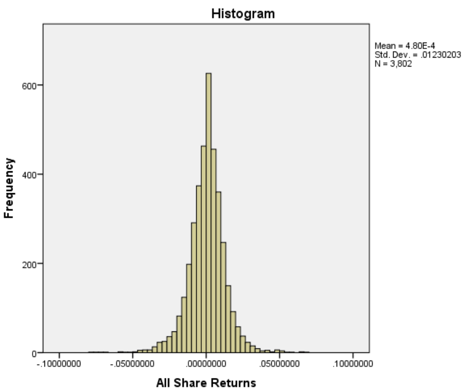
It is obvious from the plots that the returns are now stationary with a mean value which hovers around zero. However, the most important point to note about the plots is the pattern of volatility clustering as well as heteroscedasticity which is highly expected as we are dealing with financial returns. This pattern (of volatility clustering) is more predominant in the gold mining returns as well as All shares returns compared to the other two returns. The leptokurtic behavior of these returns is also evident from the histogram plots below; which shows a dense distribution of the returns along the upper distribution than along the tails.



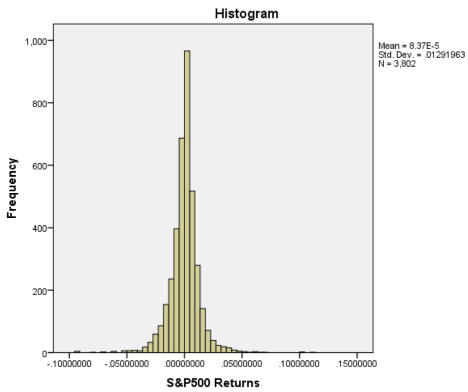
(a) USD/ZAR Exchange Returns Histogram.



(b) Gold Mining Returns Histogram.



(c) All shares Returns Histogram.



(d) S&P 500 Returns Histogram.

Figure 4.3: *Time series plots of (a) Daily USD/ZAR Exchange Returns Histogram, (b) Daily Gold Mining Returns Histogram, (c) All shares Returns Histogram, and (d) S&P* 500 *Returns Histogram for the Period* 03/01/2000 *to* 03/09/2014

A further confirmation of this leptokurtic behavior is shown in Table 4.1, which summarizes the descriptive statistics of our returns. It is evident from the table that all returns have an excess kurtosis from the normal distribution. Most importantly is the remarkable excess of about 13 for the USD/ZAR returns, followed by about 9.5 of the S&P 500 returns. Hence, even though our histogram may display an almost bell shaped distribution like that of the normal distribution, the high kurtosis suggest otherwise.

Table 4.1 Descriptive Statistics for daily returns

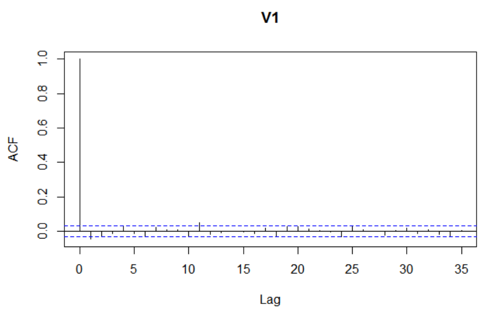
Descriptive Statistic	USD/ZAR	Gold Mining	All shares	S&P 500
Minimum Returns	-0.077362	-0.153505	-0.078968	-0.094828
Std. Dev.	0.011384	0.024727	0.012302	0.012920
Mean Returns	0.000147	0.000117	0.000480	0.000084
Skewness	0.967601	0.271464	-0.199373	-0.146497
Excess Kurtosis	13.378371	3.937869	3.853070	9.455758
Jarque-Bera Statistic (P-value)	28983.83 0.000000	2507.828 0.000000	2381.49 0.000000	14197.5 0.000000
Maximun Returns	0.157251	0.163562	0.068340	0.109572
Number of Observations	3802	3802	3802	3802

Moreover, a formal test for normality, the Jarque-Bera test, also rejects the null hypothesis of returns being normally distributed with all P-values below 0.05, at the 5% level of significance in this case. Another important observation that can be made from the table is the skewness of the returns, with the USD/ZAR returns being highly positively skewed. However, the All shares and S&P 500 returns are both negatively skewed. The positive mean of all returns indices also indicate an overall increase in our financial indices during our time period.

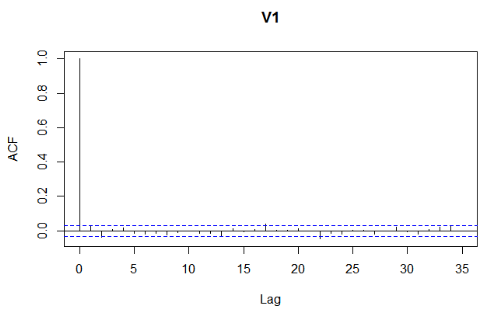
We may also proceed to assess the nature of stationarity of the returns using first the autocorrelation function (ACF) and later two formal tests.

4.2.3 Test for stationarity and autocorrelation

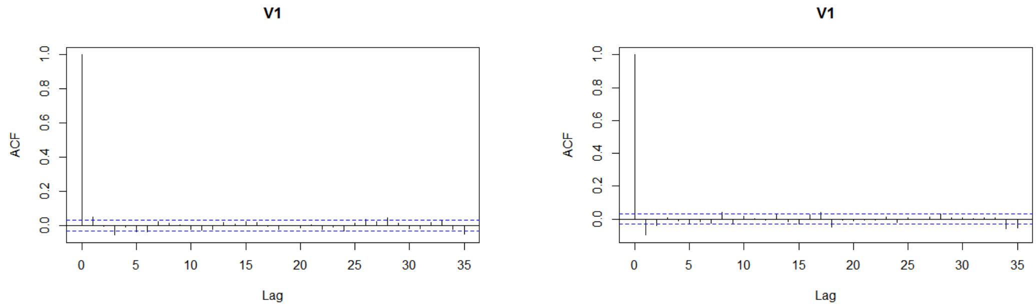
The analyzis of financial returns usually is done on the bases that the returns under consideration are stationary and are not autocorrelated. In this regards, we investigate the nature of stationarity of our series by first considering the autocorrelation function (ACF). Then, we verify our results with formal tests for stationarity; the Augmented Dickey Fuller (ADF) and Phillips-Perron (PP) tests. The ACF plots are summarized below.



(a) USD/ZAR Exchange Returns ACF.



(b) Gold Mining Returns ACF.



(c) All shares Returns ACF. (d) S&P 500 Returns ACF.

Figure 4.4: *Autocorrelation Function (ACF) of (a) Daily USD/ZAR Exchange Returns , (b) Daily Gold Mining Returns , (c) All shares Returns, and (d) S&P*500 *Returns for the Period* 03/01/2000 *to* 03/09/2014

With no significant spike in the plots above, the suggestion is that the returns plots are all stationary (non stationarity will imply some spikes extending from the lag axes up above the horizontal line in the plots). However, we may carry out formal tests to verify these results. The ADF and P-P tests in Table 4.2 suggest that the null hypothesis of returns having a unit root (non stationary) is rejected at the 5% level of significance; since all p-values are below 0.05. Thus, we conclude the returns are stationary.

Table 4.2 *Unit root tests for stationarity for daily Returns.*

Index	USD/ZAR	Gold Mining	All shares	S&P 500
ADF Statistic	-15.7288	-16.52	-15.5489	-15.6396
(p-value)	0.01	0.01	0.01	0.01
PP Statistic	-3854.966	-3545.054	-3358.184	-3899.515
(p-value)	0.01	0.01	0.01	0.01
PSR p-value	0.00	0.00	0.00	0.00

However, the stationary described above is only of first order. We may also comment about second order stationarity by considering the Priestley-Subba Rao (PSR) test, which we apply here without going into the details. The test principally focuses on testing if a series is second order stationary after all the trend and seasonality have been removed from the series. In this case thus, using the returns, Table 4.2 strongly suggests that the null hypothesis of stationary is rejected (all p-values are 0.00 for PSR test) in all the returns. Hence, although the returns are first order stationary, they are not second order stationary.

One of the properties that characterizes financial data is the lack of autocorrelations, except in very small time scales when external factors come into play. This property is hence evident from Table 4.3, where the null hypothesis of no autocorrelations is rejected at 5% level of significance.

Table 4.3 Box-Ljung Test for Autocorrelation at some Lags Values

Returns	Lag 1	Lag 2	Lag 3	Lag 11
USD/ZAR	8.3063 (0.003951)	11.9071 (0.002597)	12.5553 (0.005704)	35.7247 (0.000188)
Gold Mining	3.5991 (0.05781)	8.7551 (0.01256)	9.088 (0.02814)	16.8093 (0.1136)
All shares	8.7093 (0.003166)	8.7761 (0.01242)	20.935 (0.0001086)	39.524 (4.314e-05)
S&P 500	36.247 (1.738e-09)	41.9608 (7.733e-10)	42.1723 (3.688e-09)	58.6604 (1.643e-08)

4.3 Fitting daily USD/ZAR exchange rate to univariate GHDs and model comparison

In this section, we consider the fit of the univariate generalized hyperbolic distributions relative to that of the normal distribution to daily USD/ZAR returns. We also investigate which member of these distributions fits the return more accurately.

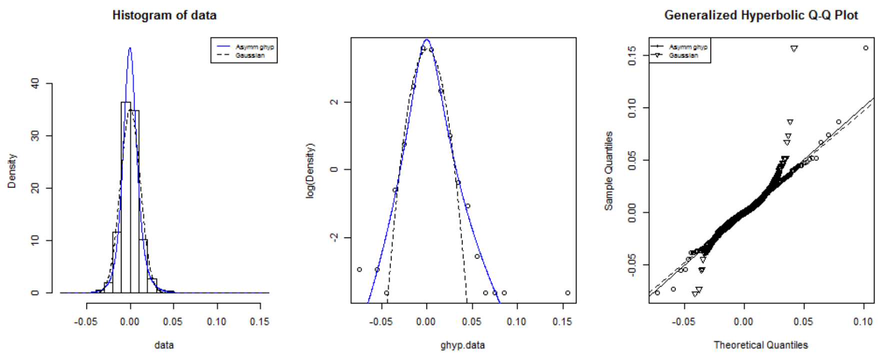


Figure 4.5: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of USD/ZAR returns using the Generalized Hyperbolic distribution.*

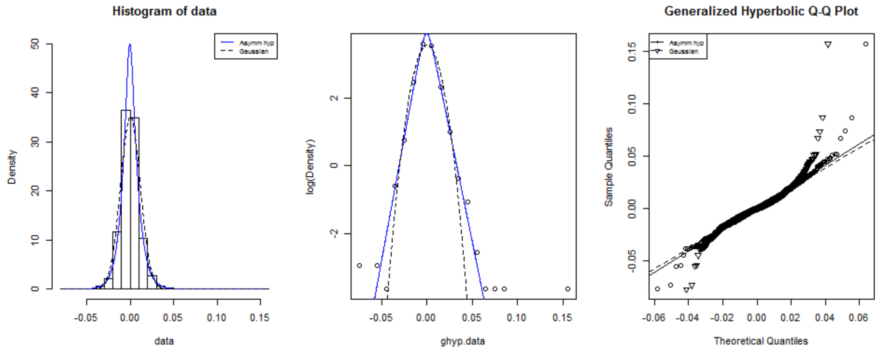


Figure 4.6: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of USD/ZAR returns using the hyperbolic distribution.*

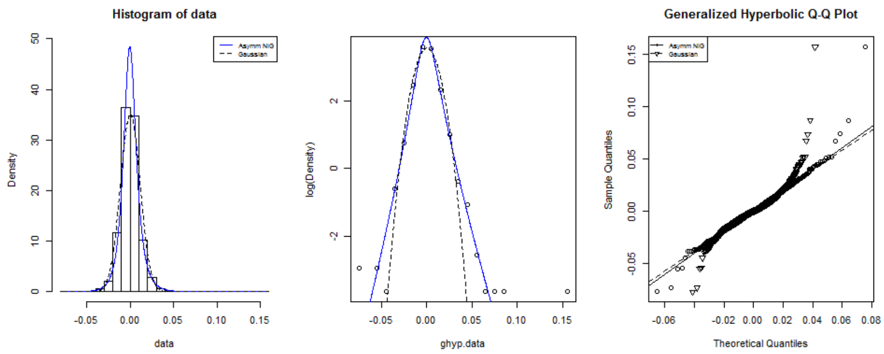


Figure 4.7: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of USD/ZAR returns using the Normal Inverse Gaussian distribution.*

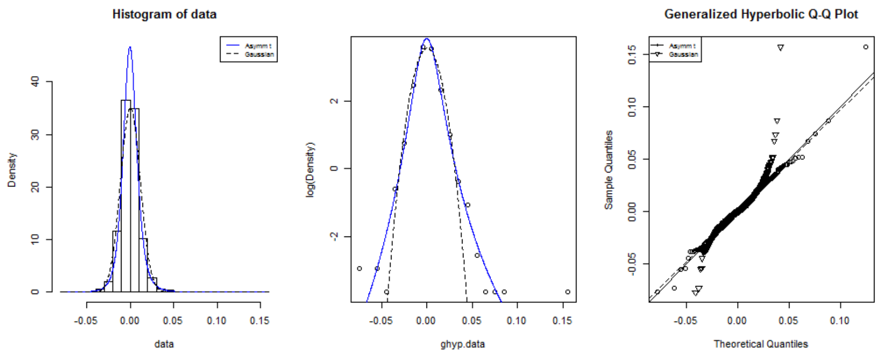


Figure 4.8: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of USD/ZAR returns using the Skew t distribution.*

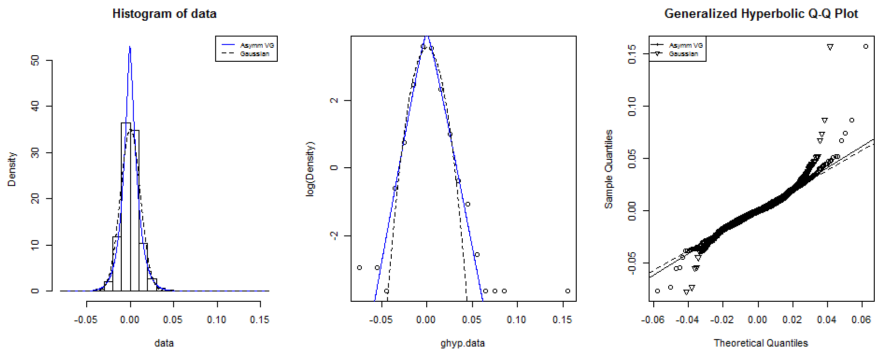


Figure 4.9: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of USD/ZAR returns using the Variance Gamma distribution.*

Even though the histogram in Figure 4.5 suggest that the normal distribution provides a better fit, the GHYP distribution is more accurate especially along the tails. This is seen from the density and Q-Q plots. The GHYP captures more points and hence the better distribution compared to the normal distribution.

Hence, in all cases, the GHDs portrays a better fit compared to the normal distribution. In this case then, the parameters of the GHDs can be estimated in Table 4.4.

Table 4.4 *Parameter Estimates of the GHDs. for daily USD/ZAR exchange rate.*

Dist./Par.	α	δ	β	μ	λ
GHYP	27.40321	0.01542617	8.674391	-0.0008933703	-1.819227
HYP	138.6539	0.00385099	6.941406	-0.0006869236	1
NIG	83.78933	0.01027212	7.627808	-0.0007922295	-0.5
Skew t	7.559395	0.01615183	7.559395	-0.0008098588	-2.019361
VG	148.7071	0	5.590765	-0.0005273284	1.331552

4.3.1 Comparison of the estimated GHDs

Having the parameters however, we have to assess the goodness of fit of these distributions. In this case, we utilized the Anderson-Darling (A-D) goodness of fit test, the combined Q-Q plot, and finally the loglikelihood and AIC values for model selection. However, we will also make use of the value-at-risk (VAR) measure using the Kupiec test to assess the fits of the models around the tails.

The A-D test is particularly significant when assessing the goodness of fit relative to a particular hypothesized distribution (in this case the GHDs), in which more emphasis is placed along the tails. This test measures the distance between the empirical distribution of the data and the hypothesized distribution (GHDs). In addition, the loglikelihood and AIC will be employed to determine the overall best model in terms of general performance or fit.

Table 4.5 A-D goodness of fit test for daily USD/ZAR exchange returns

Test	GHYP	HYP	NIG	Skew t	VG
A-D Statistic	0.6138	0.7372	0.533	0.599	0.9913
(P-Value)	0.635	0.5285	0.7134	0.6489	0.3619

Table 4.5, clearly suggests that the data fits adequately with the GHD and its sub classes. This is evident from the high p-values (of the A-D and K-S tests) which clearly do not reject the null hypothesis of the data being sampled from the GHDs. The NIG and skew t distributions having the lowest A-D statistics suggest that they fit better compared to other members with emphasis along the tail. However, the AIC and loglikelihood statistics from Table 4.6 suggest that the skew t distribution provides the best overall fit for the data with the smallest AIC value of -23887.15, and highest loglikelihood of 11947.57. It should also be noted that these statistics are very close to each other for the different subclasses of GHDs which is an indication of the goodness of the GHDs to fit the data.

Table 4.6 *AIC and log likelihood estimates of GHDs to Daily USD/ZAR returns*

Model	AIC	Log likelihood
GHYP	-23884.87	11947.44
HYP	-23859.06	11933.53
NIG	-23878.52	11943.26
Skew t	-23887.15	11947.57
VG	-23850.87	11929.44

Even though the AIC as well as loglikelihood statistics suggest that the skew t provides the best fit of the data, further analysis of the extreme left and right tails are considered. To this end, we use the Value-at-Risk (using the Kupiec likelihood ratio) which is conventionally used in financial statistics and econometrics to assess the fit of data at the tails. In our case, the left and right tails fit of the GHDs as well as the normal distribution are considered in table 4.7.

Table 4.7 *p-values for the Kupiec test for each distribution at different levels of significance of GHDs for Daily USD/ZAR returns*

Distribution	0.1%	0.5%	1%	99%	99.5%	99.9%
Normal	**0.00001**	0.05355	0.06252	**0.00020**	**0.00000**	**0.00000**
GHYP	0.91976	0.99817	0.86737	0.50464	0.81473	0.91976
HYP	0.55782	0.82143	0.40252	0.26888	0.37454	0.14238
NIG	0.91976	0.81473	0.23717	0.61783	0.65275	0.91976
Skew t	0.91976	0.81473	0.73974	0.61783	0.99817	0.91976
VG	0.5578247	0.9981653	0.31311	0.20796	0.18905	**0.023629**

At the 5% level of significance, Table 4.7 strongly suggests that the normal distribution provides a very poor fit to the daily returns, characterized by the small p-values. This is highly expected as the returns are leptokurtic (see section 4.2.2). Interestingly, the table also shows the VG distribution does not adequately depict the uppermost tail of the data with a low p-value of 0.023629. On the other hand, while the GHYP distribution accurately captures the lower tail of the USD/ZAR returns, the skew t is shown to capture the upper tail best. This is also evident from the combined Q-Q plot in figure 4.10 which displays these fits.

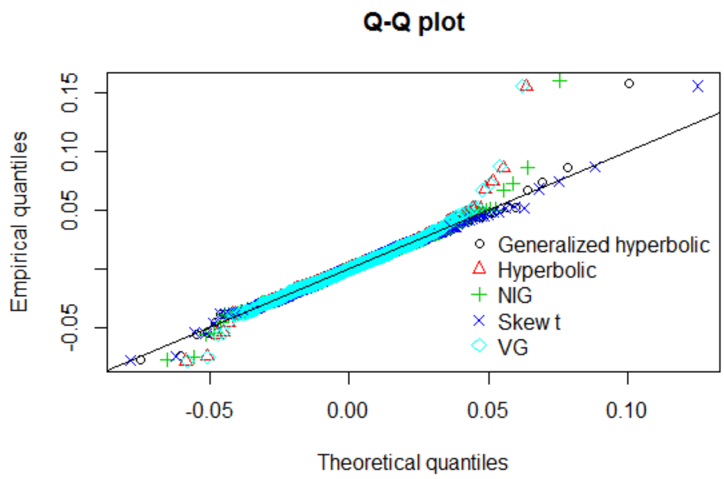


Figure 4.10: *Comparing the Fit Between the GHDs for Daily USD/ZAR Returns.*

Hence, based on the AIC and loglikelihood, but more importantly the VAR and combined Q-Q plot, we can conclude that the skew t distribution is the best member of the GHDs family to model daily USD/ZAR exchange returns for the period 03/01/2000 to 03/09/2014.

4.4 Fitting daily All shares returns to univariate GHDs and model comparison

In a similar manner like the USD/ZAR returns above, we will fit the GHDs to daily All shares returns and find out which member of the GHDs provides the best fit.

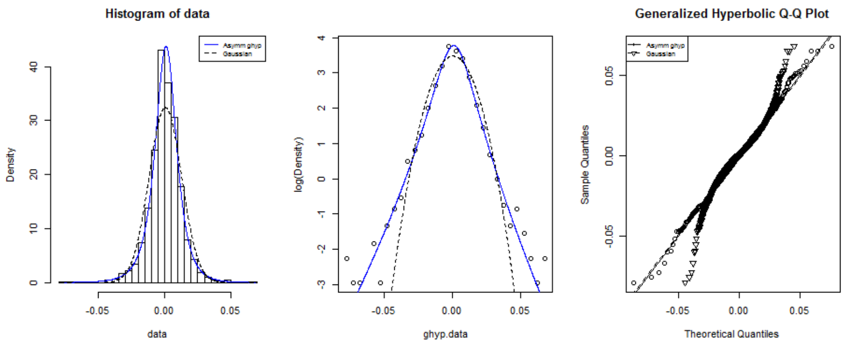


Figure 4.11: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of All shares returns using the Generalized Hyperbolic distribution.*

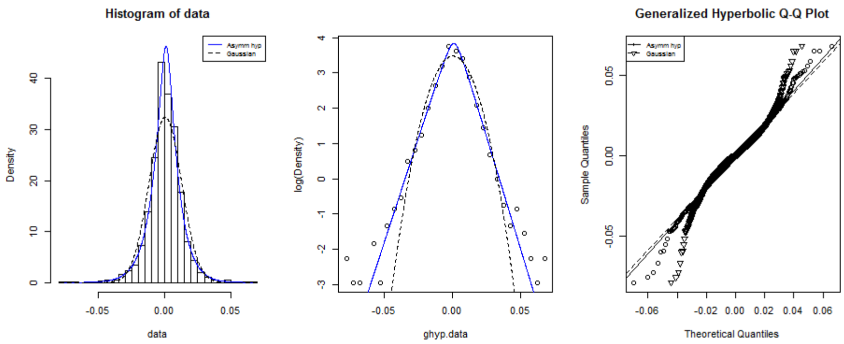


Figure 4.12: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of All shares returns using the hyperbolic distribution.*

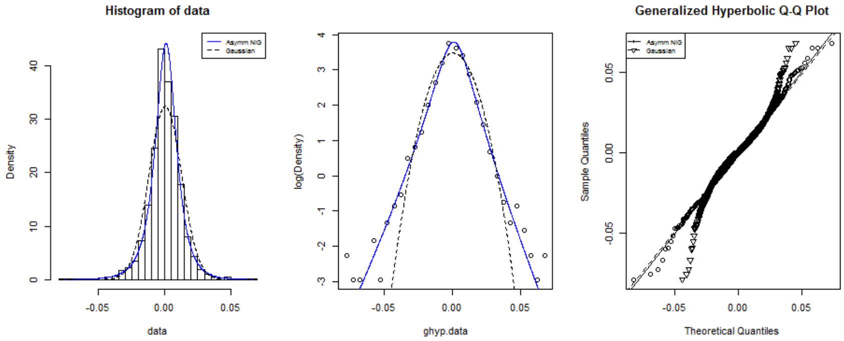


Figure 4.13: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of All shares returns using the Normal Inverse Gaussian distribution.*

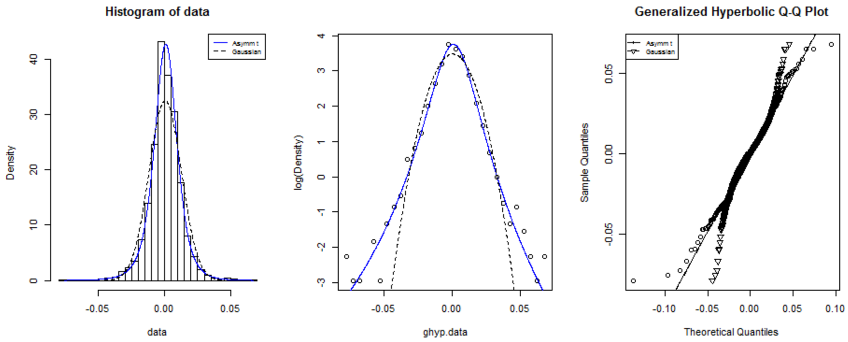


Figure 4.14: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of All shares returns using the Skew t distribution.*

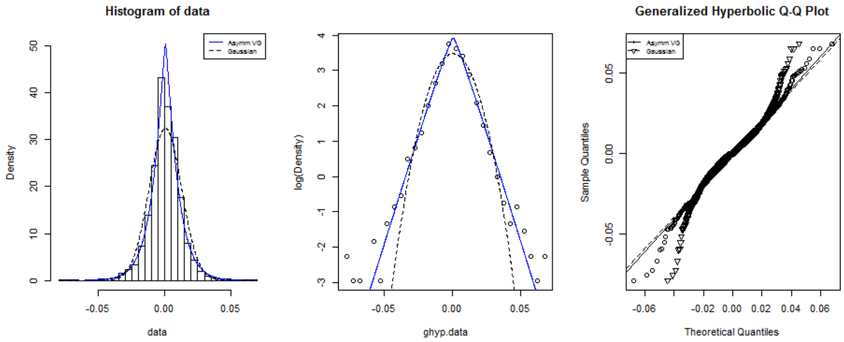


Figure 4.15: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of All shares Returns using the Variance Gamma distribution.*

The leptokurtic behavior of the returns is well captured by the GHDs in all five cases as is evident from the histogram plots. Secondly, the log(density) plots also show the accuracy of the GHDs compared to the normal distribution in which more points especially along the tails are captured by the GHDs. This is also clearly evident from the Q-Q plots. Hence, in all five cases, the GHDs display a better fit compared to the normal distribution with parameter estimated in Table 4.8.

Table 4.8 *Parameter Estimates of the GHDs. for Daily All shares Returns.*

Dist./Par.	α	δ	β	μ	λ
GHYP	61.83416	0.01273446	-5.419622	0.001306813	-0.8691459
HYP	123.7031	0.003598984	-4.727231	0.001174448	1
NIG	74.50546	0.0111189	-5.538746	0.001309138	-0.5
Skew t	4.868821	0.0172241	-4.868821	0.001235366	-1.933045
VG	128.0757	0	-0.5926151	0.0005699299	1.21221

4.4.1 Comparison of the estimated GHDs

Having estimated the parameters of the respective GHDs, we may assess the goodness of fit of these distributions. Later on, we compare the estimated models with one another. Table 4.9 summarizes the results of the A-D and K-S goodness of fit tests.

Table 4.9 *A-D goodness of fit for daily All shares Returns*

Test	GHYP	HYP	NIG	Skew t	VG
A-D Statistic	0.9055	1.1171	0.9063	1.0478	2.0447
(P-Value)	0.4109	0.3013	0.4104	0.3331	0.08678

Being a goodness of fit test with emphasis laid on the tails, the A-D test in Table 4.9 suggests that the GHDs provide a good fit to the data (with all p-values above 0.05). The NIG distribution provides the minimum distance between the actual distribution of the data and the hypothesized distribution (in this case the GHDs distribution). Thus, utilizing the AIC and loglikelihood to determine the best general fit, we obtain from Table 4.10 that the GHYP as well as the NIG distributions. They both provide the best possible fits compared to the other members. However an unusual observation is made from the table where even though the GHYP distribution has the maximum loglikelihood (11580.52), the lowest AIC value is obtained from the NIG (-23152.4).

Table 4.10 *AIC and log likelihood estimates of GHDs to Daily All shares returns*

Model	AIC	Log likelihood
GHYP	-23151.04	11580.52
HYP	-23142.19	11575.09
NIG	-23152.4	11580.2
Skew t	-23147.04	11577.52
VG	-23140.41	11574.21

A further and very insightful analyzis of these returns around the tails is carried out using the Kupiec likelihood ratio test (Table 4.11). In this case, the normal distribution is rejected at all levels of significance, justifying the fact that the All shares return is not normally distributed (or is heavy tailed).

Table 4.11 *p-values for the Kupiec test for each distribution at different levels of significance of GHDs for Daily All shares returns*

Distribution	0.1%	0.5%	1%	99%	99.5%	99.9%
Normal	**0.00000**	**0.00000**	**0.00000**	0.01429	**0.00001**	**0.00000**
GHYP	0.91976	0.47672	0.40252	0.99710	0.27038	0.55782
HYP	0.14238	0.50245	0.74896	0.74896	0.18904	0.02363
NIG	0.29868	0.47672	0.50464	0.99710	0.27038	0.55782
Skew t	0.66908	0.33832	0.31311	0.74896	0.27038	0.66908
VG	0.06095	0.37454	0.63148	0.61783	0.27038	0.14238

With the exception of the HYP distribution rejected at the right uppermost tail (99.9%), the other members prove to be adequate. However, we cannot identify a particular member as fitting the extreme left nor right tail better compared to the other members. This is because they all provide adequate fits at different levels of significance. This observation can also be evident from the combined Q-Q plot in fig 4.16. Nonetheless, we can select the GHYP as the model presenting the best fit, as it has more high p-values for the Kupiec test and maximum loglikelihood.

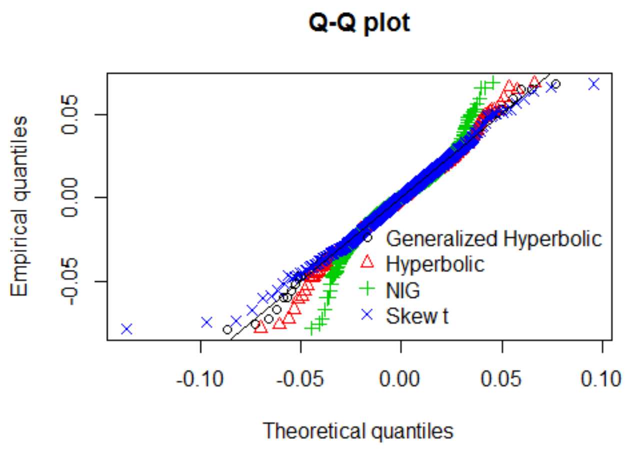


Figure 4.16: *Comparing the Fit Between the GHDs for Daily All Share Returns.*

4.5 Fitting daily gold mining returns to univariate GHDs and model comparison

In a similar manner as before, the gold returns are fitted with the GHDs and compared with the normal distribution. The different subclasses are illustrated from Figure 4.17 down to Figure 4.21.

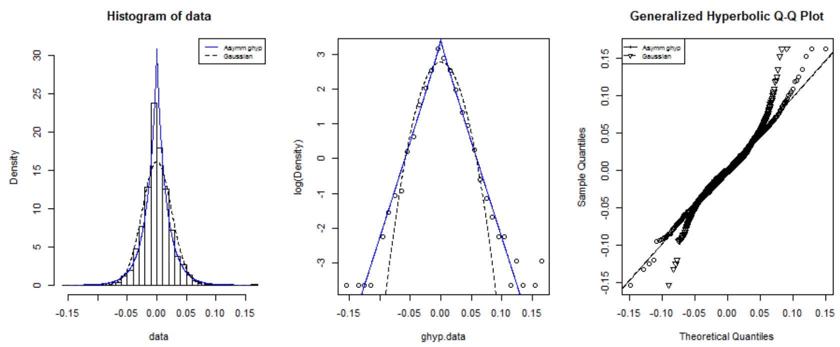


Figure 4.17: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily gold mining returns using the Generalized Hyperbolic distribution.*

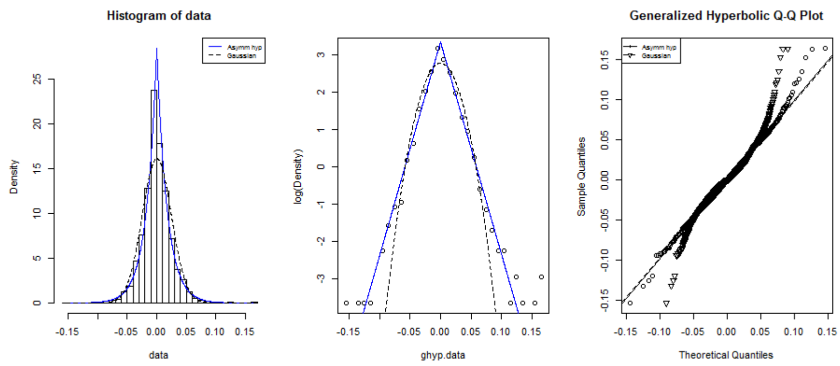


Figure 4.18: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily gold mining returns using the Hyperbolic distribution.*

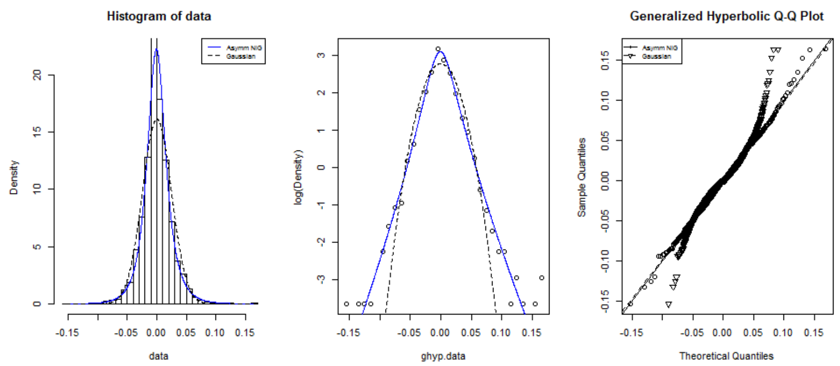


Figure 4.19: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily gold mining returns using the Normal Inverse Gaussian distribution.*

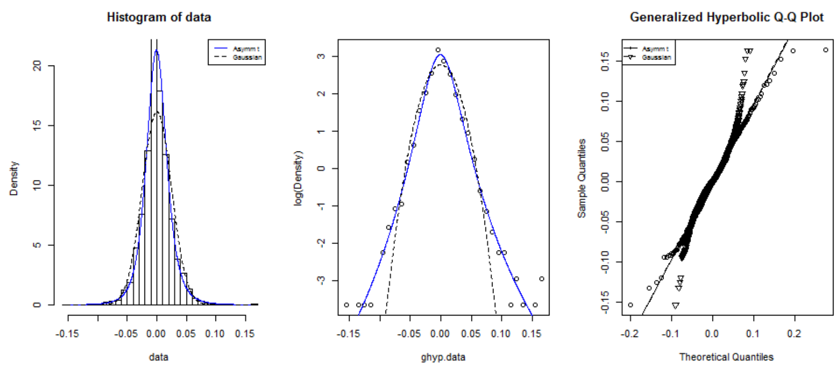


Figure 4.20: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily gold mining returns using the Skew t distribution.*

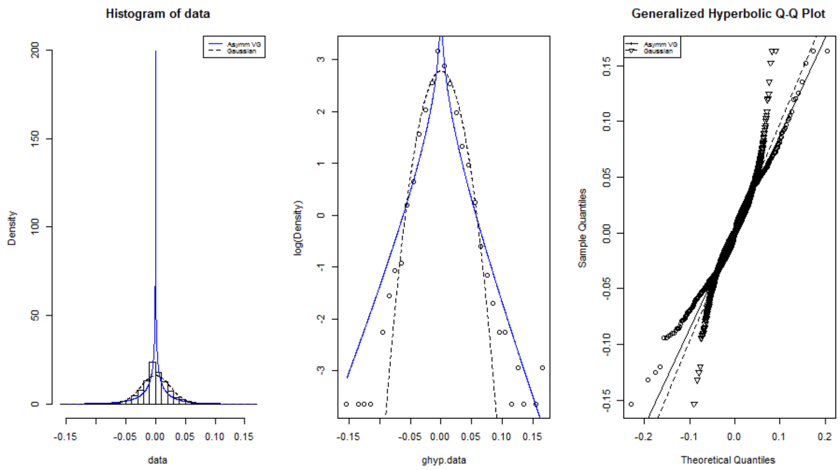


Figure 4.21: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily gold mining returns using Variance Gamma distribution.*

The histogram plots for the GHYP, NIG, HYP, and Skew t distributions show that these distributions provide a better fit compared to the normal distribution. This observation is well displayed in the log(density) and Q-Q plots, which show these GHDs capturing more points than the normal distribution. This is highly due to the leptokurtic nature of the GHDs. However, unlike the other members, the VG distribution shows some degree of over fitting especially from the histogram plot. The log(density) and Q-Q plots suggest that this distribution is not adequate as too many points are left out by the VG. Thus, to verify these observations, we will carry out a goodness of fit test. Before we proceed to the test, we consider the parameters estimation of the GHDs.

Table 4.12 *Parameter Estimates of the GHDs. for daily Gold Mining Returns.*

Dist./Par.	α	δ	β	μ	λ
GHYP	54.02455	3.494136e-07	0.01881145	-8.94528e-07	0.9100622
HYP	56.9853	1.402822e-06	0.2180549	-7.664291e-07	1
NIG	35.6466	0.02174905	2.017429	-0.001113406	-0.5
Skew t	1.916438	0.03425139	0.03425139	-0.001100534	-1.900404
VG	29.49628	0	-1.569527	-6.710295e-11	0.334238

4.5.1 Comparison of the estimated GHDs

Just as speculated above, the A-D goodness of fit test suggests that the VG provides a poor fit for the data, with a very low p-value of $1.578e-07$. However, the other members of the GHDs provide a good fit.

Table 4.13 *A-D goodness of fit for daily Gold mining Returns*

Test	GHYP	HYP	NIG	Skew t	VG
A-D Statistic	2.4573	1.801	1.0303	1.6535	63.4337
(P-Value)	0.05216	0.1185	0.3418	0.1437	1.578e-07

Given that the A-D test excludes the VG distribution to model the daily gold returns, we then find out which member of the GHDs provides the best fit. This is done using the AIC and loglikelihood statistics. Table 4.14 presents the results of the tests.

Table 4.14 *AIC and log likelihood estimates of GHDs to Daily Gold Mining returns*

Model	AIC	Log likelihood
GHYP	-17852.04	8931.019
HYP	-17853.14	8930.568
NIG	-17833.28	8920.642
Skew t	-17818.22	8913.11

Similar observations made about the daily All shares returns is made about the daily gold returns; even though the GHYP distribution has the maximum loglikelihood of 8931.019, the lowest AIC statistics (of -17853.14) is obtained from the HYP distribution. Hence as above our best model is chosen as the hyperbolic distribution.

We then shift our interest on the distribution of the gold returns along the tails. Just like in the previous cases, the Kupiec likelihood ratio test is used and the results summarized in table 4.15

Table 4.15 *p-values for the Kupiec test for each distribution at different levels of significance of GHDs for Daily Gold Mining returns*

Distribution	0.1%	0.5%	1%	99%	99.5%	99.9%
Normal	**0.00000**	**0.00001**	**0.00618**	**0.00020**	**0.00001**	**0.00000**
GHYP	0.91976	0.33832	0.12494	0.50464	0.82143	0.14238
HYP	0.91976	0.47672	0.23717	0.61783	0.47672	0.14238
NIG	0.91976	0.63789	0.86737	0.31311	0.22674	0.29868
Skew t	0.66908	0.47672	0.87362	0.31311	**0.02267**	0.66908

Table 4.15 highly suggests that the NIG distribution is particularly adequate for modeling the lower tail of the daily gold returns while the upper tail is in general best fitted by the GHYP distribution. A combined Q-Q plot may also be used to visualize the fits. This is presented in figure 4.28.

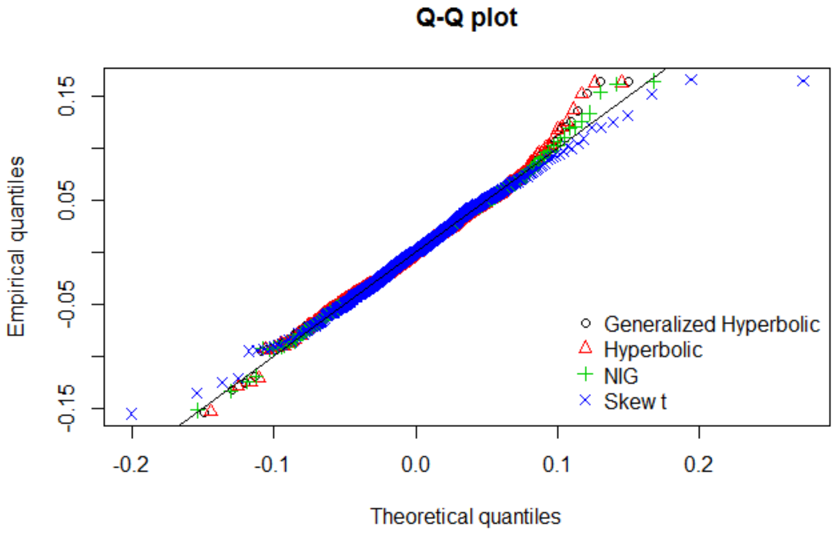


Figure 4.22: *Comparing the Fit Between the GHDs for Daily Gold Mining Returns.*

Thus in summary, even though the GHYP distribution has the largest loglikelihood, the best overall fit is obtained from the HYP distribution with the smallest AIC value. However, more analyzis at the tails using the Kupiec likelihood ratio test suggests that the lower tail is best fitted by the NIG distribution and the upper tail by the GHYP distribution in general.

4.6 Fitting daily S&P 500 returns to univariate GHDs and model comparison

Finally, we complete these univariate analyzis with the S&P 500 daily returns. As before, the different subclasses of the GHD is fitted.

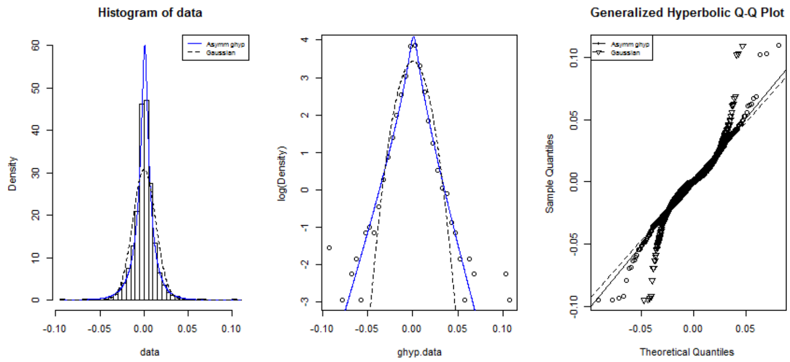


Figure 4.23: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily S&P 500 returns using the Generalized Hyperbolic distribution.*

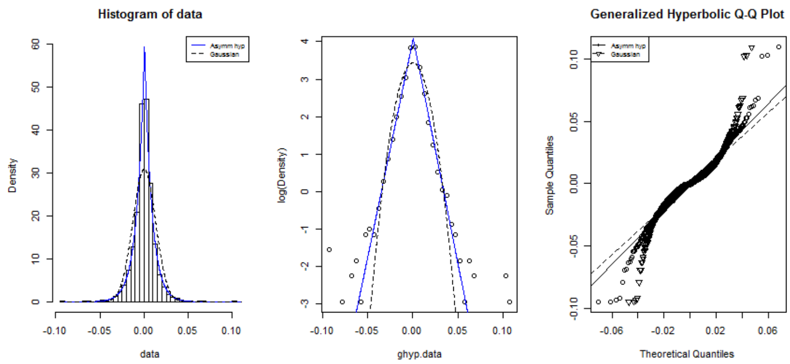


Figure 4.24: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily S&P 500 returns using the Hyperbolic distribution.*

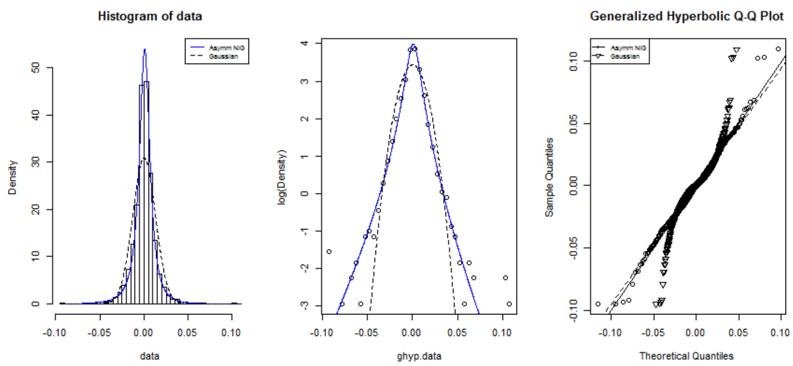


Figure 4.25: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily S&P 500 returns using the Normal Inverse Gaussian distribution.*

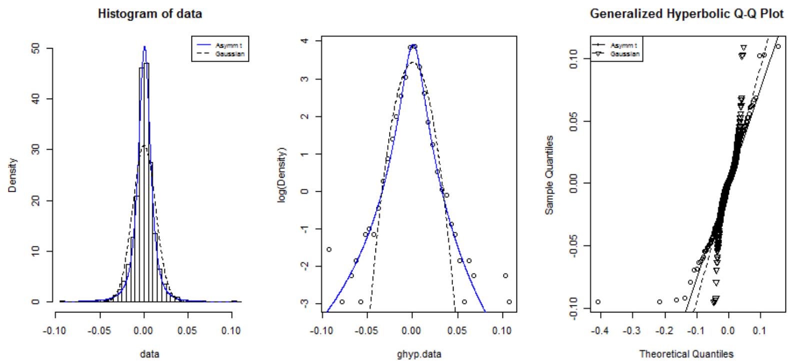


Figure 4.26: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily S&P 500 returns using the Skew t distribution.*

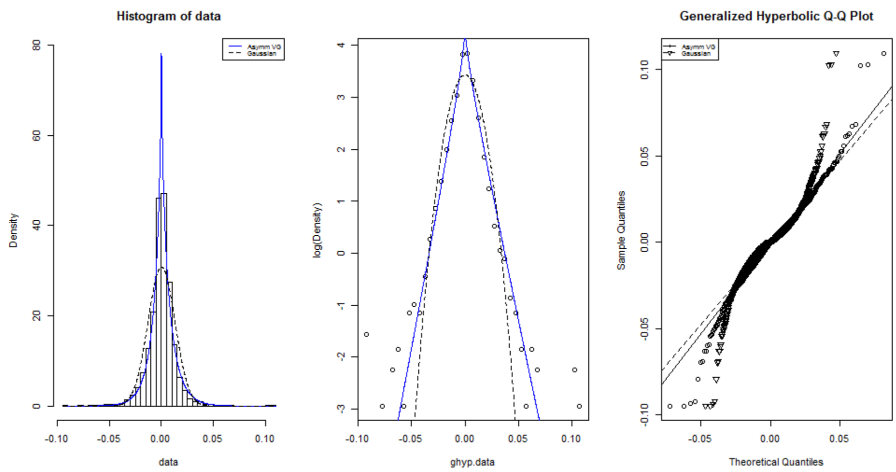


Figure 4.27: *Histogram(left), log(density) plot(middle), Q-Q plot(right) of daily S&P 500 returns using Variance Gamma distribution.*

From a graphical perspective, one will suggest that the GHDs seem to be provide a good fit, more particularly compared to the normal distribution. However, the goodness of fit test presented in table 4.17 suggests that only the GHYP as well as the NIG distribution seems to fit the data. The parameters estimates for the different fits is given in table 4.16

Table 4.16 *Parameter Estimates of the GHDs. for daily S&P 500 Returns.*

Dist./Par.	α	δ	β	μ	λ
GHYP	78.01076	0.003029988	-4.7603	0.0008442193	0.3142801
HYP	118.722	8.473528e-07	-2.672532	0.0004644861	1
NIG	45.07823	0.007441975	-4.999903	0.000915102	-0.5
Skew t	3.619791	0.01129065	-3.619791	0.0007807623	-1.236657
VG	100.2427	0	5.264963	5.672193e-12	0.7505957

4.6.1 Comparison of the estimated GHDs

One more important point to note about the goodness of fit of the GHYP and NIG distributions is that although they are the only sub class that seem to fit the returns, the small p-values is an indication that the fit is not the best. The fit of all the other subclasses is rejected from Table 4.17.

Table 4.17 *A-D goodness of fit for daily S&P 500 Returns*

Test	GHYP	HYP	NIG	Skew t	VG
A-D Statistic	1.1201	3.6235	1.6991	3.104	5.2993
(P-Value)	0.3	0.01335	0.1353	0.02424	0.002064

Hence, based on these results, we utilize the loglikelihood and AIC to find out which member provides the better fit. Table 4.18 illustrates these statistics.

Table 4.18 *AIC and log likelihood estimates of GHDs to Daily S&P 500 returns*

Model	AIC	Log likelihood
GHYP	-23477.1	11743.55
NIG	-23467.18	11737.59

Hence, with the maximum loglikelihood and smallest AIC value, the GHYP distribution seem to provide a better fit compared to the NIG distribution in general. However, more analyzis of the fits along the tails is carried out using the Kupiec likelihood ratio test. This is carried out in Table 4.19.

Table 4.19 *p-values for the Kupiec test for each distribution at different levels of significance of GHDs for Daily S&P 500 returns*

Distribution	0.1%	0.5%	1%	99%	99.5%	99.9%
Normal	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
GHYP	0.14238	0.65275	0.61783	0.63214	0.37454	0.06095
NIG	0.00581	0.00003	NAN	0.00000	NAN	0.00581

Table 4.19 clearly shows that the GHYP distribution provides the best fit com-

pared to the NIG and normal distribution both along the lower and the upper tail as well. We can hence assess the combined Q-Q plot to visualize the fit along the tails.

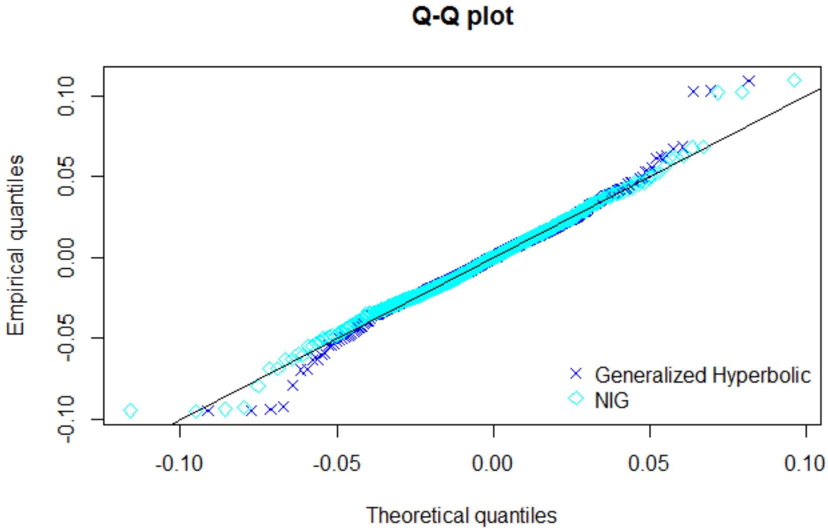


Figure 4.28: *Comparing the Fit Between the GHDs for S&P* 500 *Returns.*

Contrary to the AIC and Kupiec test, Figure 4.28 suggests that the NIG distribution provides a better fit along the tails. However, based on VAR and AIC, our best model is selected as the GHYP distribution.

4.7 Fitting returns to MAGHDs

So far in this research, we have focused on fitting returns to univariate GHDs. Henceforth, we will consider the multivariate fit; by first considering the multivariate affine GHDs, and secondly, the MGHDs without any affine transformation of the data.

As noted earlier, the MAGHDs provide more flexibility to the parameters; allowing each ω_i to be calculated for each marginal. This gives more accurate fits as each margin can now be fitted independently with a GHD. However, affine transformations do not allow to model dependencies when returns are dependent. The procedure discussed in section 1.4 using our four indices from the JSE is then implemented in *R* using the ghyp package and the parameters are estimated for each member of the MAGHD family. The estimated models are summarized from Table 4.20 through to Table 4.24.

Table 4.20 *MAGHYP Parameters Estimates*

Parameter	Index			
	Alsi	USD/ZAR	Gold Mining	S&P 500
α	0.7408859	0.1850786	1.134276	0.4017248
β	-0.06743362	0.08137277	0.08313861	-0.02503719
λ	-0.9050494	-1.966545	0.1830644	-1.2558
μ	8.737187	-4.008628	-7.255324	-3.707923
	2.309348e-04	-3.587517e-05	1.126206e-04	5.395335e-05
	-3.587517e-05	1.873834e-04	1.716552e-05	-3.424877e-05
Σ	1.126206e-04	1.716552e-05	3.068459e-04	-2.174626e-05
	5.395335e-05	-3.424877e-05	-2.174626e-05	1.420914e-04

Table 4.21 *MAHYP Parameters Estimate*

Parameter	Index			
	Alsi	USD/ZAR	Gold Mining	S&P 500
α	1.527548	1.527548	1.446441	1.50671
β	-0.05888042	-0.05888042	0.04852812	-0.03332752
λ	1.00000	1.00000	1.00000	1.00000
μ	7.8098903	9.7474070	-6.0509377	0.6885622
	5.905211e-05	-8.345255e-06	2.098133e-05	8.293274e-06
Σ	-8.345255e-06	4.806149e-05	3.000592e-06	-5.264445e-06 3
	2.098133e-05	3.000592e-06	5.638157e-05	-3.342659e-06 4
	8.293274e-06	-5.264445e-06	-3.342659e-06	2.184114e-05

Table 4.22 *MANIG Parameters Estimate*

Parameter	Index			
	Alsi	USD/ZAR	Gold Mining	S&P 500
α	0.9176057	0.9666203	0.8575087	0.7906841
β	-0.06842726	0.08477563	0.08329543	-0.0288432
λ	-0.50000	-0.50000	-0.50000	-0.50000
μ	8.662508	-4.160716	-7.365237	-3.427161
	2.253549e-04	-2.059482e-05	1.712650e-04	3.973218e-05
Σ	-2.059482e-05	1.231063e-04	2.135420e-05	-2.522139e-05
	1.712650e-04	2.135420e-05	4.477567e-04	-1.601432e-05
	3.973218e-05	-2.522139e-05	-1.601432e-05	1.046385e-04

Table 4.23 *MAST Parameters Estimate*

Parameter	Index			
	Alsi	USD/ZAR	Gold Mining	S&P 500
α	0.0599036	0.07955098	0.07955098	0.01962255
β	-0.0599036	0.07955098	0.07955098	-0.01962255
λ	-1.931949	-2.047566	-2.047566	-1.661972
μ	0.10061640	-0.05946072	-0.05946072	0.01568998
Σ	3.600800e-04	-3.122074e-05	2.929987e-04	6.184773e-05
	-3.122074e-05	1.930243e-04	3.584497e-05	-3.926001e-05
	2.929987e-04	3.584497e-05	7.632863e-04	-2.492814e-05
	6.184773e-05	-3.926001e-05	-2.492814e-05	1.628820e-04

Table 4.24 *MAVG Parameters Estimate*

Parameter	Index			
	Alsi	USD/ZAR	Gold Mining	S&P 500
α	1.323829	1.725265	1.467032	1.534956
β	0.03288869	0.0786658	0.01890113	-0.02507115
λ	0.8996854	1.385961	1.057008	1.095641
μ	2.038255e-08	-4.824887	-1.104842	2.630542e-01
Σ	2.575898e-09	-2.378413e-10	2.019783e-09	5.170912e-10
	-2.378413e-10	1.352923e-09	2.572827e-10	-3.282417e-10
	2.019783e-09	2.572827e-10	5.302174e-09	-2.084170e-10
	5.170912e-10	-3.282417e-10	-2.084170e-10	1.361810e-09

4.7.1 Goodness of fit of MAGHDs

Having estimated the parameters using the algorithm described by Schmidt *et al.* (2006), the goodness of this fit can now be investigated. Keeping in mind that the multivariate generalized hyperbolic distributions are obtained from affine transformed univariate variables, we will first carry out a univariate Anderson and Darling goodness of fit on these transformed variables.

Table 4.25 Anderson and Darling goodness of fit test for affine transformed returns.

Returns	Distribution	AD statistic	P-value
USD	GHYP	0.2255	0.9818
	HYP	0.432	0.8165
	NIG	0.2127	0.9865
	Skew t	0.242	0.9745
	VG	0.6765	0.5787
All shares	GHYP	0.9111	0.4074
	HYP	1.1039	0.3071
	NIG	0.9055	0.4109
	Skew t	1.0471	0.3335
	VG	7.0297	**0.000321**
Gold mining	GHYP	0.8555	0.4427
	HYP	1.2241	0.2585
	NIG	0.9047	0.4114
	Skew t	1.4057	0.2007
	VG	2.003	0.09149
S&P 500	GHYP	0.5872	0.6601
	HYP	0.4787	0.4787
	NIG	0.4787	0.4787
	Skew t	0.4787	0.4787
	VG	0.8566	0.442

Table 4.25 shows that these affine transformed variables present a very good fit with the generalized hyperbolic distributions. These results are even better than the univariate marginal distributions of these returns with very low distances between the actual and hypothesis distributions as well as the probability values. However, the All shares returns rejects the fit with the VG distribution. It is also important to note that even though the VG distribution is good, its provides an overall bad fit compared to the other members of the GHDs. This is perhaps due to the parameter restriction placed on the scale parameter, δ. Moving to the four dimensional multivariate affine distributions, we propose the kernel smoothing technique. As we may recall, the affine transformed variables are independent. As such, we use the following procedure (for example for the MANIG);

- We carry out the necessary transformations described by Schmidt *et al.* 2006. Then, we fit each of the returns with the univariate NIG distribution.
- We generate independent samples from the fit of each NIG distribution, since the samples are independent.
- We bring these samples together to form our multivariate distribution (four dimensional). This is possible since the affine variables are independent.

- This multivariate sample is then compared with our returns using the kernel smoothing.

Table 4.26 Four dimensional kernel smoothing goodness of fit test of MAGHDs.

Distribution	kernel smoothing distance
MAGHYP	0.01165438
MAHYP	0.01074652
MANIG	0.01010627
MAST	0.01270463
MAVG	0.01020197

Table 4.26 shows that the kernel distances are very close together, showing how similar the fits of these distributions are to one another. However, the MANIG has the smallest distance and hence can be regarded as the model producing of best fit.

4.8 Fitting returns to four dimensional GHDs and assessing goodness of fit

This section is devoted to the fit of multivariate generalized hyperbolic distributions on which no particular transformations have been carried out on the data. As before, the Multi-Cycle Expectation Conditional Maximization (MCECM) algorithm as well as the kernel smoothing goodness of fits are used in R with the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-parametrization. However, we present the four dimensional fit here while the two dimensional fit (mainly because we will be analyzing the correlation between the variables later) is documented in the appendix. The goodness of fit test is done in such a way that we compare the fit of the multivariate GHDs with the fits depicted by the multivariate normal distribution.

For the MGHYP:

$$\lambda = 0.6485490 \quad \hat{\alpha} = 0.9923645$$
$$\mu = (0010014882, -0.0004755705, -0.0003145183, 0.0006408145),$$
$$\gamma = (-0.0005410421, 0.0006458176, 0.0004480307, -0.0005777333)$$

$$\Sigma = \begin{pmatrix} 1.468555e-04 & -1.512471e-05 & 1.151816e-04 & 4.716599e-05 \\ -1.512471e-05 & 1.266780e-04 & -5.242250e-06 & -2.805251e-05 \\ 1.151816e-04 & -5.242250e-06 & 6.037810e-04 & 1.402439e-06 \\ 4.716599e-05 & -2.805251e-05 & 1.402439e-06 & 1.448873e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 2.5 \quad \hat{\alpha} = 0.3748423$$
$$\mu = (0.0010665156, -0.0005576768, -0.0003651900, 0.0006896263)$$
$$\gamma = (-0.0006151161, 0.0007389395, 0.0005060401, -0.0006352074)$$

$$\Sigma = \begin{pmatrix} 1.419720e-04 & -1.494141e-05 & 1.104862e-04 & 4.628980e-05 \\ -1.494141e-05 & 1.214818e-04 & -4.810835e-06 & -2.797471e-05 \\ 1.104862e-04 & -4.810835e-06 & 5.831227e-04 & 3.913272e-07 \\ 4.628980e-05 & -2.797471e-05 & 3.913272e-07 & 1.419618e-04 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5 \quad \hat{\alpha} = 0.9643238$$
$$\mu = (0.0010279593, -0.0004914882, -0.0003335669, 0.0006534718)$$
$$\gamma = (-0.0005577392, 0.0006497949, 0.0004589355, -0.0005796815)$$

$$\Sigma = \begin{pmatrix} 1.499574e-04 & -1.508108e-05 & 1.184344e-04 & 4.760852e-05 \\ -1.508108e-05 & 1.296561e-04 & -5.453881e-06 & -2.791123e-05 \\ 1.184344e-04 & -5.453881e-06 & 6.179283e-04 & 2.024664e-06 \\ 4.760852e-05 & -2.791123e-05 & 2.024664e-06 & 1.463836e-04 \end{pmatrix}$$

For the MVG:

$$\lambda = 1.211898 \quad \hat{\alpha} = 0.0000$$
$$\mu = (4.540003e-04, -8.618414e-05, -9.803782e-05, 2.587738e-04)$$
$$\gamma = (2.558839e-05, 2.315635e-04, 2.138884e-04, -1.737041e-04)$$

$$\Sigma = \begin{pmatrix} 1.530859e-04 & -1.603322e-05 & 1.198265e-04 & 4.911009e-05 \\ -1.603322e-05 & 1.342963e-04 & -5.547762e-06 & -2.908930e-05 \\ 1.198265e-04 & -5.547762e-06 & 6.246844e-04 & 2.440836e-06 \\ 4.911009e-05 & -2.908930e-05 & 2.440836e-06 & 1.502529e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.787308 \quad \nu = 3.574616$$
$$\mu = (0.0010053447, -0.0004634910, -0.0003164931, 0.0006207394)$$
$$\gamma = (-0.0005256595, 0.0006107702, 0.0004340736, -0.0005371098)$$

$$\Sigma = \begin{pmatrix} 1.604177e-04 & -1.566209e-05 & 1.274539e-04 & 5.038383e-05 \\ -1.566209e-05 & 1.388777e-04 & -5.840071e-06 & -2.898898e-05 \\ 1.274539e-04 & -5.840071e-06 & 6.623309e-04 & 2.720059e-06 \\ 5.038383e-05 & -2.898898e-05 & 2.720059e-06 & 1.548777e-04 \end{pmatrix}$$

For the multivariate Normal:

$$\mu = (4.797857e-04, 1.471623e-04, 1.174975e-04, 8.373226e-05)$$

$$\Sigma = \begin{pmatrix} 1.513398e-04 & -2.139560e-05 & 1.124750e-04 & 5.451118e-05 \\ -2.139560e-05 & 1.295920e-04 & -5.754663e-06 & -3.935084e-05 \\ 1.124750e-04 & -5.754663e-06 & 6.114319e-04 & -3.025136e-06 \\ 5.451118e-05 & -3.935084e-05 & -3.025136e-06 & 1.669168e-04 \end{pmatrix}$$

The parameter ν determines the degree of freedom of the distribution and is only defined for the MST distribution. By definition, $\nu = -2 \times \lambda$. Moreover, the location vector, μ, dispersion matrix, Σ, as well as the skewness vector, γ, are also summarized. It should be noted that the dispersion matrix is obtained with the returns listed in the order; All shares, USD/ZAR, S&P 500, Gold Mining.

4.8.1 Assessing the goodness of fit of four dimensional GHDs

Once the parameters estimated, we can proceed to assess the fit of the MGHDs. However, as mentioned earlier, we compare these fits with those of the four dimensional normal distribution and find out which distribution has smallest distance (the distance here estimated by kernel smoothing which compares the distribution of the data to the hypothesized distributions). Table 4.27 summarizes the fits.

Table 4.27 *Goodness of fit of MGHDs*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	2.451889	45010.98	-89981.95
MHYP	6.56303	44875.97	-89713.94
MNIG	4.21383	45042.92	-90047.84
MVG	4.305984	44971.08	-89904.15
MST	4.508078	45039.85	-90041.69
Multivariate normal	25.22783	43475.96	-86923.92

With a kernel distance of 25.22783, Table 4.27 strongly shows the poor fit of the multivariate normal distribution relative to the MGHDs which have significantly smaller distances. This suggests the better fit that MGHDs portray. However, even though the MGHYP, has the smallest distance, it does not have the smallest AIC or highest loglikelihood value. This does not however contradict the better fit indicated by the MNIG, as these MGHDs are fitted with some parameter initialization. Thus, changing these initializations will slightly influence the loglikelihood and AIC. Hence, based on logliklihood and AIC, we can conclude that the MNIG (with highest loglik-lihood and smallest AIC) presents the better fit compared to the other members of the MGHD distribution.

4.9 Fitting returns to Archimedean copulas

Earlier in this chapter, the marginal distributions of the daily returns were fitted with the different subclasses of the GHDs and the best fits were obtained. Now, we utilize copulas in order to describe the dependence structure between pairwise marginals. Before we go into these models, it is necessary to look at the scatter plots of associated with these bivariate returns.

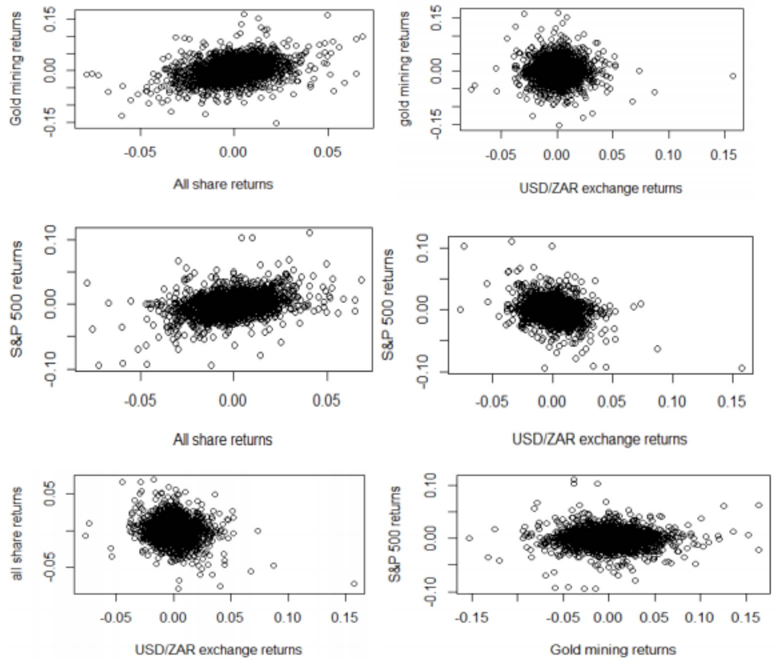


Figure 4.29: *Bivariate scatter plots.*

Figure 4.29 shows some form of negative dependency between the USD/ZAR and all other returns except the gold mining returns. Some negative dependency is also exhibited by the S&P 500 and the gold mining returns. While on the other hand, the gold mining and All shares as well as S&P 500 and All Share returns seem to be positively dependent. Table 4.28 further illustrates these observations.

Table 4.28 Kendall tau and Spearman rho correlations of the bivariate returns

Returns	Kendall tau	Spearman rho
USD/ZAR and All shares	-0.06676407	-0.09848042
UDS/ZAR and Gold mining	-0.01040614	-0.0149537
USD/ZAR and S&P 500	-0.1216815	-0.1779649
All shares and Gold mining	0.2635787	0.3723275
All shares and S&P 500	0.2113006	0.3025654
Gold mining and S&P 500	0.0006230341	-0.0001758794

As discussed earlier, Clayton copulas are known for their ability to capture lower tail dependencies. However, they cannot account for negative dependencies between marginal. On the other hand, Gumbel copulas are suitable to capture upper tails dependencies and cannot also account for negative dependencies. This means that only marginals that are positively dependent can be accounted for by the Clayton and Gumbel copulas. The Frank copula even though accounts for positive as well as negative dependencies, unfortunately does not model any tail dependencies. Thus, the ability of the kolmogorov-Smirnov goodness of fit implemented in *R*, using Kendall process will provide us with a preliminary idea of whether the observations from Figure 4.29 seem correct.

Table 4.29 Kolmogorov-Smirnov goodness of fit based on Kendall procedure for bivariate copulas

Bivariate returns	Clayton copula	Gumbel copula	Frank copula
USD/ZAR and All shares	NAN	NAN	0.7768835
USD/ZAR and gold mining	NAN	NAN	1.152871
USD/ZAR and S&P 500	NAN	NAN	0.8683853
All shares and gold mining	2.419068	1.960445	2.918488
All shares and S&P 500	1.472056	2.315072	1.897947
gold mining and S&P 500	NAN	NAN	0.9152758

- **NAN here means the statistic is unavailable.**

The lack of fit depicted by Table 4.29 between USD/ZAR and all other marginals as well as between the gold mining and the S&P 500 suggest that these bivariate returns are all negatively dependent on one another. This is because this test (Kolmogorov-Smirnov) applies only to positively dependent bivariate returns. Thus, in these cases, we will use the Frank copula to assess the negative dependencies. When assessing the fit between the All shares and gold mining returns, Table 4.29 also suggests that the Gumbel copulas provides the best fit with the minimum distance compared to other copula models. A similar observation is also made between the All shares and S&P 500, where the Clayton copula is shown to provide the best fit with minimum distance of 1.472056. Thus, with the estimated copula models in Table 4.29, we may calculate the copulas based dependency measures. This is given in Table 4.30.

Table 4.30 Estimated copula dependency measures

Bivariate returns (copula)	copula parameter	Kendall tau	Spearman rho	lower tail	upper tail
USD/ZAR and All shares (Frank copula)	-0.62209	-0.068856	-0.10315	0	0
USD/ZAR and gold mining (Frank copula)	-0.095579	-0.010619	-0.015928	0	0
USD/ZAR and S&P 500 (Frank copula)	-1.1416	-0.12523	-0.18704	0	0
All shares and gold mining (Gumbel copula)	1.3274	0.24667	0.36067	0	0.3143
All shares and S&P 500 (Clayton copula)	0.46233	0.18776	0.27807	0.2233	0
gold mining and S&P 500 (Frank copula)	0.001874	0.000208	0.0003124	0	0

As expected, Table 4.30 suggests some degree of negative dependence between the USD/ZAR exchange rate and all other returns. However, this negative dependence (indicated by Kendall tau as well as Spearman rho) is very close to 0 especially with the All shares and gold mining indices, indicating almost independence (the Frank copula parameter is close to 0, is an indication of independence). On the other hand, some positive dependence is exhibited between the All shares and gold mining as well as S&P 500. Thus, the Gumbel copula providing the best fit between the All shares and gold mining with parameter close to one is also an indication of almost independence. However, we have some upper tail dependence. We may thus conclude that even though the dependency between the two returns is not very strong, it however exists especially along the upper tails. Similar remarks are also made with the Clayton copula when fitting the All shares and S&P 500 returns. In this case along the lower tails. This is largely one of the disadvantages of these Archimedean copulas, as none is able to capture both upper and lower tails of bivariate returns at same time. One last point to note about these measures of dependency (Kendall tau and Spearman rho) is how close these values are to their sample counterparts in Table 4.28. As these lead to more flexibility when modeling dependencies with copulas.

Chapter 5

Conclusion and recommendations

5.1 Conclusion

This research shows that the GHDs provide an alternative class (to the multivariate normal and t) of distributions for modeling the heavy tails properties as well as volatility of financial returns. Most importantly, these distributions can be extended to model multivariate returns. This is very important as financial indices usually do not vary independently, but may depend on other returns. Hence, the need to model all these sequence of returns with multivariate distributions.

We started by modeling the individual univariate returns with each member of the GHD. These distributions proved to provide a very good fit based on the Anderson-Darling goodness of fit test. The value-at-risk using Kupiec test was then utilized to further assess the extend to which these distribution were adequate and we measured the level of risk. The results obtained were similar to those obtained with the Anderson-Darling test. Based on these results, as well as the AIC and loglikelihood, the best models for univariate returns were selected.

Table 5.1 Best model for univariate returns

Return	Distribution
USD/ZAR	Skew t
All shares	GHYP
Gold mining	HYP
S&P 500	GHYP

The joint distribution of these four returns was then analyzed. Firstly, the returns were transformed using affine transformations as described in Section 1.4. Then, each individual series resulting from the transformation is fitted with the univariate GHD in order to obtain the MAGHD. As a goodness of fit measure, the kernel smoothing was used. This is due to its flexibility in handling higher dimensional returns(especially in practical applications). Based on this goodness of fit test, the MANIG was selected as the best model for affinely transformed variables. It is also worth noting that the univariate A-D test was used in order to assess the goodness of fit of the transformed returns as these returns represent the model for MAGHD. These distributions proved adequate, except for the VG used in the All shares transformed returns.

Secondly, using the MGHD, the returns were fitted to all five subclasses of this distribution. Without any form of affine transformation carried out on these returns, the kernel density approach to multivariate goodness of fit showed that the MGHYP has the smallest statistic between the hypothesized and actual distribution. However, the AIC as a model selection criterion showed that the MNIG is the best model for these multivariate returns.

Furthermore, the bivariate MGHDs were also fitted and documented in the appendix. This is because the bivariate dependencies between these returns were also relevant in our study. To this end, Archimedean copulas were used. Interestingly, it was found that there is a very small negative dependency between the USD/ZAR returns and other three returns. This is however unusual as one will expect the USD/ZAR returns to be correlated to the other returns. Some degree of positive correlation was also observed between the All shares and other two returns (gold mining and S&P 500). Only in these two last cases were the upper as well as lower tail dependencies calculated with the Gumbel and Clayton copulas respectively.

5.2 Recommendations

Once these multivariate returns are modeled, it is important to look at some of the practical applications of these multivariate models.

- These models can be used to simulate future returns.
- They can also serve in the construction of efficiency frontier (Fajardo & Farias, 2009). That is, finding the model that represents the best possible combination of returns to produce the maximum returns for a given level of risk.
- In risk analyzis, they also are use for value at risk purposes. Finally, However, this last application has not been done yet in any research using MGHDs. Hence, can be a good reference for future research.
- Furthermore, analysis of multivariate models usually involves analysis of linear combinations of the returns under investigation. This is done in order to predict or observe co-movements between returns (Lettau & Ludvigson, 2001).

Bibliography

Aas, K. (2004). Modelling the dependence structure of financial assets: A survey of four copulas. *Norwegian Computing Center*.

Aas, K. & Haff, I. H. (2005). Nig and skew student t: Two special cases of the generalised hyperbolic distribution. *Norwegian Computing Center*.

Aas, K. & Haff, I. H. (2006). The generalised hyperbolic skew student's t-distribution. *journal of Financial Econometrics*, 4(2), 275–309.

Abramowitz, M. & stegun, I. K. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables. *Tenth edition. National Bureau of Standards Applied Mathematical Series*, (55).

Bagnold, R. (1954). *The Physics of Blown Sand and Desert Dunes*. Mathuen, London.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674), 401–419.

Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modeling. *Scandinavian Journal of Statistics*, 24(1), 1–13.

Barndorff-Nielsen, O. & Blæsild, P. (1981). *Statistical Distributions in Scientific Work*, volume 4, chapter Hyperbolic Distributions and Ramifications: Contributions to Theory and Application, (pp. 19Ũ44). Reidel.

Barndorff-Nielsen, O., Blæsild, P., & Seshadri, V. (1992). Multivariate distributions with generalised inverse gaussian distributions and associated poisson mixtures. *The Canadian Jornal of Statistics*, 20(2).

Barndorff-Nielsen, O. E. & Stelzer, R. (2005). Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse gaussian lévy processes. *Scandinavian Journal of Statistics*, 32(4), 617Ũ637.

Beling, P., Overstreet, G., & Rajaratnam, K. (2010). Estimation error in regulatory capital requirements: Theoretical implications for consumer bank profitability. *The Journal of the Operational Research Society*, 61(3), 381–392.

Bibby, B. & Sorensen, M. (1997). A hyperbolic diffusion model for stock prices. *Finance and Stochastics*, 1, 25–41.

Blæsild, P. & Jensen, J. L. (1981). Multivariate distributions of hyperbolic type in: C. taillie, g. patil, and b. baldessari (eds. *Statistical Distributions in Scientific Work*, 4.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). Copulas for finance. a reading guide and some applications. *Groupe de Recherche Opérationnelle, Crédit Lyonnais Paris*.

Breymann, W. & Lüthi, D. (2013). ghyp: A package on generalised hyperbolic distributions.

Campbell, J. Y. & Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. In S. F. Olivier Jean Blanchard (Ed.), *NBER Macroeconomics Annual 1991*, volume 6 (pp. 141 – 220).: PRINCETON UNIVERSITY AND NBER/PRINCETON UNIVERSITY AND CRDE MIT Press.

Cavanaugh, J. E. (2012). Lecture v: The bayesian information criterion. Department of Statistics and Actuarial Science, The University of Iowa.

Chacón, J. & Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *TEST*, 19, 375Ű398.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Qualitative Finance*, 1, 223–236.

Duong, T., Goud, B., & Schauer, K. (2012). Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceeding of the National Academy of Sciences of the United States of America*, 109(22), 8382–8387.

Eberlein, E. & Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1(3), 281–299.

Embrechts, P., McNeil, A., & Straumann, D. (1999). Correlation: Pitfalls and alternatives. risks. *preprint, ETH Zürich*.

Fajardo, J. & Farias, A. (2009). Multivariate affine generalized hyperbolic distributions: An empirical investigation. *International Review of Financial Analysis*, 18, 174–184.

Fajardo, J. & Farias, A. (2010). Derivative pricing using multivariate affine generalized hyperbolic distributions. *Journal of Banking and Finance*, 34, 1607–1617.

Fajardo, J., Farias, A., & Haas, J. R. (2005). Analysing the use of generalised hyperbolic distributions to value at risk calculations. *Brazilian Journal of Applied Economics*, 18, 174–184.

Fama, E. (1965). The behaviour of stock-market prices. *Journal of Business, University of Chicago Press*, 38(1), 34–105.

Fasano, G. & Franceschini, A. (1987). A multidimensional version of the kolmogorov-smirnov test. *Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711)*, 225, 155–170.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley and Sons, 2nd edition.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Universite de Lyon, Science, serie*, (14), 53Ũ77.

Frees, E. & Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1), 1–25.

Genest, C. & Mackay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American statistician*, (40), 280–283.

Hu, W. (2005). *Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk*. PhD thesis, Florida State University.

Hu, W. & Kercheval, A. (2007). Risk management with generalised hyperbolic dsitributions. *Proceedings of the Fourth IASTED International Conference on Financial Engineering and Applications*.

Hwang, J.-N., , Lay, S.-R., & Lippman, A. (1994). Nonparametric multivariate density estimation: A comparative study. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 42(10).

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability, 73.

Joe, H. & James, J. X. (1996). The estimation method of inference functions for margins for multivariate models. *Department of Statistics, University of British Columbia, Technical Report*, (166).

Jorion, P. (2006). *Value at risk: The new benchmark for managing financial risk*. McGraw-Hill, 3rd edition.

Jowaheer, V. & Ameerudden, N. Z. B. (2012). Modelling the dependence structure of mur/us dollar and mur/inr exchange rates using copula. *International Journal of Economics and Financial Issues*, 2(1).

Konlack, V. S. & Wilcox, D. (2014). A comparison of generalized hyperbolic distribution models for equity returns. *Journal of Applied Mathematics*, vol. 2014.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84.

Lettau, M. & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The journal of finance*, 56, 815–850.

Madan, D. B. & Seneta, E. (1990). The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63(4), 511Ű524.

Maddala, G. & Kim, I. (1998). *Unit Roots Cointegration and Structural Change*. Cambridge University Press.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394–419.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Mohadeva, L. & Robinson, P. (2004). *Unit Root Testing to Help Model Building*. Central Banking, Center for Central Banking Studies, Bank of England, London, EC2R8AH.

Necula, C. (2009). Modeling heavy-tailed stock index returns using the generalised hyperbolic distribution. *Romanian Journal of Economic Forcasting*, 2.

Nelsen, R. B. (1992). On measures of association as measures of positive dependence. *Statistics and Probability Letters*, 14, 269–274.

Nelsen, R. B. (2003). Properties and applications of copulas: A brief survey. In J. Dhaene, N. Kolev, & P. Morettin (Eds.), *In Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance* (pp. 10–28).: University Press USP:São Paulo.

Peacock, J. A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202, 615Ű627.

Prause, K. (1999). The generalised hyperbolic model: Estimation, financial derivatives and risk measures. *University of Freiburg*.

Protassov, R. S. (2003). Em- based maximun likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing*, 14(67-77).

Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engeneering*. Springer.

Schmidt, R., Hrycej, T., & Stützle, E. (2006). Multivariate distribution models with generalized hyperbolic margins. *Computational Statistics and Data Analysis*, 50, 2065–2096.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley and Sons INC.

Umberto Cherubini, Elisa Luciano, W. V. (2004). *Copula Methods in Finance*. John Wiley & Sons.

Unnikrishnan, J., Meyn, S., & Veeravalli, V. (2010). On thresholds for robust goodness-of-fit tests. *Presented in IEEE Information Theory Workshop, Dublin*.

Appendix

Analyzing the fit of the USD/ZAR and All shares returns

For the MGHYP:

$$\lambda = 0.8459007, \quad \bar{\alpha} = 1.0147183,$$
$$\mu = (-0.0003292698, 0.0008205440),$$
$$\gamma = (0.0004861215, -0.0003476884)$$
$$\Sigma = \begin{pmatrix} 1.222358e-04 & -1.593618e-05 \\ -1.593618e-05 & 1.465895e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 1.5, \quad \bar{\alpha} = 0.9291942,$$
$$\mu = (-0.0003534590, 0.0008336916)$$
$$\gamma = (0.0005121962, -0.0003620885)$$
$$\Sigma = \begin{pmatrix} 1.206342e-04 & -1.581632e-05 \\ 1.206342e-04 & -1.581632e-05 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9972694,$$
$$\mu = (-0.0003554551, 0.0008373171)$$
$$\gamma = (0.0005077875, -0.0003612092)$$
$$\Sigma = \begin{pmatrix} 0.0001254576 & -0.0000160079 \\ -0.0000160079 & 0.0001504665 \end{pmatrix}$$

For the MVG:

$$\lambda = 1.126088, \quad \bar{\alpha} = 0.9291942,$$
$$\mu = (-0.0003534590, 0.0008336916)$$
$$\gamma = (0.0005121962, -0.0003620885)$$
$$\Sigma = \begin{pmatrix} 1.206342e-04 & -1.581632e-05 \\ -1.581632e-05 & 1.447984e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.779561, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.559121,$$
$$\mu = (-0.0003408462, 0.0008235871)$$
$$\gamma = (0.0004880818 - 0.0003438530$$
$$\Sigma = \begin{pmatrix} 1.366124e-04 - 1.691123e-05 \\ 1.366124e-04 - 1.691123e-05 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (0.0001471623, 0.0004797857)$$
$$\Sigma = \begin{pmatrix} 1.29592e-04 & -0.0000213956 \\ -2.13956e-05 & 0.0001513398 \end{pmatrix}$$

Table 5.2 *Goodness of fit of bivariate GHDs for USD/ZAR and All shares returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	1.412004	23554.66	-47091.33
MHYP	-0.003910404	23538.93	-47061.85
MNIG	-0.2786012	23570.56	-47125.11
MVG	0.6072332	23538.93	-47061.85
MST	1.432617	23565.65	-47115.29
Multivariate Normal	16.67545	22993.53	-45977.06

Analyzing the fit of the USD/ZAR and gold mining returns

For the MGHYP:

$$\lambda = 0.8512479, \quad \bar{\alpha} = 1.0005726,$$
$$\mu = (-0.0002527845, -0.0003474230),$$
$$\gamma = (0.0004085910.000474969)$$
$$\Sigma = \begin{pmatrix} 1.224338e-04 & -4.726820e-06 \\ -4.726820e-06 & 5.927193e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 1.5, \quad \bar{\alpha} = 0.9126469,$$
$$\mu = (-0.0002733085, -0.0003688009)$$
$$\gamma = (0.0004306993, 0.0004981282)$$
$$\Sigma = \begin{pmatrix} 1.208487e-04 & -4.628216e-06 \\ -4.628216e-06 & 5.857365e-04 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9901852,$$
$$\mu = (-0.0002804200, -0.0003655655)$$
$$\gamma = (0.0004322565, 0.0004883437)$$
$$\Sigma = \begin{pmatrix} 1.256373e-04 & -4.881202e-06 \\ -4.881202e-06 & 6.079826e-04 \end{pmatrix}$$

For the MVG:

$$\lambda = 1.093149, \quad \bar{\alpha} = 0.0000,$$
$$\mu = (-0.0003534590, 0.0008336916)$$
$$\gamma = (0.0005121962, -0.0003620885)$$
$$\Sigma = \begin{pmatrix} 1.206342e-04 & -1.581632e-05 \\ -1.581632e-05 & 1.447984e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.775367, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.550734,$$
$$\mu = (-0.0002704096, -0.0003534263)$$
$$\gamma = (0.0004176346, 0.0004709946)$$
$$\Sigma = \begin{pmatrix} 1.367810e-04 & -5.306861e-06 \\ -5.306861e-06 & 6.623423e-04 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (0.0001471623, 0.0001174975)$$
$$\Sigma = \begin{pmatrix} 1.295920e-04 & -5.754663e-06 \\ -5.754663e-06 & 6.114319e-04 \end{pmatrix}$$

Table 5.3 *Goodness of fit of bivariate GHDs for USD/ZAR and gold mining returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	1.180364	20871.34	-41724.69
MHYP	3.524017	20853.99	-41691.98
MNIG	0.6535838	20886.84	-41757.68
MVG	1.302692	20855.3	-41694.61
MST	0.8819064	20879.86	-41743.72
Multivariate Normal	18.36181	18.36181	-40580.21

Analyzing the fit of the USD/ZAR and S&P 500 returns

For the MGHYP:

$$\lambda = 0.7039140, \quad \bar{\alpha} = 0.7039140,$$
$$\mu = (-0.0003392368, 0.0005862007),$$
$$\gamma = (0.0005050174, -0.0005217019)$$
$$\Sigma = \begin{pmatrix} 1.231128e-04 & -2.865619e-05 \\ -2.865619e-05 & 1.453043e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 1.5, \quad \bar{\alpha} = 0.8627276,$$
$$\mu = (-0.0003617394, 0.0005991104)$$
$$\gamma = (0.0005289171, -0.0005356483)$$
$$\Sigma = \begin{pmatrix} 1.212781e-04 & -2.859505e-05 \\ -2.859505e-05 & 1.440560e-04 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9527169,$$
$$\mu = (-0.0003641028, 0.0006014738)$$
$$\gamma = (0.0005251991, -0.0005318520)$$
$$\Sigma = \begin{pmatrix} 0.0001263767 & -0.0000285861 \\ -0.0000285861 & 0.0001473452 \end{pmatrix}$$

For the MVG:

$$\lambda = 1.064079, \quad \bar{\alpha} = 0.0000,$$
$$\mu = (6.774818e-05, 1.508599e-04)$$
$$\gamma = (8.034915e-05, -6.791799e-05)$$
$$\Sigma = \begin{pmatrix} 1.326392e-04 & -3.039094e-05 \\ -3.039094e-05 & 1.538599e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.753040, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.506081,$$
$$\mu = (-0.0003423536, 0.0005655998)$$
$$\gamma = (0.0004895895, -0.0004819400)$$
$$\Sigma = \begin{pmatrix} 1.374985e-04 & -3.009278e-05 \\ -3.009278e-05 & 1.582403e-04 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (1.471623e-04, 8.373226e-05)$$
$$\Sigma = \begin{pmatrix} 1.295920e-04 & -3.935084e-05 \\ -3.935084e-05 & 1.669168e-04 \end{pmatrix}$$

Table 5.4 *Goodness of fit of bivariate GHDs for USD/ZAR and S&P 500 returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	4.144898	23721.12	-47424.25
MHYP	6.864506	23682.9	-47349.8
MNIG	3.184267	23757.86	-47499.72
MVG	4.336816	23710.79	-47405.59
MST	3.6726	23776.19	-47536.38
Multivariate Normal	21.68993	22903.6	-45797.2

Analyzing the fit of the All shares and gold mining returns

For the MGHYP:

$$\lambda = 0.8336270, \quad \bar{\alpha} = 0.9751478,$$
$$\mu = (0.0008425585, -0.0004491674),$$
$$\gamma = (-0.0003711640, 0.0005797724)$$
$$\Sigma = \begin{pmatrix} 0.0001449808 & 0.0001127419 \\ 0.0001127419 & 0.0005843908 \end{pmatrix}$$

For the MHYP:

$$\lambda = 1.5, \quad \bar{\alpha} = 0.8732382,$$
$$\mu = (0.0008526149, -0.0004684544)$$
$$\gamma = (-0.0003824607, 0.0006010891)$$
$$\Sigma = \begin{pmatrix} 0.0001432612 & 0.0001111331 \\ 0.0001111331 & 0.0005777306 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9699907,$$
$$\mu = (0.0008706557, -0.0004654669)$$
$$\gamma = (-0.0003949053, 0.0005889830)$$
$$\Sigma = \begin{pmatrix} 0.0001484233 & 0.0001162599 \\ 0.0001162599 & 0.0005987365 \end{pmatrix}$$

For the MVG:

$$\lambda = 0.9583156, \quad \bar{\alpha} = 0.0000,$$
$$\mu = (2.196377e - 04, -5.776083e - 05)$$
$$\gamma = (0.0002592507, 0.0001746538)$$
$$\Sigma = \begin{pmatrix} 0.0001534141 & 0.0001181439 \\ 0.0001181439 & 0.0006120017 \end{pmatrix}$$

For the MST:

$$\lambda = -1.764637, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.529275,$$
$$\mu = (0.0008569230, -0.0004384037)$$
$$\gamma = (-0.0003771939, 0.0005559848)$$
$$\Sigma = \begin{pmatrix} 0.0001612016 & 0.0001272048 \\ 0.0001272048 & 0.0006510066 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (0.0004797857, 0.0001174975)$$
$$\Sigma = \begin{pmatrix} 0.0001513398 & 0.0001124750 \\ 0.0001124750 & 0.0006114319 \end{pmatrix}$$

Table 5.5 *Goodness of fit of bivariate GHDs for All shares and gold mining returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	10.8543	20889.35	-41760.7
MHYP	13.11752	20865.41	-41714.83
MNIG	10.34514	20908.2	-41800.4
MVG	7.53518	20989.35	-41962.7
MST	8.081153	20899.23	-41782.46
Multivariate Normal	29.30679	20278.86	-40547.72

Analyzing the fit of the All shares and S&P 500 returns

For the MGHYP:

$$\lambda = 0.7005579, \quad \bar{\alpha} = 0.9771152,$$
$$\mu = (0.0009848865, 0.0005280862),$$
$$\gamma = (0.0009848865, 0.0005280862)$$
$$\Sigma = \begin{pmatrix} 1.460388e-04 & 4.800569e-05 \\ 4.800569e-05 & 1.443437e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 0.7005579, \quad \bar{\alpha} = 0.9771152,$$
$$\mu = (0.0009848865, 0.0005280862)$$
$$\gamma = (-0.0005247864, -0.0004616719)$$
$$\Sigma = \begin{pmatrix} 1.460388e-04 & 4.800569e-05 \\ 4.800569e-05 & 1.443437e-04 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9392296,$$
$$\mu = (0.0010119386, 0.0005301148)$$
$$\gamma = (-0.0005459903, -0.0004579897)$$
$$\Sigma = \begin{pmatrix} 0.0001495728 & 0.0000486828 \\ 0.0000486828 & 0.0001461658 \end{pmatrix}$$

For the MVG:

$$\lambda = 1.021717, \quad \bar{\alpha} = 0.0000,$$
$$\mu = (0.0004133312, 0.0001922438)$$
$$\gamma = (6.753222e-05, -1.102713e-04)$$
$$\Sigma = \begin{pmatrix} 1.571832e-04 & 5.099958e-05 \\ 5.099958e-05 & 1.530909e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.745295, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.49059,$$
$$\mu = (0.0009843794, 0.0004937140)$$
$$\gamma = (-0.0005046695, -0.0004100434)$$
$$\Sigma = \begin{pmatrix} 1.624563e-04 & 5.232955e-05 \\ 5.232955e-05 & 1.566982e-04 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (4.797857e-04, 8.373226e-05)$$
$$\Sigma = \begin{pmatrix} 1.513398e-04 & 5.451118e-05 \\ 5.451118e-05 & 1.669168e-04 \end{pmatrix}$$

Table 5.6 *Goodness of fit of bivariate GHDs for All shares and S&P* 500 *returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	5.701713	23539.76	-47061.51
MHYP	7.804611	23539.76	-47061.51
MNIG	4.858352	23577.28	-47138.56
MVG	3.422572	23570.28	-47124.55
MST	3.347802	23591.75	-47167.5
Multivariate Normal	23.96084	22705.38	-45400.76

Analyzing the fit of the gold mining and S&P 500 returns

For the MGHYP:

$$\lambda = 0.7403650, \quad \bar{\alpha} = 0.9633387,$$
$$\mu = (-0.0002668656, 0.0005543552),$$
$$\gamma = (0.0003983104, -0.0004877003)$$
$$\Sigma = \begin{pmatrix} 5.990734e-04 & -1.158626e-06 \\ -1.158626e-06 & 1.445869e-04 \end{pmatrix}$$

For the MHYP:

$$\lambda = 1.5, \quad \bar{\alpha} = 0.8274115,$$
$$\mu = (-0.0002838368, 0.0005661987)$$
$$\gamma = (0.0004164772, -0.0005006705)$$
$$\Sigma = \begin{pmatrix} 5.909923e-04 & -1.423322e-06 \\ -1.423322e-06 & 1.434485e-04 \end{pmatrix}$$

For the MNIG:

$$\lambda = -0.5, \quad \bar{\alpha} = 0.9395966,$$
$$\mu = (-0.0002895067, 0.0005663870)$$
$$\gamma = (0.0004162768, -0.0004936507)$$
$$\Sigma = \begin{pmatrix} 6.145014e-04 & -7.480313e-07 \\ -7.480313e-07 & 1.463384e-04 \end{pmatrix}$$

For the MVG:

$$\lambda = 0.96772, \quad \bar{\alpha} = 0.0000,$$
$$\mu = (-4.281344e-06, 3.483396e-05)$$
$$\gamma = (1.235734e-04, 4.961888e-05)$$
$$\Sigma = \begin{pmatrix} -3.155260e-08 & 1.539208e-04 \\ -3.155260e-08 & 1.539208e-04 \end{pmatrix}$$

For the MST:

$$\lambda = -1.746636, \quad \bar{\alpha} = 0.0000, \quad \nu = 3.493273,$$
$$\mu = (-0.0002791661, 0.0005361986)$$
$$\gamma = (0.0003967233, -0.0004525343)$$
$$\Sigma = \begin{pmatrix} 6.693127e-04 & -4.058751e-07 \\ -4.058751e-07 & 1.568585e-04 \end{pmatrix}$$

For the Multivariate Normal:

$$\mu = (1.174975e-04, 8.373226e-05)$$
$$\Sigma = \begin{pmatrix} 6.114319e-04 & -3.025136e-06 \\ -3.025136e-06 & 1.669168e-04 \end{pmatrix}$$

Table 5.7 *Goodness of fit of bivariate GHDs for gold mining and S&P 500 returns*

Multivariate distribution	Kernel distance	Loglikelihood	AIC
MGHYP	5.254211	20613.19	-41208.38
MHYP	5.8566	20573.33	-41130.66
MNIG	5.543554	20644.96	-41273.92
MVG	4.643629	20691.58	-41367.16
MST	4.346161	20650.55	-41285.09
multivariate Normal	26.04928	19813.33	-39616.66

Bessel function

In this section, we present the Bessel function of the third kind, which forms and integral part of generalized hyperbolic functions.

Let $\lambda \in \mathbb{R}$, *the modified Bessel function of the third kind* with index λ is defined by the equation (Abramowitz & stegun, 1972)

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty u^{\lambda-1}e^{-\frac{1}{2}x(u+u^{-1})}du \quad x > 0.$$

Using the function above, the following results follow (Barndorff-Nielsen & Blæsild (1981) and Abramowitz & stegun (1972)):

$$\begin{aligned} K_{-\lambda}(x) &= K_\lambda(x), \\ K_{\lambda+1}(x) &= \frac{2\lambda}{x}K_\lambda(x) + K_{\lambda-1}(x), \\ K'_\lambda(x) &= -\frac{\lambda}{x}K_\lambda(x) - K_{\lambda-1}(x) \end{aligned}$$

Indeed, for the first one, we have

$$\begin{aligned} K_{-\lambda}(x) &= \frac{1}{2}\int_0^\infty u^{-\lambda-1}e^{-\frac{1}{2}x(u+u^{-1})}du, \\ &= -\frac{1}{2}\int_\infty^0 v^{\lambda+1}e^{-\frac{1}{2}x(v^{-1}+v)}v^{-2}dv, \quad (v = \frac{1}{u}, \quad dv = -\frac{1}{u^2}du), \\ &= \frac{1}{2}\int_0^\infty v^{\lambda-1}e^{-\frac{1}{2}x(v+v^{-1})}dv, \\ &= K_\lambda(x). \end{aligned}$$

Secondly, we obtain

$$\begin{aligned}K_{\lambda+1}(x)-K_{\lambda-1}(x) &= \frac{1}{2}\int_0^\infty (u^\lambda - u^{\lambda-2})e^{-\frac{1}{2}x(u+u^{-1})}du \\ &= \frac{1}{2}\int_0^\infty u^\lambda(1-u^{-2})e^{-\frac{1}{2}x(u+u^{-1})}du.\end{aligned}$$

Thus using integration by parts and letting

$$\begin{cases} v_1 = u^\lambda, & \Rightarrow v_1' = \lambda u^{\lambda-1} \\ v_2' = (1-u^{-2})e^{-\frac{1}{2}x(u+u^{-1})}, & \Rightarrow v_2 = -\frac{2}{x}e^{-\frac{1}{2}x(u+u^{-1})}, \end{cases} \tag{5.1}$$

we obtain

$$\begin{aligned}K_{\lambda+1}(x)-K_{\lambda-1}(x) &= \frac{1}{2}\left(\left[-\frac{2u^\lambda}{x}e^{-\frac{1}{2}x(u+u^{-1})}\right]_0^\infty + \frac{2\lambda}{x}\int_0^\infty u^{\lambda-1}e^{-\frac{1}{2}x(u+u^{-1})}du\right), \\ &= \frac{2\lambda}{x}\left(\frac{1}{2}\int_0^\infty u^{\lambda-1}e^{-\frac{1}{2}x(u+u^{-1})}du\right), \\ &= \frac{2\lambda}{x}K_\lambda(x).\end{aligned}$$

Lastly, we have

$$\begin{aligned}K_\lambda(x)' &= -\frac{1}{2}\int_0^\infty u^{\lambda-1}\frac{1}{2}(u+u^{-1})e^{-\frac{1}{2}x(u+u^{-1})}du, \\ &= -\frac{1}{4}\int_0^\infty u^\lambda(1+u^{-2})e^{-\frac{1}{2}x(u+u^{-1})}du, \\ &= -\frac{1}{4}\int_0^\infty u^\lambda e^{-\frac{1}{2}x(u+u^{-1})}du - \frac{1}{4}\int_0^\infty u^{\lambda-2}e^{-\frac{1}{2}x(u+u^{-1})}du \\ &= -\frac{1}{2}K_{\lambda+1}(x) - \frac{1}{2}K_{\lambda-1}(x) \\ &= -\frac{1}{2}\left(2\frac{\lambda}{x}K_\lambda(x)+K_{\lambda-1}(x)\right) - \frac{1}{2}K_{\lambda-1}(x) \quad \text{from above,} \\ &= -\frac{\lambda}{x}K_\lambda(x) - K_{\lambda-1}(x).\end{aligned}$$

It should also be noted that when $\lambda = n+\frac{1}{2}, n = 0, 1, 2, \ldots$, we have that (Barndorff-Nielsen & Blæsild, 1981)

$$K_{n+\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}e^{-x}\left\{1+\sum_{i=1}^n \frac{(n+i)!}{(n-i)!i!}(2x)^{-i}\right\}.$$

It should also be noted that if $\lambda > 0$, then for very small values of it's argument ($x \downarrow 0$), this function can be approximated by

$$K_\lambda \sim \Gamma(\lambda)2^{\lambda-1}x^{-\lambda},$$

and for large values of the arguments, we have

$$K_\lambda = \sqrt{\frac{\pi}{2x}} e^{-x} \left(1 + \frac{4\lambda^2 - 1}{8x} + \frac{(4\lambda^2-1)(4\lambda^2-9)}{2!(8x)^2} + \frac{(4\lambda^2-1)(4\lambda^2-9)(4\lambda^2-25)}{3!(8x)^3} + \dots\right)$$

Likelihood equations

In this appendix, we present the log likelihood equations of the GHD. AS we know, the pdf of the GHD is given by

$$gh(x;\lambda,\alpha,\beta,\delta,\mu) = a_\lambda(\delta^2 + (x-\mu)^2)^{(\lambda-1/2)/2} K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2}) \times \exp\beta(x-\mu),$$

where $a_\lambda = a(\lambda;\alpha,\beta,\delta) = \frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}, \quad x,\mu \in \mathbb{R}$, and K_λ is the modified Bessel function of the third kind. According to Prause (1999), the log likelihood of this distribution is given by

$$L = n\log(a_\lambda) + \left(\frac{\lambda}{2} - \frac{1}{4}\right)\sum_{i=1}^{n}\log(\delta^2 + (x_i-\mu)^2) + \sum_{i=1}^{n}\log K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x_i-\mu)^2}) + \sum_{i=1}^{n}\beta(x_i-\mu).$$

It should be noted that we have considered the general case. Particular cases for the HYP NIG and VG can be obtained similarly, by setting $\lambda = 1, \lambda = -1/2$ and $\delta = 0$ respectively. Hence, differentiating this equation with respect to the different parameters of the GHD, we obtain the following equations (Prause, 1999)

$$
\begin{aligned}
\frac{\partial L}{\partial \lambda} &= n\left[\frac{1}{2}\ln\frac{\alpha^2-\beta^2}{\alpha\delta} - \frac{k_\lambda(\delta\sqrt{\alpha^2-\beta^2})}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}\right] \\
&\quad + \sum_{i=1}^{n}\left[\frac{1}{2}\ln(\delta^2+(x_i-\mu)^2) + \frac{k_{\lambda-1/2(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}}{K_{\lambda-1/2(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}}\right], \\
\frac{\partial L}{\partial \alpha} &= n\frac{\delta\alpha}{\sqrt{\alpha^2-\beta^2}}R_\lambda(\delta\sqrt{\alpha^2-\beta^2}) \\
&\quad - \sum_{i=1}^{n}\sqrt{\delta^2+(x_i-\mu)^2}R_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x_i-\mu)^2}) \\
\frac{\partial L}{\partial \beta} &= -n\left[\frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}R_\lambda(\delta\sqrt{\alpha^2-\beta^2})\mu\right] + \sum_{i=1}^{n}X_i, \\
\frac{\partial L}{\partial \delta} &= n\left[-2\frac{\lambda}{\delta} + \sqrt{\alpha^2-\beta^2 R_\lambda(\delta\sqrt{\alpha^2-\beta^2})}\right] \\
&\quad + \sum_{i=1}^{n}\left[\frac{(2\lambda-1)\delta}{\delta^2+(x_i-\mu)^2} - \frac{\alpha\delta R_\lambda(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}{\sqrt{\delta^2+(x_i-\mu)^2}}\right] \\
\frac{\partial L}{\partial \mu} &= -n\beta + \sum_{i=1}^{n}\frac{x_i-\mu}{\sqrt{\delta^2+(\mu-x_i)^2}}\left[\frac{2\lambda-1}{\sqrt{\delta^2+(\mu-x_i)^2}}\right. \\
&\quad \left. - \alpha R_{\lambda-1/2}(\alpha\sqrt{\delta^2+(\mu-x_i)^2})\right].
\end{aligned}
$$

where we have used the following properties of the Bessel function (Barndorff-Nielsen & Blæsild, 1981);

$$
\begin{aligned}
\frac{dK_\lambda}{d\lambda} &= k_\lambda, \\
R_\lambda &= \frac{K_{\lambda+1}(x)}{K_\lambda(x)}, \\
S_\lambda &= \frac{K_{\lambda+2}(x)K_\lambda(x) - K^2_{\lambda+1}(x)}{K^2_\lambda(x)}, \qquad x > 0.
\end{aligned}
$$

Thus setting the above equations to zero and solving, we obtain the estimates of the parameters.

Tail behavior of GHDs

In this section, we will give details about the functions that describe the tail behavior of the GHDs. In general, the equation governing the tail behavior of the GHD is given by

$$f(x) \sim const|x|^{\lambda-1}\exp((\mp\alpha+\beta)x), \qquad x \to \pm\infty. \tag{5.2}$$

However, we recall from *Section* 1.1 that the tail behavior of the NIG distribution is given by

$$f(x) \sim const|x|^{-3/2}\exp(-\alpha|x|+\beta x), \quad \text{for } x \to \pm\infty.$$

More precisely, this means that as x changes between $\pm\infty$, the tails behave as follows; the heaviest tails of the distribution decays according to (Aas & Haff, 2005)

$$f(x) \sim const|x|^{-3/2}\exp(-\alpha|x|+|\beta||x|), \qquad \text{as} \begin{cases} \beta < 0, & \text{and } x \to -\infty; \\ \beta > 0, & \text{and } x \to +\infty. \end{cases}$$

while the lightest tails decay according to

$$f(x) \sim const|x|^{-3/2}\exp(-\alpha|x|-|\beta||x|), \qquad \text{as} \begin{cases} \beta < 0, & \text{and } x \to +\infty; \\ \beta > 0, & \text{and } x \to -\infty. \end{cases}$$

Thus, given that the tails of our data are semi-heavy, we see that the heaviest and lightest tails behave differently (Aas & Haff, 2005).

Similarly, for the GH student t distribution, we had that the tail decay according to the equation

$$f(x) \sim const|x|^{-\nu/2-1}\exp(-|\beta||x|+\beta x).$$

More precisely, the heaviest tails decay according to the equation

$$f(x) \sim const|x|^{-\nu/2-1}, \qquad \text{as} \begin{cases} \beta < 0, & \text{and } x \to -\infty; \\ \beta > 0, & \text{and } x \to +\infty. \end{cases}$$

while the lightest tails decay according to

$$f(x) \sim const|x|^{-\nu/2-1}\exp(-2|\beta||x|) \qquad \text{as} \begin{cases} \beta < 0, & \text{and } x \to +\infty; \\ \beta > 0, & \text{and } x \to -\infty. \end{cases}$$

From this behavior comes the property that the GH skew student t distribution has one heavy and one semi heavy tail, and it is the only subclass of the GHDs with this property.