Long Range Dependence in South African Platinum Prices under Heavy Tailed Error Distributions

By

Sihle Kubheka

submitted in accordance with the requirements for
the degree of

MASTER OF SCIENCE

In the subject

STATISTICS

at the

University of South Africa

Supervisor : Dr. E. Ranganai

November, 2016

Student Number: 3630-6886

I declare that **Long Range Dependence in South African Platinum Prices under Heavy Tailed Error Distributions** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

----------	----------
Signature	Date

Acknowledgements

I would like to thank my supervisor Dr E. Ranganai for helping me in this thesis and further pushing me in doing more what is required. I also thank my family for giving me time and support.

Abstract

South Africa is rich in platinum group metals (PGMs) and these metals are important in providing jobs as well as investments some of which have been seen in the Johannesburg Securities Exchange (JSE). In this country this sector has experienced some setbacks in recent times. The most notable ones are the 2008/2009 global financial crisis and the 2012 major nationwide labour unrest. Worrisomely, these setbacks keep simmering. These events usually introduce jumps and breaks in data which changes the structure of the underlying information thereby inducing spurious long memory (long range dependence). Thus it is recommended that these two phenomena must be addressed together. Further, it is well-known that financial returns are dominated by stylized facts. In this thesis we carried out an investigation on distributional properties of platinum returns, structural changes, long memory and stylized facts in platinum returns and volatility series. To understand the distributional properties of the returns, we used two classes of heavy tailed distributions namely the alpha-Stable distributions and generalized hyperbolic distributions. We then investigated structural changes in the platinum return series and changes in long range dependence and volatility. Using Akaike information criterion, the ARFIMA-FIAPARCH under the Student distribution was selected as the best model for platinum although the ARCH effects were slightly significant, while using the Schwarz information criteria the ARFIMA-FIAPARCH under the Normal distribution. Further, ARFIMA-FIEGARCH under the skewed Student distribution and ARFIMA-HYGARCH under the Normal distribution models were able to capture the ARCH effects. The best models with respect to prediction excluded the ARFIMA-FIGARCH model and were dominated by ARFIMA-FIAPARCH model with non-Normal error distributions which indicates the importance of asymmetry and heavy tailed error distributions.

Contents

List of Figures

List of Tables

List of Abbreviations

ACF Autocorrelation Function

AIC Akaike Information Criteria

AR Autoregressive

ARCH Autoregressive Conditional Heteroskedasticity

ARFIMA Autoregressive Fractionally Integrated Moving Average

ARIMA Autoregressive Integrated Moving Average

ARMA Autoregressive Moving Average

BIC Bayesian Information Criteria

DFT Disrete Fourier Transform

EGARCH Exponential Generalized Autoregressive Conditional Heteroskedasiticy

ETF Exchange Traded Funds

FIAPARCH Fractional Integrated Power Autoregressive Conditional Heteroskedasticity

FIEGARCH Fractional Integrated Exponential generalized autoregressive conditional heteroskedasticity

FIGARCH Fractional Integrated Generalized Autoregressive Conditional Heteroskedasticity

GARCH Generalized Autoregressive Conditional Heteroskedasticity

GDP Gross Domestic Product

GED Generalised Error Distribution

GPH Geweke Porter-Hudak

HYGARCH Hyperbolic Generalized Autoregressive Conditional Heteroskedasticity

JSE Johannesburg Securities Exchange

KPSS KwiatkowskiPhillipsSchmidtShin

LLH Log likelihood

LM Long Memory

LW Local Whittle

MAE Mean Absulute Error

MAE Moving Average

MSE Mean Square Error

MZ Mincer-Zarnowitz

NIG Normal Inverse Gaussian

PACF Partial Autocorrelation Function

PGM Platinum Group Metal

SBC Schwartz Bayesian Criteria

TIC Theil Inequality Constraint

List of Symbols

$\rho(k)$	Autocorrelation function at lag k.
$\gamma(k)$	Autocovariance function at lag k.
Φ_{kk}	Partial autocorrelation function.
$\bar{X}$	Sample Mean.
s^2	Sample variance.
μ	Mean.
σ^2	Variance.
L	Backshift lag operator.
$\theta(L)$	MA polynomial.
$\phi(L)$	AR polynomial.
ω_k	Fourier frequency.
$I(\omega_k)$	The periodogram.
$f(\omega_k)$	The spectrum.
d	Long range dependence parameter.
F_{t-1}	Information set of the past history to time $t-1$.
r_t	Log returns.
(R/S)	Rescaled Range.
H	Hurst coefficient.

Chapter 1

Introduction

As a result of economic growth in emerging economies such as the then BRIC (Brazil, Russia, India and China) and advanced economies, the South African mining sector has experienced a boom that started globally in 2001 up until 2008. Within the sector is the platinum group of metals (PGMs) (Matthey, 2013a), and which mainly extract platinum, palladium, rhodium, ruthenium, and iridium. From the later point in time, the sector has experienced major setbacks/challenges, namely the global financial crisis of unprecedented proportions in 2008 followed by the recurring labour unrest since the 2012 Marikana incident which resulted in lost productivity as a result of lost production during labour unrest. This occurred mainly in three major PGMs producers, which comprise Impala Platinum, Lonmin, and Anglo American Platinum, and the loss was estimated to be around 750 000 oz (Matthey, 2013a). Further, the declining demand of platinum as a result of lower market share of diesel cars in Europe which uses platinum in autocatalysts and the weak rand to dollar exchange exacerbated the situation.

Despite these setbacks and challenges faced by the industry, the minerals resources sector remains the largest employment sector. On average, from 2008 to 2013, the percentage contribution to the South African GDP from this sector was 2.3% with a yearly increase of 3.3% and an employment head count of 191 781 with R27 826 as the average employee salary (Matthey, 2013a). Due to the abundance of the PGMs, mining is the mainstay of the South African economy. South Africa holds more than 80% of platinum and 80% of the resources of the PGMs. It is rich in PGMs particularly platinum and palladium. These types of metals have attracted investments to the country. Also, it is noteworthy that South Africa is the largest producer of platinum and second largest

when it comes to palladium with Russia supplying more. Table 1.1 shows the demand and supply of platinum. The rise in demand is solely because of difficulty in finding a substitute for their main application in catalytic converters. Platinum is predominantly used in diesel cars autocatalysts and palladium used in gasoline cars. Europe being the only manufacturer of diesel autos, this suggests that the demand for platinum is mostly derived from European diesel auto sales.

Table 1.1: Platinum Supply and Demand

Supply	2011	2012	2013
South Africa	4 860	4 090	4 120
Russia	835	800	780
Others	790	760	840
Demand	**2011**	**2012**	**2013**
Autocatalyst	3 185	3 190	3 125
Jewellery	2 475	2 780	2 740
Industrial	1 975	1 605	1 790
Investment	460	455	765

PGMs in South Africa are not only important in providing jobs but investments in PGMs is also seen in the stock market's exchange traded funds (ETF). ETFs are funds that track an underlying asset like indexes, commodities, forex, equities and so on. The value of ETF's are derived from these underlying assets. Two palladium funds, Standard Bank's AfricaPalladium ETF and Absa Capital's newPaladium ETF were launched in March of 2014 in the JSE. These exchange traded funds are backed by the physical palladium metal. With such an environment of investments and economic contribution by PGMs, it is important to understand the distribution and volatility dynamics of the PGM prices. In this thesis, we will only analyze platinum prices and its volatility.

Challenges in PGMs sector have attracted a lot of interest from researchers to analyse the risk and volatility of these prices. A lot of researchers have studied related topics, this includes a paper by Huang et al. (2014) where they investigated the South African mining index and found that the returns of the South African Mining Index are not Guassian but explained by a heavy tailed distribution. Socgina and Wilcox (2014) compared the generalized hyperbolic distributions for equity returns on the JSE mining stocks and they found evidence of departure from the Guassian assumption, specifically, the long memory phenomenon has been thoroughly investigated in the PGMs. Areas where this volatility

phenomenon has been studied exist in the literature. Such papers include Holton (2004), Hammoudeh et al. (2011), Lane et al. (2012). Arouri et al. (2012b) investigated long memory spot futures prices of precious metals using GARCH type models and found that the FIGARCH model explains conditional volatility spot and futures prices better. Other papers include Arouri et al. (2012a) who investigated the potential of structural changes and long memory in return of precious metals and concluded that there is evidence of long memory in precious metals price using an ARFIMA-FIGARCH model. In analysis and modelling of financial returns, there are stylized facts that we must address. In the following section, we will discuss these phenomena further.

1.1 Stylized Facts of Returns

In econometrics and financial market studies, stylized facts refer to the presentation of some empirical findings during a study. These findings are then summarized. The main issue with such generalizations is that they may have inaccuracies and some spurious information. A simple way to explain stylized facts is taking the common denominator among properties observed in studies of different markets and instruments and generalizing it as a rule.

It is widely accepted that the distribution of financial returns are not normal. This is true for some returns structure, that is the reason each researcher needs to check this before analysis is done. The following is a list of stylized facts that are common for financial returns:

- Autocorrelations of asset returns are often insignificant.
- The distribution of returns have a Pareto-like tail with a finite tail index.
- As the time scale increases over which returns are calculated, their distributions look more and more like a normal distribution.
- Different measure of volatility display a positive autocorrelation over several days, which suggest that high volatility events tend to cluster in time, this is known as volatility clustering.
- Residual time series exhibit heavy tails, even after conditional heteroscedastic models corrections.

- The autocorrelation function of absolute returns decays slowly as a function of time lag. This suggests long-memory in data.
- Most measures of volatility of an asset are negatively correlated with the returns of that asset.

This is a summary of the stylized facts in literature. A detailed list can be found in Cont (2001).

1.2 Research Problem and Objectives

Price volatility of precious metals is still a major interest in financial economics (Arouri et al., 2012a). This is because this measure is used mostly in risk management and valuation of assets. The price data that we will use to study the evolution of the volatility process of platinum prices is from Matthey (2013b). The research seeks to develop a model which will identify and capture the evolution of the volatility of platinum prices in light of platinum prices.

In this study, we will use different long memory tests which will be able to detect whether the long memory is inherent in platinum prices and hence assist us in selecting the best suitable model which will explain the volatility evolution of the price data. The objectives of this study are as follows:

- Analyse the structure of platinum price returns,
- Perform different long memory tests in identifying tests that better explain the returns,
- Determine the best fit for return error distributions using heavy tailed distributions,
- Model price returns using different ARFIMA models to identify the best model,
- Compare other GARCH type models with FIGARCH type models for analysis of volatility.

1.3 Thesis Structure

In chapter 2, we discuss discrete time series process under the Box-Jenkins methodology. We first discuss stationary and then non stationary processes, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) and further variance stabilizing transformations. The chapter ends with a discussion of model selection methods.

In chapter 3, we discuss spectral analysis theory methods that are used in long memory models. These methods are in the frequency domain. The chapter starts by discussing Fourier analysis as it forms the basis of spectral analysis. Then this theory is used in the discussion of the spectrum which is similar to the ACF when working in time domain. The periodogram, which estimates the spectrum, is then elaborated on with its properties and methods proposed in literature to smooth it as it is an inconsistent estimator of the spectrum. Theory and methods of long memory are discussed in chapter 4. We start the chapter by discussing self-similar processes, this is because these processes have attractive properties that assists in deriving some properties on long memory. We then discuss ARFIMA processes which are normally used as a mean equation when modeling long memory. There-after we discuss discrete and semi-parametric long memory estimation methods. This then leads us to discussing structural breaks in long memory and spurious long memory.

In chapter 5 we discuss associated heavy tailed distributions. The first property of heavy tailed distributions we discuss is the infinite divisibility property. We will then discuss Levy processes which can also be used in explaining infinite divisibility as it has increments that have this property. Distributions that we consider are the alpha-stable distributions and generalized hyperbolic distribution. In chapter 6, we discuss conditional heteroscedastic models and estimation methods. We start by discussing short-term dependency models which are ARCH and GARCH models. We then continue to discuss FIGARCH type models which are FIGARCH, FIEGARCH, FIAPARCH and HYGARCH processes.

In chapter 7, we discuss different forecasting methods but we place more focus on the Mincer-Zarnowitz regression as this is the test we will use for testing models forecasting performance. Using methods discussed in the preceding chapters we will then analyse South African platinum return series in chapter 8. In chapter 9 we summarise the methods

and findings of this thesis.

Chapter 2

Time Series Processes

A stochastic process time series is a collection or a family of random variables indexed by time t. There are two types of such stochastic processes, namely discrete and continuous time processes. A stochastic process X_t is a discrete time process if it is defined only for a set of time $t = 1, 2, ..., T$, else X_t is a continuous process defined on $t \geq 0$. In this study we consider discrete time stochastic processes. The basic assumption in classical time series analysis is stationarity. In the next section, we discuss stationary time series processes and their characteristics.

2.1 Stationary Processes

There are two concepts of stationarity, namely strict and weak stationarity. A time series $\{X_t\}$ is said to be strictly stationary if the joint distribution of $\{X_{t_1}, ..., X_{t_r}\}$ is identical in distribution to $\{X_{t_1+k}, ..., X_{t_r+k}\}$ for all time t_r. A weak stationarity condition requires that the mean and the covariance do not depend on time. Thus, for a stationary process $\{X_t\}$, the mean $E(X_t)$ and the variance $Var(X_t)$ are constant and the covariance $\gamma(t, s)$ is a function of the lag $|t - s|$ only. In univariate time series analysis linear dependence of the variable X_t to its past values X_{t-i}, for $i = 1, 2, ..., t - 1$, is of interest as this reveals some of the properties of the time series process. The basic diagnostic tool for linear dependence is the autocorrelation defined by

$$\rho(k) = \frac{\gamma(k)}{\sqrt{Var(X_t)Var(X_t)}}, \tag{2.1}$$

where

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}), \tag{2.2}$$

is the autocovariance function. The autocovariance and autocorrelation function has some interesting properties which are,

1. $\gamma(0) = Var(X_t)$ and $\rho(0) = 1$.
2. $|\gamma(k)| \leq \gamma(0)$ and $|\rho(k)| \leq 1$.
3. $\gamma(k) = \gamma(-k)$ and $\rho(k) = \rho(-k)$ for stationary processes.

After removing the linear mutual dependence of the intervening variables $\{X_{t+1}, ..., X_{t+k-1}\}$, the autocorrelation of X_t and X_{t+k} is explained by the partical autocorrelation function (PACF) defined by

$$\Phi_{kk} = \text{Corr}(X_t, X_{t+k}|X_{t+1}, ..., X_{t+k-1}). \tag{2.3}$$

Another important process in time series analysis is the white noise process. By definition, a process $\{a_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with a constant mean, constant variance, i.e. $\gamma(k) = Cov(a_t, a_{t+k}) = 0$ for all $k \neq 0$. Then the autocovariance of this process is given by

$$\gamma(k) = \sigma_a^2, \quad k = 0, \tag{2.4}$$

where σ_a^2 is the variance of the process. The autocorrelation function and partial autocorrelation functions are respectively given by $\rho(k)$=1 and $\Phi(kk) = 1$ for $k = 0$ and both equal to zero for $k \neq 0$. If $E(a_t) = 0$, the white noise process is referred to as zero mean Gaussian process.

Since a time series is characterised by its mean, variance, autocorrelation and patial autocorelation, it is important to understand their estimations. In this Section we will only consider single realization estimation. The sample mean of a time series process is

$$\begin{aligned} \bar{X} &= E(X_t) \\ &= \frac{1}{T}\sum_{t=1}^{T} X_t. \end{aligned}$$

This is an unbiased estimator of the mean since

$$\begin{aligned} E(\bar{X}) &= \frac{1}{T}\sum_{t=1}^{T} E(X_t) \\ &= \frac{1}{T}.T.\mu \\ &= \mu, \end{aligned} \tag{2.5}$$

and it can be shown that this estimator is also a consistent estimator since

$$\begin{aligned} \lim_{T\to\infty} \bar{X} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} X_t \\ &= \mu. \end{aligned} \tag{2.6}$$

Thus a time series with this property is said to be ergodic. This is very important as a ergodic process has a property that $\rho_t \to 0$ as $t \to \infty$, that is

$$\lim_{t\to\infty} \frac{1}{T} \sum_{t=-(T-1)}^{T-1} \rho_t = 0 \tag{2.7}$$

and hence

$$\lim_{t\to\infty} Var(X_t) = 0. \tag{2.8}$$

This property tells us that if we have two observations that are far apart, then they are almost uncorrelated. The sample autocovariance function can be calculated as

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{k=1}^{T-t}(X_t - \bar{X})(X_{t+k} - \bar{X}). \tag{2.9}$$

The sample autocovariance $\hat{\gamma}(k)$ is an asymptotically unbiased estimator of $\gamma(k)$, hence for the process to be ergodic, the autocovariances must be summable, i.e.

$$\sum_{t=-\infty}^{\infty} |\gamma(t)| < \infty. \tag{2.10}$$

The sample autocorrelation function is then be defined as

$$\begin{aligned}\hat{\rho}(k) &= \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \\ &= \frac{\sum_{k=1}^{T-t}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}.\end{aligned} \tag{2.11}$$

The sample autocorrelation function can be used to plot a correlogram by plotting $\hat{\rho}(k)$ against k. To calculate the partial autocorrelation, a recursive method

$$\hat{\Phi}(t+1, t+1) = \frac{\hat{\rho}(t+1) - \sum_{j=1}^{t} \hat{\Phi}(t_j)\hat{\rho}(t+1-j)}{1 - \sum_{j=1}^{t} \hat{\Phi}(t_j)\hat{\rho}(j)}, \tag{2.12}$$

is often used where

$$\hat{\Phi}(t+1, j) = \hat{\Phi}(t_j) - \hat{\Phi}(t+1, t+1)\hat{\Phi}(t, t+1-j), j = 1, 2, ..., k. \tag{2.13}$$

In the analysis a time series $\{X_t\}$ can be represented as a moving average (MA) or an autoregressive (AR) process. In MA representation, we write the process X_t as a linear combination of a sequence of uncorrelated random variables as

$$\begin{aligned}X_t - \mu &= a_t + \theta_1 a_{t-1} + ... \\ &= \sum_{j=0}^{\infty} \theta_j a_{t-j}, \quad \theta_0 = 1,\end{aligned} \tag{2.14}$$

where $\{a_t\}$ is a white noise process and

$$\sum_{j=0}^{\infty} \theta_j^2 < \infty. \tag{2.15}$$

A compact form of the representation is useful and for this, we use the backshift operator $LX_t = X_{t-1}$ to present the MA process as follows

$$\begin{aligned}X_r - \mu &= a_t + \theta_1 L^1 a_t + \theta_2 L^2 a_t + ... \\ &= (1 + \theta_1 L^1 + \theta_2 L^2 + ...)a_t \\ &= \theta(L)a_t,\end{aligned} \tag{2.16}$$

where $\theta(L)$ is called the MA polynomial. The expectation and variance of this process is

$$\begin{aligned} E(X_t) &= \mu, \\ Var(X_t) &= \sigma_a^2 \sum_{j=0}^{\infty} \theta_j^2, \end{aligned}$$

and the covariance function

$$\begin{aligned} \gamma(k) &= E(X_t X_{t+k}) \\ &= E\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_i \theta_j a_{t-i} a_{t+k-j} \right) \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+k}. \end{aligned} \tag{2.17}$$

The ACF is then given by

$$\rho(k) = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+k}}{\sum_{i=0}^{\infty} \theta_i^2}. \tag{2.18}$$

This shows us that the mean and variance are constant and the autocovariance and autocorrelation depends only on lag k. From the above properties, $\sum \theta_j^2 < \infty$ and hence this process is stationary. This is a linear process which any stationary process can be represented. In an AR representation, the process X_t is of the form

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + a_t \\ &= \phi_1 L X_t + \phi_2 L^2 X_t + ... + a_t \\ \left(1 - \sum_{j=1}^{\infty} \phi_j L^j \right) X_t &= a_t \\ \phi(L) X_t &= a_t, \end{aligned} \tag{2.19}$$

and $1 + \sum_{\forall j} |\phi_j| < \infty$. This process is invertible and this property is useful for forecasting. For a process to be invertible, the roots of $\phi(L) = 0$ must lie outside the unit circle, and thus, for the AR process to be stationary, it must be presentable as an MA process

$$\begin{aligned} X_t &= \frac{1}{\phi(L)} a_t \\ &= \theta(L) a_t. \end{aligned} \tag{2.20}$$

The shortcoming of the MA and AR representations are that they produce infinite parameters that needs to be estimated using finite data series. To overcome this, we let the AR process be of order p

$$\phi_p(L)X_t = a_t, \tag{2.21}$$

and the MA process be of order q

$$X_t = \theta_q(L)a_t. \tag{2.22}$$

In order to describe a variety of time series models, a class of ARMA(p, q) processes is used. This is a combination of AR and MA processes of order p and q respectively. The ARMA(p, q) is given by

$$\phi_p(L)X_t = \theta_q(L)a_t, \tag{2.23}$$

where $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ and $\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$ are polynomials of respective orders p and q. For this process to be invertible and stationary we require the roots of $\theta_q(L) = 0$ lie outside the unit circle and the roots of $\Phi_p(L) = 0$ to lie outside the unit circle, respectively.

2.2 Nonstationary Processes

The study of nonstationary processes is of utmost importance in econometrics. This is because most time series data in practice are not stationary. This can be as a result of non-constant means or time-varying variance functions. Processes that are nonstationary in mean posses a challenge in model estimation and forecasting. They however can be transformed into a stationary process. If a series is nonstationary in mean, the most used method is to integrate the series through differencing.

2.2.1 Autoregressive Integrated Moving Average (ARIMA) Processes

Autoregressive integrated moving average (ARIMA) models are used to describe nonstationary ARMA models. This is done by differencing the nonstationary time series using the differencing operator $\Delta^d = (1 - L)^d$. An ARIMA model is of the form

$$\phi_p(L)(1 - L)^d X_t = \theta_q(L)a_t, \tag{2.24}$$

where $d = 1, 2, ...$ is the order of difference parameter. This process can be presented as an ARMA process by letting

$$W_t = (1 - L)^d X_t.$$

Then, the resulting process becomes the usual ARMA process

$$\phi_p(L)W_t = \theta_q(L)a_t. \tag{2.25}$$

This process is only useful in the case where a time series process is nonstationary in mean. There are separate methods proposed in literature that are used for processes that are nonstationary in variance. The following subsection explains these phenomena and their transformation to stationarity.

2.2.2 Variance-Stabilising Transformations

Differencing a time series is useful when the process in nonstationary in mean. However, this method is not useful when the time series is nonstationary in variance. Variance can change as its levels changes, and this can be represented by

$$Var(X_t) = cf(\mu_t). \tag{2.26}$$

For this times series to be stationary in variance, we need a transformation function, $T(X_t)$, that would result in the transformed series having a constant variance. For this

type of nonstationarity, we could use the logarithmic transformation function

$$T(X_t) = \ln(X_t). \tag{2.27}$$

The transformed series $T(X_t)$ will have a constant variance. If the variance of the series is proportional to the levels, such that

$$Var(X_t) = c\mu_t, \tag{2.28}$$

then a square root transformation

$$T(X_t) = \sqrt{X_t}, \tag{2.29}$$

is used which will result in a constant variance. If the standard deviation of the series is proportional to the square of the level, such that

$$Var(X_t) = c\mu_t^4, \tag{2.30}$$

then the reciprocal function transforms the series to be stationary in variance. In general a power transformation which belongs to the Box and Cox (1964) class of transformations is of the form

$$T(X_t) = \frac{X_t^\lambda - 1}{\lambda}. \tag{2.31}$$

This transformation class incorporates all the above discussed transformations. The parameter λ plays an important role in deciding the transformation function to be used. In applications and practice, λ is selected such that it minimises the residual mean squared error. Seasonal nonstationarity will be considered in the next section.

2.3 Seasonal Time Series Models

Consider an additive seasonal time series model

$$\begin{aligned} X_t &= P_t + S_t + e_t \\ &= \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it} + \sum_{j=1}^{k} \beta_j V_{jt} + e_t, \end{aligned} \tag{2.32}$$

where $P_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it}$ is the trend-cycle, $S_t = \sum_{j=1}^{k} \beta_j V_{jt}$ is the seasonal component and e_t is the irregular component. Normally, S_t is represented by using sinusoidal functions as

$$S_t = \sum_{j=1}^{[s/2]} \left[\beta_j \sin\left(\frac{2\pi jt}{s}\right) + \gamma_j \cos\left(\frac{2\pi jt}{s}\right) \right], \tag{2.33}$$

and hence, the above additive seasonal time series is then given by

$$X_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{[s/2]} \left[\beta_j \sin\left(\frac{2\pi jt}{s}\right) + \gamma_j \cos\left(\frac{2\pi jt}{s}\right) \right] + e_t. \tag{2.34}$$

This logic can be applied to ARIMA models

$$\phi_p(L)(1-L)^d X_t = \theta_q(L) b_t \tag{2.35}$$

and in this case, $\{b_t\}$ is not white noise as it will contain seasonal correlations. To overcome this, a Box-Jenkins multiplicative seasonal ARIMA model is used. This model is given by

$$\Phi_P(L^s)\phi_p(L)(1-L)^d(q-L^s)^D X_t = \theta_q(L)\Theta_Q(L^s)a_t \tag{2.36}$$

where $\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - ... - \Phi_P L^{Ps}$ and $\Theta_Q(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - ... - \Theta_Q L^{Qs}$ with the sub-index s referring to the seasonal period. This model is of the form of an ARIMA with extra parameters for seasonality and is denoted as an ARIMA$(p,d,q)X(P,D,Q)_s$

2.4 Forecasting

One of the main objectives of modelling in time series is forecasting future values of the series. In forecasting, the objective is to produce optimal forecasts that have little or no error. The criteria that is normally used is the minimum mean square error (MSE) criterion. There are many other criteria used in literature to find optimal forecasts but for simplicity and ease of use, the MSE criterion is normally used.

The optimal forecasts of X_{n+l} is given by its conditional expectation

$$E(X_{n+l}|X_n, X_{n-1}, ...). \tag{2.37}$$

For the ARIMA(p, d, q) model, this is expressed as ,

$$\begin{aligned} \hat{X}_n(l) &= E(X_{n+l}|X_n, X_{n-1}, ...) \\ &= \phi_1\hat{X}_n(l-1) + ... + \phi_{p+d}\hat{X}_n(l-p-d) \\ &\quad +\hat{a}_n(l) - \theta_1\hat{a}_n(l-1) - ... - \theta_q\hat{a}_n(l-q), \end{aligned} \tag{2.38}$$

where

$$\hat{a}_n(l) = X_{n+l} - \hat{X}_{n+l-1}, \quad and \quad \hat{a}_n(j) = 0, \quad \forall l \geq 1. \tag{2.39}$$

2.5 Model Identification

Model identification refers to the methodology in identifying the required transformations, such as variance stabilising transformations and differencing transformations and polynomial orders of AR(p) and MA(q). Model identification steps are as follows

1. Plot the time series data and choose proper transformations.

2. Compute and examine the sample ACF and the sample PACF of the original series to further confirm a necessary degree of differencing so that differenced series is stationary. If the sample ACF decays very slowly and the sample PACF cuts off after lag k, then it indicates that differencing is needed.

3. Compute and examine the sample ACF and PACF of the properly transformed

and differenced series to identify the orders of p and q. If the ACF tails off in an exponential decay or damped sine wave fashion and PACF cuts off after lag p, then the process is AR(p). If the ACF cuts off after lag q and PACF tails off as exponential decay or damped sine wave fashion, then the process is MA(q).

This modelling methodology is referred to as Box-Jenkins methodology.

2.6 Model Estimation

In the ARIMA(p, d, q) estimation, the maximum likelihood estimation (MLE) has been widely used because of ease of use and some interesting properties of asymptotic normality that can be derived from the estimates. Box et al. (1994) suggested maximizing the unconditional log-likelihood function

$$\ln L(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2}\ln[2\pi\sigma_a^2] - \frac{S(\phi, \mu, \theta)}{2\sigma_a^2}, \tag{2.40}$$

where $S(\phi, \mu, \theta)$ is the unconditional sum of squares function given by

$$S(\phi, \mu, \theta) = \sum_{t=M}^{n} \left[E(a_t|\phi, \mu, \theta, X)\right]^2. \tag{2.41}$$

where $E(a_t|\phi, \mu, \theta, X)$ is the conditional expectation of a_t. After this estimation process, the model needs to be checked for adequacy.

2.7 Diagnostic Checking

In estimation, we have assumed that $\{a_t\}$ is a white noise process. However, after estimation, we need to check the residual series $\{\hat{a}_t\}$ if this assumption is met. A test popularly used in literature is the portmanteau lack of fit test. The null hypothesis is then

$$H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0, \tag{2.42}$$

where the test statistic is given by

$$Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}_k^2. \tag{2.43}$$

Under the null hypothesis, Q statistic approximately follows the $\chi^2(k-p-q)$ distribution. This test is also known as the Lung-Box statistic.

2.8 Statistical Tests

Tests that we will use for Normality in the thesis are the Kolmogorov-Sminorv(KS)(Chakravarti et al., 1967), Jarque-Bera(JB)(Jarque and Bera, 1987) and Phillips-Peron(PP)(Phillips and Perron, 1988), we chose these tests to compare with similar results in literature. The KS test is used for testing the null hypothesis that the cumulative distribution function $F(x)$ equals a hypothesized distribution $F_1(x)$. The test statistic is

$$D_n = \sup_x \|F_n(x) - F_1(x)\|. \tag{2.44}$$

We reject the null hypothesis is D_n is too large. The JB test statistic is defined as

$$\text{JB} = \frac{N}{6}\left(S^2 + \frac{(K-3)^2}{4}\right), \tag{2.45}$$

where S is the skewness, K is the kurtosis, and N is the sample size. The JB statistic follows χ^2 distribution with 2 degrees of freedom. The PP test involves fitting the regression

$$y_i = \alpha + \rho y_{i-1} + \epsilon_i. \tag{2.46}$$

There are two statistics

$$\begin{aligned} Z_p &= n(\hat{\rho}_n - 1) - \frac{1}{2}\frac{n^2\hat{\rho}^2}{s_n^2}\left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}\right) \\ Z_\tau &= \sqrt{\frac{\hat{\gamma}_{0,n}}{\hat{\lambda}_n^2}} - \frac{1}{2}(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n})\frac{n\hat{\sigma}}{\hat{\gamma}_n s_n} \end{aligned} \tag{2.47}$$

where

$$\begin{aligned}\hat{\gamma}_{j,n} &= \frac{1}{n}\sum_{i=j+1}^{n}\hat{u}_i\hat{u}_{i-j} \\ \hat{\lambda}_n^2 &= \hat{\lambda}_{0,n} + 2\sum_{j=1}^{q}\left(1-\frac{j}{q+1}\right)\hat{\gamma}_{j,n} \\ s_n^2 &= \frac{1}{n-k}\sum_{i=1}^{n}\hat{u}_i^2 \end{aligned} \tag{2.48}$$

where u_i is the OLS residual, k is the number of covariates in the regression, q is the number of Newey and West (1994) lags.

2.9 Model Selection

Model selection criteria are used in selecting the best model to be used. There are several methods proposed in literature. The most widely used criteria for selection are the Akaike's information criterion (AIC) and the Bayesian information criteria (BIC). The AIC is defined as

$$\text{AIC}(M) = -2\ln[\text{maximum} - \text{likelihood}] + 2M, \tag{2.49}$$

where M is the number of parameters in the model. The disadvantage of this method, seen by Shibata (1976) is that it overestimates the orders of autoregression, as a result the BIC was introduced. The BIC is defined as

$$\text{BIC}(M) = n\ln(\hat{\sigma}_a^2) - (n-M)\ln(1-\frac{M}{n}) + M\ln(n) + M\ln\left[\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_a}^2\right)/M\right], \tag{2.50}$$

where $\hat{\sigma}_X^2$ is the sample variance of the time series. Schwarz (1978) introduced the Schwartz's Bayesian criterion and is defined as

$$\text{SBC}(M) = n\ln(\hat{\sigma}_a^2) + M\ln(n). \tag{2.51}$$

It is advisable to use all of the above mentioned criteria to check the best model as AIC overestimates the order of autoregression.

2.10 Chapter Summary

In this chapter, we discussed fundamental time series proceses that build sophisticated models. We started by discussing stationarity which is important in any time series analysis. This is due to some interesting properties that are derived when a series is stationary. ARMA models are stationary and invertible but when financial data is fitted to this model, it wont' work as there is nonstationary in financial returns. As a consequence, different methods were discussed which would assist. If the series is nonstationary in mean, then a series would be differenced until it is stationary. If the series in nonstationary in variance, then other methods are applied. Methods discussed in this chapter are the variance-stabilising transformations which consists of the following methods

- Logarithmic transformation
- Square-root transformation
- Reciprocal transformation, and in general the
- Power transformation

Seasonal models were also considered to address seasonality and from then, we discussed a simple forecasting method for ARIMA models, a method that could be used to identify an ideal parsimonious ARIMA model. The maximum likelihood estimation was discussed, diagnostic checks using the portmanteau lack of fit test is used in literature for goodness of fit. For model selection, we discussed three model selection methods, the AIC, BIC and SBC.

Chapter 3

Spectral Analysis Theory in Long Memory

Spectral analysis plays a major role in searching for periodicities in data. By definition, a periodic function repeats its values over intervals of a fixed length. The range between these intervals is called the period. Trigonometric functions are known to be periodic and hence are useful in the study of periodicity. A sinusoidal function is a good example of a periodic function. For instance a sinusoidal function

$$\sin(2\pi ft) \tag{3.1}$$

has a period $T = 1/f$ where f is a parameter that is measured in cycles per unit time. This then becomes the frequency of the function such that $sin(2\pi ft) = sin(\omega t)$. The parameter $\omega = 2\pi f$ is called the angular frequency of the sinusoidal function which is measured in radians per unit time. This measurement however does not give meaning in practice and instead normalized angular frequencies

$$f = \omega_t/2\pi, \quad t = 1, 2, ..., T$$

are used. This is measured in cycles per unit time. In the next section, we will discuss Fourier analysis which forms the fundamentals in spectral analysis.

3.1 Fourier Analysis

The study of Fourier analysis is the basis for constructing arbitrary functions using sinusoids. For a finite sequence $\{X_t\}$ for $t = 1, 2, ..., T$, it can be represented as a linear combination of trigonometric functions

$$X_t = \sum_{k=0}^{[T/2]} \left[a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right) \right] + \epsilon_t, \tag{3.2}$$

for $t = 1, 2, ..., T$, and this is called the Fourier series of the sequence $\{X_t\}$. The values a_k and b_k are called Fourier coefficients. From the orthogonal properties of trigonometric functions

$$a_k = \frac{1}{T} \sum_{t=1}^{T} X_t \cos\left(\frac{2\pi kt}{T}\right), \quad k = 1, 2, ..., [(T-1)/2], \tag{3.3}$$

where [.] refers to the integer part if T is even

$$a_k = \frac{2}{T} \sum_{t=1}^{T} X_t \cos\left(\frac{2\pi kt}{T}\right), \quad k = 1, 2, ..., [(T-1)/2], \tag{3.4}$$

if T is odd. The representation of b_k is given as

$$b_k = \frac{2}{T} \sum_{t=1}^{T} X_t \sin\left(\frac{2\pi kt}{T}\right), \quad k = 1, 2, ..., [(T-1)/2]. \tag{3.5}$$

Fourier frequencies are then defined

$$\omega_k = \left(\frac{2\pi kt}{T}\right) \quad t = 1, 2, ..., [(T-1)/2].$$

When we use complex exponential, the Fourier series can be represented as

$$X_t = \sum C_k e^{i\omega_k t} + \epsilon_t, \quad \forall k, \tag{3.6}$$

where the inverse is given by

$$C_k = \frac{1}{T} \sum_{t=1}^{T} X_t e^{-i\omega_k t}, \tag{3.7}$$

and C_k are Fourier coefficients. Hence the sequence can be represented as

$$X_t = \sum_{k=0}^{T/2} \left(a_k \cos(\omega_k t) + b_k \sin(\omega_k t) \right). \tag{3.8}$$

Then it can be seen that a periodic function is uniquely determined by its pattern in a range of one period and hence the series in equation 3.8 is periodic with n periods. This forms the fundamental in understanding spectral analysis. The frequency domain is based on Fourier analysis using the periodogram or the sample spectrum. In the following sections, we will discuss the spectrum and the periodogram, it's estimator.

3.2 The Spectrum

By using the Fourier theory, for a given time series X_t with a period of T can be expressed as a sum of sinusoids with frequencies

$$\omega_k = k/T, \quad k = 1, 2, ...,$$

In the literature, there are two types of models for sinusoidal functions, the real sinusoidal model given by

$$X_t = \sum_{k=1}^{q} \left[A_k \cos(\omega_k t) + B_k \sin(\omega_k t) \right], \tag{3.9}$$

for $A_k, B_K \in R$ are amplitude parameters satisfying the conditions

$$A_k^2 + B_k^2 > 0.$$

The complex sinusoidal model is given by

$$X_t = \sum_{k=1}^{q} C_k \cos(\omega_k t + \phi_k), \tag{3.10}$$

where $C_k > 0$ for $k = 1, 2, ...$ are the amplitude parameters and $\phi_k \in (-\pi, \pi)$ are the phase parameters. The frequencies are assumed to be in the order

$$0 < \omega_1 < \omega_2 < ... < \omega_q < \pi.$$

In time domain inference, the ACF is used to explain the properties of the time series. The ACF can also be represented in frequency domain as a linear combination of many sinusoidal functions with different frequencies. If $\gamma(h)$ is the ACF of a stationary process $\{X_t\}$, then there exists a unique non-decreasing right continuous function $F(\omega)$, with $F(-\pi) = 0$, such that

$$\gamma(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega h} dF(\omega). \tag{3.11}$$

$F(\omega)$ is the spectral distribution function. If $\gamma(h)$ is absolutely summable such that

$$\sum_{i=1}^{\infty} |\gamma(h)| < \infty, \tag{3.12}$$

then there exists a unique uniformly continuous non-negative function $f(\omega)$ called the spectral density function such that

$$\gamma(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{i\omega h} d\omega, \tag{3.13}$$

and

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}. \tag{3.14}$$

In frequency domain analysis of long memory, inference is based on the spectral density $f(\omega)$, and it's estimator the periodogram. If a time series $\{X_t\}$ has autocovariance $\gamma(h)$ satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{3.15}$$

then there is a spectral density defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}. \tag{3.16}$$

The spectral density gives an alternative view of a stationary series. One method of estimating the spectral density of a time series $\{X_1, X_2, , , X_T\}$ is by replacing the $\gamma(.)$ by the sample autocovariance

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{T-|h|} \left(X_{t+|h|} - \bar{X}\right) \left(X_t - \bar{X}\right). \tag{3.17}$$

Another method that is widely used in literature is by computing the periodogram, which is defined to be the square of the discrete Fourier transform (DFT) modulus at frequency ω_j. An advantage of using the periodogram is that its asymptotic expectation is the spectral density $f(\omega)$, which shows that the periodogram is an unbiased estimator of the spectral density. In the following section, we discuss the periodogram method and its properties.

3.3 The Periodogram

In time series, a periodogram is used to identify the periodicity of a time series. This can be helpful for identifying the dominant cyclic behavior in a series, particularly when cycles are not related to the commonly encountered seasonality. The periodogram is the square of the modulus of the DFT. A time series $\{X_t\}$ with T observations can be represented using a Fourier representation as

$$X_t = \sum_{k=0}^{T/2} \left[a_k \cos(\omega_k t) + b_k \sin(\omega_k t)\right]. \tag{3.18}$$

where ω_k are Fourier frequencies and a_k and b_k are Fourier coefficients. Fourier coefficients are estimates of fitting the regression model

$$X_t = \sum_{k=1}^{q} \left[A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\right] + e_t. \tag{3.19}$$

For a sequence $\{X_1, X_2, , , X_T\}$, the DFT is given by $\{W(\omega_0), W(\omega_1), ..., W(\omega_{T-1})\}$ where

$$W(\omega_k) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t e^{-2\pi i \omega_k t}, \tag{3.20}$$

and the Fourier frequencies

$$\omega_k = \frac{2\pi k}{T}, \quad k = 0, 1, 2, ..., T/2.$$

The periodogram is then defined as

$$\begin{aligned} I(\omega_j) &= |W(\omega_j)|^2 \\ &= \frac{1}{\pi T} \left| \sum_{t=1}^{T} X_t e^{-i\omega_j t} \right|^2 . \end{aligned} \tag{3.21}$$

The periodogram is asymptotically an unbiased estimator of the spectral density $f(\omega)$. This means that

$$E(I(\omega)) = f(\omega).$$

The periodogram however is an inconsistent estimator of the spectral density.

3.4 Properties of the Periodogram

In order to understand the properties of the periodogram, an analysis of the Fourier coefficients can be used. For an iid $\mathrm{N}(0, \sigma^2)$ time series

$$E(a_k) = \frac{2}{T} \sum_{t=1}^{T} E(X_t) \cos(\omega_k t) = 0, \tag{3.22}$$

and

$$\begin{aligned}\text{Var}(a_k) &= \frac{4}{T^2}\sum_{t=1}^{T}\sigma^2\cos^2(\omega_k t) \\ &= \frac{4\sigma^2}{T^2}\sum_{t=1}^{T}\cos^2(\omega_k t) \\ &= \frac{4\sigma^2}{T^2}.\frac{T}{2} \\ &= \frac{2\sigma^2}{T}. \end{aligned} \tag{3.23}$$

similarly,

$$E(b_k) = 0,$$

and

$$\text{Var}(b_k) = \frac{2\sigma^2}{T}.$$

Hence, a_k and b_k are $iidN(0, 2\sigma^2/T)$ so that

$$\frac{Ta_k^2}{2\sigma^2} \quad \text{and} \quad \frac{Tb_k^2}{2\sigma^2} \tag{3.24}$$

follows a chi-square distribution with 1 degree of freedom. For $k \neq j$,

$$\begin{aligned}\text{Cov}(a_k, b_j) &= \frac{4}{T^2}E\left(\sum_{t=1}^{T}X_t\cos(\omega_k t).\sum_{u=1}^{T}X_u\sin(\omega_j u)\right) \\ &= \frac{4}{T^2}\sum_{t=1}^{T}\left(E\left(X_t\right)\cos(\omega_K t)\sin(\omega_j t)\right) \\ &= \frac{4\sigma^2}{T^2}\sum_{t=1}^{T}\left(\cos(\omega_k t)\sin(\omega_j t)\right) \\ &= 0, \quad \forall k, j \end{aligned} \tag{3.25}$$

and hence, the periodogram ordinates

$$\frac{I(\omega_k)}{\sigma^2} = \frac{T}{2\sigma^2}\left(a_k^2 + b_k^2\right) \tag{3.26}$$

follows a chi-square distribution with two degrees of freedom. The shortfall however of this estimator is that it is an inconsistent estimator of the spectral density. This is as a result of the variance of the periodogram not depending on T, such that

$$\lim_{T\to\infty} \text{Var}(I(\omega)) \neq 0 \tag{3.27}$$

This then suggests that some form of averaging is needed for the periodogram to be consistent. In the literature, smoothing techniques are used as a form of averaging to remove some of the inconsistency in the periodogram estimate.

Consider the sample spectrum

$$\hat{f}(\omega_k) = \frac{1}{4\pi} I(\omega k). \tag{3.28}$$

We can therefore deduce that

$$\hat{f}(\omega_k) \approx f(\omega_k)\frac{\chi^2(2)}{2} \tag{3.29}$$

and we can use equation 3.29 to find moments of the sample spectrum. The first moment is given by

$$\begin{aligned} E(\hat{f}(\omega_k)) &= E\left[f(\omega_k)\frac{\chi^2(2)}{2}\right] \\ &= f(\omega_k) \end{aligned} \tag{3.30}$$

with the variance is

$$\begin{aligned} \text{Var}(\hat{f}(\omega_k)) &= Var\left[f(\omega_k)\frac{\chi^2(2)}{2}\right] \\ &= \left[f(\omega_k)\right]^2. \end{aligned} \tag{3.31}$$

Thus we can see that the variance does not depend on T. This shows the inconsistency of the sample spectrum and the peridogram. The periodogram ordinates are then given by

$$\text{Cov}(\hat{f}(\omega_k)), \hat{f}(\omega_j)) = 0, \quad k \neq j.$$

A natural way of reducing this variation is by smoothing the sample spectrum in the vicinity of the target frequency. A smoothed sample spectrum obtained from weighted average of M values at a target frequency ω_k,

$$\hat{f}_W(\omega_k) = \sum_{j=-M}^{M} W_T(\omega_j)\hat{f}(\omega_k - \omega_j), \tag{3.32}$$

where M, the number of frequencies used in the smoothing is the function of T such that $M \to \infty$ as $n \to \infty$. $W_T(\omega_j)$ is the weighting function such that

1.

$$\sum_{j=-M}^{M} W_T(\omega_j) = 1. \tag{3.33}$$

2.

$$W_T(\omega_j) = W_T(-\omega_j). \tag{3.34}$$

3.

$$\lim_{T\to\infty} \sum_{j=-M}^{M} W_T(\omega_j) = 0. \tag{3.35}$$

This weighting function $W_T(\omega_j)$ is called the spectrum window. Calculating moments of the sample spectral in equation 3.32 yields

$$\begin{aligned} E[\hat{f}_W(\omega_k)] &= \sum_{j=-M}^{M} W_T(\omega_j) E(\hat{f}(\omega_k - \omega_j)) \\ &\approx f(\omega_k) \sum_{j=-M}^{M} W_T(\omega_j) \\ &= f(\omega_k) \end{aligned} \tag{3.36}$$

and

$$\begin{aligned} Var[\hat{f}_W(\omega_k)] &\approx \sum_{j=-M}^{M} W_T^2(\omega_j)(f(\omega_k))^2 \\ &\approx (f(\omega_k))^2 \sum_{j=-M}^{M} W_T^2(\omega_j) \end{aligned} \tag{3.37}$$

and thus, using property in equation 3.35, the variance of this sample spectrum decreases as m_T increases. This smoothing procedure, like any other statistical smoothing procedures, suffers from the balance between variance reduction and bias.

Since the spectrum is the Fourier transform of the autocovariance function $\gamma(k)$, an alternative smoothing can be applied to sample autocovariances $\hat{\gamma(k)}$ using the weighting function $W(k)$. This is represented by

$$\hat{f}_W(\omega) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{(T-1)} W(k)\hat{\gamma}(k)e^{-i\omega k} \tag{3.38}$$

and due to the symmetric nature of the autocovariance, this can also be given as

$$\hat{f}_W(\omega) = \frac{1}{2\pi} \sum_{k=-M}^{M} W(k)\hat{\gamma}(k)e^{-i\omega k} \tag{3.39}$$

where M is the value that depends on T. The weighting function $W_T(k)$ for the autocovariance is called the lag window. The lag window and the spectral window has the following relationship;

1.

$$W_T(\omega) = \frac{1}{2\pi} \sum_{k=-M}^{M} W_T(k)e^{-i\omega k} \tag{3.40}$$

is the spectral window, and

2.

$$W_T(k) = \int_{-\pi}^{\pi} W_T(\omega)e^{i\omega k}d\omega \tag{3.41}$$

is the lag window.

The above relationship shows that the spectral window is the Fourier transform of the lag window and the lag window is the inverse Fourier transform of the spectral window. This kind of relationship forms a Fourier transform pair. In the following section, we discuss some of the commonly used lag and spectral window and properties thereof.

3.5 Windows of the Periodogram

Smoothing techniques of the estimated spectrum using lag or spectral windows is discussed in the sections. One of the most basic and simple windows is the rectangular window. This lag window is defined as

$$W_T^R(k) = \begin{cases} 1 & |k| \leq M \\ 0 & |k| > M \end{cases}, \tag{3.42}$$

where $M < T - 1$ is a truncation point, and the corresponding spectral window is given by

$$\begin{aligned} W_T^R(\omega) &= \frac{1}{2\pi} W_T^R(k) e^{-i\omega k} \\ &= \frac{1}{2\pi} \frac{\sin(\omega[M + 1/2])}{\sin(\omega/2)}. \end{aligned} \tag{3.43}$$

The bandwidth of the rectangular estimator is

$$\frac{2\pi}{2M+1},$$

and hence this bandwidth has an inverse relationship with the truncation point M. This simply means that as M increases, the bandwidth decreases and the variance of the smoothed spectrum increases and the bias decreases. When M decreases, the variance will also decrease and the bias increases. This displays the trade-off between the variance and the bias problem.

The advantage of using the rectangular lag window is its simplicity and mathematical ease of use. The disadvantage however is that this window may lead to negative values for some frequencies ω. This results violates the properties of the spectrum that it is a nonnegative function.

Bartlett's proposed the lag window

$$W_n^B(k) = \begin{cases} 1 - |k|/M & |k| \leq M \\ 0 & |k| > M \end{cases}, \tag{3.44}$$

and the corresponding spectral window given by

$$\begin{aligned} W_n^B(\omega) &= \frac{1}{2\pi} \sum_{k=-M}^{M} \left(1 - \frac{|k|}{M}\right) e^{-i\omega k} \\ &= \frac{1}{2\pi M} \left[\frac{\sin(\omega M/2)}{\sin(\omega/2)}\right]^2. \end{aligned} \tag{3.45}$$

From these results, the Bartletts spectrum estimator is nonnegative. It can be shown that some of the shortcomings of the rectangular lag window are corrected when using the Bartletts triangle lag window. This includes a phenomena called the leakage which results from big side lobes of a window which then results in functions that has sharp corners. These type of functions are always avoided in the estimation.

Blackman and Tuckey introduced the lag window

$$W_n^T(k) = \begin{cases} 1 - 2a + 2a\cos(\pi k/M) & |k| \leq M \\ 0 & |k| > M \end{cases} \tag{3.46}$$

and the corresponding spectral window is given by

$$\begin{aligned} W_n^T(\omega) &= \frac{1}{2\pi} \sum_{k=-M}^{M} \left[1 - 2a + 2a\cos\left(\frac{\pi k}{M}\right)\right] e^{-i\omega k} \\ &= \frac{a}{2\pi} \left[\frac{\sin[(\omega - \pi/M)(M + 1/2)]}{\sin[(\omega - \pi/M)/2]} \frac{\sin[(\omega + \pi/M)(M + 1/2)]}{\sin[(\omega + \pi/M)/2]}\right] \\ &+ \frac{(1 - 2a)}{2\pi} \frac{\sin[\omega(M + 1/2)]}{\sin(\omega/2)}. \end{aligned} \tag{3.47}$$

This is a weighted linear combination of the spectral windows for the rectangular lag function at frequencies $(\omega - \pi/M)$, ω, and ,$(\omega + \pi/M)$. Thus, there is a possibility that the Blackman-Tukey spectrum may also be negative at some frequencies ω. When

$a = 0.25$, this window is given by

$$W_n^T(k) = \begin{cases} 1/2[1 + \cos(\pi k/M)] & |k| \leq M \\ 0 & |k| > M \end{cases} \tag{3.48}$$

and is called the Tukey-Hanning window. Parzen introduced a lag window given by

$$W_n^P(k) = \begin{cases} 1 - 6(k/M)^2 + 6(|k|/M)^3 & |k| \leq M/2 \\ 2(1 - |k|/M)^3 & M/2 < |k| \leq M \\ 0 & |k| > M \end{cases} \tag{3.49}$$

and the corresponding spectral window for an even and large M is given by

$$W_n^P(\omega) \approx \frac{3}{8\pi M^3} \left[\frac{\sin(\omega M/4)}{1/2\sin(\omega/2)} \right]^4. \tag{3.50}$$

It must be noted that the bandwidth of the Tukey-Hanning window is narrower that the bandwidth of the Parzen spectral window. This suggests that the Tukey-Hanning estimator will contain lesser bias that the Parzen estimator. This also shows the challenge in time series that is faced by researchers, which is the selection of the optimal bandwidth. It is known in theory that wider bandwidths produce smoother spectrums with smaller variance of the estimators and narrower bandwidths lead to smaller bias and smudging. This needs a trade-off between high stability and high resolution. In literature, the window closing algorithm is normally applied to minimise the error of selecting a non-optimal bandwidth. The window closing algorithm is as follows:

1. Chose a spectral window with an acceptable and desirable shape.

2. Initially calculate the spectral estimates using a large bandwidth and then re-calculate the estimates by gradually reducing the bandwidth until optimal stability and resolution is achieved.

3.6 Summary

This chapter discusses spectral theory and methods which forms the fundamentals in long memory analysis in frequency domain. We started the chapter with Fourier analysis, this is mainly because spectral analysis is based on transforming a series into a sinusoidal

function. This can be done using Fourier transforms and to go back to the initial function using the inverse Fourier transform. With this information, we then discussed the spectrum. The spectrum is as important in frequency domain as much as the ACF is in time domain. Using Fourier transforms, we constructed general spectrum and we then discussed the periodogram which is used to estimate the spectrum. The periodogram is shown to be an unbiased but inconsistent estimator of the spectrum. An inconsistent estimator resulted from the periodogram being independent of the period, T. This was further shown using the sample spectrum.

Inconsistency is estimating the spectrum has been dealt with in the literature. In this chapter, we discussed lag and spectral windows which could be used to smooth the periodogram or the sample spectrum and reduce the inconsistency in variation. Lag and spectral windows rely on bandwidths which are used in a trade-off between high stability and high resolution. The main challenge that we found is that of selecting the optimal bandwidth does not have a direct method. We only discussed the *closing-window* algorithm, which is simple to use and can assist in selecting an optimal bandwidth. In the next chapter, we will discuss long memory processes. The theory that was studied in the chapter will be useful when discussing long memory models under frequency domain.

Chapter 4

Theory and Methods of Long Memory Processes

A stationary process X_t is a long memory process if there exists a real number $0 < H < 1$ such that the autocorrelation function (ACF,) $\rho(\tau)$ has the following hyperbolic decay

$$\lim_{\tau \to \infty} \rho(\tau) = C^{2H-1}, \quad C > 0, \tag{4.1}$$

where C is a finite constant and H is the Hurst exponent. A covariance stationary process $\{X_t\}$ with autocovariance function $\gamma(k)$ is said to have long memory if

$$\sum_{\forall k} |\gamma(k)| = \infty. \tag{4.2}$$

This displays the property of unsummability of the autocovariance in long memory processes. The autocovariance decays slowly at a hyperbolic rate overtime as given by equation 4.1. In the literature of long memory models, the long range parameter d is often used and sometimes used interchangeably with the Hurst exponent, H. The long range dependence parameter d of a stationary process with long memory varies in the interval

$$0 < d < \frac{1}{2}, \tag{4.3}$$

and has the following relationship with the Hurst exponent

$$d = H - \frac{1}{2}. \tag{4.4}$$

By using this equality, we can deduce the equivalent interval for the Hurst exponent which is $1/2 < H < 1$. For a stationary long memory process satisfying the above discussed conditions, the spectral density, if it exists, satisfies

$$f(\omega) = C|\omega|^{-2d}, \quad 0 < d < 1/2. \tag{4.5}$$

This shows that a decay rate of C^{2H-1} is similar to the decay rate of $|\omega|^{-2d}$ of the spectral density $f(\omega)$ near the frequency $\omega = 0$. Long memory processes have a relationship with self-similar processes and sometimes can be better understood using these processes. In the next section, we discuss self-similar processes and their properties thereof.

4.1 Self-Similar Processes

By definition, self-similar processes are non-stationary processes whose increments may generate stationary long memory processes under certain conditions. A continuous stochastic process $\{X_t\}$ is said to be self-similar of index $H > 0$ if, for all $a > 0$, the finite-dimensional distributions of $\{X(at)\}$ are the same as those of $\{a^H X(t)\}$. A good example of a self similar process is a Brownian motion with $H = 1/2$. Let

$$r_H(s,t) = \frac{1}{2}|s|^{2H} + |t|^{2H} - |s-t|^{2H}, \tag{4.6}$$

for a Gaussian process $\{X_j\}$. Let $\{X_t\}$ be an H-self similar stationary increment (H-sssi) process with index $H > 0$, and $X_0 = 0$, $E(X_t^2) < \infty$. Then

$$Cov(X_s, X_t) = \sigma^2 r_H(s,t), \quad \sigma^2 = E(X_1^2). \tag{4.7}$$

Moreover, if $E(X_t^2) \to 0$ as $t \to 0$, then $0 < H \leq 1$. Now define a process $Y_k = X_{k+1} - X_k$, then

$$\gamma_Y(k) = \frac{\sigma^2}{2}\left(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}\right). \tag{4.8}$$

For any $0 < H < 1$, the following hold:

i.) $r_H(s,t)$ is a covariance function

ii.) The process $\{Y_k\}$ is a covariance stationary process with an autocovariance function

$$Cov(Y_j, Y_k) = \gamma_Y(j-k) \quad \forall k, j, \tag{4.9}$$

and

$$\gamma_Y(k) = \sigma^2 H(2H-1)|k|^{2H-1} + O(k^{-2}), \quad k \to \infty. \tag{4.10}$$

Moreover, for $0 < H < 1/2$,

$$\sum_{\forall k} \gamma_Y(k) = 0. \tag{4.11}$$

An important property to note is the fact that for $H = 1/2$, $\{Y_k\}$ is an uncorrelated process, whereas for $1/2 < H < 1$, $\sum_{\forall k} \gamma_Y(k) = \infty$ and hence $\{Y_k\}$ has long memory with parameter $d = H - 1/2$. A Brownian motion is a Gaussian process $\{B(t)\}$ with $B(0) = 0$, $E(B(t)) = 0$ and covariance function

$$\begin{aligned} \gamma_B(s,t) &= E(B(s)B(t)) \\ &= \min(s,t). \end{aligned} \tag{4.12}$$

Further, a Brownian motion process is non-stationary since Var$(B(t))$=t but has stationary increments $B(t+h) - B(t)$ which follows $N(0,h)$ and are independent. The processes $\{B(t)\}$ is $H-$self similar with parameter $H = 1/2$, and for any $a > 0$, $\{B(at)\}$=$a^{1/2}\{B(t)\}$. Fractional Brownian motion is a classical example of H-self similar processes. Let $0 < H < 1$, a Gaussian process $B_H = \{B_H(t)\}$ with $B_H(0) = 0$, $E(B_H(t) = 0)$ and covariance function $r_H(s,t)$ is called the fractional Brownian motion with parameter $0 < H < 1$. This implies that for $0 < H < 1$, B_H is an $H-$sssi process. Thus, for any $a > 0$, the Gaussian process $\{B_H(at)\}$ and $\{a^H B_H(t)\}$ have equal covariances, $r_H(at, as) = a^{2H} r_H(s,t)$, which imply $H-$self similar property. From the above discussion, it is clear that self-similar processes are non-stationary processes whose

increments may generate stationary long memory processes under certain conditions. In the following sections, we discuss a widely used process in long memory modelling.

4.2 ARFIMA Processes

In long memory processes, one of the mostly used and understood parametric models is the autogregressive fractionally integrated moving average (ARFIMA) model. This model was introduced by Granger and Joyeux (1980) to model processes which have unsummable autocovariances and hence presenting hyperbolic decay of autocorrelations. These models have an advantage in that they are similar to autoregressive integrated moving average (ARIMA) processes which are also widely used in econometrics. The only difference in these models is that ARFIMA processes generalise ARIMA processes by allowing non-integer values for d.

To further define d, consider the differencing operator $\Delta^d = (1-L)^d$ with non-integer values $d > -1$ are used. A Taylor expansion is used to evaluate $(1-L)^d$ as follows:

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \frac{\Gamma(d+1)L^k}{\Gamma(k+1)\Gamma(d-k+1)} \\ &= 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - ... \\ &= 1 - \sum_{k=0}^{\infty} C_k(d)L^k, \end{aligned} \tag{4.13}$$

and it is highlighted that $C_k = \Gamma(k-d)/\Gamma(k+1)\ k^{-(d+1)}$ is large and hence the coefficients of the polynomial will decay hyperbolically. The process $\{X_j\}$ is said to be an ARFIMA$(0,d,0)$ process with $-1/2 < d < 1/2$ if $\{X_j\}$ is a stationary solution with zero mean of the difference equations

$$\Delta^d X_j = \epsilon_j, \quad \forall j, \tag{4.14}$$

where $\epsilon_j\ WN(0,\sigma_\epsilon^2)$ is a white-noise sequence. A linear process of $\{\epsilon_j\}$ can then be found

by impulse response weights as

$$\begin{aligned} X_j &= \Delta^{-d}\epsilon_j \\ &= \sum_{k=0}^{\infty} \pi_{k,-d}\epsilon_{j-k}, \end{aligned} \tag{4.15}$$

where $\pi_{k,d} = \binom{k-d-1}{k}$ and $\pi_{0,-d} = 1$. This linear process is a unique stationary solution to equation 4.14. To study the properties of this process, the following theorems are made use of.

Theorem 4.1. *Let*

$$X_j = \sum_{k=0}^{\infty} \pi_{k,-d}\epsilon_{j-k} \tag{4.16}$$

with

i.) $\pi_{0,d} = 1$ *and*

$$\begin{aligned} \pi_{k,-d} &= \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \\ &= \prod_{s=1}^{k} \frac{s-1-d}{k}, \quad k = 1, 2, ... \\ &= \frac{1}{\Gamma(d)} k^{-1+d}(1 + O(k^{-1})), \quad k \to \infty. \end{aligned} \tag{4.17}$$

Moreover, for $-1/2 < d < 0$,

$$\sum_{k=0}^{\infty} \pi_{k,-d} = 0, \quad \sum_{\forall} \gamma_X(j) = 0 \tag{4.18}$$

ii.) The spectral density is

$$\begin{aligned} f_X(\omega) &= \frac{\sigma_\epsilon^2}{2\pi} \left|1 - e^{i\omega}\right|^{-2d} \\ &= \frac{\sigma_\epsilon^2}{2\pi} \left|2\sin\left(\frac{\omega}{2}\right)\right|^{-2d} \\ &\approx \frac{\sigma_\epsilon^2}{2\pi} |\omega|^{-2d} \end{aligned} \tag{4.19}$$

iii.) For $j = 1, 2, ...$

$$\begin{aligned} \gamma_X(0) &= \sigma_\epsilon^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)} \\ \gamma_X(j) &= \sigma_\epsilon^2 \frac{\Gamma(j+d)}{\Gamma(j-d+1)} \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \\ r_X(j) &= \frac{\Gamma(j+d)}{\Gamma(j-d+1)} \frac{\Gamma(1-d)}{\Gamma(d)} \\ &= \prod_{k=1}^{j} \frac{k-1+d}{k-d} \end{aligned} \tag{4.20}$$

and as $j \to \infty$

$$\gamma_X(j) \; C_\gamma j^{-1+2d}, \quad C_\gamma = \frac{\Gamma(1-2d)}{\Gamma(d)\gamma(1-d)} \sigma_\epsilon^2 \tag{4.21}$$

Theorem 4.2. *Suppose* $-1/2 < d < 1/2$, $d \neq 0$ *and* $\{\epsilon_j\}$ *is an ergodic white-noise process. Then* $\{X_j\}$, *where*

$$X_j = \sum_{k=0}^{\infty} \pi_{k,-d} \epsilon_{j-k},$$

is a unique stationary ergodic zero-mean solution of

$$\Delta^d X_j = \epsilon_j, \quad \forall j.$$

Moreover, $\{X_j\}$ *is invertible:*

$$\begin{aligned} \epsilon_j &= \Delta^d X_j \\ &= \sum_{k=0}^{\infty} \pi_{k,d} X_{j-k}, \quad \pi_{0,d} = 1, \end{aligned} \tag{4.22}$$

where

$$\pi_{k,d} = \frac{1}{\Gamma(-d)} k^{-1-d} (1 + O(k^{-1})), \quad k \to \infty. \tag{4.23}$$

Keeping the above theorems for ARFIMA(0,d,0) process in mind, we will discuss the ARFIMA(p,d,q) process. The stationary zero mean process $\{X_j\}$ is said to be an

ARFIMA(p, d, q) process with parameter d, $-1/2 < d < 1/2$ if it satisfies the model

$$\phi(L)\Delta^d X_j = \theta(L)\epsilon_j, \tag{4.24}$$

where $\{\epsilon_j\}$ is a $WN(0, \sigma_\epsilon^2)$ sequence and $\phi_p(L)$ and $\theta_q(L)$ are polynomials of order p and q respectively.

The solution $\{X_j\}$ of an ARFIMA(p, d, q) can be causal or invertible. It is said to be causal if it can be represented as a moving average

$$X_j = \sum_{k=0}^{\infty} a_k \epsilon_{j-k}, \quad \sum_{k=0}^{\infty} a_k^2 < \infty \tag{4.25}$$

and if invertible is represented as

$$\epsilon_j = \sum_{k=0}^{\infty} \nu_k X_{j-k}, \quad \sum_{k=0}^{\infty} \nu_k^2 < \infty \tag{4.26}$$

which holds for real numbers $\{\nu_k\}$. The following theorem establishes square summability and existence of $E(X_j) = 0$, $Var(X_j) = \sigma_\epsilon^2 \sum_{k=0}^{\infty} a_{k,-d}^2$, and

$$\gamma_X(j) = \sigma_\epsilon^2 \sum_{k=0}^{\infty} a_{k,-d} a_{k+j,-d}.$$

Theorem 4.3. *Suppose that $|d| < 1/2$ and assume $|\phi(z)| > 0$,$|\theta(z)| > 0$. Then the process $\{X_j\}$ represented by*

$$\begin{aligned} X_j &= A(L)\epsilon_j \\ &= \sum_{k=0}^{\infty} a_{k,-d}\epsilon_{j-k}, \quad a_{0,-d} = 1 \end{aligned} \tag{4.27}$$

has the following properties

i.) if d=0, then for some $\alpha > 0$ and $C > 0$:

$$|a_{k,0}| \leq Ce^{-\alpha k}, \quad k \geq 1$$

$$|\gamma_X(j)| \leq Ce^{-\alpha j}, \quad j \geq 1.$$

ii.) If $d \neq 0$ then with $\nu = \frac{\theta(1)}{\phi(1)}$ and $C_a = \frac{\nu}{\Gamma(d)}$,

$$\begin{aligned} a_{k,-d} &= \nu\pi_{k,-d} + O(k^{-2+d}) \\ &= C_a k^{d-1} + O(k^{-2+d}) \\ \gamma_X(j) &\approx \nu^2 \gamma_Y(j) \\ &\approx C_\gamma j^{-1+2d}, \quad C_\gamma = C_a^2 L(d, 1-2d)\sigma_\epsilon^2, \end{aligned}$$

and in addition, if $-1/2 < d < 0$, then

$$\sum_{k=0}^{\infty} a_{k,-d} = 0, \quad \sum_{\forall} \gamma_X(j) = 0.$$

iii.) The spectral density of $\{X_j\}$ is

$$\begin{aligned} f_X(\omega) &= \frac{\sigma_\epsilon^2}{2\pi} \left| 2\sin\left(\frac{\omega}{2}\right) \right|^{-2d} \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2 \\ &\approx \frac{\sigma_\epsilon^2}{2\pi} \left| \frac{\theta(1)}{\phi(1)} \right|^2 |\omega|^{-2d}. \end{aligned} \tag{4.28}$$

The proof of the above theorems can be found in Giraitis et al. (2012). The following theorem establishes the existence and uniqueness of an invertible moving average solution to ARFIMA equations and demonstrates the existence of its infinite order autoregressive (AR) representation

$$\epsilon_j = \sum_{k=0}^{\infty} b_{k,d} X_{j-k}, \tag{4.29}$$

where AR weights $b_{k,d}$ are defined by the expansion

$$(1-\alpha)^d \frac{\phi(z)}{\beta(z)} = \sum_{k=0}^{\infty} b_{k,d} z^k, \quad |z| < 1, \quad b_{0,d} = 1 \tag{4.30}$$

and for $d \neq 0$ have the property

$$\begin{aligned} b_{k,d} &= (\nu' + O(1))\pi_{k,d} \\ &\approx \frac{\nu'}{\Gamma(-d)} k^{-1-d}, \quad k \to \infty, \quad \nu' = \frac{\phi(1)}{\theta(1)}. \end{aligned}$$

Theorem 4.4. *Suppose* $-1/2 < d < 1/2$, $\{\epsilon_j\}$ *is an ergodic process and assume* $|\phi(z)| > 0$, $|\theta(z)| > 0$, $|z| \leq 1$. *Then the process*

$$X_j = A(L)\epsilon_j$$

is a unique invertible stationary ergodic zero mean solution of the ARFIMA equation

$$\phi(L)\Delta^d X_j = \theta(L)\epsilon_j.$$

4.3 Discrete Long Memory Methods

In long memory testing, a widely used method in estimating the Hurst exponent is the rescaled range estimator $(R/S)(n)$ that was developed by Hurst (1951). This is mainly due to its simplicity and easy to estimate and interpret. In the following sections, we discuss discrete methods of estimating the Hurst exponent. The first method that we elaborate on is the rescaled range estimator, followed by the KPSS statistic which is used to differentiate between a long memory and a short memory stationary process. Lastly we will discuss the rescaled variance estimator which is based on centering the KPSS statistic.

4.3.1 Rescaled Range estimator

A rescaled range static, $(R/S)(n)$, was formerly introduced by Mandel (1971) in finance. The $(R/S)(n)$ is used to estimate the Hurst exponent. This estimator can characterize a series, and examples of this was done in the paper by Kale and Butar (2010). Consider a time series $\{X_t\}$ of length T and divide it into N adjacent sub-periods of length $\nu = \frac{T}{N}$. Then for each of the sub-periods generated, calculate the average values and the series of accumulated standard deviations from the arithmetic mean values. This is known in literature as a profile. The range is then calculated as the difference of the maximum and the minimum values of the profile divided by the standard deviation for each sub-period. Each range (R_i) is standardised by the corresponding standard deviation (S_i) and forms a rescaled range $(R/S)(\nu)$ for the sub-period (Kale and Butar, 2010). Below we further discuss this statistic and further analysis is found in Teverovsky et al. (1999). For a time

series $\{X_t\}$ with a partial sum

$$Y(n) = \sum_{i=1}^{n} X_i, \quad n \geq 1, \tag{4.31}$$

and a sample variance

$$S^2(n) = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \frac{1}{n} Y(n) \right)^n, \quad n \geq 1. \tag{4.32}$$

The R/S static is defined

$$(R/S)(n) = \frac{1}{S(n)} \left\{ \max_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0 \leq t \leq} \left(Y(t) - \frac{t}{n} Y(n) \right) \right\}, \quad n \geq 1, \tag{4.33}$$

and this can further be related to

$$E\left[(R/S)(n)\right] \sim C_1 n^H \quad \text{as} \quad n \to \infty, \tag{4.34}$$

where H is in the interval $(0, 0.5)$, C_1 is finite positive constant independent of n. Then

$$ln\left(E\left[(R/S)(n)\right]\right) = lnC_1 + Hln(n), \tag{4.35}$$

which can be re-parameterised by letting $\alpha = lnC_1$ and $\beta = H$ and hence resulting in the following

$$\eta = \alpha + \beta ln(n). \tag{4.36}$$

The above log equation is used to graphically estimate H, where we equate the estimator of H with the slope of the graph. If the data process is white noise then the plot is a straight line with slope of 0.5. If the process is persistent the slope is greater than 0.5. And if the data process is anti-persistent, the slope is less than 0.5. The values that are used for the expected values of the R/S estimate are approximated by Granero et al. (2008) using

$$E\left[(R/S)(n)\right] = \begin{cases} \frac{n-\frac{1}{2}}{n} \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} & \text{for} \quad n \leq 340 \\ \frac{n-\frac{1}{2}}{n} \frac{1}{\sqrt{n\frac{\pi}{2}}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} & \text{for} \quad n \geq 340 \end{cases}, \tag{4.37}$$

and we must note that the above was calculated using the normality assumption in Teverovsky et al. (1999). The results that were found by Granero et al. (2008) using the above formula was not consistent with their data and thus they opted to using the linear regression of the classical R/S analysis with an extra term

$$lnH_n = ln(R/S)(n) - lnE(R/S)(n) + ln(n)/2,25 \tag{4.38}$$

where the expectation is defined above. To find H, they then used the regression (Granero et al., 2008)

$$ln(H_n) = ln(C_1) + Hln(n). \tag{4.39}$$

If the time series is short term, then the relation becomes

$$E\left[(R/S)(n)\right] \sim C_1 n^{\frac{1}{2}} \quad \text{as} \quad n \to \infty. \tag{4.40}$$

The difference between the two estimates above is known as the *Hurst Effect*.

The above R/S statistic has been said to be too sensitive to short-term memory thus Lo (1991) introduced a modified statistic. Consider the above estimator obtained from 4.36, instead of normalising R with $S^2(n)$, normalise it with the weighted sum of auto-covariances

$$S_q(n) = \left(S^2(n) + 2\sum_{j=1}^{q} \omega_j(q)\hat{\gamma}_j \right), \tag{4.41}$$

where $\hat{\gamma}_j$ are sample autocovariances and weights $\omega_j(q)$ are such that

$$\omega_j(q) = 1 - \frac{j}{q+1}, \quad q < n, \tag{4.42}$$

where $q = 0$ is similar to the classical R/S statistic and it produces the sample variance which is used to normalise the range. The modified R/S statistic is defined in Teverovsky et al. (1999) as

$$V_q(n) = \frac{1}{\sqrt{n}} \left[\frac{R}{S_q(n)} \right]. \tag{4.43}$$

The main challenge outlined in Teverovsky et al. (1999) is to find an appropriate choice

of q. A data driven approach used is estimating the optimum q by the formula

$$q_{opt} = \left[\left(\frac{3n}{2} \right)^{1/3} \left(\frac{2\hat{\rho}}{1-\hat{\rho}}^2 \right)^{2/3} \right], \tag{4.44}$$

where n is the length of data, $\hat{\rho}$ is the estimated first-order auto-correlation coefficient and [.] is the greatest integer function.

R/S have been further tested by Kristoufek (2012) on how is it affected by different memory and distributional properties. It was found that the modified R/S is biased if there is presence of short memory in the time series, the classical R/S was also tested and found that it is biased toward short-range dependence. The conclusions of Ladislav and Petra (2013) agreed with what other authors in this field have recommended that this estimate must not be used in isolation, it must be used in conjunction with other tests.

4.3.2 The KPSS Statistic

The KPSS (Kwiatkowski et al., 1992) statistic is used to test stationarity against unit root. This test is mostly used to distinguish between short and long memory processes. This statistic was used by Lee and Schmidt (1996) to test long memory in stationary time series. The KPSS statistic is defined as:

$$\eta = \frac{1}{T^2 \hat{\sigma_T}^2(q)} \sum_{\forall t} S_t^2, \tag{4.45}$$

where S_t is the partial sum $\sum \hat{e_i}$, $\{\hat{e_i}\}$ are the residuals of the regressions and $\hat{\sigma_T^2}(q)$ is the estimate of the variance of the residuals. A similar statistics to the KPSS and the rescaled range statistic is the rescaled variance statistic introduced by Giraitis et al. (2001).

4.3.3 The Rescaled Variance Statistic

The Rescaled Variance (V/S) statistic is based on centering η the KPSS statistic and is defined as

$$(V/S) = \frac{1}{T} \hat{V} \left(S_1^*, \ldots, S_T^* \right) * \frac{1}{\hat{\sigma_T^2}(q)}, \tag{4.46}$$

where $S_j^* = \sum_{i=1}^{j} (X_i - \bar{X}_N)$, the (V/S) is the sample variance of the series of the partial sum. The (V/S) has uniformly higher power that the KPSS statistic and is less sensitive

to the choice of the order q than the modified (R/S) statistic. For a level series $2 \leq q \leq 10$ detects the existence of long memory. The properties of this test are further discussed in Giraitis et al. (2003). The R/S, V/S and KPSS statistic are time domain tests. Other time domain methods that could be used are the Aggregated variance method, Differenced aggregated variance method, Aggregated absolute value method, Fractal (Higuchi) dimension method and variance of residuals (Peng) method among other. As an alternate to these tests are frequency domain estimation methods which we discuss in the following sub section. As an alternate to these tests, frequency domain estimation methods have been proposed in literature.

4.4 Semi-Parametric Long Memory Methods

In frequency domain analysis, there are two widely used statistical procedures in the class of semi-parametric estimation which are based on the log periodogram namely, (1) Geweke, Porter and Hudak (GPH) estimator (Geweke and Porter-Hudak, 1983) and (2) the Local Whittle Estimator (LW). LW known to be more effecient than GPH regression in the stationary case (Shimotsu and Phillips, 2005).

4.4.1 Log Periodogram Regression

Geweke and Porter-Hudak (1983) proposed a semi-parametric approach to the testing of long memory. For a given a fractionally integrated process, its spectral density is given by

$$f(\omega) = [2\sin(\omega/2)]^{-2d} f_u(\omega), \tag{4.47}$$

where ω is the Fourier frequency, $f_u(\omega)$ is the spectral density and u_t is a stationary short memory disturbance with a zero mean. The log periodogram regression is based on

$$\ln[f(\omega_j)] = \ln[f_u(0)] - d\ln\left[4\sin^2\left(\frac{\omega_j}{2}\right)\right] + \ln\left[\frac{f_u(\omega)}{f_u(u)}\right]. \tag{4.48}$$

This then becomes

$$\ln[I(\omega)] = a + d\ln\left[4\sin^2\left(\frac{\omega_j}{2}\right)\right] + \zeta_\omega, \tag{4.49}$$

which we can re-parameterise as

$$y_j = a + dx_j + \zeta_j, \tag{4.50}$$

where $y_j = \ln[I(\omega_j)]$ and $x_j = \ln\left[4\sin^2\left(\frac{\omega_j}{2}\right)\right]$. The long range dependence parameter is estimated as

$$\hat{d} = \frac{N\bar{x}\bar{y} - \sum_{j=1}^{m} y_i x_i}{\left(\sum_{j=1}^{m} x_i^2 - n\bar{x}^2\right)}, \tag{4.51}$$

where $m = g(T)$ and this estimator is asymptotically normal in large samples, i.e.

$$\hat{d} \sim N\left(d, \frac{\pi^2}{6\sum_{j=1}^{m}(X_j - \bar{X})}\right). \tag{4.52}$$

As $T \to \infty$ we get

$$\sqrt{m}(\hat{d} - d) \sim N\left(0, \frac{\pi^2}{6\sum_{j=1}^{m}(X_j - \bar{X})}\right), \tag{4.53}$$

and the parameter m must be selected such that $m = T^{\nu}$, $0 < \nu < 1$. This formulation assumes OLS and hence an OLS estimate is derived with error terms being independent and identically Guassian distributed. In the following Section we discuss the Whittle estimator.

4.4.2 Whittle Estimator

The Whittle estimator is the second semi-parametric estimator based on the periodogram (Taqqu and Teverovsky, 1996). It is the estimator which minimizes the function

$$Q(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} \frac{I(\omega)}{f(\omega;\boldsymbol{\theta})} d\omega, \tag{4.54}$$

where $I(\omega)$ is the periodogram

$$I(\omega) = \frac{1}{2\pi N}\left|\sum_{i=1}^{N} X_j e^{ij\omega}\right|^2, \tag{4.55}$$

and $f(\omega;\theta)$ is the spectral density at frequency ω and $\boldsymbol{\theta}$ denotes the vector of unknown parameters. Under fractional integrated model, θ is the Hurst exponent (Shimotsu and

Phillips, 2005). This means that the Whittle estimator of θ_0 is

$$\hat{\theta_T} = \arg\min_{\theta \epsilon \Theta} Q(\boldsymbol{\theta}), \tag{4.56}$$

where $Q(\boldsymbol{\theta})$ is

$$Q(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left[\ln(f(\omega_t;\theta)) + \frac{I(\omega_t)}{f(\omega_t;\theta)}\right], \quad \omega_y = \frac{2\pi t}{T}. \tag{4.57}$$

The Local Whittle estimator of θ is known to have the limiting distribution (Baillie and Kapetanios, 2007; Zaffaroni, 2009)

$$m^{\frac{1}{2}}\left(\hat{d} - d_0\right) \rightarrow N\left(0, \frac{1}{4}\right), \tag{4.58}$$

where d_0 denotes the true value of d and m represents the choice of bandwidth such that $m \leq T^{4/5}$. Long memory at times results from occassional mean breaks which introduces structural breaks in data. Such breaks may result in spurious long memory. Thus in the estimation of long memory, we need to cater for structural breaks that might arise and we discuss this phenomena in the following Section.

4.5 Long Memory and Structural Breaks

A number of methods have been introduced in literature to address structural breaks in time series processes. In this section we discuss the ICSS algorithm which is used to identify breaks in the mean and for modeling long range dependence breaks, we apply the split methodology and visual inspection of cumulative samples methodology.

4.6 Iterative Cumulative Sum of Squares Algorithm

The ICSS algorithm due to Inclan and Tiao (1994) is one of the most widely used algorithm in testing for structural breaks. To calculate the ICSS, we first need to calculate the Inclan-Tiao(IT) statistic which is given by

$$IT(k) = \sup_k \left| \left(\frac{T}{2}\right)^{\frac{1}{2}} D_k \right|,$$

where T is the length of the series, $D_0 = D_T = 0$, and $D_k = \left[\frac{C_k}{C_T}\right] - \frac{k}{T}$ for $k = 1, 2, ..., T$ with $C_k = \sum_{t=1}^{k} r_t^2$. The null hypothesis when using this test is that the unconditional variance is constant against the alternative that there is a break at some point in the data. The value of k that maximise $IT(k)$ is the estimate of the break time point in the data series and the factor $(T/2)^{0.5}$ is used to standardise the distribution of D_k. If there are no breaks in the time series, the value of D_k revolves around zero. If breaks are detected then D_k revolves away from zero. Under the null hypothesis, D_k is asymptotically a Brownian bridge, i.e. $\hat{W}_r = W(r) - rW(1)$ where $W(r)$ is a standard Brownian motion.

Since the $IT(k)$ statistic was designed for independent and identically distributed $(i.i.d)$ processes, it has been found to be oversized when the underlying process is not $i.i.d$, therefore non-parametric Bartlett kernel adjustment is applied. The adjusted $IT(k)$ is given by

$$AIT(k) = \sup_k \left| T^{1/2} G_k \right|,$$

where $G_k = \hat{\omega}^{-1/2} \left[C_k - \left(\frac{k}{T}\right) C_T \right]$,$\hat{\omega} = \hat{\gamma}_0 + 2\sum_{l=1}^{m} \left[1 - l(m+1)^{-1}\right] \hat{\omega}_l$,$\hat{\omega}_l = \frac{1}{T}\sum_{t=l+1}^{T}(r_t^2 - \hat{\sigma}^2)(r_{t-l}^2 - \hat{\sigma}^2)$ and $\hat{\sigma}^2 = T^{-1}C_T$. The lag truncation parameter m is selected under the Newey and West (1994) methodology.

Inclan and Tiao (1994) computed the critical value of 1.358 being the 95th percentile of the asymptotic distribution of D_k. Thus, a breakpoint in variance will be identified if it exceeds the boundary ± 1.358 in the D_k plot. We must note that the function D_k alone is not powerful when a series contains multiple breaks. To address this shortcoming, Inclan and Tiao (1994) developed an algorithm that uses the D_k function to systematically check for change points in the series. This algorithm evaluates the function D_k over different time periods. In the next section, we discuss structural breaks diagnosis in long memory.

4.6.1 Sample split and differencing methodology

When long memory is due to structural changes in data, it is known as spurious long memory. A method that is used in literature to detect spurious long memory is due to Shimotsu (2006). In this method, the series of returns is split into b sub-samples and for each sub-sample, the long memory parameter is estimated using the local Whittle estimator.

Let the process of a split series be $\{X_t : t = (a-1)T/b + 1, ..., aT/b\}$ where $a = 1, ..., b$. We assume the periodogram ordinates $2\pi/(T/b), ..., 2\pi(m/b)/(T/b)$ as calculated in Shimotsu (2006). Thus, for each of the samples we get

$$\hat{d}^{(a)} = \arg \min_{d \in [\alpha_1, \alpha_2]} R^{(a)}(d), \tag{4.59}$$

where

$$R^{(a)}(d) = \ln[\hat{G}^{(a)}(d)] - 2d\frac{b}{m}\sum_{j=1}^{m/b} \ln(\omega_j),$$

$$\hat{G}^{(a)}(d) = \frac{b}{m}\sum_{j=1}^{m/b} \tilde{\omega}_j^{2d} I^{(a)}(\tilde{\omega}_j), \quad \tilde{\omega}_j = \frac{2\pi j}{T/b}, \quad j = 1, ..., T/b, \tag{4.60}$$

and

$$I^{(a)}(\tilde{\omega}_j) = \frac{1}{2\pi T}\left| \sum_{t=(a-1)T/b+1}^{aT/b} X_t e^{it_j} \right|^2. \tag{4.61}$$

The assumptions of this methodology can be found in Robinson (1995). A first test that can be used is the visual inspection. This test however can only be used to detect the structure of the data. In testing, the null hypothesis is

$$H_0 : d = d^{(1)} = d^{(2)} = ... = d^{(k)}$$

against the alternative of structural change hypothesis, where $\hat{d}^{(a)}$ is the value of d from the ath sample. The sample that is split into b sub-samples has

$$\hat{\boldsymbol{d}}_b = \begin{pmatrix} \hat{d} - d_0 \\ \hat{d^{(1)}} - d_0 \\ \vdots \\ \hat{d^{(b)}} - d_0 \end{pmatrix},$$

and

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & ... & 0 \\ . & . & . & . \\ 1 & 0 & ... & -1 \end{pmatrix},$$

where d_0 is the true parameter and $\hat{d}$ is the estimate of the total sample. We can define

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{J}_b' \\ \boldsymbol{J}_b & b\boldsymbol{I}_b \end{pmatrix},$$

where $\boldsymbol{I}_i$ is a $b \times b$ identity matrix, and $\boldsymbol{J}_i$ is a vector of ones. The Wald statistic and the adjusted Wald statistic under H_0 are

$$W = 4m\boldsymbol{A}\hat{\boldsymbol{d}}_b \left(\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\right)^{-1} \left(A\hat{\boldsymbol{d}}_n\right)'$$

and

$$W_c = 4m \left(\frac{c_{m/b}}{(m/b)}\right) \boldsymbol{A}\hat{\boldsymbol{d}}_b \left(\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\right)^{-1} \left(A\hat{\boldsymbol{d}}_n\right)'$$

respectively, where the correction c_m is given by

$$c_m = \sum_{j=1}^{m} v_j^2 \quad v_j = \log j - \frac{1}{m}\sum_{j=1}^{m} \log j \quad for \quad m < T.$$

Under H_0, W_c has an asymptotic χ^2 distribution with $n-1$ degrees of freedom.

Another test that is used by Shimotsu (2006) is based on the behavior of the differenced series. If an $I(d)$ series is differenced $\hat{d}$ times, where $\hat{d}$ is the consistent estimator of d, the resulting series should be $I(0)$. Two tests that are used on the differenced series are the KPSS test statistic and the Phillips Perron test statistic. The first step in this method is to demean the series

$$X_t - \mu_0 = (1-L)^{-d_0} u_t 1_{\{t \geq 1\}}, \tag{4.62}$$

where $1_{\{t \geq 1\}}$ is an indicator function. The mean of the process X_t is estimated by the sample average $\bar{X}$ when $d_0 < 1$. When $d_0 > 1$, the mean can be estimated by X_1 and to

cater for both of these conditions, an adjustment $\hat{\mu}(\hat{d})$ can be used as an estimator of the mean of the process. The dth differenced series is then defined as

$$\hat{u}_t = (1-L)^d(X_t - \hat{\mu}(\hat{d})), \tag{4.63}$$

and then we apply the unit roots test to $\hat{u}_t$. The adjustment $\hat{\mu}(\hat{d})$ is

$$\hat{\mu}(\hat{d}) = w(d)\bar{X} + (1-w(d))X_1, \tag{4.64}$$

where $w(d)$ is a smooth weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq 3/4$. When $d \in (1/2, 3/4)$, then the value of $w(d)$ that could be used is $1/2[1 + \cos(4\pi d)]$.

4.6.2 Cumulative Samples Methodology

In the literature, most authors use sub-sample splitting methodology, as outlined in the previous section, to investigate breaks in long memory using the long range dependence parameter, d This method is attractive as it is simple to implement. However, it is difficult to identify the dynamics of long memory in data overtime. A method that can assist is the use of cumulative samples which can identify the structure of the long range dependence parameter estimates and possible regimes that might exist overtime.

For a sample of size T, we can break it into b sub-samples each of size t_i, $i = 1, 2, ..., b$ such that the total sample size is $T = t_1 + t_2 + ... + t_b$, for all $t_i > t_{i-1}$. For the purpose of defining cumulative sub-samples, let $T(i) = t_1 + t_2 + ... + t_i$ and define $\tilde{X}_i$ as a cumulative sub-sample of size $T(i)$. Then, we can define the $i-$th cumulative sample as

$$\begin{aligned} \tilde{X}_i &= \{X_j\}_{j=1}^{T(i)} \\ &= \{X_j\}_{j=1}^{t_1+t_2+...+t_i} \\ &= \{X_j\}_{j=1}^{\sum_{k=1}^{i} t_k}. \end{aligned} \tag{4.65}$$

For each of the cumulative sub-sample, we can then calculate the long range dependence parameter estimate, $\hat{d}$

$$\hat{d}_{(i)} = \arg \min_{d \in [\alpha_1, \alpha_2]} R^{(i)}(d), \tag{4.66}$$

where d_i is the parameter estimate of the $i-$th cumulative sub-sample. Plot the estimates to visually view the evolution of long range dependence in data overtime. However, we must note that this method is only a visual inspection of the long range dependence parameter estimates overtime.

4.7 Chapter Summary

Similar to short memory time series processes, stationarity in long memory processes play an integral role. To summarise this chapter, we started by definitions of covariance stationary processes. This property is important as covariance stationary processes exhibit long memory for unsummable autocovariances. The hurst exponent is then used to measure the degree of long memory.

Self-similar processes are stochastic processes that are closely related to long memory processes. This is as a result of self-similar process increments generating long memory under certain conditions. From the idea of self-similar processes, we then discussed a parametric long memory model, ARFIMA. ARFIMA is widely used in econometrics literature to model return data which exhibit long memory. There are discrete and non-parametric methods which are used in estimating the long range dependence parameter. Discrete methods in time domain discussed are

- Rescaled range statistic,
- KPSS statistic,
- Rescaled variace statistic,
- Aggregated variance method,
- Differenced aggregated variance method,
- Aggregated absolute value method,
- Fractal dimension method, and
- Variance of residuals method,

and the semi-parametric methods in frequency domain are the

- Log periodogram regression, and
- Whittle estimator.

Because of jumps and breaks in data, these methods produce uncertain results. In order to overcome this, structural breaks methods are used to test level of breaks and how it affects the memory of data series. The first method we discussed is the Inclan-tiao which uses an iterative cumulative sum of squares algorithm to identify breaks. The second method compares the equality of long range dependent parameters derived from samples.

Chapter 5

Heavy Tailed Distributions

Heavy tailed error distributions have been widely used in the econometrics literature to model financial asset returns, an example would be a paper by Wright (2000) and Huang et al. (2014). Recently generalized hyperbolic distributions have been used to model heavy tailed data since they are able to capture information in the tails of empirical distribution. This family of distributions include hyperbolic distribution, the student-t distribution, the Laplace distribution, the generalized inverse Gaussian distribution, exponential distribution, the variance-gamma distribution and the skewed generalised hyperbolic student-t distribution.

In the theory of heavy tailed distributions, the upper tail of the distribution is important as this is studied to get the tail behavior and properties of the distribution. The upper tail of a random variable X with distribution function $F(X)$ is defined

$$S_X(x) = P(X > x). \tag{5.1}$$

In life analysis and linear models applications this is known as the survival function. The tail behavior can be studied by observing $S_X(x)$ as x increases. Tail behavior has been associated with Pareto-type tails. A distribution has Pareto-type tails if it decays like a power function as x becomes large. This suggests that $S_X(x)$ can be explained by

$$S_X(x) = x^{-\alpha}Y(x), \quad 0 < \alpha < 2, \tag{5.2}$$

where $Y > 0$ is a slowly varying function which is slower than any power function and

$-\alpha$ is an index. A slow varying function has a property

$$\lim_{x\to\infty} \frac{Y(tx)}{Y(x)} = 1, \quad t > 0 \tag{5.3}$$

hence, for any $\alpha > 0$,

$$\lim_{x\to\infty} x^{\alpha} Y(x) = \infty \tag{5.4}$$

and

$$\lim_{x\to\infty} x^{-\alpha} Y(x) = 0. \tag{5.5}$$

For large x, α is called the tail index as it controls the rate of tail decay and provides a measure of the fatness of the tails. For any $\lambda > 0$, the distribution of a random variable X with distribution function $F(x)$ is said to have a heavy right tail if

$$\lim_{x\to\infty} e^{\lambda x} S_X(x) = \infty, \quad \forall \lambda > 0. \tag{5.6}$$

The above definition uses a class of distributions called subexponential distributions. For this type of distributions, the decay rate is slower than the exponential. There have been a number of distributions that have been proposed in literature to model heavy tails. Alpha-stable distributions are said to be the origin of heavy tailed distributions in finance. The use of this type of distributions have been widely reviewed in the econometrics literature. This is because it is an alternative to the Gaussian law and contains some interesting properties such as the stability property. This property is attractive in that the index of the alpha-stable remains the same under scaling and addition of different stable random variables with the same shape parameter. This type of phenomenon raises an interesting feature of these type of distributions which is referred to as *infinite divisibility*. In the next section, we expand on infinite divisibility and example of processes which have this property.

5.1 Infinite Divisibility

In modeling of financial returns, infinitely divisible distributions are attractive and appealing as they make modeling of returns simpler. A random variable X is said to be infinitely divisible if for every $n \in \mathbb{N}$ it can be written, in distribution, as

$$X \stackrel{d}{=} X_1 + ... + X_n, \tag{5.7}$$

where $X_1, ..., X_n$ are independent and identically distributed random variables. An alternative definition of infinite divisibility is by the use of the characteristic function. A random variable X is said to be infinitely divisible, if for all $n \in \mathbb{N}$, there exist a random variable $X^{(1/n)}$ such that

$$\phi_X(u) = (\phi_{X^{(1/n)}}(u))^n, \tag{5.8}$$

where $\phi_X(u)$ denotes the characteristic function. By definition a distribution F is infinitely divisible if for each n, there exists a distribution function F_n such that F is the n-fold convolution $F_n * ... * F_n$ of F_n. For any finite non-negative measure

$$\phi_X(u) = \exp\left\{dx \int_{\mathbb{R}} (e^{iux} - 1 - iux)\frac{1}{x^2}\mu(dx)\right\} \tag{5.9}$$

is the characteristic function of infinite divisibility distribution with mean 0 and variance $\mu(R)$. Every infinite divisible distribution with mean 0 and finite variance is the limiting law of $S_n = X_{n,1} + ... + X_{n,r_n}$ for some independent triangular array satisfying

1. $E(X_{n,k}) = 0$,
2. $\lim_{n\to\infty} \max_{1\leq k\leq r_n} E(X_{n,k}^2) = 0$,
3. $\sup_{n\geq 1} S_n^2 < \infty$, where $S_n^2 = \sum_{k=1}^{r_n} E(X_{n,k}^2)$.

Levy processes are stochastic processes that are closely related to infinite divisibility. To understand this, we will discuss these processes below. A stochastic process $\{X_t, t \geq 0\}$ is a Levy process if

i.) $X_0 = 0$ almost surely,

ii.) $X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ are independent for any $a \leq t_1 < t_2 < ... < t_n < \infty$,

iii.) $X_t - X_s$ is equal in distribution to X_{t-s} for any $s < t$, and

iv.)

$$\lim_{h \to 0} P(|X_{t+h} - X_t| > \epsilon) = 0 \tag{5.10}$$

for any $\epsilon > 0$ and $t \geq 0$.

v.) X_t is right continuous with left limits almost surely.

The Levy-Khintchine formula is fundamental in the study of Levy processes. Let X be a Levy process in $\mathbb{R}^d$. Then there exists a triplet (A, γ, ν) of A a symmetric non-negative definite $d \times d$ matrix (the Gaussian covariance), γ a constant in $\mathbb{R}^d$ and ν a measure on $\mathbb{R}^d$ with ν_0=0 and $\int_{\mathbb{R}^d}(|y|^2 \wedge 1)d(\nu(y)) < \infty$ (the Levy measure). Which in that case is uniquely determined such that for all $u \in \mathbb{R}^d$ and $t \geq 0$, $E(e^{iux(t)}) = e^{t\psi(u)}$, where

$$\psi(u) = -1/2 < u, Au > +i < u, \gamma > + \int_{\mathbb{R}^d}(e^{i<u,x>} - 1 - 1_{(|x|\leq 1)} i < u, x >)d\nu(x). \tag{5.11}$$

If $\gamma_0 = \gamma - \int_{|x|\leq 1} x d\nu(x)$ is well defined and finite, then we may re-write the Levy-Khintchine formula with a new triplet $(A, \gamma_0, \nu)_0$ (the drift) as

$$\psi(u) = -1/2 < u, Au > +i < u, \gamma > + \int_{\mathbb{R}^d}(e^{i<u,x>} - 1)d\nu(x). \tag{5.12}$$

If $\gamma_1 = \gamma + \int_{|x|\leq 1} x d\nu(x)$ is well defined and finite, the we can re-write the Levy-Khintchine formula with a new triplet $(A, \gamma_1, \nu)_1$ (the center) as

$$\psi(u) = -1/2 < u, Au > +i < u, \gamma > + \int_{\mathbb{R}^d}(e^{i<u,x>} - 1 - i < u, x >)d\nu(x). \tag{5.13}$$

Thus a random variable X is infinitely divisible if and only if its characteristic function $\phi_X(u)$ is given by

$$log\phi_X(u) = i\beta u - \frac{\sigma^2\mu^2}{2} + \int_{-\infty}^{\infty}(e^{iux} - 1 - iux1_{|x|\leq 1})\nu d(x) \tag{5.14}$$

for some $\beta \in \mathbb{R}$, $\sigma \geq 0$ and a Levy measure ν such that $\int(x^2 \wedge 1)\nu(dx) < \infty$.

There are two widely used Levy processes that are applicable to financial returns. These are the Brownian motion and the Poisson process.

If X_t is a Brownian motion process (or Weiner Process), the probability of the increments $X_t - X_s$ have a normal distribution with an expected value of 0 and the variance $t - s$. If X_t is a Poisson process, the probability of increments $X_t - X_s$ have a Poisson distribution with an expected value of $\lambda(t - s)$.

From this discussion, we can now show that stable distributions are infinitely divisible distributions.

Theorem 5.1. *If a distribution of X is stable, then the distribution is infinitely divisible.*

Proof. let $n \in \mathbb{N}$ and $\{X_t\}$ be a sequence of independent variables, each with the same distribution as X. By the stability property, there exists $a_n \in \mathbb{R}$ and $b_n \in (0, \infty)$ such that $\sum X_i$ has the same distribution as $a_n + b_n X$. But

$$\frac{1}{b_n}\left(\sum_{i=1}^{n} X_i - a_n\right) = \sum_{i=1}^{n} \frac{X_i - a_n/n}{b_n} \tag{5.15}$$

has the same distribution as X. Thus, it can then be deduced that $\{(X_i - a_n/n)/b_n\}$ is an $i.i.d$ sequence and hence has the same distribution as X and is infinite divisible. □

Examples of distributions that are stable and infinitely divisible are the normal and the Cauchy distribution. If X has a normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in (0, \infty)$, then for $n \in \mathbb{N}$, X has the same distribution as $X_1 + X_2 + ... + X_n$ where $\{X_i\}$ are independent, and X_i has a normal distribution with mean μ/n and standard deviation $\sigma/\sqrt{n}$ for $i = 1, 2, ..., n$. If X has a Cauchy distribution with location parameter $a \in \mathbb{R}$ and scale parameter $b \in (0, \infty)$, then for $n \in \mathbb{N}$, X has the same distribution as $X_1 + X_2 + ... + X_n$ where $\{X_i\}$ are independent, and X_i has a Cauchy distribution with location parameter a/n and scale parameter b/n for $i = 1, 2, ..., n$.

5.2 Stable Distributions

A class of heavy tailed distributions in common use are stable distributions. A random variable X is stable if two random variables X_1 and X_2 are independent copies of X for some positive constants a and b, the distribution of $aX_1 + bX_2$ is equal to the distribution of $cX + d$. X is then considered strictly stable if $d = 0\ \forall a, b$. X is symmetric stable if the distribution of X is equal to the distribution of $-X$. In other words, if random variables

$X_1, X_2, ..., X_T$ are *i.i.d*, they are then considered to be α-stable if the distribution of $X_1 + X_2 + ... + X_T$ is the same as the distribution of $c_T X + \alpha_T$, where c_T is of the form $T^{1/\alpha}$. The α-stable distribution is characterised by parameters:

- Characteristic exponent α, $0 < \alpha \leq 2$,
- Skewness parameter β, $-1 \leq \beta \leq 1$,
- Scale or gamma parameter $\gamma > 0$, and
- Location parameter δ.

The characteristic exponent α is the rate at which the tails decay. Then, a random variable X with the above parameters is denoted as $X \sim S(\alpha, \beta, \sigma, \mu)$. Only three α-stable distributions have closed form density functions and that is the Normal, Cauchy and Levy distribution. If X follows $N(\mu, \sigma^2)$ then its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

and if $X \sim Cauchy(\gamma, \delta)$ then it has a density

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\delta)^2}, \quad -\infty < x < \infty$$

and if $X \sim Levy(\gamma, \delta)$ then it has a density

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\delta)}\right), \quad \delta < x < \infty.$$

In some α-stable distributions, it is a bit difficult to express the density function, as an alternative, the characteristic function is used. Any probability distribution can be defined by its characteristic function as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) e^{-ixt} dt.$$

The characteristic function of X assumes

$$\ln \phi_X(t) = -\sigma^\alpha |t|^\alpha (1 - i\beta \text{sgn}(t) \tan(\pi\alpha 2) + i\mu t).$$

When $\alpha \neq 1$ and when $\alpha = 1$ it assumes

$$\ln\phi_X(t) = -\sigma|t|(1 - i\beta\frac{\pi}{2}\text{sgn}(t)ln|t|) + i\mu t).$$

Another form of the characteristic function is

$$\ln\phi_X(t) = -\sigma_2^\alpha|t|^\alpha(1 - i\beta_2\text{sgn}(t)\frac{\pi}{2}K(\alpha) + i\mu t).$$

where $K(\alpha) = \alpha$ for $\alpha \leq 1$ and $K(\alpha) = \alpha - 2$ otherwise, this is further discussed in **?** . $\sigma_2 = \sigma(1 + \beta^2\tan^2(\frac{\pi\alpha}{2}))^{1/2\alpha}$ and $\tan(\beta_2\frac{\pi K(a\alpha)}{2}) = \beta\tan(\frac{\pi\alpha}{2})$ for $\alpha \neq 1$, and for $\alpha = 1$

$$\ln\phi_X(t) = -\sigma|t|(\frac{\pi}{2} + i\beta_2\frac{\pi}{2}\text{sgn}(t)\ln|t|))$$

Some useful properties of stable distributions are as follows:

1. If X_1 and X_2 are independent stable random variables with $X_i \sim S(\alpha_i, \beta_i, \sigma_i, \mu_i)$ for $i = 1, 2, ..., n$, then $X_1 + X_2 \sim S(\alpha, \beta, \sigma, \mu)$ with

$$\begin{aligned}
\sigma &= \left(\sum_{i=1}^n \sigma_i^\alpha\right)^{1/\alpha}, \\
\beta &= \frac{\sum_{i=1}^n \beta_i\sigma_i^\alpha}{\sum_{i=1}^n \sigma_i^\alpha}, \\
\mu &= \sum_{i=1}^n \mu_i.
\end{aligned}$$

2. $X + a$ $S(\alpha, \beta, \sigma, \mu + a)$ if X $S(\alpha, \beta, \sigma, \mu)$ and some constant a.

3. The tail of the density function decays like a power function

$$P(|X| > x) \approx Cx^{-\alpha}, \quad x \to \infty$$

4. Moments satisfy

$$\begin{aligned}
&E(|X|)^p < \infty, \quad 0 < p < \alpha, \\
&E(|X|) = \infty, \quad p \geq \infty.
\end{aligned}$$

5. The expectation of an α - Stable distribution is

$$\begin{aligned} E(X) &= \mu, \quad \alpha > 1, \\ E(X) &= \infty, \quad , \alpha \leq 1. \end{aligned} \tag{5.16}$$

The following method is used to generate stable distribution random numbers. The parameters used in the algorithm are that $U \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$, E is a standard Exponential random variable independent of U, and $U_0 = -\frac{\pi}{2}\beta_2 \frac{K(\alpha)}{\alpha}$ and the random variable is defined

$$\begin{aligned} X &= \frac{\sin\alpha(U - U_0)}{(\cos U)^{1/\alpha}} \left\{ \frac{\cos(U - \alpha(U - U_0))}{E} \right\}^{\frac{1-\alpha}{\alpha}} \quad \alpha \neq 1 \\ X &= \left(\frac{\pi}{2} + \beta_2 U\right) \tan(U) - \beta_2 \ln\left(\frac{E\cos(U)}{\pi/2 + \beta_2 U}\right) \quad , \alpha = 1 \end{aligned}$$

and the algorithm steps to generate stable numbers is

1. Generate U_1 $U(-\frac{\pi}{2}, \frac{\pi}{2})$ and U_2 $U(0, 1)$.

2. E=-lnU_2,

3. If $\alpha \neq 1$ then $B(\alpha, \beta) = \frac{\tan^{-1}(\beta \tan(\pi\alpha/2))}{\alpha}$, $S(\alpha, \beta) = (1 + \beta^2 \tan^2(\pi\alpha/2)^{1/2\alpha}$ and then

$$X = S(\alpha, \beta) \frac{\sin\alpha(U + B(\alpha, \beta))}{(\cos(U))^{1/\alpha}} \left(\frac{\cos(U - \alpha(U + B(\alpha, \beta)))}{E} \right)^{\frac{1-\alpha}{\alpha}}$$

   and if $\alpha = 1$,

$$X = \frac{2}{\pi} \left(\frac{\pi}{2} + \beta U\right) \tan(U) - \beta \ln\left[\frac{\frac{\pi}{2} E\cos(U)}{\frac{\pi}{2} + \beta U} \right],$$

4. Set

$$\begin{aligned} Y &= \sigma X + \mu \quad \text{for} \quad \alpha \neq 1, \\ Y &= \sigma X + \frac{2}{\pi} \beta\sigma \ln(\sigma) + \sigma \quad \textit{for} \quad \alpha = 1. \end{aligned}$$

5.3 Generalised Hyperbolic Distributions

The generalised hyperbolic (GH) distribution is given by

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

with

$$a(\lambda,\alpha,\beta,\delta,\mu)=\frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})},$$

where $K_\lambda(.)$ is the modified Bessel function of the third kind, $\delta \geq 0$ and $0 < |\beta| < \alpha$. If $\beta = 0$, it is symmetric and elliptical and if $\alpha \to \infty$ and $\delta \to \infty$ such that $\delta/\alpha \to \sigma^2$ then the $gh(x)$ is the normal distribution with mean $\mu + \beta\sigma^2$ and variance σ^2.

If $\lambda = 1$, the resulting distribution is the mostly used hyperbolic distribution. The hyperbolic distribution is given by

$$hyp(x|\alpha,\beta,\delta,\mu)=a(\alpha,\beta,\delta,\mu)exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\right\},$$

where

$$a(\alpha,\beta,\delta,\mu)=\sqrt{\alpha^2-\beta^2}\left[2\delta\alpha K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)\right]^{-1}$$

for $\delta > 0$ and $0 \leq \beta < \alpha$. In the literature, the (ξ,χ) parameterization is normally used. then the hyperbolic distribution is defined as

$$hyp(x|\xi,\chi,\delta,\mu) =$$
$$a(\xi,\chi,\delta,\mu)exp\left\{-\frac{1-\xi^2}{\sqrt{\xi^2-\chi^2}}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}-\frac{\chi(1-\xi^2)}{\xi^2\sqrt{\xi^2-\chi^2}}\left(\frac{x-\mu}{\delta}\right)\right\} \quad (5.17)$$

where

$$a(\xi,\chi,\delta,\mu)=\left[2\delta\frac{\xi}{\xi^2-\chi^2}K_1\left(\frac{1-\xi^2}{\xi^2}\right)\right]^{-1}.$$

If $\beta \to 0$, the resulting distribution is the widely known and used Student-t distribution.

The limits for ξ and χ are $0 \leq |\chi| < \xi < 1$ which provides a useful presentation of the parameters known as the hyperbolic shape triangle. This can be summarised as follows:

- If $\xi \to 0$, the resulting distribution is a Normal distribution. If $\xi \to 1$, we get the Laplace distribution,
- If $\chi \to \pm\xi$ we get the generalised Inverse Gaussian distribution and if $|\chi| \to 1$ the resulting distribution is the Exponential distribution.

The Normal Inverse Gaussian distribution results if $\lambda = -\frac{1}{2}$ and is given by

$$\begin{aligned} &nig(x|\alpha,\beta,\delta,\mu) = \\ &a(\alpha,\beta,\delta,\mu)\left(\delta^2+(x-\mu)^2\right)^{-1/2} K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})exp(\beta(x-\mu)), \end{aligned} \quad (5.18)$$

where

$$a(\alpha,\beta,\delta,\mu) = \pi^{-1}\delta\alpha(\delta\sqrt{\alpha^2-\beta^2}), \quad \delta > 0, \quad 0 \leq |\beta| < \alpha$$

and the density function is obtained from the GH density. If $\beta > 0$ and $\delta = 0$, this results in a limiting case known as the variance-gamma distribution and is given by

$$vg(x|\alpha,\beta,\delta,\mu) = \frac{(\sqrt{\alpha^2-\beta^2})^{2\lambda}|x-\mu|^{\lambda-1/2}K_{\lambda-1/2}(\alpha|x-\mu|)}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-1/2}}, \quad (5.19)$$

where $\Gamma(.)$ is the gamma function. When $\lambda = -\frac{\nu}{2}$ and $\alpha \to |\beta|$, the skewed generalised hyperbolic Student-t is obtained and is given by

$$\begin{aligned} &sgh(x|\nu,\beta,\delta,\mu) = \\ &\frac{\delta^\nu|\beta|^{(\nu+1)/2}K_{(\nu+1)/2}\left(\beta\sqrt{\delta^2+(x-\mu)^2}\right)e^{\beta(x-\mu)}}{2^{(\nu+1)/2}\Gamma(\nu/2)\sqrt{\pi}\left(\sqrt{\delta^2+(x-\mu)^2}\right)}, \end{aligned} \quad (5.20)$$

where $\delta > 0, \nu > 0, \beta \neq 0$.

5.4 Generalized Error Distributions

Another class of distributions used in volatility modeling is that of the generalized error distribution (GED). This comes from the exponential family and in this thesis we will

use parameterization by Box and Tiao (1973). The GED(μ, σ, γ) distribution is given by

$$f(y) = \omega(\gamma)\sigma^{-1}\texttt{exp}\left[-c(\gamma)\left|\frac{y-\mu}{\sigma}\right|\right] \tag{5.21}$$

where

$$\omega(\gamma) = \frac{(\Gamma(3(1+\gamma)/2))^{1/2}}{(1+\gamma)\Gamma(1+\gamma)/2)^{3/2}} \tag{5.22}$$

and

$$c(\gamma) = -\left(\frac{\Gamma(3(1+\gamma)/2)}{\Gamma((1+\gamma)/2)}\right)^{1/(1+\gamma)} \tag{5.23}$$

where $y \in (-\infty, \infty)$, μ is the location parameter, and σ is the scale parameter. γ is the shape parameter that is related to the kurtosis of the distribution. The GED is symmetric with zero skewness and kurtosis is defined by

$$k_3 = \frac{\Gamma(5(1+\gamma)/2)\Gamma((1+\gamma)/2)}{(\Gamma(3(1+\gamma)/2)^2}. \tag{5.24}$$

There are special cases of the GED. For $\gamma = 0$ the GED is the Normal distribution. if γ=1 it becomes the double exponential distribution and if $\gamma \rightarrow$-1 it becomes the uniform distribution.

5.5 Chapter Summary

In summary, heavy tailed distributions have been shown to explain financial returns better than the normal distributions. The tails of financial return data is said to behave like Pareto-tails and have infinite divisible distributions. In this chapter, we started by discussing infinite divisibility. This property is important in explaining heavy tails. In order to understand infinite divisibility, we studied Levy processes. Levy processes are stochatic processes which have the infinite divisibility property. From the definitions of Pareto tails and infinite divisibility, we discussed alpha-stable distributions and a class of generalised hyperbolic distributions which have been widely used in literature to explain financial returns.

Chapter 6

Conditional Heteroskedastic Models and Estimation

In time series analysis, especially in regression models, one of the basic assumptions is that the variance of residuals is constant. Literature and applications have shown that this assumption does not hold. Financial data is a good example where it has been shown in many research papers that in most cases the assumption of constant variance does not hold. This has opened up interest in volatility modeling and hence there are a number of models that can be used to model volatility. A class of heteroskedastic models is one class that is widely used to explain conditional volatility evolution, behaviour and properties.

6.1 ARCH and GARCH models

Consider a regression model

$$\mathbf{Y}_t = \mathbf{X'}_t + a_t, \tag{6.1}$$

where a_t are uncorrelated errors but have variances that are not constant. This error term can be modeled by

$$a_t = \sigma_t \epsilon_t, \tag{6.2}$$

where ϵ_t are *iid* random variables with mean 0 and variance 1, and σ_t is the variance of errors. The evolution of the variance can be described by

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_s a_{t-1}^2, \tag{6.3}$$

and hence, we can find the conditional variance of a_t which is given by

$$Var(a_t|\mathsf{F}_{t-1}) = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_s a_{t-1}^2, \tag{6.4}$$

where F_{t-1} is the filtration of the process $\{a_t\}$. This representation is known as the autoregressive conditional heteroscedastic model developed by Engle (1982) and it shows an important property in volatility analysis which is volatility clustering. Volatility clustering is defined when a large error tends to be followed by another large error. This model is helpful in modeling conditional volatility and deriving some facts but has disadvantages. Some of the disadvantages of this model are

- The model assumes that positive and negative shocks have the same effects on volatility as it depends on the square of the previous shocks.
- Parameter constraints becomes complicated for higher order ARCH models.
- The ARCH model does not provide information about the source of variation in financial time series.
- The ARCH model respond slowly to large isolated shocks and hence are likely to over-predict volatility.
- This model often requires many parameters to adequately describe the volatility process.

In light of the above shortcomings, Bollerslev (1986) proposed a generalisation of the ARCH model known as the generalised ARCH(GARCH) model. The GARCH representation of volatility is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-1}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2. \tag{6.5}$$

For ease of use, this formula is often represented in ARMA form by letting $\eta_t = a_t^2 - \sigma_t^2$ and then

$$a_t^2 = \alpha_0 + \sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i)\, a_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}, \tag{6.6}$$

where

$$\sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) < 1. \tag{6.7}$$

To incorporate the break points in the series, the above standard GARCH(1,1) model is then defined

$$\sigma_t^2 = \alpha_0 + d_1 D_{1t} + ... + d_n D_{nt} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{6.8}$$

where D_{it}, for $i = 1, 2, ..., n$ are indicator variables associated with each break point in the variance for all time, t. Incorporating such regime shifts, the persistence in volatility, $(\alpha_1 + \beta_1)$, is expected to decrease than when using the standard GARCH model. The main advantage is using this model over the ARCH model is that fewer parameters are required to model the time series. If the above restriction is not met, such that

$$\sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) = 1, \tag{6.9}$$

then the model is integrated and the resulting model is called an integrated GARCH(IGARCH). When using this model, the effect of errors on the conditional variance is symmetric in a sense that for the same magnitude, negative and positive shocks produces the same effect. A simple IGARCH(1,1) model is of the form

$$\sigma_t^2 = \alpha_0 + \beta \sigma_{t-1}^2 + (1 - \beta) a_{t-1}^2, \tag{6.10}$$

where $0 < \beta < 1$. To accommodate for symmetric relationship between financial variables and volatility changes, Nelson (1991) proposed as model

$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 L + ... + \beta_s L^s}{1 - \alpha_1 L - ... - \alpha_m L^m} g(\epsilon_{t-1}), \tag{6.11}$$

and

$$f(x) = \begin{cases} (\theta+\gamma)\epsilon_t - \gamma E(|\epsilon_t|) & : if \quad \epsilon_t \geq 0 \\ (\theta-\gamma)\epsilon_t - \gamma E(|\epsilon_t|) & : if \quad \epsilon_t < 0, \end{cases}$$

where α_0 is a constant and $g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)]$ and $1 - \alpha_1 L - ... - \alpha_m L^m$ are polynomials with zeros outside the unit circle. In this model, the function $g(\epsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of a_t. In some instances, the returns of a security may depend on its volatility such that

$$r_t = \mu + c\sigma_t^2 + a_t. \tag{6.12}$$

This phenomenon is addressed by the GARCH-In-Mean (GARCH-M) model. A simple GARCH-M(1,1) is of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{6.13}$$

6.1.1 Estimation

For a sample of observations $\{x_1, x_2, ..., x_n\}$ that are independently and identically distributed with a distribution function $f(x)$ and parameters θ, the joint density function is defined

$$\begin{aligned} f(x_1, x_2, ..., x_n|\theta) &= f(x_1|\theta)\times, f(x_2|\theta) \times ... \times f(x_n|\theta) \\ &= \prod_{\forall i} f(x_i|\theta) \end{aligned}$$

and this can be seen as a likelihood function

$$L(x_1, x_2, ..., x_n|\theta) = \prod_{\forall i} lnf(x_i|\theta)$$

and hence, the log-likelihood function is then defined

$$l(x_1, x_2, ..., x_n|\theta) = \sum_{\forall i} \ln f(x_i|\theta).$$

For a GARCH(p,q) model of the conditional variance. Without loss of generality, let $h_t = \sigma_t^2$. The conditional variance is then

$$a_t = \sqrt{h_t}\epsilon_t, \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j-1}^{q} \beta_j \sigma_{t-j}^2.$$

A basic case would be to assume that a_t is normally distributed with zero mean and conditional variance h_t, and then the distribution will be

$$p(a_t|a_{t-1}, ..., a_0) = \frac{1}{\sqrt{2\pi h_t}} exp\left(-\frac{a_t^2}{2h_t}\right)$$

The parameters of the log-likelihood function would then be

$$\theta = (\alpha_0, ..., \alpha_q, \beta_1, ..., \beta_p)^T$$

and hence

$$\begin{aligned} l(\theta) &= \sum_{t=q+1}^{n} l_t(\theta) \\ &= \sum_{t=q+1}^{n} \left(-\frac{1}{2} ln(2\pi) - \frac{1}{2}\ln h_t - \frac{a_t^2}{2h_t}\right). \end{aligned}$$

Then we can find

$$\frac{\partial l_t(\theta)}{\partial \theta} = \left(\frac{a_t^2}{2h_t^2} - \frac{1}{2h_t}\right)\frac{\partial h_t}{\partial \theta},$$

and

$$\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} = \left(\frac{a_t^2}{2h_t^2} - \frac{1}{2h_t}\right)\frac{\partial^2 h_t}{\partial \theta \partial \theta^T} + \left(\frac{1}{2h_t^2} - \frac{a_t^2}{h_t^3}\right)\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta^T},$$

where

$$\frac{\partial h_t}{\partial \theta} = (1, a_{t-1}^2, ..., a_{t-q}^2, h_{t-1}, ..., h_{t-p})^T + \sum_{i=1}^{p} \beta_i \frac{\partial h_{t-i}}{\partial \theta}.$$

Thus, the gradient and Fisher information is then defined by

$$G(\theta) = \frac{1}{2}\sum_{t=q+1}^{n}\left(\frac{a_t^2}{2h_t^2} - \frac{1}{2h_t}\right)\frac{\partial h_t}{\partial \theta}$$

and

$$\begin{aligned} J(\theta) &= \sum_{t=q+1}^{n} E\left[\left(\frac{a_t^2}{2h_t^2} - \frac{1}{2h_t}\right)\frac{\partial^2 h_t}{\partial\theta\partial\theta^T} + \left(\frac{1}{2h_t^2} - \frac{a_t^2}{h_t^3}\right)\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^T}\right] \\ &= -\frac{1}{2}\sum_{t=q+1}^{n} E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^T}\right). \end{aligned}$$

For a simple GARCH(1,1) model

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1},$$

the gradient and Fisher information is then

$$G(\theta) = \frac{1}{2}\sum_{t=2}^{n}\left(\frac{a_t^2}{h_t^2} - \frac{1}{h_t}\right)\frac{\partial h_t}{\partial \theta}$$

and

$$J(\theta) = -\frac{1}{2}\sum_{t=2}^{n} E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^T}\right),$$

where

$$\frac{\partial h_t}{\partial\theta} = (1, a_{t-1}^2, h_{t-1})^T + \beta_1\frac{\partial h_{t-1}}{\partial\theta}.$$

We can then use Newtons iterative methods

$$\theta_{k+1} = \theta_k + J^{-1}(\theta_k)G(\theta_k),$$

where the initial guess is

$$\theta_0 = (\alpha_0, ..., \alpha_q, \beta_1, ..., \beta_p)^T \tag{6.14}$$

is required and estimation of $\frac{\partial h_t}{\partial \theta}$. The iterations are repeated until θ converges.

6.1.2 Forecasting

For forecasting using a simple GARCH(1,1) model, we use

$$\sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2$$

for a 1-step ahead forecast

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2$$

while for a 2-step ahead forecast

$$\sigma_h^2(2) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(1)$$

and in general,

$$\sigma_h^2(l) = \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{l-1}]}{1 - (\alpha_1 + \beta_1)} + \alpha_1 + \beta_1^{l-1}\sigma_h^2(1),$$

where

$$\begin{aligned} \lim_{l\to\infty} \sigma_h^2(l) &= \lim_{l\to\infty} \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{l-1}]}{1 - (\alpha_1 + \beta_1)} + \alpha_1 + \beta_1^{l-1}\sigma_h^2(1) \\ &= \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \end{aligned}$$

which is the unconditional variance of a_t.

6.2 FIGARCH Processes

Sometimes the return series contains long memory, then its correlations are not summable with lags declining hyperbolically. In this case, the fractional IGARCH (FIGARCH) model is then used. The FIGARCH model is characterised by a volatility persistence shorter than an IGARCH model but longer than the GARCH model. The FIGARCH model is then obtained by extending the IGARCH model and allowing the integration

factor to be fractional. The FIGARCH(p,d,q) is defined

$$\begin{aligned} r_t &= \mu_t + a_t \\ \epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \omega(1-\beta(L))^{-1} + \left\{1-(1-\beta(L))^{-1}\phi(L)(1-L)^d\right\} a_t^2, \end{aligned} \tag{6.15}$$

where z_t is $i.i.d$ N(0,1). This can further be defined as

$$\begin{aligned} \sigma_t^2 &= \omega\beta(L)\sigma_t^2 + (1-\beta(L))a_t^2 - \phi(L)(1-L)^d a_t^2 \\ &= \omega(1-\beta(L))^{-1} + \left\{1-(1-\beta(L))^{-1}\phi(L)(1-L)^d\right\} a_t^2 \\ &= \omega(1-\beta(L))^{-1} + \lambda(L)a_t^2 \\ &= \omega * + \sum_{i=1}^{\infty} \lambda_i L^i a_t^2. \end{aligned} \tag{6.16}$$

A Taylor expansion is used to study $(1-L)^d$ as follows:

$$\begin{aligned} (1-L)^d &= \sum_{k=0}^{\infty} \frac{\Gamma(d+1)L^k}{\Gamma(k+1)\Gamma(d-k+1)} \\ &= 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - ... \\ &= 1 - \sum_{k=0}^{\infty} C_k(d)L^k \end{aligned} \tag{6.17}$$

and it is highlighted that $\Gamma(k-d)/\Gamma(k+1)\ k^{-(d+1)}$ is large and hence the coefficients of the polynomial with decay hyperbolically.

Baillie and Morana (2009) introduced an adaptive FIGARCH model which captures long memory and sructural changes in the volatility process. The conditional variance is then defined as

$$[1-\beta(L)](h_t - \omega_t) = [1-\beta(L)-\phi(L)(1-L)^d]\epsilon_t^2,$$

where the intercept parameter ω_t is allowed to be time varying and defined as

$$\omega_t = \omega_0 + \sum_{j-1}^{k} \left[\gamma_j \sin\left(\frac{2\pi jt}{T}\right) + \delta_j \cos\left(\frac{2\pi jt}{T}\right)\right] \tag{6.18}$$

This model allows for breaks, cycles and changes to the drift of the time series process. Ben Nasr et al. (2010) introducted the fractionally integrated time-varying GARCH model to capture both long memory and structural changes in the volatility process. This model is defined as

$$[1 - \phi(L)](1 - L)^d a_t^2 = \omega_t + [1 - \beta_t(L)]\eta_t,$$

where $a_t = \epsilon_t h_t^{1/2}$ and ϵ_t is $N(0, 1)$. The conditional variance is decomposed into two components. The first component follows the standard FIGARCH process and the second is a time varying component. It is noted that this model is not strictly stationary. Because of this, a re-parameterization of this model was then given as

$$[1 - \beta(L)]h_t = \omega_t + [1 - \beta_t(L) - (1 - \phi_t(L))(1 - L)^d]a_t^2,$$

where the conditional variance of the model is

$$\begin{aligned} h_t &= \omega_t[1 - \beta_t(L)]^{-1} + \left(1 - [1 - \beta_t(L)]^{-1}(1 - \phi_t(L))(1 - L)^d\right) a_t^2 \\ &= \tilde{\omega}_t + \sum_{j=1}^{\infty} \psi_{j,t} a_{t-j}^2. \end{aligned}$$

To address non-linearity in the conditional volatility process, Kilic (2007) introduced a smooth transition FIGARCH. This model is able to accommodate smooth changes both in the amplitude of the volatility clusters as well as asymmetry in the conditional volatility. The first order of this model is defined as

$$(1 - \phi(L))(1 - L)^d a_t^2 = \alpha_0 + [1 - \beta(1 - G(z_{t-s,\gamma,c}))L - \beta^* G(z_{t-s,\gamma,c})]\eta_t,$$

where $\eta_t = a_t^2 - h_t$, $0 < d < 1$, β and β^* are the volatility dynamics parameters and

$$G(z_{t-s,\gamma,c}) = [1 + exp(-\gamma(z_{t-s-c}))]^{-1}$$

is the logistic transition function with transition variable s and period lagged z. s is called the delay parameter and c is the location parameter. γ is assumed to be positive and characterise the speed of transition between extreme regimes. To address asymmetric

effects in volatility, Hwang (2001) introduced the asymmetric FIGARCH model defined as

$$a_t = \sqrt{h_t} z_t, \quad z_t \sim N(0,1)$$

where

$$\sqrt{h_t^{\lambda}} = \frac{\alpha_0}{1-\beta_1} + \left[1 - \frac{(1-\phi_1 L)(1-L)^d}{1-\beta_1 L}\right] f^{\nu}(a_t) h_t^{\lambda/2}$$

with

$$f(a_t) = \left|\frac{a_t}{\sigma_t} - b\right| - c\left(\frac{a_t}{\sigma_t} - b\right), \quad |c| \leq 1,$$

and z_t are iid with mean 0 and variance 1. The values of b and c represent the shift and the rotation of the news impact curve. A modified version by Ruiz and Prez (2003) is defined by

$$(1-\phi_1)(1-L)^d \frac{h_t^{\lambda/2} - 1}{\lambda} = \alpha_0 * + \alpha(a - \phi_1 L) h_{t-1}^{\lambda/2} [f^{\nu}(z_{t-1}) - 1].$$

In the following subsection, we discuss inequality constraints of FIGARCH models which are necessary and sufficient conditions to ensure nonnegativity of the FIGARCH conditional variance.

6.2.1 Inequality Constraints of the FIGARCH(1,d,1) Model

Inequality constraints are necessary and sufficient conditions that ensure the nonnegativity of the conditional variance of the FIGARCH model. Various conditions are available in literature but most used conditions are of Baillie et al. (1996) constraints defined by

$$0 \leq \beta_1 \leq \phi_1 + d \quad \text{and} \quad 0 \leq d \leq 1 - 2\phi_1.$$

Bollerslev and Mikkelsen (1996) inequality constraints defined by

$$\beta_1 - d \leq \phi_1 \leq \frac{2-d}{3} \quad \text{and} \quad d\left[\phi_1 - \frac{1-d}{2}\right] \leq \beta_1(\phi_1 - \beta_1 + d).$$

Chung (1999) constrains are given by

$$0 \leq \phi_1 \leq \beta_1 \leq d < 1.$$

Conrad and Haag (2006) inequality constraints uses an $ARCH(\infty)$ representation of the $FIGARCH(1, d, 1)$ using ψ_i defined by

$$\begin{aligned}
\psi_1 &= d + \phi_1 - \beta_1, \\
\psi_i &= \beta_1 \psi_{i-1} + (f_i - \phi_1)(-g_{i-1}) \quad \forall i \geq 2 \quad or \\
\psi_i &= \beta_1^2 \psi_{i-2} + [\beta_1 (f_{i-1}\phi_1) + (f_i - \phi_1) f_{i-1}] (-g_{i-2}), \quad \forall i \geq 3.
\end{aligned}$$

The inequality constraints are summarised in the corollary below found in Conrad and Haag (2006).

Corollary 1

The conditional variance of the $FIGARCH(1, d, 1)$ in nonnegative almost surely iff:

Case 1:$0 < \beta_1 < 1$.

Either $\psi_i \geq 0$ and $\phi_1 \leq f_2$ or for $k > 2$ with $f_{k-1} < \phi_1 \leq f_k$, it holds that $\psi_{k-1} \geq 0$.

Case 2: $-1 < \beta_1 < 0$.

Either $\psi_1 \geq 0$,$\psi_2 \geq 0$ and $\phi_1 \leq f_2(\beta_1 + f_3)/(\beta_1 + f_2)$, or

For $k > 3$ $f_{k-2}(\beta_1 + f_{k-1})/(\beta_1 + f_{k-2}) < \phi_1 \leq f_{k-1}(\beta_1 + f_k)/(\beta_1 + f_{k-1})$, it holds that $\psi_{k-1} \geq 0$ and $\psi_{k-2} \geq 0$,

where $F_i = f_i - \phi_i$, $f_{=}\frac{j-1-d}{j}$ and $g_j = f_j.g_{j-1} = \prod_{i=1}^{j} f_i$, and the initial conditions are $g_0 = 1$, $f_1 = -d < 0$, $f_2 = (1-d)/2 > 0$, and $f_i > 0$, $\forall j > 2$. The FIGARCH model is one of the most understood and easier models to work with when modeling fractional integration of conditional heteroscedastic models. This model has some drawbacks that are not ideal for modeling financial returns. One drawback that is mostly found in almost all applications of this models is that we must have $d \geq 0$ and the polynomial coefficients must satisfy the above mentioned constraints for the conditional variance to be positive.

In the following sections, we discuss more models which could be used to avoid this shortcoming of the model.

6.3 FIEGARCH Processes

The exponential FIGARCH(FIEGARCH) model is defined by

$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 - \sum_{i=1}^{p} \alpha_i L^i}{1 - \sum_{j=1}^{q} \beta_j L^j}(1 - L)^{-d} g(\epsilon_{t-1}), \tag{6.19}$$

where the function

$$g(\epsilon_{t-1}) = \theta\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E(|\epsilon_{t-1}|)] \quad \forall t \in \mathbb{Z}. \tag{6.20}$$

FIEGARCH processes models volatility clusters, asymmetry and model long memory in volatility. These models offer better modeling as they don't suffer from FIGARCH drawbacks since the variance under FIEGARCH is defined in terms of the logarithm functions. It is easier to identify stationarity since this process is weakly stationary iff $d < 0.5$. This process has variations of parameterizations, we will define it as given by Lopes and Prass (2014). Let $\{X_t, t \in \mathbb{Z}\}$ be a stochastic process defined as

$$X_t = \sigma_t^2 \epsilon_t, \tag{6.21}$$

$$\ln(\sigma_t^2) = \alpha_0 + \frac{\alpha(L)}{\beta(L)}(1 - L)^{-d} g(\epsilon_{t-1}), \quad \forall t \in \mathbb{Z}, \tag{6.22}$$

where $\alpha_0 \in \mathbb{R}$ and $g(.)$ is defined as

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)] \tag{6.23}$$

for all $t \in \mathbb{Z}$, with $\theta, \gamma \in \mathbb{R}$ then $\{X_t, t \in \mathbb{Z}\}$ is a FIEGARCH process denoted by FIEGARCH(p, d, q). A much more defined criterion for the FIEGARCH model to be stationary is stated in the following theorem.

Theorem 6.1. *Define* $\{\sigma_t^2\}$,$\{X_t\}$ *and* $\{\epsilon_t\}$ *for all* $t \in \mathbb{Z}$ *by*

$$\begin{aligned} X_t &= \sigma_t \epsilon_t, \quad \epsilon_i \; iid(0,1) && (6.24) \\ \ln(\sigma_t^2) &= \alpha_t + \sum_{k=1}^{\infty} \lambda_k g(\epsilon_{t-k}), \quad \lambda_1 = 1, \quad and && (6.25) \\ g(\epsilon_t) &= \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)], && (6.26) \end{aligned}$$

where $\{\alpha_t, t \in \mathbb{Z}\}$ *and* $\{\lambda_t, t \in \mathbb{N}\}$ *are real, nonstochastic, scalar sequances and assume that* θ *and* γ *do not both equal zero. Then* $\{e^{-\alpha_t}\sigma_t^2\}$, $\{e^{-\alpha_t}X_t^2\}$ *and* $\{ln(\sigma_t^2) - \alpha_t\}$ *are strictly stationary and ergodic and* $\{ln(\sigma_t^2) - \alpha_t\}$ *is covariance stationary if and only if* $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$.
If $\sum_{k=1}^{\infty} \lambda_k^2 = \infty$,*then* $|ln(\sigma_t^2) - \alpha_t| = \infty$ *almost surely. If* $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$ *then for* $k > 0$,

$$Cov(\epsilon_{t-k}, \ln(\sigma_t^2)) = \lambda_k(\theta + \gamma E(X_t|X_t|)) \tag{6.27}$$

and

$$Cov(\ln(\sigma_t^2), \ln(\sigma_{t-k}^2)) = Var(g(\epsilon_t)) \sum_{j=1}^{\infty} \lambda_j \lambda_{j=k} \tag{6.28}$$

In order to gain more modelling flexibility, fractional integrated asymmetric power ARCH (FIAPARCH) models have been suggested in the literature.

6.4 FIAPARCH Processes

The fractional integrated asymmetric power ARCH (FIAPARCH) process increases the flexibility of the conditional variance specification by allowing

- An asymmetric response of volatility to positive and negative shocks.
- The data to determine the power of returns for which the predictable structure in the volatility pattern is strongest, and
- Long range volatility dependence

A simple FIAPARCH(1,d,1) model is given by

$$(1 - \phi L)(1 - L)^d f(a_t) = \alpha_0 + (1 - \beta L)a_t \tag{6.29}$$

where

$$f(a_t) = [|a_t| - \gamma a_t]^{\delta}, \tag{6.30}$$

where γ is the leverage parameter defined in $-1 < \gamma < 1$, δ is the parameter for the power term, $|\phi| < 1$,$\alpha_0 > 0$ and $0 \leq d \leq 1$. This process would reduce to the FIGARCH process for $\gamma = 0$ and $\delta = 2$.

6.5 HYGARCH Processes

The hyperbolic GARCH (HYGARCH) model introduced by Davidson (2004) has the GARCH and FIGARCH as special cases. It is covariance stationary, similar to the GARCH model and has hyperbolic decay impulse reposed coefficients similar to the FIGARCH. The HYGARCH process is obtained by

$$\phi(L)\left((1-\tau) + \tau(1-L)^d\right) a_t^2 = \alpha_0 + \beta(L)\eta_t \tag{6.31}$$

When $\tau = 0$ and $d = 0$, the model is GARCH and when $\tau = 1$, the model is FIGARCH. To further understand this model, we can re-write is as

$$\begin{aligned} \sigma_t^2 &= \frac{\alpha_0}{\beta(1)} + \Psi^{HY}(L)a_t^2 \\ &= \frac{\alpha_0}{\beta(1)} + \sum_{i=1}^{\infty} \psi_i^{HY} a_{t-i}^2 \end{aligned} \tag{6.32}$$

with

$$\Phi^{HY}(L) = \tau\Psi^{FI}(L) + (1-\tau)\Psi^{GA}(L) \tag{6.33}$$

and thus it follows that

$$\psi_i^{HY} = \tau\psi_i^{FI} + (1-\tau)\psi_i^{GA}. \tag{6.34}$$

6.5.1 Inequality Constraints of the HYGARCH(1,d,1)

A restrictive sufficient condition from equation 6.33 for non-negativity of the ψ_i^{HY} coefficients is the non-negativity of the ψ_i^{FI} and ψ_i^{GA} coefficients, for $0 < \tau < 1$ (Conrad, 2010). A recursive representation for the ARCH(∞) coefficients of the HYGARCH(p, d, q) is vital in understanding the constraints.

Lemma 6.1. *In the HYGARCH(p,d,q), the sequence* $\{\psi_i^{HY}, i = 1, 2, ...\}$ *can be written as*

$$\psi_i^{HY} = \psi_i^{HY(p)}, \tag{6.35}$$

where

$$\psi_i^{HY(r)} = \lambda_{(r)}\psi_{i-1}^{HY(r)} + \psi_i^{HY(r-1)}, \quad 1 < r \leq p, \quad and \quad i \geq 1 \tag{6.36}$$

with starting values $\psi_0^{HY(r)} = -\tau$, *for* $r = 2, ..., p$ *and* $\{\psi_i^{HY(1)}\}$ *is given by*

$$\begin{aligned} \psi_i^{HY(1)} &= \tau\left(-c_i + \sum_{j=1}^{i}\psi_i c_{i-j}\right) + (1-\tau)\sum_{i=1}^{j}\lambda_{(1)}^{i-j}(\phi_j - \beta_j), \quad for \quad i = 1, 2, ..., q \\ &= \lambda_{(1)}\psi_{i-1}^{HY(1)} + \tau F_i(-g_{i-q}) \quad i \geq q+1 \end{aligned} \tag{6.37}$$

and

$$c_i = \sum_{j=0}^{i}\lambda_{(1)}^{i-j}g_j. \tag{6.38}$$

For the HYGARCH(1,d,1) model, the recursive representation of the ARCH(∞) coefficients in Lemma 6.1 simplifies to

$$\begin{aligned} \psi_1^{HY} &= \tau d + \phi_1 - beta_1, \\ \psi_i^{HY} &= \beta_1\psi_{i-1}^{HY} + \tau(f_i - \phi_1)(-g_{i-1}) \quad , for \quad i \geq 2, \\ \psi_i^{HY} &= \beta_1^2\psi_{i-2}^{HY} + \tau(\beta_1(f_{i-1} - \phi_1) + (f_i - \phi_1)f_{i-1})(-g_{i-2}), \quad for \quad i \geq 3. \end{aligned} \tag{6.39}$$

Theorem 6.2. *The conditional variance of the HYGARCH(1,d,1) is non-negative almost*

surely if and only if

Case 1 $0 < \beta < 1$.

> *Either* $\psi_1^{HY} \geq 0$ *and* $\phi_1 \leq f_2$ *or for* $k > 2$ *with* $f_{k-1} < \phi_1 \leq f_k$ *it holds that* $\psi_{k-1}^{HY} \geq 0$,

Case 2 $-1 < \beta < 0$.

> *Either* $\psi_1^{HY} \geq 0$,$\psi_2^{HY} \geq 0$ *and* $\phi_1 \leq f_2(\beta_1 + f_3)/(\beta_1 + f_2)$ *or for* $k = 3$ *with* $f_{k-2}(\beta_1 + f_{k-1})/(\beta_1 + f_{k-2}) < \phi_1 \leq f_{k-1}(\beta_1 + f_k)/(\beta_1 + f_{k-1})$ *it holds that* $\phi_{k-1}^{HY} \geq 0$ *and* $\psi_{k-2}^{HY} \geq 0$.

These are the sufficient conditions for the HYGARCH(1,d,1), more conditions for higher order HYGARCH are found in Conrad (2010).

6.6 Residual Models

Recently in econometrics literature, there have been interest on residual analysis. This analysis is specifically on residuals of ARFIMA models. Models that are currently used for this purpose are the models discussed above. An ARFIMA model is of the form

$$\phi(L)\Delta^d X_j = \theta(L)\epsilon_j, \tag{6.40}$$

where the series $\{X_j\}$ in this thesis are the log returns and the residuals are $\{\epsilon_t\}$. In order to better understand these residuals, GARCH-type models can be applied but this will only show us short-term dependencies. To better capture long term dependencies and other properties that are not given by GARCH-type models, fractional integrated models are used. The residual process is then allowed to follow: (1) The FIGARCH(p, d, q) model as given by

$$\sigma_t^2 = \alpha_0(1 - \beta(L))^{-1} + 1 - (1 - \beta(L))^{-1}\phi(L)(1 - L)^d\epsilon_t^2 \tag{6.41}$$

or, (2) the FIEGARCH(p, d, q) model as given by

$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 - \sum_{i=1}^{p} \alpha_i L^i}{1 - \sum_{j=1}^{q} \beta_j L^j}(1 - L)^{-d} g(\epsilon_{t-1}), \tag{6.42}$$

where the function

$$g(\epsilon_{t-1}) = \theta\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E(|\epsilon_{t-1}|)], \quad \forall t \in \mathbb{Z} \tag{6.43}$$

or (3) the FIAPARCH(p, d, q) model as given by

$$\sigma_t^{\delta} = \alpha_0(1 - \beta(L))^{-1} + 1 - (1 - \beta(L))^{-1}\phi(L)(1 - L)^d(|\epsilon_t| - \gamma\epsilon_t)^{\delta} \tag{6.44}$$

and lastly, (4) the HYGARCH(p, d, q) model as given by

$$\phi(L)\left((1 - \tau) + \tau(1 - L)^d\right)a_t^2 = \alpha_0 + \beta(L)\epsilon_t. \tag{6.45}$$

When $\tau = 0$ and $d = 0$, the model is GARCH and when $\tau = 1$, the model is FIGARCH. All four of these models have different properties and hence they model properties that are different on the residual series. The best model that would be select will suggest that the series behaves in that manner.

6.7 Chapter Summary

Conditional heteroscedastic models are methods which are used to model conditional volatility. This arises as a result of variance of residuals not being constant and hence the assumption does not hold. The basic model is the ARCH model. This model has a couple of short-comings but the apparent weakness is in the estimation of parameters. ARCH models require a lot of parameter estimates, and as a result GARCH models can be considered. GARCH models generalise the ARCH models and often requires less parameters to explain the data series. The GARCH model has a restriction

$$\sum(\alpha_i + \beta_i) < 1 \tag{6.46}$$

but if the series is integrated, this property does not hold in a sence that

$$\sum(\alpha_i + \beta_i) = 1 \tag{6.47}$$

and in this case, IGARCH models are used. Where there is a symmetric relationship in the mean and the volatility, an EGARCH model is used as it can better explain this phenomenon. Lastly, we discussed GARCH-M which is useful if the returns of a security depends on its volatility. These models explain short-term dependencies in volatility.

To explain long memory in volatility, a FIGARCH model is used. The FIGARCH model must satisfy some constraints for the volatility to be non-negative. These constraints introduce a weakness in these models.

In order to overcome this short-coming, a FIEGARCH model may be used. This is because the variance in the FIEGARCH model is determined in term of the logarithmic function. The FIAPARCH process may be used if we need to determine the power of returns for which the predictable structure in the volatility pattern is strongest. Lastly, we discussed the HYGARCH model. The advantage of the HYGARCH models is that it is covariance stationary similar to the GARCH model and has hyperbolic decay impulse response coefficients similar to the FIGARCH model. The HYGARCH model also suffers from the constraints which assures non-negativity of volatility. Last section of the chapter discusses how the ARFIMA models, which models the return series, is used with volatility models that are used to model volatility of residuals.

Chapter 7

Forecasting and Estimation Methods

Volatility is unobservable, this property of volatility makes it difficult to evaluate it forecast. In the literature, squared returns are normally used as a measure of volatility. The use of this measure does not guarantee the same result as if the true volatility was used. In other cases, extreme observations, even if few, may lead to large impacts on the outcome of forecast evaluations. In this case, loss functions may be utilised as they can be less sensitive to extreme observations than the widely used square error function. In evaluating volatility forecast, when the objective of the analysis is the forecasting accuracy of a single model, the quality of the model can be measured by the correlations between prediction and realizations. A common method used in this case is the Mincer-Zarnowitz(MZ) regression, which involves regressing the realization of a variable on a constant and its forecast. When the forecastor is interested in comparing two or more models, the Diebold and Mariano test could be used.

There are many methods proposed in literature that could be used to evaluate volatility forecasts. The hypothesis that we would like to examine in the forecasting of volatility is

$$\begin{aligned} H_0: \quad & h_t = \sigma_t^2 \\ & vs \\ H_1: \quad & h_t \neq \sigma_t^2 \end{aligned} \tag{7.1}$$

where h_t is the volatility forecast and σ_t^2 is the true conditional variance at time t.

7.1 Mincer-Zarnowitz Regression

In the literature, evaluating forecasts can be done by using the Mincer-Zarnowitz regression which is a simple linear regression of the realisation of the conditional volatility on the forecasts. The Mincer-Zarnowitz regression is given by

$$\hat{\sigma}_t^2 = \beta_0 + \beta_1 h_t + e_t$$

which yields an unbiased estimate of β_0 and β_1. We will then test the hypothesis

$$\begin{aligned} \bar{H}_0 : \quad & \beta_0 = 0 \quad and \quad \beta_1 = 1 \\ & vs \\ \bar{H}_1 : \quad & \beta_0 \neq 0 \quad or \quad \beta_1 \neq 0 \end{aligned} \tag{7.2}$$

There have been variants of this test using different volatility measures. More on the other volatility measures can be found Jorion (1995) and Wright and Bollerslev (1999). Transformation of volatility done by these authors is mainly due to reduce the impact of large returns.

For the Mincer-Zarnowitz regression under univariate case,

$$\hat{\sigma}_t^2 = \beta_0 + \beta_1 h_t + e_t$$

generalised least squares can be used to evaluate the residuals for the regression. We have seen that under the null hypothesis,

$$\begin{aligned} e_t &= \hat{\sigma}_t^2 - h_t \\ &= \hat{\sigma}_t^2(\eta_t - 1) \end{aligned}$$

and that $E(e_t|F_{t-1}) = 0$, then

$$\begin{aligned} V(e_t|F_{t-1}) &= \hat{\sigma}_t^4 V(\eta_t - 1) \\ &= \omega_t^2. \end{aligned}$$

But ω_t^2 is unknown. If we could assume squared returns to be our volatility measure, the

least squares could be defined as

$$\frac{\sigma_t^2}{h_t} = \beta_0 \frac{1}{h_t} + \beta_1 + \hat{e}_t,$$

where $\sigma_t^2 = r_t^2$. Volatility forecasting evaluation can also be used when comparing competing volatility forecasts. The MZ regression can also be viewed as a test of effeciency in the sense that

$$E[h_t(\hat{\sigma}_t^2 - h_t)] = 0. \tag{7.3}$$

If the forecast and forecast errors are correlated then it would be possible to produce superior forecasts by the relationship

$$\hat{\sigma}_t^2 - h_t = \alpha + (\beta - 1)h_t + e_t, \tag{7.4}$$

and thus

$$E[h_t(\hat{\sigma}_t^2 - h_t)] = \alpha E(h_t) + (\beta - 1)E(h_t^2) + E(h_t e_t) = 0 \tag{7.5}$$

for $\alpha = 0$ and $\beta = 1$.

7.2 Loss Functions

Another approach that could be used in evaluating forecast performance is by comparing expected loss evaluated with respect to the true variance. Define the precision measure in terms of expected loss of some forecast, $h_{k,t}$ with respect to the true $E(L(\sigma_t, h_{k,t}))$. The aim is to seek conditions that ensure consistency of the ranking between any two forecasts k and j when a conditional unbiased proxy is substituted to the true variance such that

$$E(L(\sigma_t, h_{k,t})) \leq E(L(\sigma_t, h_{j,t})) \leftrightarrow E(L(\hat{\sigma}_t, h_{k,t})) \leq E(L(\hat{\sigma}_t, h_{j,t})), \tag{7.6}$$

A loss functions L is robust if the ranking of any two, possible imperfect volatility forecasts, h_{1t} and h_{2t}, by expected loss is the same whether the ranking is done using the

true conditional variance, σ_t^2, or some conditional unbiased volatility proxy, $\hat{\sigma_t^2}$. That is

$$E(L(\sigma_t, h_{1t})) \leq E(L(\sigma_t, h_{2t})) \leftrightarrow E(L(\hat{\sigma}_t, h_{1t})) \leq E(L(\hat{\sigma}_t, h_{2t})) \tag{7.7}$$

for any $\hat{\sigma}_t^2$ such that

$$E[\hat{\sigma}_t^2 | \mathbf{F}_{t-1}] = \sigma_t^2. \tag{7.8}$$

A sufficient condition for the definition to hold is for

$$\frac{\partial^2 L(\sigma_t, h_t)}{(\partial \sigma_t)^2} \tag{7.9}$$

to exist and be independent on h_t. The squared errors are normally accepted by this condition. A homogeneous class of consistent loss functions due to the work of Hansen and Lunde (2006) is given by

$$L(\hat{\sigma}_t^2, h_t) = \begin{cases} \frac{1}{\xi - 1}(\hat{\sigma}_t^{\xi} - h_t^{\xi}) - \frac{1}{\xi - 1}(\hat{\sigma}_t - h_t), & : \xi \notin (0, 1) \\ h_t - \hat{\sigma}_t + \hat{\sigma}_t \ln(\hat{\sigma}_t / h_t), & : \xi = 1 \\ \frac{\hat{\sigma}_t}{h_t} - \ln(\hat{\sigma}_t / h_t) - 1, & : \xi = 0. \end{cases}$$

This loss functions can take variety of shapes. For instance, for $xi = 0, 1, 2$ the function can be alternatively derived from the objective functions of the Gaussian, Poisson and Gamma densities respectively.

7.3 Predictive Ability : Pair-Wise Test

This test is based on the comparison of the mean square error(MSE) of a pair of forecasts with the target. Define $d_{k,t} = \sigma_t - h_{k,t}$ as the forecast error and $MSE(L_{k,t}) = T^{-1} \sum_{\forall t} e_{k,t}^2$ the mean square forecast error of model k with respect to σ_t. Then, if we seek to compare model k to model j, we then need to evaluate

$$\text{MSE}(L_{k,t}) - \text{MSE}(L_{j,t}) = (Var(e_{k,t} - Var(e_{j,t}))) + (\bar{e}_k^2 - \bar{e}_j^2) \tag{7.10}$$

where $\bar{e}_i = T^{-1} \sum_{\forall t} e_{i,t}$. Then define $D_t = e_{k,t} - e_{j,t}$, $S_t = e_{k,t} - e_{j,t}$, and let $\bar{D}$ and $\bar{S}$ be their respective means. This then allows us to re-write the above evaluation as

$$\text{MSE}(L_{k,t}) - \text{MSE}(L_{j,t}) = \text{Cov}(D_t, S_t) + \bar{D}\bar{S}. \tag{7.11}$$

Then by using the hypothesis

$$H_0 : Cov(D_t, S_t) = 0 \tag{7.12}$$

we test equal predictive ability. This can further be expressed as a linear regression of D_t and S_t as

$$D_t = \alpha + \beta(S_t - \bar{S}) + \epsilon_t \tag{7.13}$$

which then the null hypothesis becomes

$$H_0 : \alpha = 0 \tag{7.14}$$

and under H_0, the test is

$$GN - T = \frac{\rho}{\sqrt{(T_1)^{-1}(1-\rho)^2}} \sim t_{T-1}, \tag{7.15}$$

where $\rho = Cov(D_t, S_t)/\sqrt{Var(D_t)Var(S_t)}$ and t_{T-1} is the student-t distribution with $T-1$ degrees of freedom. Another approach (explained in the next sections) of testing pair-wise predictive ability is a test proposed by Diebold and Mariano (1994).

7.4 Diebold-Mariano Test

The Diebold-Mariano test uses loss functions and under the null hypothesis, predictive accuracy of competing forecasts is equal versus an alternative of either lesser or more accurate. This test is used when we need to compare different volatility forecasts against a specified volatility measure. Suppose we have a series

$$\{\sigma_t : t = 1, 2, ..., T\}$$

of the actual values and

$$\{\sigma_{it} : i = 1, 2 \quad t = 1, 2, ..., T\}$$

as forecast values. We will use two forecasts in this case. A forecast error would then be found by

$$e_{it} = \hat{\sigma}_{it} - \sigma_t, \quad i = 1, 2.$$

If we define a loss function $g(e_{it})$ to be associated with forecast i at time t, then it is a function of the forecasting error. Loss functions that are considered are zero or positive and increase with an increase in e_{it}. $g(e_{it})$ can take many forms but in literature, squared returns, absolute returns, realised volatility and modified squared returns have been used. This test utilises sample realisation of a loss differential series, $\{d_t\}_{t=1}^{T}$. The loss differential is a difference between two forecasts

$$d_t = g(e_{it}) - g(e_{2t})$$

and the accuracy of the forecasts can be seen in $E(d_t) = 0$. Hence, the null hypothesis then becomes

$$\begin{aligned} H_0 : \quad & E(d_t) = 0 \\ H_1 : \quad & E(d_t) \neq 0. \end{aligned}$$

The asymptotic distribution is then given as

$$\sqrt{T}(\bar{d} - \mu)^d \rightarrow N(0, 2\pi f_d(0))$$

where

$$\bar{d} = \sum_{t=1}^{T} d_t$$

is the sample mean,

$$\mu = E(d_t) \tag{7.16}$$

is the population mean, and

$$f_d(0) = \frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty} \gamma_d(k)\right)$$

is the spectral density of the loss differential at frequency 0, where $\gamma_d(k)$ is the autocovariance of the loss differential at lag k. Then, under H_0,

$$\frac{\bar{d} - \mu}{\sqrt{\frac{2\pi f_d(0)}{T}}} \to N(0,1)$$

Under H_0 for an h-step ahead forecast, where $h \geq 2$, the Dieblom-Mariano statistic is then defined as

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}},$$

where

$$\hat{f}_d(0) = \frac{1}{2\pi}\sum_{k=-(T-1)}^{T-1} l\left(\frac{k}{h-1}\right)\hat{\gamma}_d(k)$$

for

$$\hat{\gamma}_d(k) = \frac{1}{T}\sum_{t=|k|+1}^{T}(d_t - \bar{d})(d_{t-|k|} - \bar{d})$$

and

$$l\left(\frac{k}{h-1}\right) = \begin{cases} 1 & : \left|\frac{k}{h-1} \leq 1\right| \\ 0 & : Otherwise. \end{cases}$$

We know that under covariance stationary, $\hat{\gamma}_d(k) = \hat{\gamma}_d(-k)$ and hence, the consistent estimator of $f_d(0)$ is then

$$\hat{f}_d(0) = \frac{1}{2\pi}\left(\hat{\gamma}_d(0) + 2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)\right).$$

Then the Dieblom-Mariano(DM) test is

$$\begin{aligned} DM &= \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}} \\ &= \frac{\bar{d}}{\sqrt{\frac{\hat{\gamma}_d(0)+2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)}{T}}}. \end{aligned}$$

To get an adequate estimate, we can use

$$\sum_{k=-M}^{M}\hat{\gamma}_d(k),$$

where $M = T^{1/3}$, and thus,

$$DM = \frac{\bar{d}}{\sqrt{\frac{\sum_{k=-M}^{M}\hat{\gamma}_d(k)}{T}}}.$$

Under the null hypothesis, we reject equal accuracy if

$$|DM| > z_{\alpha/2}.$$

Other methods that are used to evaluate forecasting performance of models, error functions are used. Mostly used functions in literature are

1. The Mean Square Error(MSE)

$$\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{\sigma}_t - \sigma_t)^2}.$$

2. The Mean Absolute Error(MAE)

$$\frac{1}{N}\sum_{t=1}^{N}|\hat{\sigma}_t - \sigma_t|.$$

3. The Mean Absolute Percentage Error (MAPE)

$$\frac{1}{N}\sum_{t=1}^{N}\frac{|\hat{\sigma}_t - \sigma_t|}{\sigma_t}.$$

4. The Mean Error (ME)

$$\frac{1}{N}\sum_{t=1}^{N}\left(\hat{\sigma}_t - \sigma_t\right).$$

5. The Mean Square Error (MSE)

$$\frac{1}{N}\sum_{t=1}^{N}\left(\hat{\sigma}_t - \sigma_t\right)^2.$$

6. The Mean Log-absolute Error (MLAE)

$$\frac{1}{N}\sum_{t=1}^{N}ln|\hat{\sigma}_t - \sigma_t|.$$

where $\hat{\sigma_t}$ is the volatility predicted by the model at time t, and σ_t is the actual volatility of the out-of-sample observations.

7.5 Estimation Methods

Bayesian methods are good estimation tools because of their simplicity. Considers a time series $\{X_i\}, i = 1..., N$ with the parameter θ having a prior distribution $\pi(\theta)$. Let $\pi(X|\theta)$ be the likelihood function of the model. The posterior distribution of θ given the X is

$$\pi(\theta|X) \propto \pi(X|\theta)\pi(\theta).$$

What is of interest in finding the Bayes estimator for θ? A loss function $L(\theta, X)$ can be considered, that is

$$\hat{\theta} = argmin \int L(\theta, X)\pi(\theta|X)dX.$$

The evaluation of these integrals is complex and hence introduce a challenge to researchers. A simpler method can be used to evaluate these integrals are the Markov chain Monte Carlo (MCMC) algorithms. Let $\theta \in \Theta$ be a parameter vector for data X. To make inference, we need information about the distribution $\pi(\theta|X)$. To be successful in this method, we need to create a Markov process whose stationary transition probability is specified by $\pi(\theta|X)$. In the literature, there are two well known procedures for this purpose, the Gibbs sampling and the Metropolis-Hastings algorithm.

7.5.1 Gibbs Sampling

This algorithm has the advantage to decompose a high dimensional estimation problem into several lower dimensional problems. Suppose the parameter vector consists of $\theta_1, \theta_2, \theta_3$ and X is the collection of data. The goal of Gibbs sampling is to estimate the parameters so that the fitted model can be used to make inference. Further, suppose three conditional distributions of a single parameter given other parameters are available. That is,

$$f_1(\theta_1|\theta_2, \theta_3, X),$$
$$f_2(\theta_2|\theta_1, \theta_3, X),$$
$$f_3(\theta_3|\theta_1, \theta_2, X)$$

are available, where f_1, f_2 and f_3 denote the condition distributions of parameters θ_1, θ_2 and θ_3 respectively. What is required in applications in the ability to draw a random number from these conditional distributions. Suppose $\theta_{2,0}$ and $\theta_{3,0}$ are starting values for θ_2 and θ_3 respectively. This method proceeds

1.) Draw a random sample from $f_1(\theta_1|\theta_{2,0}, \theta_{3,0}, X)$ and denote by $\theta_{1,1}$,

2.) Draw a random sample from $f_2(\theta_2|\theta_{1,1}, \theta_{3,0}, X)$ and denote by $\theta_{2,1}$.

3.) Draw a random sample from $f_3(\theta_3|\theta_{1,1}, \theta_{2,1}, X)$ and denote by $\theta_{3,1}$.

For the next iteration, we use the new parameters as starting values and repeat the above iteration. Under regularity conditions, it can be shown that for sufficiently large value of m, the parameters $(\theta_{1,m}, \theta_{2,m}, \theta 3, m)$ are approximately equivalent to a random draw from the joint distribution $f(\theta_1, \theta_2, \theta_2|X)$.

Under general conditions, suppose θ can be decomposed as $\theta = (\theta_1, ..., \theta_r)$ and we are able to simulate from the conditional densities

$$f_i(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_r), \quad i = 1, 2, ..., r.$$

To sample from the joint distribution of θ, we proceed as follows. Given a sample $(\theta_{1,m}, ..., \theta_{r,m})$, generate

1. $\theta_{1,m+1} \sim f_1(\theta_1|\theta_{2,m}, ..., \theta_{r,m})$.
2. $\theta_{2,m+1} \sim f_1(\theta_2|\theta_{1,m}, ..., \theta_{r,m})$.
3. $\vdots$
4. $\theta_{r,m+1} \sim f_1(\theta_r|\theta_{r,m}, ..., \theta_{r-1,m})$.

The steps above are repeated with the new parameters. Another method that is used in the Metropolis-Hasting algorithm.

7.5.2 Metroplis-Hastings Algorithm

This algorithm is used when the conditional posterior distribution is known. Suppose we need to draw a random sample from the distribution $\pi(\theta|X)$ which contains a complicated normalization constant such that a direct draw is either too time consuming or infeasible. But there exists an approximate distribution for which random draws can be sampled. This algorithm generates a sequence of random draws from the approximate distribution whose distributions converge to $\pi(\theta, X)$. This algorithm proceeds as follows.

1.) Draw a random sample starting value θ_0 such that $\pi(\theta_0|X) > 0$.

2.) For t=1,2,.....

a.) Draw a candidate sample θ_* from a known distribution at iteration t given the previous draw θ_{t-1}. Denote the known distribution by $J_t(\theta_t|\theta_{t-1})$ which is called a jumping distribution. The jumping distribution must be symmetric such that $J_t(\theta_i|\theta_j)$=$J_t(\theta_j|\theta_i)$.

b.) Calculate the ratio

$$r = \frac{\pi(\theta_*|X)}{\pi(\theta_{t-1}|X)}$$

and under the Metropolis-Hasting algorithm

$$r = \frac{\pi(\theta_*|X)J_t(\theta_{t-1}|\theta_*)}{\pi(\theta_{t-1}|X)J_t(\theta_*|\theta_{t-1})},$$

Under this algorithm, the jumping distribution does not have to be symmetric.

c.) Set

$$\theta = \begin{cases} \theta_*, & \min(r,1) \\ \theta_{t-1}, & \text{Else} \end{cases},$$

under regularity conditions, the sequence $\{\theta_t\}$ converges in distribution to $\pi(\theta|X)$.

7.6 Chapter Summary

In time series analysis, forecasting is imperative as models are designed to produce future forecasts. Since volatility is unobservable, there are a number of volatility measures, that are available for use. The squared returns have been widely used as a volatility measure. In producing forecasts, the evaluation of the forecasts is important and in literature there are methods proposed. The basic method that is used is the MZ regression which involves regressing the realization of a variable and a constant and its forecast. With this method, R^2 could be compared. Diebold and Mariano test is used for comparing two or more forecasts from different models. The use of loss functions provides a flexible form of pair-wise comparison. The pair-wise predictive ability is measured by the difference of MSE's. Except for these methods, there are other functions that could be used like the MAE, MAPE, ME, MLAE among others.

Chapter 8

Empirical Analysis of South African Platinum Returns

In this chapter, we use methods discussed in the previous chapters to investigate the distributional properties of platinum return series and inherent volatility dynamics. In the literature it has been found that financial returns are explained by heavy tailed distributions.These include generalised hyperbolic distributions and alpha stable distributions introduced by Mandel (1971). In the analysis of distributional properties, we use alpha stable distributions and then we apply recently used generalized hyperbolic distributions.

Another platinum price return phenomenon we investigate in this chapter are structural breaks that occur as a result of major events that affect the structure of the data. We first use the iterative cumulative sum of squares (ICSS) algorithm to capture mean breaks and then use a method by Shimotsu (2006) to examine any breaks found in long memory of data.

In modelling long memory in log mean returns, we apply the widely used ARFIMA model as it is able to capture long memory dynamics and any form of persistence that might be inherent in the data series. This model is attractive as it is simpler to estimate and hence conclude from it. With this model, we can be able to tell if log returns of platinum are random walk, persistent or anti-persistent. For volatility modelling we use fractional integrated GARCH type models which capture long memory and other properties of financial returns volatility. A good example of such properties is the symmetry that has been found in some financial returns especially commodities (Arouri et al., 2012a; Diaz, 2016). In the next section we discuss data aspects of platinum price series and any

formulation of return series we apply going forward.

8.1 Data Series

The data used are daily and monthly closing platinum prices from February 1994 to June 2014. This data is sourced from Matthey (2013b). In the analysis of platinum returns, we use log returns as defined by

$$r_t = \ln\left[\frac{X_t}{X_{t-1}}\right], \tag{8.1}$$

where X_t is the price of the underlying asset at time t. To measure volatility of these returns, we use squared log returns as the volatility measure.

Descriptive statistics in Table 8.1 shows that the log returns are skewed with high kurtosis which is greater than the kurtosis of a Normal distribution.

Table 8.1: Descriptive Statistics of Returns

Returns Statistic	Platinum Returns
Q1	-3.5970
Q2	0.4938
Q3	4.4950
Mean	0.0249
Kurtosis	14.9697
Skewness	0.6575

Table 8.2: Statistical tests of Returns

Test	Test-Statistic	P-value
Kolmogorov-Sminov	0.3655	0.0001
Jarque-Beta	31640.87	0.0001
Phillips-Perron (lag=10)	-3348.85	0.01

Jarque-Bera and Kolmogorov-Smirnov tests in Table 8.2 illustrates that the series are not Normally distributed. Figure 8.1 shows the Normal Q-Q plot, this illustrates that the tails of the data series are heavier than the tails of the Normal distribution.

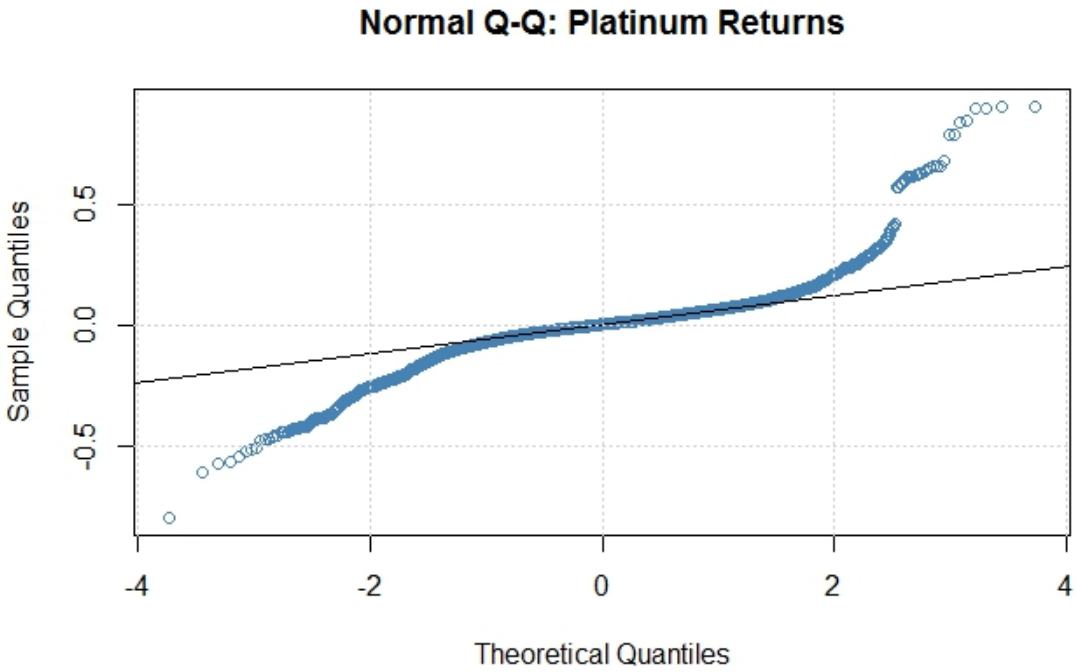

Figure 8.1: Normality Q-Q Plots

To test for unit roots, Phillips-Perron test in Table 8.2 shows that platinum return series are stationary in mean, we use the Phillips-Perron test since it is robust to the presence of serial correlation and heteroskedasticity.

Figure 8.2: Platinum prices from February 1994 to June 2014

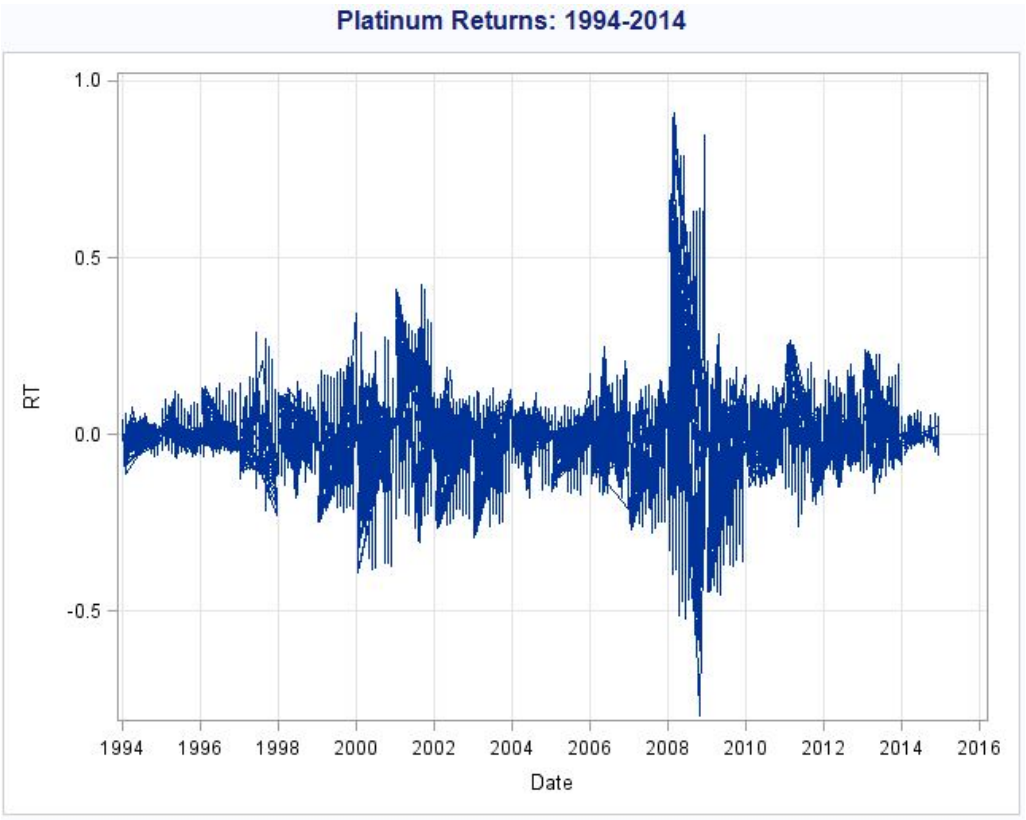

Figure 8.3: Platinum returns from February 1994 to June 2014

Figure 8.2 shows platinum prices and Figure 8.3 shows platinum return series. An erratic behaviour in prices is seen and can be further seen in platinum returns. This volatility is seen around the period of 2008/2009 where we experienced a global recession. Further, volatility clustering, a phenomenon explained by heteroskedastic models is exhibited in both figures. Jumps seen in prices might suggest that the data may contain structural breaks. In the next section, we discuss distributional properties of the return series.

8.2 α-Stable Distributions

In order to explain the returns, we fitted α-Stable distributions to platinum returns series. We fitted the data to four parameter α-stable distributions using Nolan's S_1 parameterization. The estimated parameters are shown in Table 8.3, and the value of $\hat{\alpha}$ is in the region $(0, 2]$ which illustrates that the estimated distribution is in the Stable domain. By definition, the value of $\hat{\alpha}$ close to zero signifies that a distribution has thicker tails and if it is close to 2 then a distribution has lighter tails.

Table 8.3: Stable Distribution Parameters

Parameters	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
Estimated	1.287	-0.254	0.04075	0.008417

The value of the tail estimate, $\hat{\alpha}$, is not far from 2 when we look at the range [0-$\hat{\alpha}$,$\hat{\alpha}$-2] and this suggests intermediate heavy tails.

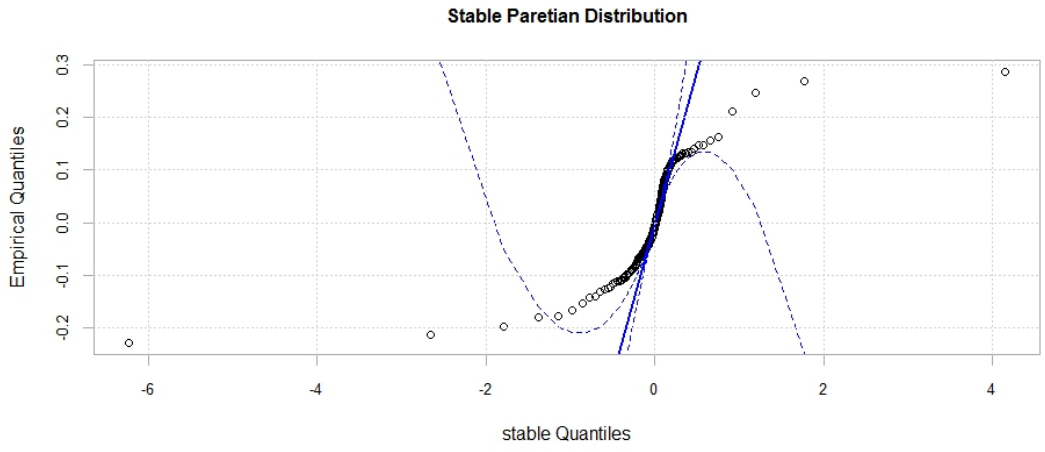

Figure 8.4: Stable Empirical Quantile Distribution

However, the quantile distribution of the data with Stable quantiles indicate that the tails of the data series distribution is heavier than the estimated tail value. By inspecting the sample quantile distribution in Figure 8.4, it is also evident that this distribution cannot explain the tails of platinum return series. Stable quantiles and empirical quantiles do not align and thus it appears that the tails of the data can be explained by a tail heavier than the estimated tail. The value of $\hat{\beta} < 0$ indicates that the fitted series is left skewed.

Table 8.4: Statistical Tests

Statistical Test	Test Statistic	p-value
Kolmogorov-Smirnov	0.0234	0.1219
Anderson-Darling	2.9840	0.0272

Table 8.4 shows the results of statistical tests, Kolmogorov-Sminorv test has a high p-value suggesting that we cannot reject the hypothesis of α-Stable returns, however the Anderson-Darling test which emphasizes tail fits, rejects the hypothesis of α-Stable returns. Further the tails of return series data and 1.287-Stable distribution, as shown by figure 8.4 are far apart and thus this suggest another class of distributions that could be used to explain the heavy tails of this data. In the following section, we fit the generalized hyperbolic distributions that has recently been widely used in the econometrics literature to explain financial returns.

8.3 Generalized Hyperbolic Distribution

From the above section, it is evident that the tails of the return series could not be explained by Normal and Stable distributions. In this section, we fit different generalized hyperbolic distributions and select the best model based on information criteria. The results of fitting generalized hyperbolic distributions to platinum return series are displayed in Table 8.5. We fit symmetric and asymmetric models to capture as much of the distributional properties as possible. When using the AIC criteria, the asymmetric normal inverse Gaussian(NIG) distribution is selected as the best model.

Table 8.5: Distributions of Returns

Model	Symmetric	Lambda	alpha.bar	mu	sigma	gamma	AIC	LLH	Converged	n.iter
NIG	FALSE	-0.5000000	0.2086312751	0.83394358	10.995948	-0.8077912	37453.93	-18722.97	TRUE	271
ghyp	FALSE	-0.5828097	0.1968024719	0.83391304	11.069800	-0.8051073	37455.02	-18722.51	TRUE	494
NIG	TRUE	-0.5000000	0.2023381382	0.57269517	11.113126	0.0000000	37470.84	-18732.42	TRUE	158
ghyp	TRUE	-0.5969093	0.1874861471	0.57272763	11.217219	0.0000000	37471.54	-18731.77	TRUE	143
t	FALSE	-1.0049419	0.0000000000	0.74294782	71.880684	-23.7454144	37495.35	-18743.68	FALSE	501
t	TRUE	-1.0000471	0.0000000000	0.55949109	733.518552	0.0000000	37505.90	-18749.95	TRUE	178
VG	FALSE	0.6734416	0.0000000000	0.85106897	10.027476	-0.6322782	37670.10	-18831.05	TRUE	309
VG	TRUE	0.6826808	0.0000000000	0.48543783	10.071785	0.0000000	37688.42	-18841.21	TRUE	157
hyp	FALSE	1.0000000	0.0005853805	0.95216858	9.548469	-0.9278388	37769.72	-18880.86	TRUE	213
hyp	TRUE	1.0000000	0.0001113495	0.49364723	9.594733	0.0000000	37792.71	-18893.36	TRUE	182
gauss	TRUE	NA	Inf	0.02493965	11.102176	0.0000000	40077.36	-20036.68	TRUE	0

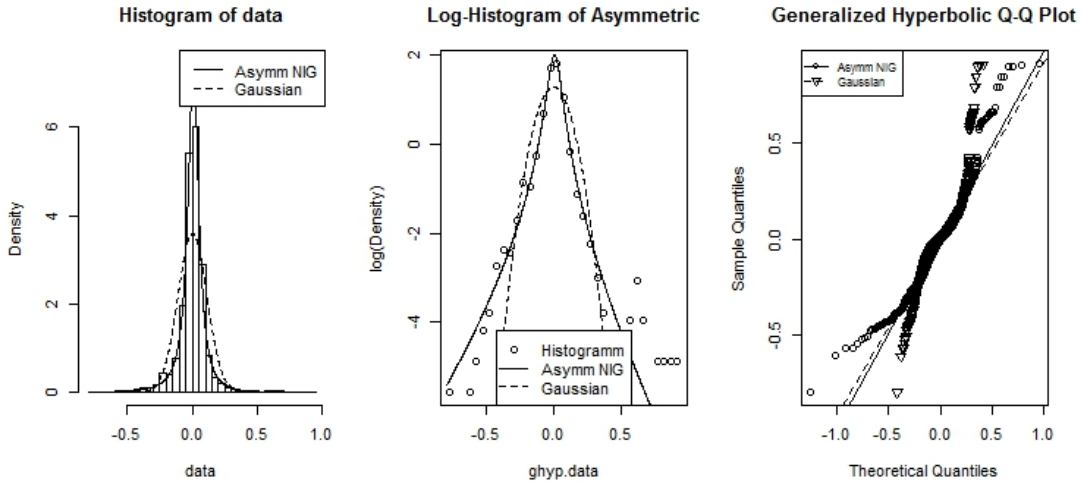

Figure 8.5: Platinum Returns under Asymmetric Normal Inverse Gaussian Distribution

Table 8.6: Statistical Tests

	Kolmogorov-Smirnov		Anderson-Darling	
Distribution	Statistic	p-value	Statistic	p-value
Normal Inverse Gaussian	0.0250	0.0819	2.5440	0.04716
Asymmetric Normal Invese Gaussian	0.0024	0.1540	1.958	0.0969
Generalized hyperbolic p	0.0180	0.3786	1.7420	0.1280
Asymmetric Generalized hyperbolic	0.0197	0.2741	1.282	0.2382
Student-t	0.0239	0.1089	3.7390	0.0117
Asymmetric Student-t	0.0311	0.0141	4.1800	0.0072
Variance Gamma	0.0373	0.0016	5.5890	0.0015
Asymmetric Variance Gamma	0.0331	0.0074	5.5230	0.0016
Hyperbolic	0.0419	0.0003	9.727	0.0001
Asymmetric Hyperbolic	0.0325	0.0090	6.8280	0.0004

Figure 8.5 shows the histogram of the Normal Inverse Gaussian fit, the log-histogram and quantile-quantile plot for the fitted data, fitted plots for other distributions are shown in the appendix. From the figures produced, it is evident that the Gaussian distribution does not explain the data better that any of the generalised hyperbolic distributions. We must note that as much as the asymmetric NIG fits the returns better, it does not perfectly fit the tails of the platinum returns.

Statistical tests are shown in Table 8.6. At 5% level of significance, the Kolmogorov-Smirnov test rejects the null hypothesis of the asymmetric Student t, Variance Gamma, asymmetric Variance Gamma, Hyperbolic and asymmetric Hyperbolic distribution. The Anderson-Darling additionally rejects the Student t distribution. This illustrates that the generalized hyperbolic distributions can be used to explain platinum return series.

8.4 Structural Breaks Diagnosis

In the structural breaks diagnosis we start by using the mean breaks algorithm which is the ICSS algorithm to identify breaks that might arise from sudden changes in data. The second algorithm we use is due to Shimotsu (2006) which investigates long memory parameter consistency and lastly, we use the sample accumulation algorithm that assists in identifying trends and regimes in long memory.

8.4.1 Mean Breaks

To identify breaks in mean, we used the ICSS algorithm to determine break dates. Table 8.7 lists the dates of breaks found in platinum return series and respective averages per regime identified. The 2008/2009 breaks can be attributed to the global financial crises that resulted in global economy disintegrating, and the 2012/2013 breaks can be associated with the South African labour unrest in the mining sector. Such breaks may introduce spurious results in persistence and long memory.

It can be seen from Figure 8.6 that platinum return series is separated into different regimes. This indicates that persistence observed in the return series might be spurious and hence arise from breaks and shifts. In testing persistence using the results from the ICSS algorithm, simple GARCH models are used and a comparison is done by including dummy variables in one GARCH model to the simple GARCH model without dummy variables. Using this method, for every break or regime experienced a dummy variable is assigned. This is done to account for regimes in estimating data persistence. We compared two GARCH models, the standard GARCH(1,1) model which does not contain dummy variables and the GARCH(1,1) model with dummy variables and the results are shown in Table 8.8.

Table 8.7: ICSS Algorithm Results of platinum squared returns

Break-Dates	Averages
02JAN1994	0.000452094
02FEB1994	0.0003475636
24NOV1994	0.0000734619
13JAN1995	-0.000026251
06NOV1996	-0.000818846
02APR1997	0.0010452113
17OCT1997	-0.000356536
02NOV1999	0.0014604989
02MAY2001	-0.001052794
02JUN2001	0.0004234699
24DEC2003	0.0007842299
02MAY2008	-0.001498914
01AUG2008	-0.004645594
01FEB2009	0.0009802843
06JUL2011	-0.000785104
07MAR2012	-0.00026438
13NOV2013	-0.000072072

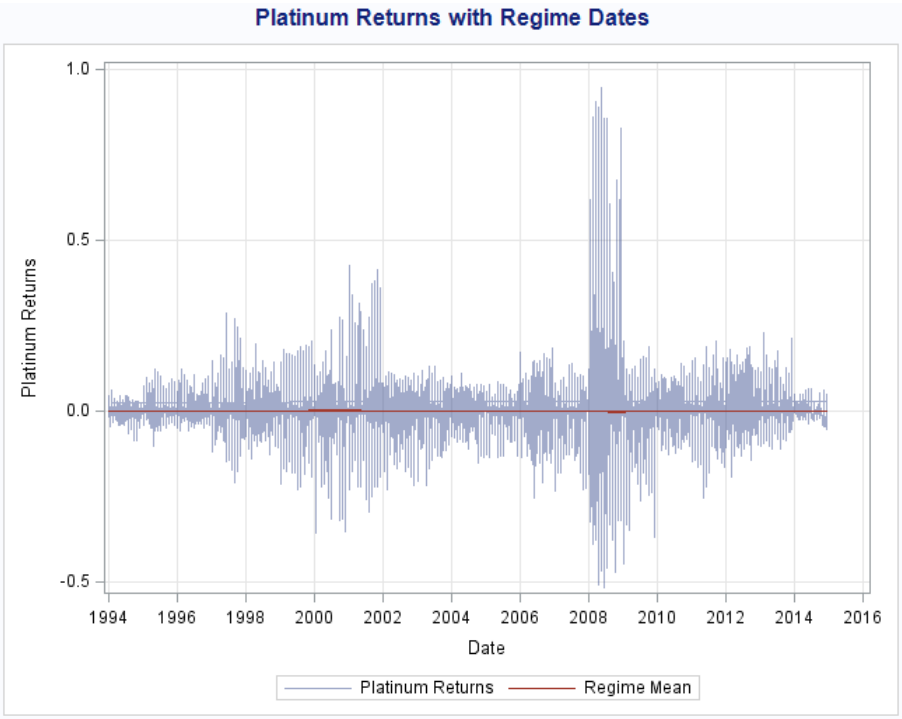

Figure 8.6: Platinum Returns with Regime Averages

Table 8.8 shows the results of the standard GARCH(1,1) model and GARCH(1,1) model with dummy variables (D-GARCH(1,1)). The persistence of the model with dummy variables, in theory, should be less than the persistence of the model without

dummy variables.

Table 8.8: GARCH(1,1) results and D-GARCH(1,1) the GARCH with dummy variables

Parameter	DF	Estimate	St.Err	t-value	P-value
GARCH(1,1)					
ARCH0 (α_0)	1	5.63E-6	2.56E-6	2.20	0.0275
ARCH1 (α_1)	1	0.0411	0.003165	12.98	<0.0001
GARCH1 (β_1)	1	0.9596	0.002948	325.55	<0.0001
Persistence		1.0007			
D-GARCH(1,1)					
ARCH0 (α_0)	1	6.12E-6	2.57E-6	2.38	0.0173
ARCH1 (α_1)	1	0.0406	0.003151	12.88	<0.0001
GARCH1 (β_1)	1	0.9597	0.002946	325.73	<0.0001
Persistence		1.0003			

Notes: The persistence is measured by the GARCH parameters (α_1+β_1)

The difference in persistence between the two models is very small and this suggests that the persistence in platinum prices is not affected by the breaks in data. Hence, the high persistence observed might come from other sources other than structural breaks. As it has been investigated in literature that long memory and structural breaks can be confused since one can arise from another, in the next subsection we investigate breaks in the long range dependence parameter estimation.

8.4.2 Breaks in Long Memory

In structural breaks diagnosis of long memory, we used a method introduced by Shimotsu (2006) which tests parameter consistency using sub-samples methodology. For this method, we split the sample into sub-samples and for each of the sub-samples selected, we obtain estimates of d, the long range dependence parameter for the sub-samples, $\hat{d}_2$ and $\hat{d}_4$ which are the averages of splitting the sub-sample into 2 and 4 samples respectively. We used the Wald test statistic on $\hat{d}_2$ and $\hat{d}_4$ to test parameter consistency in long range dependence parameters. Chi-square statistics $\chi^2_{0.95}(1)$=3.84 and $\chi^2_{0.95}(3)$=7.82 were used as cut off values for testing the significance of d_2 and d_4 at the 5% level of significance, respectively.

From Table 8.9, it is evident that the platinum return series contain breaks. It can be seen that the long range dependence parameter is not consistent between sub-samples. This is further shown by the rejection of parameter consistency by W_2 and W_4 tests. The

KPSS test statistic does not reject the presence of long memory. The results from Table 8.9 shows that sub-sample parameter estimates are not consistent using the Wald test statistic and the KPSS test does not reject the presence of long memory. This illustrates that not all long memory presence is due to structural breaks.

Table 8.9: Test Results of platinum squared returns

m	$\hat{d}$	$\hat{d}_2$	$\hat{d}_4$	W_2	W_4	KPSS	p-value(KPSS)
500	0.0125	0.0095	-0.0494	0.6151	5.6580	0.0077	0.1000
1000	0.1275	0.0125	0.0095	7.9680	19.4500	0.0184	0.1000
1500	0.0763	-0.0016	0.0091	0.0001	10.9500	0.0185	0.1000
2000	0.0781	0.1275	0.0125	22.6700	31.5300	0.0089	0.1000
2500	0.0670	0.1156	0.0120	38.5800	15.0100	0.0077	0.1000
3000	0.0722	0.0763	-0.0016	3.0550	11.5700	0.0131	0.1000
3500	0.0672	0.1054	0.1547	16.3000	46.6100	0.0250	0.1000
4000	0.1085	0.0781	0.1275	2.8320	30.0600	0.0093	0.1000
4500	0.1076	0.0712	0.1175	5.3880	33.0400	0.0153	0.1000
5000	0.1076	0.0670	0.1156	8.0600	53.4300	0.0145	0.1000

From Table 8.9 we have seen that different sub samples produce different long memory results. Following from this, we analyzed the effect of each of the sub-samples to the long range dependence parameter overtime using a method of sample accumulation and estimation. With this method, we accumulate sub-samples and estimate the long-range dependence parameter for each of the accumulations. The results of this method is shown in Figure 8.7. Thus from Figure 8.7, it can be seen that between sample 30 and sample 40, there was a period where the long range dependence estimate jumped.

Long range dependence (LRD) parameter estimates up to the 20th cumulative sub-sample are volatile. Data contained are from February 1994 to September 2004. The total observations used up to this point are 2800 and hence, the volatility seen is due to small data. From the 21st cumulative sample, the LRD parameter estimates are less volatile up to sample 34. The introduction of sample 35 to the total sample created a jump in estimation. Data added is from April 2013 to August 2013. This suggests that there was a break or a shift in volatility experienced during this period. As a result, from the 36th sample on-wards, LRD parameter estimates are higher than the initial estimate. This illustrates that the LRD estimate in this period might be contaminated by breaks information and thus might not be a true reflection of the memory inherent in data or another cause might be that LRD of this return series is time varying as shown in Figure

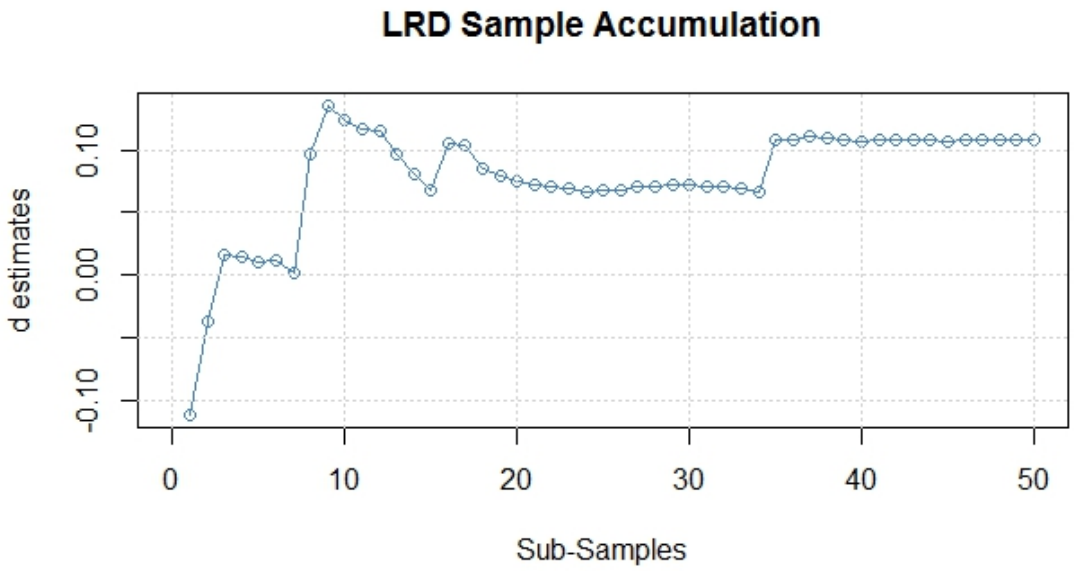

Figure 8.7: long range dependence parameter of sample accumulation

8.7.

With this information at hand, we can split the data series into 2 regimes as shown in Figure 8.7. The LRD estimate for the first regime (d_{R1}) and the LRD estimate for the second regime (d_{R2}), that is d_{R1}=0.07 and d_{R2}=0.11 as defined by the averages. In modelling of long memory, the first regime can be modelled using d_{R1} and the second regime can be modelled using d_{R2}. To confirm the trend shown, a backward estimation of sub-sampling can be done which in theory should return the same results as shown in Figure 8.7. Thus, using these results, we can use April 2013 as a date where there was a shift in long memory inherent in the data series. A formulation can be expressed as

$$\hat{d} = \begin{cases} \hat{d_{R1}}(t) & t < April - 2013 \\ \hat{d_{R2}}(t) & t \geq April - 2013 \end{cases},$$

in modelling of log returns for the platinum return series. In the next section, we test for LM and estimate the long range dependence parameter using different estimation methods.

8.5 Long Memory Tests

In long memory testing, we fitted different long memory tests to the squared log returns of platinum prices. The Hurst exponent results of long memory tests are shown in Table 8.10 for platinum log squared returns.

Table 8.10: Platinum Log Squared Returns Long Memory Tests

Method	Hurst	Standard Error	t-value	p-value
Aggregated Variance Method	0.9358(0.4358)	0.0421	22.2319	<0.0001
Differenced Aggregated Variances	1.1907(0.6907)	0.1813	6.5684	<0.0001
Aggregated Absolute Value Method	0.9909 (0.4909)	0.0253	39.2360	<0.0001
Higuchi Method	0.9739 (0.4739)	0.0358	27.1544	<0.0001
Peng Method	0.6836 (0.1836)	0.1127	6.0681	<0.0001
R/S Method	0.6667 (0.1667)	0.0754	8.8486	<0.0001
Periodogram Method (GPH)	0.9665 (0.4665)	0.0382	25.2618	<0.0001
Boxed(Modified) Periodogram Method	0.8313 (0.3313)	0.0463	17.9453	<0.0001
Wavelet Estimator	0.5248 (0.0248)	0.1005	5.2199	0.0034
Whittle Estimator	0.6080 (0.1080)	0.0089	68.2000	<0.0001

Note: The value of $\hat{d}$ is shown in brackets

All of the tests used suggest long memory as all the p-values are less than 0.01. Note that the differenced aggregated variances method violates the condition $0 < H < 1$. This should not be a concern as its main purpose is to distinguish nonstationarity due to jumps ($H \cong 0.5$) to that due to actual trend ($H \gg 0.5$). So in this case, trend is not due to jumps in the data. It is clear that platinum squared returns have high persistence and it appears they could be explained by a fractionally integrated model. In the next section, we fit long memory mean models and conditional volatility models on platinum return series to investigate dual long memory of mean returns and volatility.

8.6 Empirical Results of Volatility Models

To explain the dual long memory of the mean and volatility of platinum return series, we fitted ARFIMA-FIGARCH type models under heavy tailed error distirubtions. We used an ARFIMA model for modelling log returns and for volatility we used FIGARCH, FIEGARCH, FIAPARCH and HYGARCH under heavy tailed error distributions. Distributions considered are the Normal, Student, Generalized extreme distribution (GED), and the skewed Student distribution.

Table 8.11: ARFIMA-FIGARCH Parameter Estimation of Models

Parameters	Normal	Student	GED	Skewed Student
Cst(M)	0.0002**	0.0001***	0.0001**	0.0003***
d_m	-0.1027***	-0.0893**	-0.0838	-0.5276***
AR(1)	0.4062***	0.4158***	0.4193***	0.1356***
MA(1)	-0.9327***	-0.9353***	-0.9361***	0.0699*
Cst(V)	0.0013**	0.0011*	0.0011*	0.0006
d_v	0.6492***	0.6486***	0.6473***	0.6545***
ARCH(α_1)	0.3686***	0.3567***	0.3543***	0.2138***
GARCH(β_1)	0.8721***	0.8730***	0.8720***	0.8680***
Akaike	-2.4717	-2.4720	-2.4715	-2.4017
Schwarz	-2.4591	-2.4578	-2.4574	-2.3860
ARCH-LM	4.0565**	4.7917***	4.6751***	4.5570**

Notes: $*$,$**$ and $***$ represent the significant level at 10%, 5% and 1% levels respectively.

Table 8.12: ARFIMA-FIEGARCH Parameter Estimation of Models

Parameters	Normal	Student	GED	Skewed Student
d_m	-0.4668***	-0.4012***	-0.3903***	-0.0488
AR(1)	-0.2576***	-0.0012	0.0300	0.4407***
MA(1)	0.3411***	0.1230	0.0863	-0.9404***
d_v	0.8975***	0.9173***	0.9173***	0.6623***
ARCH(α_1)	1.1344***	1.2337***	1.2557***	-0.6664***
GARCH(β_1)	-0.8946***	-0.8977***	-0.9047***	0.8725***
EGARCH(θ_1)	0.0234***	0.0645***	0.0433***	0.0414
EGARCH(θ_2)	0.2187***	0.2279***	0.2122***	0.1367
Akaike	-2.2483	-2.2809	-2.2885	-2.4602
Schwarz	-2.2358	-2.2668	-2.2743	-2.4413
ARCH-LM	0.8213	14.202***	14.175***	1.6987

Notes: $*$,$**$ and $***$ represent the significant level at 10%, 5% and 1% levels respectively.

Parameter estimation results are shown from Tables 8.11 to 8.14. Let d_m denote LM paramater in the mean model. For all the models, the long range dependence parameter of the ARFIMA model is negative ($-1/2 < d_m < 0$) indicating anti-persistence (intermediate persistence) of log returns. This illustrates that log returns of platinum are mean reverting and hence will revert to the mean overtime. Let d_v denote LM paramater in the volatility model. For volatility the long range dependence parameter is positive ($0 < d_v < 1$) and shows strong long memory. This confirms the results by other authors (Arouri et al., 2012a), platinum shows high persistence.

Based on the Akaike information criterion, the best model is ARFIMA-FIAPARCH under the Student distribution. However, the ARCH-effect is slightly significant (*). Based on the Schwarz information criterion, the best model is ARFIMA-FIAPARCH un-

Table 8.13: ARFIMA-FIAPARCH Parameter Estimation of Models

Parameters	Normal	Student	GED	Skewed Student
Cst(M)	0.0001	0.0001	0.0001	0.0003***
d_m	-0.0793***	-0.0556	-0.0410	-0.5176***
AR(1)	0.3920***	0.4053***	0.4137***	0.1308***
MA(1)	-0.9327***	-0.9361***	-0.9381***	0.0687*
Cst(V)	0.0041	0.0043*	0.0046*	0.0005
d_v	0.6512***	0.6462***	0.6378***	0.6432***
ARCH(α_1)	0.3974***	0.3929***	0.3919***	0.2325***
GARCH(β_1)	0.8743***	0.8744***	0.8718***	0.8607***
APARCH(γ_1)	-0.2928***	-0.3716***	-0.4169*	-0.2012*
APARCH(δ)	1.8067***	1.7569***	1.7309***	2.0904***
Akaike	-2.4773	-2.4783	-2.4781	-2.4030
Schwarz	-2.4616	-2.4610	-2.4608	-2.3841
ARCH-LM	2.2572	2.6098*	2.7374*	2.8892*

Notes: $*$,$**$ and $***$ represent the significant level at 10%, 5% and 1% levels respectively.

Table 8.14: ARFIMA-HYGARCH Parameter Estimation of Models

Parameters	Normal	Student	GED	Skewed Student
d_m	-0.0802***	-0.4125***	-0.3947***	-0.4247***
AR(1)	0.3788***	0.0254	0.0684	0.0899**
MA(1)	-0.9355***	0.0985*	0.0540	0.0587
d_v	0.8199***	0.7930***	0.7968***	0.7983***
ARCH(α_1)	0.2981***	0.1091***	0.1086***	0.0960***
GARCH(β_1)	0.9142***	0.909***2	0.9088***	0.9113***
Akaike	-2.4590	-2.3467	-2.3547	-2.3710
Schwarz	-2.4480	-2.3341	-2.3421	-2.3568
ARCH-LM	1.9986	1.4914	0.72470	0.72483

Notes: $*$,$**$ and $***$ represent the significant level at 10%, 5% and 1% levels respectively.

der the Normal distribution and has no ARCH-effect. Although the ARFIMA-FIEGARCH under the Skewed Student distribution and ARFIMA-HYGARCH under the Normal distribution were not selected based on the two information criteria they have no ARCH-effects.

This agrees with the results of Diaz (2016) who found that platinum returns volatility are characterized by asymmetric response to negative and positive shocks as explained by the FIAPARCH model. From the results for models, γ <0 which illustrates that positive shocks have relatively more impact than negative shocks on volatility. Although these metals respond similar to negative and positive news, positive news have a higher impact and hence making these metals a good investment vehicle as outlined in Arouri et al. (2012a). We discuss forecast performance of these models in the following section.

8.7 Forecast Evaluation Methods

Evaluation of forecasts for models is important as it helps us understand the forecasting accuracy of the models estimated. There are a number of forecasts evaluation measures available in literature. For our analysis, we used three measures commonly used in literature, namely the mean square error (MSE), the mean absolute error (MAE) and the Theil Inequality Coefficient (TIC). These measures are defined as

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t - \hat{\sigma}_t)^2,$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\sigma_t - \hat{\sigma}_t|,$$

$$\text{TIC} = \frac{\sqrt{1/n\sum_{t=1}^{n}(\sigma_t - \hat{\sigma}_t)^2}}{\sqrt{1/n\sum_{t=1}^{n}\sigma_t^2} + \sqrt{1/n\sum_{t=1}^{n}\hat{\sigma_t}^2}},$$

where n is the number of forecasts, σ_t is the observed volatility and $\hat{\sigma}_t$ is the predicted conditional volatility at time t. The best model must exhibit least prediction error as given by the three measures. Table 8.15 to Table 8.18 show forecast evaluation results. For platinum return series, the MSE gives low prediction errors for all models

Table 8.15: ARFIMA-FIGARCH Forecast Evaluation

Parameters	Normal	Student	GED	Skewed Student
MSE	0.0003	0.0003	0.0003	0.0003
MAE	0.0082	0.0082	0.0083	0.0087
TIC	0.6515	0.6517	0.6517	0.6331
Alpha(MZ)	0.0002	0.0002	0.0002	0.0006
Beta(MZ)	1.2062	1.1974	1.1948	1.0028
R^2(MZ)	0.0689	0.0670	0.0665	0.0551

Table 8.16: ARFIMA-FIEGARCH Forecast Evaluation

Parameters	Normal	Student	GED	Skewed Student
MSE	0.0004	0.0004	0.0004	0.0004
MAE	0.0097	0.0115	0.0107	0.0085
TIC	0.6012	0.5626	0.5773	0.6657
Alpha(MZ)	0.0034	0.0035	0.0037	0.0028
Beta(MZ)	0.5231	0.3826	0.4132	0.7664
R^2(MZ)	0.0317	0.0320	0.0307	0.0220

except ARFIMA-FIEGARCH which has slightly high errors. Based on the MAE, the ARFIMA-FIAPARCH under the Normal and Student distribution and the ARFIMA-HYGARCH under the Normal distribution gives less prediction errors .Lastly, for the TIC, the ARFIMA-FIEGARCH under Student distribution gives less prediction error. Hence this confirms the selection of ARFIMA-FIAPARCH models under Student and GED error distributions as good models since it it evident for the MAE evaluation measure.

Table 8.17: ARFIMA-FIAPARCH Forecast Evaluation

Parameters	Normal	Student	GED	Skewed Student
MSE	0.0003	0.0003	0.0003	0.0003
MAE	0.0081	0.0081	0.0082	0.0087
TIC	0.6699	0.6721	0.6727	0.6384
Alpha(MZ)	0.0004	0.0005	0.0005	-0.0009
Beta(MZ)	1.2716	1.2550	1.2415	1.1667
R^2(MZ)	0.0634	0.0584	0.0554	0.0510

Table 8.18: ARFIMA-HYGARCH Forecast Evaluation

Parameters	Normal	Student	GED	Skewed Student
MSE	0.0003	0.0003	0.0003	0.0003
MAE	0.0081	0.0088	0.0087	0.0092
TIC	0.6581	0.6319	0.6351	0.6152
Alpha(MZ)	0.0007	0.0013	0.0013	0.0014
Beta(MZ)	1.2053	0.9064	0.9233	0.8067
R^2(MZ)	0.0737	0.0491	0.0495	0.0470

A popular method used for assessing forecasting performance of volatility models is the Mincer-Zarnowitz regression defined

$$\tilde{\sigma}_t^2 = \alpha + \beta\hat{\sigma}_t^2 + u_t, \quad t = 12, ..., T,$$

where $\tilde{\sigma}_t^2$ is the observed volatility as measured by squared innovations and $\hat{\sigma}_t^2$ is the predicted volatility. If the conditional volatility model is correctly specified and $\tilde{\sigma}_t^2$ is unbiased for the true variance then the parameters will be $\alpha = 0$ and $\beta = 1$. This then suggest that the observed volatility will compleetely be explained by the predicted volatility. An R^2 value from this regression model compares predictive ability of volatility models. The Mincer-Zarnowitz regression results are shown in Table 8.15 to Table 8.18. This regression jointly tests the intercept (alpha) and slope (beta). For all the models used, the null hypothesis for zero intercept is rejected at 5% level of significance. This tells us that the models will underestimate or overestimate the volatility to some extent and thus we would need to adjust the forecasts with calculated intercept values. For the selected models the platinum model has slope of 0.0005, hence the platinum model underestimates volatility by 0.0005. The null hypothesis of a unit slope is not rejected at 5% level of significance for all models. This tells us that our forecasts from the models explains the observed values. In summary, the ARFIMA-FIGARCH type models under heavy tailed error distributions shows an improvement of forecasts as compared to the assumption of Normally distributed errors, and further ARFIMA-FIAPARCH models proved to explain platinum returns series under these error distributions.

8.8 Summary

To summarize this chapter, the return series of platinum prices was found to be not normally distributed. Different heavy tailed distributions were tested, specifically two classes, the alpha-stable distributions and generalized hyperbolic distributions. Alpha-stable distributions do not fit tails of platinum return series, generalized hyperbolic distributions were then used. In generalized hyperbolic distributions, the normal inverse gaussian distribution was selected based on the information criteria, as it performed better than other distributions. We must note however that the asymmetric NIG did not perfectly fit the returns well.

In modeling volatility, especially long memory models, structural breaks in data produce spurious long memory. To test spurious long memory, we used a method due to Shimotsu (2006). This method suggested that memory in data is not spurious. We then tested long memory in squared returns and all the tests used suggested long memory. The ARFIMA-FIGARCH type models were used in the modelling of mean and volatility. ARFIMA models were used for mean process and FIGARCH models were used to model volatility. ARFIMA-FIAPARCH model was selected under the AIC and BIC criteria, this illustrates that platinum returns volatility has an asymmetric response to negative and positive shocks.

On forecast evaluation using the MSE, MAE and TIC, ARFIMA-FIAPARCH models under heavy tailed error distributions performs well. The Student and GED distributions were chosen. To examine performance of these models, we used the Mincer-Zarnowitz regression which jointly estimates the slope and intercept of regressing the forecast value on true volatility measure. We found that these models will underestimate volatility to some extent and needs adjustment in estimation.

Chapter 9

Summary and Further Research

In this thesis, we analyzed log squared returns of platinum price series. Different aspects of squared log returns were discussed in this thesis which explain conditional volatility dynamics of the platinum return series. Our data analysis showed that the return series is skewed and has high kurtosis. This then tells us that the platinum price return series in not normal. We used different distributions to identify a distribution that best explains the return series. The first distribution was the alpha-stable distribution, this distribution could not fit the tails of the return series. The next distribution was the class of generalized distributions, the decision of the selected distribution was based on the AIC and the log likelihood.

From the six distributions of the generalized hyperbolic distributions, the NIG was selected as the distribution of choice. We then tested structural breaks in the returns series and could only identify additive breaks which occurred in June 2008. This break did not affect long memory, this was confirmed by the visual inspection of changes in long-range parameter of different samples. Long memory tests on squared returns all suggested that there is long memory in platinum squared return series.

In modeling volatility dynamics, we started by GARCH-type models which includes the ARMA-GARCH, ARMA-IGARCH and ARMA-EGARCH. The first observation that was identified was with the results of the ARMA-GARCH model,i.e. the two GARCH parameters were close to unit and this suggests that the data is integrated and hence cannot be explained by a standard GARCH model. The next model we fitted was the ARMA-IGARCH and ARMA-EGARCH model.

From the AIC criteria and the log likelihood, the ARMA-EGARCH model was se-

lected to be the best model. In modeling long range dependence, we fitted four models, namely, ARFIMA-FIGARCH, ARFIMA-FIEGARCH, ARFIMA-FIAPARCH and ARFIMA-HYGARCH. From these models, the ARFIMA-FIEGARCH was selected as the best model based on the AIC and BIC and log likelihood. When studying the Mincer-Zarnowitz regression's R^2, MAPE and AMAPE forecast evaluation, the ARFIMA-FIEGARCH has the least forecast errors. These results leave a question for further research onto time-varying fractional integrated conditional volatility model and stochastic models.

Bibliography

Arouri, M., Hammoudeh, S., Lahiani, A., and Nguyen, D. (2012a). Long memory and structural breaks in modeling the return and volatility dynamics of precious metals. *The Quarterly Review of Economics and Finance*, 52(2):207 – 218.

Arouri, M., Lahiani, A., Levy, A., and Nguyen, D. (2012b). Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Economics*, 34(1):283 – 293.

Baillie, R. and Kapetanios, G. (2007). semi parametric estimation of long memory: the holy grail or a poisoned chalice?.

Baillie, R. and Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive figarch approach. *Journal of Economic Dynamics and Control*, 33(8):1577 – 1592.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, Elsevier*, 74(1):3–30.

Ben Nasr, A., Boutahar, M., and Trabelsi, A. (2010). Fractionally integrated time varying garch model. *Statistical Methods & Applications*, 19(3):399–430.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bollerslev, T. and Mikkelsen, H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics, Elsevier*, 73(1):151–184.

Box, G. and Cox, D. (1964). An anaysis of transformations. *Journal of the Royal Statistical Society*, 26:211–252.

Box, G., Jeckins, G., and Reinsel, G. C. (1994). Time series analysis. In *Time Series Analysis*. New Jersey : Prentice Hall.

Box, G. and Tiao, G. (1973). Bayesian inference in statistical analysis. In *Bayesian Inference in Statistical Analysis*. New York : Wiley.

Chakravarti, L., Laha, R., and Roy, J. (1967). Handbook of methods of applied statistics. *John Wiley and Sons*, 1:392–394.

Chung, C. (1999). Estimating the fractional integrated garch model. *Working Paper, National Taiwan University*.

Conrad, C. (2010). Non-negativity conditions for the hyperbolic garch model. *Journal of Econometrics*, 157(2):441–457.

Conrad, C. and Haag, B. R. (2006). Inequality constraints in the fractionally integrated garch model. *Journal of Financial Econometrics*, 4(3):413–449.

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

Davidson, J. E. H. (2004). Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22(1):16–29.

Diaz, J. (2016). Do scarce precious metals equate to safe habour investments? the case of platinum and palladium. *Economics Research International*, 2016.

Diebold, F. and Mariano, R. (1994). Comparing predictive accuracy. Working Paper 169, National Bureau of Economic Research.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):pp. 987–1007.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 4:221–238.

Giraitis, L., Kokoszka, P., and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, 38(4):1033–1054.

Giraitis, L., Kokoszka, P., Leipus, R., and Teyssire, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112(2):265 – 294.

Giraitis, L., Koul, H., and Surgailis, D. (2012). Large sample inference for long memory processes. In *Large Sample Inference for Long Memory Processes*.

Granero, M., Segovia, J., and Perez, J. (2008). Some comments on hurst exponent and the long memory processes on capital markets. *Physica A: Statistical Mechanics and its Applications*, 387(22):5543 – 5551.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Hammoudeh, S., Malik, F., and McAleer, M. (2011). Risk Management of Precious Metals. KIER Working Papers 765, Kyoto University, Institute of Economic Research.

Hansen, P. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24:127–161.

Holton, G. (2004). Difining risk. *Financial Analysts Journal : The CFA Institute*, 60(6).

Huang, C., Chinhamu, K., Huang, C., and Hammujuddy, J. (2014). Generalized hyperbolic distributions and value-at-risk estimation for the south african mining index. *International Business and Economics Research Journal*, 13(2):319–328.

Hurst, E. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Engineers*, 116.

Hwang, Y. (2001). Asymmetric long memory {GARCH} in exchange return. *Economics Letters*, 73(1):1 – 5.

Inclan, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923.

Jarque, C. and Bera, A. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55:163–172.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, 50(2):507–528.

Kale, M. and Butar, F. (2010). Fractal analysis of time series and distribution properties of hurst exponent. *Journal Of Mathematical Sciences and Mathematical Education*.

Kilic, R. (2007). Conditional Volatility and Distribution of Exchange Rates: GARCH and FIGARCH Models with NIG Distribution. *Studies in Nonlinear Dynamics & Econometrics*, 11(3):1–33.

Kristoufek, L. (2012). How are rescaled range analyses affected by different memory and distributional properties: A monte carlo study. *Institute of Economic Studies, Charles University*, 4(8).

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(13):159 – 178.

Ladislav, K. and Petra, L. (2013). Long-term Memory in Electricity Prices: Czech Market Evidence. *Czech Journal of Economics and Finance (Finance a uver)*, 63(5):407–424.

Lane, G., Terblanche, M., Meyer, G., and Sasto, N. (2012). Case study on quantitative risk modelling to obtain a realistic risk-adjusted project valuation. *The Southern African Institute of Mining and Metallurgy*.

Lee, D. and Schmidt, P. (1996). On the power of the {KPSS} test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics*, 73(1):285 – 302.

Lo, A. (1991). Long-term memory in stock market prices. *Econometrica*, pages 1279–1313.

Lopes, S. and Prass, T. (2014). Theoretical results on fractionally integrated exponential generalized autoregressive conditional heteroskedastic processes. *Physica A: Statistical Mechanics and its Applications*, 401(C):278–307.

Mandel, B. (1971). Analysis of long-run dependence in economics: the r/s technique. *Econometrica*, 39:68–69.

Matthey, J. (2013a). Platinum 2013. http://www.platinum.com. Johnson Matthey:Platinum Today.

Matthey, J. (2013b). Platinum today. http://www.platinum.com/prices/price-charts. Johnson Matthey:Platinum Today.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns : A new approach. *Econometrica*, 59:347–370.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):pp. 631–653.

Phillips, P. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):pp. 1630–1661.

Ruiz, E. and Prez, A. (2003). Asymmetric long memory garch: a reply to hwangs model. *Economics Letters*, 78(3):415 – 422.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464.

Shibata, R. (1976). Selection of the order of an autoregressive model by akaike's information criterion. *Biometrika*, 63(1):117–126.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. *Queen's Economics. Working Paper: 1101.*

Shimotsu, K. and Phillips, P. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

Socgina, V. K. and Wilcox, D. (2014). A comparison of generalized hyperbolic distribution models for equity returns. *Hindawi Publishing Corporation Journal of Applied Mathematics*, 2014.

Taqqu, M. S. and Teverovsky, V. (1996). Semi-parametric graphical estimation techniques for long-memory data. In Robinson, P. and Rosenblatt, M., editors, *Athens Conference on Applied Probability and Time Series Analysis*, volume 115 of *Lecture Notes in Statistics*, pages 420–432. Springer New York.

Teverovsky, V., Taqqu, S., and Willinger, W. (1999). A critical look at lo's modified r/s statistic. *Journal of Statistical Planning and Inference*, 80(12):211 – 227.

Wright, J. (2000). Log-periodogram estimation of long memory volatility dependencies with conditionally heave tailed returns. *International Finance Discussion Papers*.

Wright, J. H. and Bollerslev, T. (1999). High frequency data, frequency domain inference and volatility forecasting. International Finance Discussion Papers 649, Board of Governors of the Federal Reserve System (U.S.).

Zaffaroni, P. (2009). Whittle estimation of {EGARCH} and other exponential volatility models. *Journal of Econometrics*, 151(2):190 – 200. Recent Advances in Time Series Analysis: A Volume Honouring Peter M. Robinson.

Appendix A

Platinum Return Distributions

A.1 Results

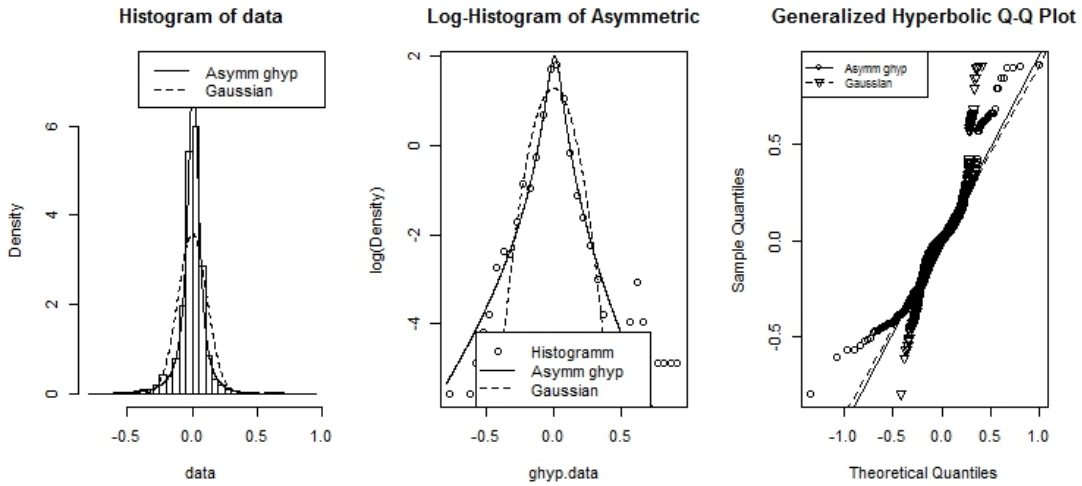

Figure A.1: Platinum Returns under Asymmetric Generalised Hyperbolic Distribution

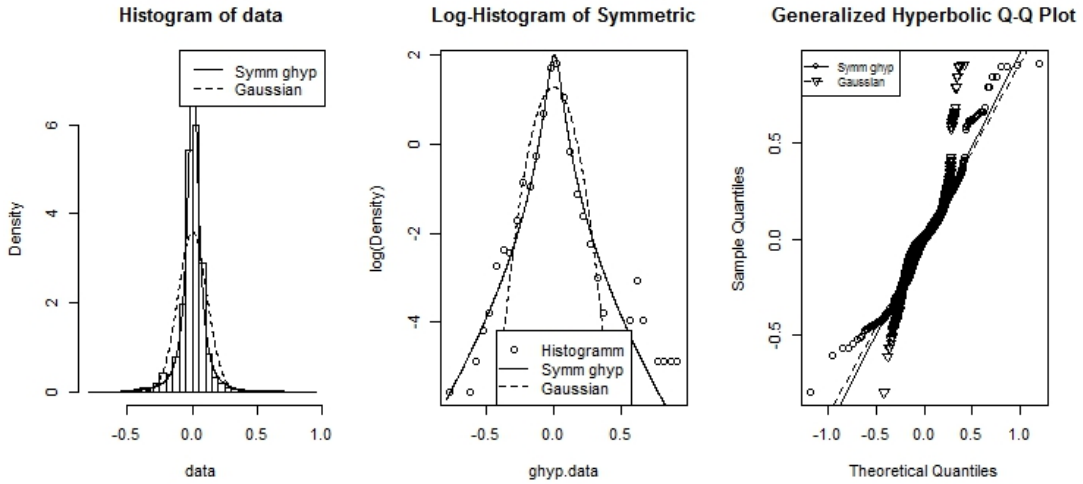

Figure A.2: Platinum Returns under Symmetric Generalised Hyperbolic Distribution

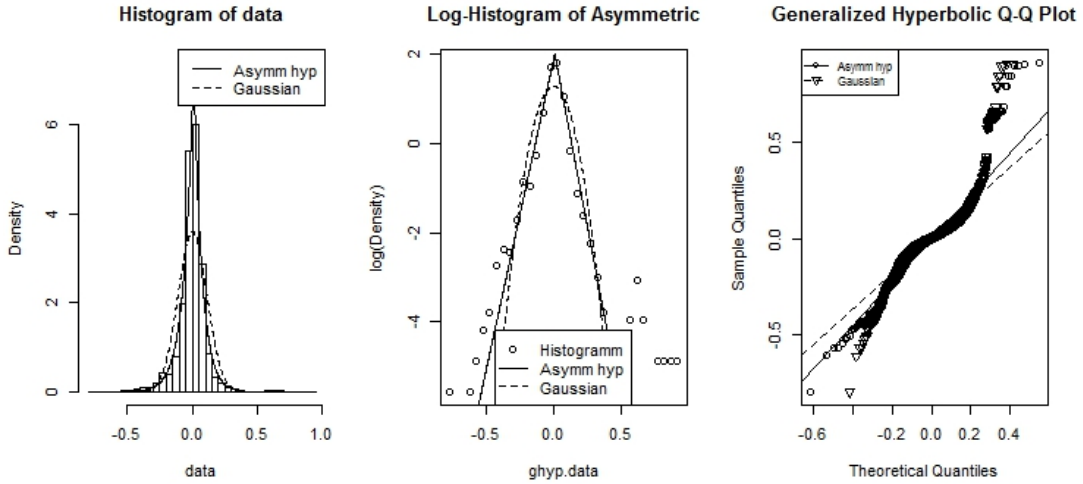

Figure A.3: Platinum Returns under Asymmetric Hyperbolic Distribution

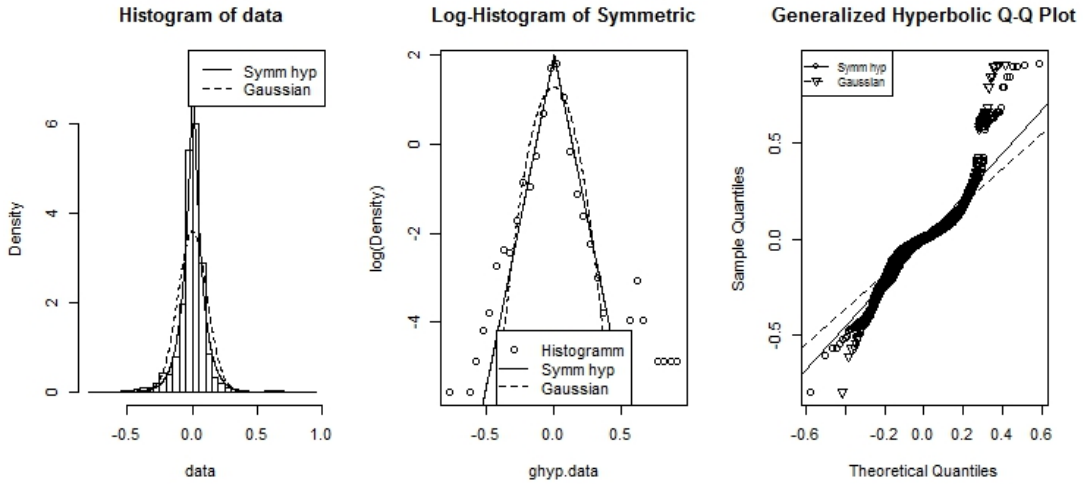

Figure A.4: Platinum Returns under Symmetric Hyperbolic Distribution

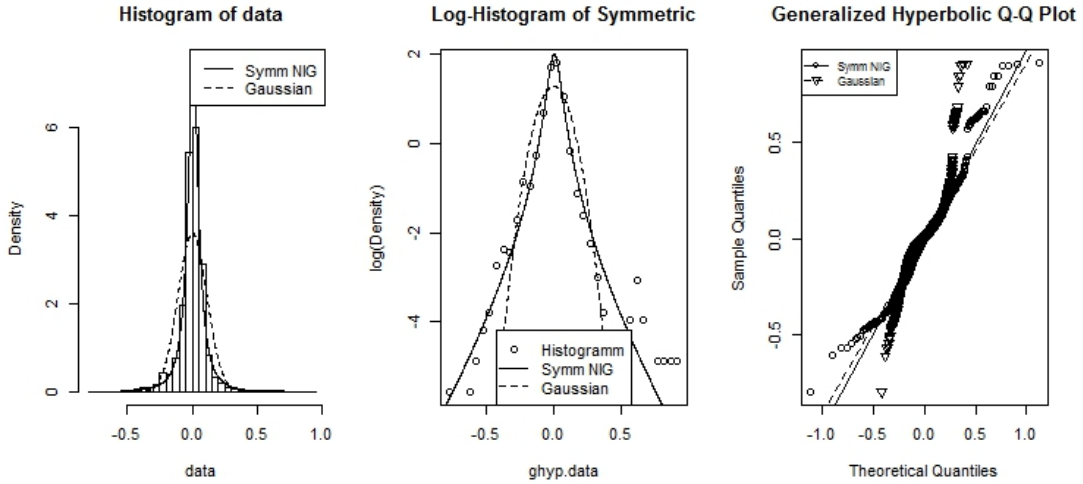

Figure A.5: Platinum Returns under Symmetric Normal Inverse Gaussian Distribution

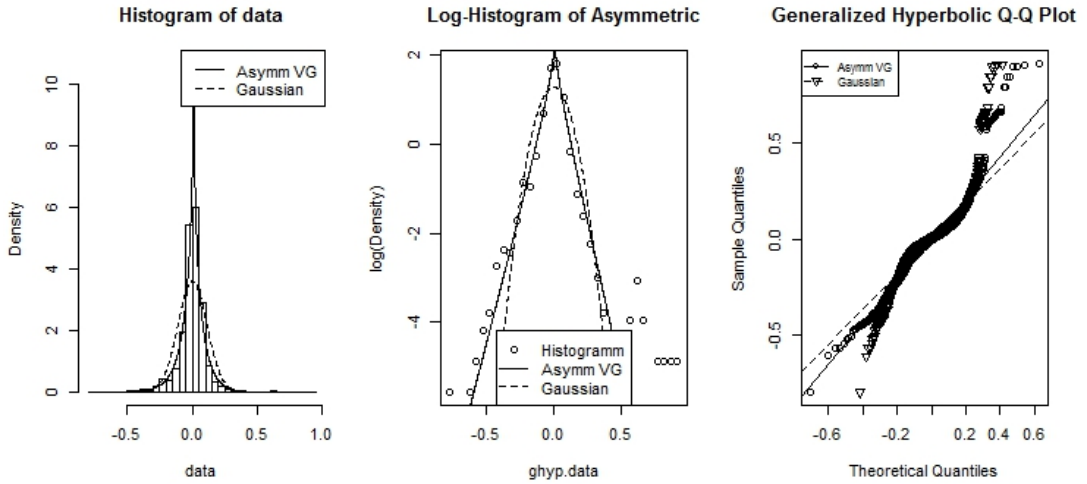

Figure A.6: Platinum Returns under Asymmetric Variance-Gamma Distribution

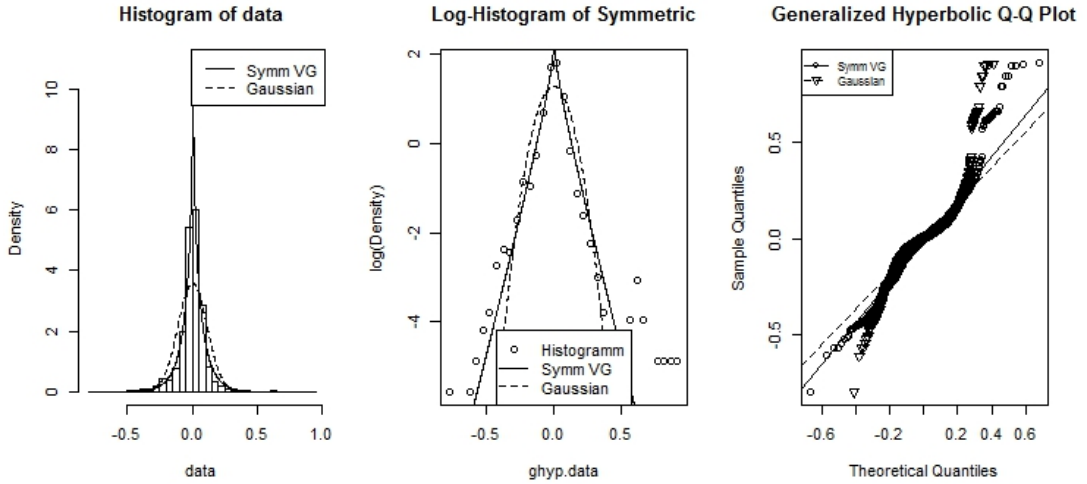

Figure A.7: Platinum Returns under Symmetric Variance-Gamma Distribution

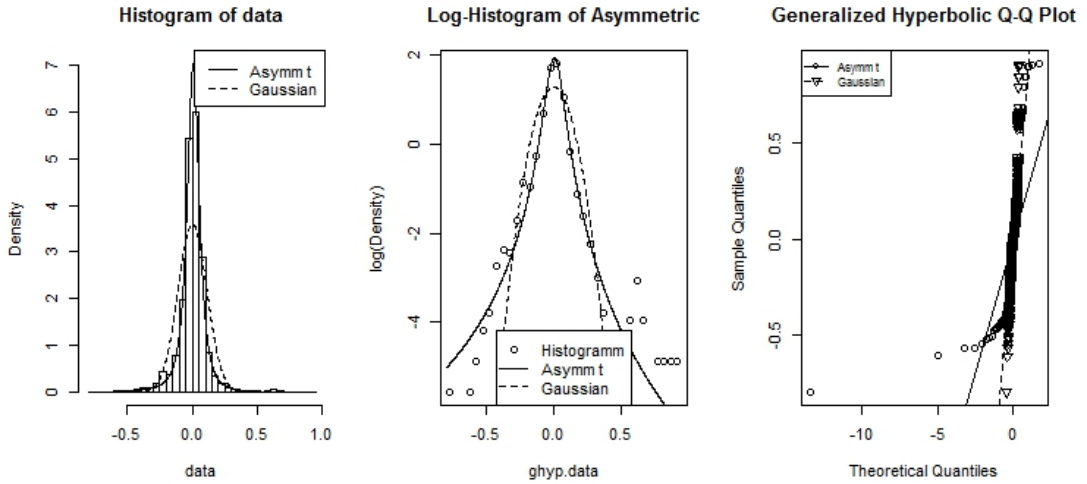

Figure A.8: Platinum Returns under Asymmetric skewed Student-t Distribution

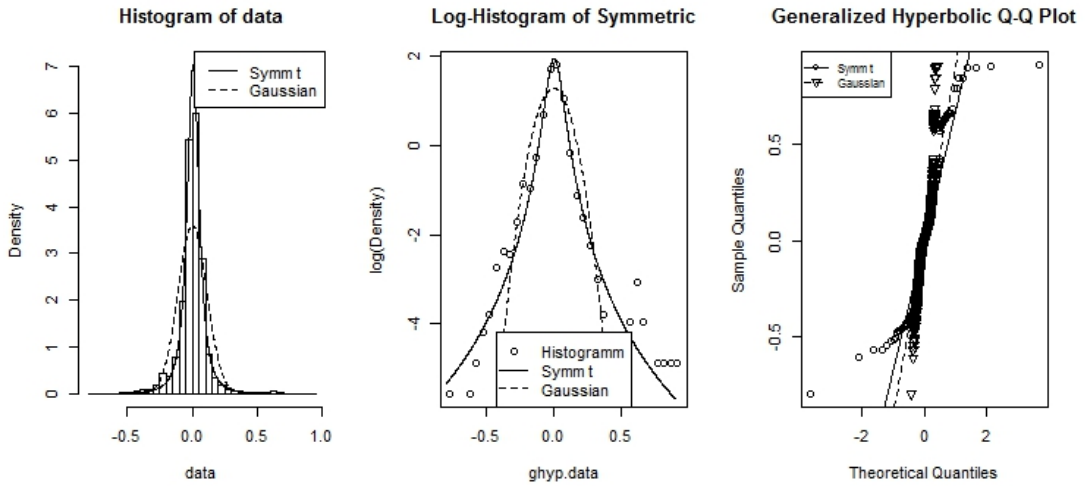

Figure A.9: Platinum Returns under Symmetric skewed Student-t Distribution

Appendix B

Programs and Algorithms

B.1 Breaks SAS Program

```
data platinum(keep=Date pt lag_pt rt rename=(Date=date_));
 set work.pt_pd;
  lag_pt=lag(pt);
  rt=log(pt/lag_pt);
 run;

data platinum;
set platinum;
where not missing(rt);
run;

proc sort data=platinum ;
by date_ rt;
run;

data platinum;
set platinum;
  week=week(date_);
  month=month(date_);
  year=year(date_);
```

```
run;

proc sql;
 create table platinum as
  select
         *,
 mean(rt) as mean_return
  from platinum
group by year, month;
quit;

proc sql;
 create table platinum as
  select distinct date_, year, month, mean_return
  from platinum
;
quit;

proc sort data=platinum nodup;
by date_;
run;


data platinum2;
set platinum;
 by year month;
 if first.month then output;
run;



proc ucm data=platinum2 ;
```

```
id date_ interval=month;
model mean_return;
irregular;
level plot=smooth checkbreak;
estimate;
forecast plot=decomp;
run;
```

B.2 Inequality Constraints R Program

```
#Inequality constrants of FIGARCH(1,d,1)

getConstraints<-function(phi1,beta1,d){
  #Baillie
  b1=phi1+d
  b2=1-2*phi1
  cat("Baillie et all [1996]: 0<= ",beta1," <= ",b1," and  0<= ",d," <= ",b2)

  #Bollerslev
  c0=beta1-d
  c1=(2-d)/3
  c2=d*(phi1-(1-d)/2)
  c3=beta1*(phi1-beta1+d)
  cat("\n Bollerslev and Ole Mikkelsen [1996]: ",c0," <= ",c1," and ",c2," <= ",c3)

  #Chung
  cat("\n Chung [1999]: 0<= ",phi1," <= ",beta1," <= ",d," < ",1)

}

f<-function(j,d){
      return((j-1-d)/j)
```

```
}

g<-function(j,d){

  sum=1
  if(i<=0){return(1)}
 else{
   for(i in 1:j){
     sum=sum*f(i,d)
    }
   return(-sum)
 }
}

psi<-function(phi1,beta1,d,i){
  if(i<=1){
    return(d+phi1-beta1)
  }
  else {
    return(beta1*psi(phi1,beta1,d,i-1)+(f(i,d)-phi1)*-g(i-1,d))
  }
}
```

B.3 ARFIMA-FIGARCH Type Models OxMetrics Program

```
#include <oxstd.h>
#include <oxdraw.h>
#import <packages/garch6/garch>
```

```
main()
{
decl garchobj;

garchobj = new Garch();

//*** DATA ***//
garchobj.Load("C:/Users/Sihle/Documents/oxProjects/examples/data/PGM Data.xlsx");
garchobj.Info();

    garchobj.Select(Y_VAR, {"rtr1",0,0});
// garchobj.Select(Z_VAR, {"rt2",0,0}); // REGRESSOR IN THE VARIANCE

garchobj.SetSelSample(-1, 1,3000, 1);

//*** SPECIFICATIONS ***//
garchobj.CSTS(0,0);  // cst in Mean (1 or 0), cst in Variance (1 or 0)
garchobj.DISTRI(1); // 0 for Gauss, 1 for Student, 2 for GED, 3 for Skewed-Student
garchobj.ARMA_ORDERS(1,1);  // AR order (p), MA order (q).
garchobj.ARFIMA(1); // 1 if Arfima wanted, 0 otherwise
garchobj.GARCH_ORDERS(1,1); // p order, q order
garchobj.ARCH_IN_MEAN(0); // ARCH-in-mean: 1 or 2 to add the variance or std. dev in
garchobj.MODEL(10);    // 0: RISKMETRICS  1:GARCH 2:EGARCH 3:GJR 4:APARCH 5:IGARCH
//  6:FIGARCH_BBM 7:FIGARCH_CHUNG   8:FIEGARCH
//  9:FIAPARCH_BBM 10: FIAPARCH_CHUNG 11: HYGARCH
garchobj.TRUNC(2000); // Truncation order (only F.I. models with BBM method)



//*** TESTS & FORECASTS ***//
garchobj.NORMALITY_TEST(1); // 1 to report the normality test, 0 otherwise
garchobj.INFO_CRITERIA(1); // 1 to report Information Criteria, 0 otherwise
```

```
garchobj.BOXPIERCE(<10;15;20>); // Lags for the Box-Pierce Q-statistics, <> otherwise
garchobj.ARCHLAGS(<2;5;10>); // Lags for Engle's LM ARCH test, <> otherwise
garchobj.NYBLOM(1); // 1 to compute the Nyblom stability test, 0 otherwise
garchobj.SBT(1); // 1 to compute the Sign Bias test, 0 otherwise
garchobj.PEARSON(<40;50;60>); // Cells (<40;50;60>) for the adjusted Pearson Chi-squa
garchobj.RBD(<10;15;20>); // Lags for the Residual-Based Diagnostic test of Tse, <> o
garchobj.FORECAST(1,15,1);      // Arg.1 : 1 to launch the forecasting procedure, 0 o
// Arg.2 : Number of forecasts
// Arg.3 : 1 to Print the forecasts, 0 otherwise

//*** OUTPUT ***//
garchobj.MLE(2); // 0 : MLE (Second derivatives), 1 : MLE (OPG Matrix), 2 : QMLE
garchobj.COVAR(0);     // if 1, prints variance-covariance matrix of the parameters.
garchobj.ITER(0); // Interval of iterations between printed intermediary results (if
garchobj.TESTS(0,1); // Arg. 1 : 1 to run tests PRIOR to estimation, 0 otherwise
// Arg. 2 : 1 to run tests AFTER estimation, 0 otherwise
garchobj.GRAPHS(0,1,"file2"); // Arg.1 : if 1, displays graphics of the estimations (
// Arg.2 : if 1, saves these graphics in a EPS file (OK with all Ox versions)
// Arg.3 : Name of the saved file.
garchobj.FOREGRAPHS(0,1,"forecast");   // Same as GRAPHS(p,s,n) but for the graphics

//*** PARAMETERS ***//
garchobj.BOUNDS(0); // 1 if bounded parameters wanted, 0 otherwise
garchobj.FIXPARAM(0,<0;0;0;0;1;0>);
// Arg.1 : 1 to fix some parameters to their starting values, 0 otherwize
// Arg.2 : 1 to fix (see garchobj.DoEstimation(<>)) and 0 to estimate the correspondi

//*** ESTIMATION ***//
garchobj.MAXSA(0,5,0.5,20,5,2,1);
// Arg.1 : 1 to use the MaxSA algorithm of Goffe, Ferrier and Rogers (1994)
//         and implemented in Ox by Charles Bos
// Arg.2 : dT=initial temperature
```

```
// Arg.3 : dRt=temperature reduction factor
// Arg.4 : iNS=number of cycles
// Arg.5 : iNT=Number of iterations before temperature reduction
// Arg.6 : vC=step length adjustment
// Arg.7 : vM=step length vector used in initial step

garchobj.Initialization(<>);
// m_vPar = m_clevel | m_vbetam |  m_dARFI | m_vAR | m_vMA |
 m_calpha0 | m_vgammav | m_dD |  m_vbetav |
// m_valphav | m_vleverage | m_vtheta1 | m_vtheta2 |
 m_vpsy | m_ddelta | m_cA | m_cV | m_vHY | m_v_in_mean

garchobj.PrintStartValues(0);   // 1: Prints the S.V. in a table form; 2: Individuall
garchobj.PrintBounds(1);
garchobj.DoEstimation(<>);
garchobj.Output();
garchobj.STORE(1,1,1,1,1,"01",0);
//  Arg.1,2,3,4,5 : if 1 -> stored. (Res-SqRes-CondV-MeanFor-VarFor)
    //  Arg.6 : Suffix. The name of the saved series will be "Res_ARG6" (or "MeanFor_
//  Arg.7 : if 0, saves as an Excel spreadsheet (.xls). If 1, saves as a GiveWin data
println("********* STARTING THE FORECASTING PROCEDURE***************");
decl forc=<>,h,yfor=<>,shape=<>;
decl number_of_forecasts=1000; // number of h_step_ahead forecasts
decl step=1;   // specify h (h-step-ahead forecasts)
decl T=garchobj.GetcT();;
println("Size-T = ",T);
decl distri=garchobj.GetDistri();  println("Distribution = ",distri);
if (distri==1 ||distri==2)
shape=garchobj.GetValue("m_cV");
else if (distri==3)
shape=garchobj.GetValue("m_cA")|garchobj.GetValue("m_cV");
println("shape = ",shape);
```

```
println("Size = ",garchobj.GetSize());
    if (T+number_of_forecasts+step-1>garchobj.GetSize())
{
number_of_forecasts+=garchobj.GetSize()-T-number_of_forecasts-step+1;
println("\nWarning: the number of forecast is incompatible with the sample size");
println("number_of_forecasts has been set to ",number_of_forecasts,"\n");
}

for (h=0; h<number_of_forecasts; ++h)
{
garchobj.FORECAST(1,step,0);
garchobj.SetSelSample(-1, 1, T+h, 1);
garchobj.InitData();
yfor|=garchobj.GetForcData(Y_VAR, step)[step-1];
forc|=garchobj.Forecasting()[step-1][];
}

decl Hfor = (yfor - meanc(yfor)).^2; //println("Hfor = ",Hfor);
decl cd=garchobj.CD(yfor-forc[][0],forc[][1],garchobj.GetDistri(),shape);
println("Density Forecast Test on Standardized Forecast Errors");
garchobj.PearsonTest(cd,20|30,garchobj.GetValue("m_cPar"));
garchobj.Auto(cd, number_of_forecasts, -0.1, 0.1, 0);
garchobj.confidence_limits_uniform(cd,30,0.95,1,4);
DrawTitle(5, "Conditional variance forecast and realized volatility");
  Draw(5, (Hfor~forc[][1])');
ShowDrawWindow();
garchobj.MZ(Hfor, forc[][1], number_of_forecasts);
garchobj.FEM(forc, yfor~Hfor);
savemat("C:/Users/Sihle/Documents/oxProjects/data/hygarch
/MeanFor_bbm1.txt",forc[][0];    // Saves the mean forecasts in an Excel file.
savemat("C:/Users/Sihle/Documents/oxProjects/data/hygarch
/VarFor_bbm1.txt",forc[][1];    // Saves the variance forecasts in an Excel file.
```

```
savemat("C:/Users/Sihle/Documents/oxProjects/data/hygarch
//HFor_bbm1.txt",Hfor);    // Saves the variance forecasts in an Excel file.
savemat("C:/Users/Sihle/Documents/oxProjects/data/hygarch
//cd.txt",cd);      // saves the distribution
savemat("C:/Users/Sihle/Documents/oxProjects/data/hygarch
//yfor.txt",yfor);      // save forcasting data

decl d=garchobj.m_dD;
decl phi1=garchobj.m_valphav;
decl beta1=garchobj.m_vbetav;



delete garchobj;
}
```

B.4 Shimotsu R code

```
library(forecast)
library(fracdiff)
library(tseries)
library(MASS)
library(afmtools)
library(fArma)

vj <-function(j,m){
  sum=log(j)
  sum1=0
  for(i in 1:m){
    sum1=sum1+log(i)
  }
```

```
  sum=sum-(1/m)*sum1
  return(sum^2)
}

crr<-function(m,b){
  sum=0;
  for(i in 1:m){
    sum=sum+vj(i,m)
  }
  sum=sum/(m/b)

  return(sum)
}

wald2<-function(d_0,d1,d2,m){
  d=matrix(cbind(c(d_0,d1,d2)))
  A=matrix(cbind(c(1,-1,0),c(1,0,-1)),2,byrow=TRUE)
  Ad=A%*%d
  O=matrix(cbind(c(1,1,1),c(1,2,0),c(1,0,2)),3,byrow=TRUE)
  G=ginv(A%*%O%*%t(A))
  w=t(Ad)%*%G%*%Ad
  #c=c(m,2)
  wald=4*m*w#*c
  return(wald)
}



wald<-function(theta,A,O,r,m){
  Ad=A%*%theta
  G=ginv(A%*%O%*%t(A))
  w=t(Ad)%*%G%*%Ad
```

```
  wald=4*m*w
  return(wald)
}




shims<-function(n,X,Y){
  data=X[1:n]
  b2=as.integer(n/2)
  b4=as.integer(n/4)
  b41=2*b4
  b42=3*b4
  b43=4*b4
  x11=X[1:b2]
  x12=X[b2:n]
  x21=X[1:b4]
  x22=X[b4:b41]
  x23=X[b41:b42]
  x24=X[b42:b43]
  d11=as.numeric(coef(WhittleEst(x11)))-0.5
  d12=as.numeric(coef(WhittleEst(x12)))-0.5
  d21=as.numeric(coef(WhittleEst(x21)))-0.5
  d22=as.numeric(coef(WhittleEst(x22)))-0.5
  d23=as.numeric(coef(WhittleEst(x23)))-0.5
  d24=as.numeric(coef(WhittleEst(x24)))-0.5
  d=as.numeric(coef(WhittleEst(X[1:n])))-0.5
  #d11=as.numeric(rsFit(x11)@hurst$H)-0.5
  #d12=as.numeric(rsFit(x12)@hurst$H)-0.5
  #d21=as.numeric(rsFit(x21)@hurst$H)-0.5
  #d22=as.numeric(rsFit(x22)@hurst$H)-0.5
  #d23=as.numeric(rsFit(x23)@hurst$H)-0.5
```

```
#d24=as.numeric(rsFit(x24)@hurst$H)-0.5
#d=as.numeric(rsFit(X[1:n])@hurst$H)-0.5
#Shimotsu Methods

#for b=4
theta=matrix(cbind(c(d,d21,d22,d23,d24)))
A=matrix(cbind(c(1,-1,0,0,0),
               c(1,0,-1,0,0),
               c(1,0,0,-1,0),
               c(1,0,0,0,-1)
),4,byrow=TRUE)
O=matrix(cbind(c(1,1,1,1,1),
               c(1,4,0,0,0),
               c(1,0,4,0,0),
               c(1,0,0,4,0),
               c(1,0,0,0,4)
),5,byrow=TRUE)

#for b=2
theta1=matrix(cbind(c(d,d11,d12)))
A1=matrix(cbind(c(1,-1,0),
                c(1,0,-1)
),2,byrow=TRUE)
O1=matrix(cbind(c(1,1,1),
                c(1,2,0),
                c(1,0,2)
),3,byrow=TRUE)
d1=mean(d11,d12)
d2=mean(d21,d22,d23,d24)
```

```
  W2=wald(theta1,A1,O1,1,n)
  CR=crr(m,2)
  WC2=W2*CR

  W4=wald(theta,A,O,1,n)
  CR=crr(m,4)
  WC4=W4*CR

  dmrt=Y[1:n]-mean(Y[1:n])
  fit.21=arfima(dmrt,max.p=0,max.q=0)
  dfit= coef(fit.21)
  y.diff=diffseries(dmrt,dfit)
  kpy=kpss.test(y.diff)

  dmrt=X[1:n]-mean(X[1:n])
  fit.21=arfima(dmrt,max.p=0,max.q=0)
  dfit= coef(fit.21)
  y.diff=diffseries(dmrt,dfit)
  kpx=kpss.test(y.diff)

  cat(n,"&",signif(d,digits=4),"&",signif(d1,digits=4),"&",signif(d2,digits=4),
  "&",signif(W2,digits=4),"&",signif(W4,digits=4),"&",signif(kpy$statistic,digits=4)
  ,"&", signif(kpy$p.value,digits=4),"\\","\\")
}

#X and Y is the same series
shims(550,X,Y)
#shims(1000,X,Y)
#shims(1500,X,Y)
#shims(2000,X,Y)
#shims(2500,X,Y)
#shims(3000,X,Y)
```

```
#shims(3500,X,Y)
#shims(4000,X,Y)
#shims(4500,X,Y)
#shims(5000,X,Y)
```

B.5 Cumulative Sampling R Code

```
library(fracdiff)
library(longmemo)
N=length(X)
t=1#Running Sample
b=50
n=as.integer(N/b)
 rdata=0
perdata=0
  i=t
  k=1
  for(i in 1:b*n){
    d_sample=X[1:i]
    perdata[k]=arfima.whittle(d_sample,nar=0,nma=0)$d#as.numeric(coef(WhittleEst(d_sa
    cat("regime[",k,"]=",perdata[k],"\n")
    k=k+1
    t=i-n
    #print(i)
    #print(t)
  }


plot(perdata[0:b], type='o',col='steelblue',main='Sample Accumulation by regime split
```

```
     xlab="Sub-Samples", ylab="d estimates",xlim=c(0,b))
grid()
```