



**UNIVERSITY OF  
KWAZULU-NATAL**

---

**INYUVESI  
YAKWAZULU-NATALI**

**Modelling South African Official Gold Reserves Position, and Foreign  
Exchange Reserves Position Using  
Time Series Models**

By

Sibusiso Gumede

Student Number: 212535942

A Thesis Submitted in Fulfilment of the Academic Requirements for the

Degree of

Master of Science

in

Statistics

Under the Supervision of

Supervisor: Dr Oliver Bodhlyera

Co-Supervisor: Prof. Henry Mwambi

School of Mathematics, Statistics, and Computer Science

University of KwaZulu-Natal

Pietermaritzburg Campus

South Africa

2020

## Declaration

The research work described in this thesis was carried out under the supervision of Dr Oliver Bodhlyera and the Co-Supervisor Prof. Henry Mwambi, School of Mathematics, Statistics and Computer Science at University of KwaZulu-Natal, Pietermaritzburg campus. The work presents the original work of the author and has not otherwise been submitted in any form for any degree or diploma to any University. Where use has been made of the work of others, the author is duly acknowledged in the text.



Mr. Sibusiso Gumede.

04/02/2020

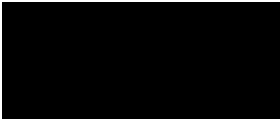
Date:



Dr Oliver Bodhlyera.

10/02/2020

Date:



Prof. Henry Mwambi.

21/02/2020

Date:

## Abstract

Every central bank of the country should hold enough reserves such as foreign exchange currency, gold, or any form of reserves to be able to help its country in times of difficulties or financial crises. This involves the process of ensuring that adequate official public sector foreign assets are readily available to meet any defined range of objectives by a country. Reserves can also play a pivotal role in supporting and maintaining confidence in the policies for monetary and exchange rate management, including the ability to intervene in the foreign market to influence the value of the local currency. It can also be used to provide proof to the market that a country can meet its current and future external obligations, limit external exposure by maintaining foreign currency liquidity to absorb shock during times of crisis, show the support of domestic currency by external assets, assist the government in meeting its foreign exchange needs and external debt obligations, and maintain sufficient reserves for national disasters or emergencies.

All this cannot be done without the understanding of all factors that affect reserves of the country, hence careful analysis of reserves in a country plays a crucial role on how the central bank should manage the reserves of such a country. This includes a wide range of social, economic, and statistical analyses. However, this study focuses more on the statistical analysis part, which is, building models to predict or forecast the trajectory of reserves positions in future. These models should be able to consider all the factors that influence the reserves, such as trend, seasonality and the variability (random variability). The Seasonal ARIMA models were used as initial models to forecast the future reserves positions. Seasonal ARIMA Generalized Autoregressive Conditional Heteroskedasticity models with Skewed Student- $t$  Distribution (SARIMA – GARCH – SSTD) were also used to forecast volatility from the foreign exchange reserves data after statistical test were carried out and the data was found to have ARCH Effects. The best volatility model that was found to produces best forecast for foreign exchange reserves data was the SARIMA (0,1,0) (2,1,0)<sub>12</sub> – GARCH (1,1) – SSTD model.

The SARIMA model developed earlier for gold reserves data was then benchmarked with the Holt-Winters' Seasonal method. The results from the analysis showed that SARIMA model outperformed Holt-Winters' Seasonal method in forecasting gold reserves positions. We found that future gold reserves positions can be better predicted using the SARIMA (1,1,0) (0,1,2)<sub>12</sub> model. The best model was selected from many other models using model diagnostics process such as comparisons of the AIC, RMSE, number of significant parameters and the evaluation of residuals to identify their flexibility. Using the forecasting methods developed in this study, the central bank can better understand what to expect in the future and decide on what measures to implement for national economic stability.

**Keywords:** *Reserves, S/ARIMA, G/ARCH, Volatility, forecasting, Seasonality, Damping Parameter.*

## Acknowledgements

I would first like to thank my thesis supervisor Dr Oliver Bodhlyera of the School of Mathematics, Statistics, and Computer Science at the University of KwaZulu-Natal. The door to his office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my work and steered me in the right direction whenever he thought I needed it.

I would also like to thank a friend of mine, Tshiololi Maduvha, for helping me with Latex; she helped me with the fundamental basics on how to write a professional thesis using Latex. I appreciate her inputs and endless efforts to ensure that this thesis was successfully completed. Without her passionate participation and inputs, this thesis would not have been successfully conducted.

I would also like to acknowledge all sponsorships who funded this thesis throughout the study period, namely University of KwaZulu-Natal Cum Laude Scholarship, NRF Block Grants Scholarship, and South African Statistical Association (SASA) Scholarship; indeed, without their financial assistance I could not have completed this thesis.

Finally, I must express my very profound gratitude to God the Almighty, who has given me the strength to carry on with this thesis through all the ups and downs I have faced. I would also like to dedicate a special thanks to my relatives for their support and encouragement throughout the period of this study. This completion would not have been possible without them.

## **Contents**

<b>Declaration.....</b>	<b>2</b>
<b>Abstract.....</b>	<b>3</b>
<b>Acknowledgements.....</b>	<b>4</b>
<b>List of Tables .....</b>	<b>9</b>
<b>List of Figures.....</b>	<b>11</b>
<b>Chapter 1: .....</b>	<b>12</b>
<b>1. Introduction.....</b>	<b>12</b>
<b>1.1 Gold and Foreign Reserves and their Importance.....</b>	<b>12</b>
<b>1.2 Literature Review.....</b>	<b>13</b>
<b>1.2.1 Reserves Economic Aspects.....</b>	<b>15</b>
<b>1.3 Volatility.....</b>	<b>19</b>
<b>1.4 Problem Statement.....</b>	<b>19</b>
<b>1.5 Aim and Objectives .....</b>	<b>19</b>
<b>1.6 Significance of the Study .....</b>	<b>20</b>
<b>1.7 Thesis Structure.....</b>	<b>20</b>
<b>Chapter 2: .....</b>	<b>21</b>
<b>2. Theory of Exploratory Data Analysis in Time Series.....</b>	<b>21</b>
<b>2.1 Introduction .....</b>	<b>21</b>
<b>2.2 Exploratory Data Analysis and Data Transformations Theory .....</b>	<b>21</b>
<b>2.2.1 Stationary and Non-Stationary Series.....</b>	<b>21</b>
<b>2.2.2 Variance Transformation.....</b>	<b>22</b>
<b>2.2.3 Moving Average .....</b>	<b>22</b>
<b>2.2.4 Data Decomposition .....</b>	<b>23</b>
<b>2.2.5 Components of Time Series.....</b>	<b>24</b>
<b>Chapter 3.....</b>	<b>26</b>

<b>3. Theory of Homoscedastic Models: S/ARIMA Models .....</b>	<b>26</b>
3.1 Introduction .....	26
3.2 Autoregressive Models.....	26
3.3 Moving Average Models .....	27
3.4 Building an ARMA Model.....	27
3.4.1Box-Jenkins Methodology .....	29
3.5 ARMA and/or ARIMA Models Limitations.....	34
<b>Chapter 4.....</b>	<b>36</b>
<b>4. Volatility Models .....</b>	<b>36</b>
4.1 Introduction .....	36
4.2 An ARCH(s) Model Overview .....	36
4.2.1 Building an ARCH(s) Model.....	37
4.2.2 An ARCH (1) Model Overview.....	41
4.2.3 ARCH Models Limitations.....	45
4.3 Exponentially Weighted Moving Average (EWMA) .....	46
4.4 The Generalized ARCH (GARCH) Model .....	47
4.4.1 A GARCH ( $r, s$ ) Model .....	47
4.4.2 A GARCH (1,1) Model .....	52
4.4.3 GARCH Models Limitations.....	59
4.5 Extensions of the GARCH Models .....	59
4.5.2 The Integrated GARCH Model .....	60
4.5.3 The Threshold GARCH Model.....	60
4.5.4 The Asymmetric Power ARCH.....	61
4.6 Testing for an ARCH Effect.....	61
4.7 An ARCH Model Selection Criteria .....	62
4.8 Chapter Summary .....	63
<b>Chapter 5.....</b>	<b>64</b>

<b>5. Application of SARIMA Models to Reserves Time Series Data .....</b>	<b>64</b>
<b>5.1 Data Description.....</b>	<b>64</b>
<b>5.2 Gold Reserves Data Analysis.....</b>	<b>65</b>
5.2.1 Data Preparation.....	66
5.2.2 Model Identification.....	70
5.2.3 Model Diagnostics and Parameter Coefficients Estimation .....	70
5.2.4 Forecasting Gold Reserves Positions Using SARIMA Models.....	73
<b>5.3 Foreign Exchange Reserves Data Analysis .....</b>	<b>74</b>
5.3.1 Data Preparation.....	75
5.3.2 Model Identification.....	79
5.3.3 Model Diagnostics and Parameter Coefficients Estimation .....	79
5.3.4 Forecasting Foreign Exchange Reserves Positions using SARIMA models .....	82
<b>5.4 Chapter Summary .....</b>	<b>82</b>
<b>Chapter 6.....</b>	<b>85</b>
<b>6. Volatility Models Application to Foreign Exchange Data .....</b>	<b>85</b>
6.1 Introduction .....	85
6.2 Model Selection Criteria.....	85
6.3 Analysis of Foreign Exchange Reserves .....	85
6.4 Chapter Summary.....	93
<b>Chapter 7.....</b>	<b>94</b>
<b>7. Holt-Winters' Method in Time Series Analysis .....</b>	<b>94</b>
7.1 Introduction .....	94
7.2 Holt-Winters' Linear Trend Method .....	94
7.3 Holt-Winters' Damped Trend Method .....	95
7.4 Holt-Winters' Seasonal Method.....	95
7.5 Holt-Winters' Damped Seasonal Method .....	97

7.6 Chapter Summary.....	97
<b>Chapter 8.....</b>	<b>98</b>
<b>8. Application of Holt-Winters Method in Time Series Data Analysis.....</b>	<b>98</b>
8.1 Introduction .....	98
8.2 Gold Reserves Data Analysis.....	98
8.3 Chapter Summary .....	103
<b>Chapter 9.....</b>	<b>104</b>
<b>9. Conclusion and Recommendations .....</b>	<b>104</b>
<b>10. References .....</b>	<b>106</b>
<b>11. Appendices .....</b>	<b>111</b>
11.1 Rstudio Code Used:.....	113



## List of Tables

Table 5.1: Descriptive statistics for gold reserves positions. ....	66
Table 5.2: Results of the ADF and KPSS tests for the stationarity of the original series of gold reserves positions. ....	67
Table 5.3: Results of the ADF and KPSS tests for the stationarity of the non- and seasonally differenced log gold reserves positions. ....	70
Table 5.4: Goodness of fit statistics of the different models fitted to the series of gold reserves positions. ....	71
Table 5.5: Ljung-Box test results for residuals from selected fitted SARIMA (1,1,0) (0,1,2) <sub>12</sub> model. ....	72
Table 5.6: Parameter estimates of the selected SARIMA (1,1,0) (0,1,2) <sub>12</sub> model for the gold reserves series. ....	73
Table 5.7: Descriptive statistics for foreign exchange reserves positions.....	75
Table 5.8: Results of the ADF and KPSS tests for the stationarity of the foreign exchange reserves positions. ....	76
Table 5.9: Results of the ADF and KPSS tests for the stationarity of the non- and seasonally differenced log foreign exchange reserves positions. ....	79
Table 5.10: Goodness of fit statistics of the different models fitted to the series of foreign exchange reserves positions. ....	80
Table 5.11: Ljung-Box test results for residuals from selected fitted SARIMA (0,1,0) (2,1,0) <sub>12</sub> model. ....	81
Table 5.12: Parameter estimates of the selected SARIMA (0,1,0) (2,1,0) <sub>12</sub> model for the foreign exchange reserves series. ....	81
Table 5.13: The gold and foreign exchange reserves serial correlation test. ....	83
Table 5.14: The gold and foreign exchange reserves ARCH Effect (heteroscedasticity) test from the SARIMA model residuals.....	83
Table 6.1: GARCH – type models fitted with normal distribution.....	86
Table 6.2 : MLE parameter estimates for fitted GARCH (1,1) model, conditional on the normal distribution. ....	86
Table 6.3: Normality test statistics with p – values for standardized residuals tests from GARCH (1,1) model under the standard normal distribution. ....	87
Table 6.4: GARCH (1,1) models fitted with student – t (std), skewed student – t (sstd) and generalized error (ged) distribution.....	89
Table 6.5: MLE parameter estimates of GARCH (1,1) model fitted with skewed student – t distribution (sstd).....	90

Table 6.6: Standardized residuals tests for GARCH (1,1) model under skewed student- $t$ distribution.	91
Table 8.1: HW additive seasonal smoothing components (Gold Reserves Data).....	100
Table 8.2: HW multiplicative seasonal smoothing components (Gold Reserves Data). ....	101
Table 8.3: HW damped multiplicative seasonal smoothing components (Gold Reserves Data).....	103
Table 9.1: Observed versus forecasts for gold reserves positions. ....	104

## List of Figures

Figure 5.1 : Gold and foreign exchange reserves positions time series plots. ....	65
Figure 5.2: Gold reserves time series, ACF and PACF plots.....	67
Figure 5.3: The time series, ACF and PACF plots of detrended ( $d=1$ ) and period 12 de-seasonalized ( $D=1$ ) log transformed gold reserves positions. ....	69
Figure 5.4: Residuals from SARIMA (1,1,0) (0,1,2) <sub>12</sub> model. ....	72
Figure 5.5: Gold reserves positions forecasts from SARIMA (1,1,0) (0,1,2) <sub>12</sub> model. ....	74
Figure 5.6: Foreign exchange reserves time series, ACF and PACF plots. ....	76
Figure 5.7: The time series, ACF and PACF plots of detrended ( $d=1$ ) and period 12 de-seasonalized ( $D=1$ ) log transformed foreign exchange reserves positions. ....	78
Figure 5.8: Residuals from SARIMA (0,1,0) (2,1,0) <sub>12</sub> model. ....	81
Figure 5.9: Forecasts from SARIMA (0,1,0) (2,1,0) <sub>12</sub> model. ....	82
Figure 6.1: Standardized and squared standardized residuals ACF for GARCH (1,1) model fitted with standard normal distribution. ....	87
Figure 6.2: Standardized residuals frequency distribution vs. standard normal distribution density. ..	88
Figure 6.3: QQ-Plots for GARCH (1,1) model fitted with different distributions governing the innovations. ....	90
Figure 6.4: Standardized residuals ACF for GARCH (1,1) model with skewed student- $t$ distribution. ....	92
Figure 6.5: The forecasts from GARCH (1,1) model fitted with skewed student- $t$ distribution governing the innovations. ....	93
Figure 8.1: Forecasting official gold reserves positions in South Africa using the Holt-Winters' method with both additive and multiplicative seasonality.....	99
Figure 8.2: Forecasts of monthly gold reserves positions from Holt-Winters' Damped method with multiplicative seasonality.....	102

# Chapter 1:

## 1. Introduction

The South African Reserve Bank (SARB) is the central bank of the Republic of South Africa (RSA). It plays a pivotal role in the South African economic system in the sense that it maintains price stability. The SARB is also involved in a supervisory of all domestic commercial banks and stock exchanges. It helps local banks that are temporarily illiquid and holds and manages reserves of the country (e.g., official gold and foreign exchange reserves). This study focuses on the management of South African official gold and foreign exchange reserves.

### 1.1 Gold and Foreign Reserves and their Importance

The reserves of the country are similar to the working capital of a company or the current balance account of an individual (Weisweiller, 1972). In most countries, the reserves are made of the foreign currency and official gold detained by the central bank. However, the reserves detained by such central bank do not indicate the total assets owned by its country. Some well-developed countries with strong economic growth in the world have small reserves, whereas some with very large reserves have also very few other assets abroad. Those assets might be in the form of stock exchange investment held by individuals, or subsidiaries owned by domestic companies abroad (Weisweiller, 1972).

In practice, “there is no unique definition of foreign exchange assets or reserves, however, for policies and operational purposes many countries use the definition suggested by International Monetary Fund which states that “ foreign exchange reserves are the external assets that are readily available and controlled by monetary policy for directly financing external payments imbalances, and for directly regulating the magnitudes of such imbalances through intervention in the exchange market to affect the currency exchange rate and/or other commitments” (Reddy, 2002). South African Reserve Bank also defines foreign exchange reserves in their document titled “Management of gold and foreign exchange reserves, 2010” as “the currency or any other financial instruments that allow one country to settle amounts owed to other foreign countries”. For instance, the amount owed by South Africa to the United Kingdom is settled in foreign exchange (i.e., in Pound/GBP). Furthermore, foreign exchange reserves are instruments used to control and manage the exchange rate, while allowing organized absorption of international money and capital flows (Reddy, 2002) . The deal made between two countries or two individuals from two different countries to purchase or sell a product from each other is known as a foreign exchange deal. A foreign exchange deal occurs in the foreign exchange market and is simply an exchanging of one currency or national money for another. The exchange rate is the price of one currency expressed in terms of the other. It simply expresses the correlation between two currencies;, whenever an exchange rate moves, this is due to a change in the value of one or the other currency (Weisweiller, 1972). Therefore, the rate of a currency is then determined by the supply of and/or demand

of that currency. For instance, the rand can be described as depreciated against the US dollar if each US dollar cost more rands to buy. If it has appreciated, then fewer rands are paid for a single US dollar.

Reserves can be considered as a self-insurance during financial crises in the country (Steiner, 2013). The International Monetary Fund (IMF, 2014) discussed the importance of reserves management in the "Revised Guidelines for Foreign Exchange Reserves" as a process of ensuring that "sufficient official public sector foreign assets are readily accessible to and controlled by the authorities for meeting a defined variety objectives for a country". Therefore, the process of reserves management involves many aspects, that is, the management information system that deals with the measurement of various risks exposure like liquidity, exchange rate and interest rate, and ensures that such risks are accurately managed [<https://www.resbank.co.za>]. According to the (IMF, 2014), central banks should have enough foreign exchange reserves in support of numerous objectives, including, but not limited to, the ones that may be specified by the country itself based on its financial state; backing and sustain confidence its country's monetary policies, including the ability to intervene in supporting the national or union currency; provide proof to the market that a country can meet its current and future external obligations; help the government in meeting its foreign exchange needs and external debt obligations; and maintain reserves for national disasters or emergencies.

Traditionally, the appropriate level of foreign exchange reserves to be held by the central banks was equal to three or four months of the country's imports (Rodrik, 2006). However, due to the unusual rise of financial crises in many emerging countries such an idea has been reviewed. It has been discovered that countries need enough reserves to cover their short-term debt (Fischer, 2001). The idea for countries to hold enough reserves to cover their short-term debt played a critical role in the Asian crisis, as it was discovered that "countries that held very large reserves were better able to avoid the devastating effects of the crisis than those which had smaller reserves" (Fischer, 2001). Therefore, the IMF imposed the rule known as Guidotti-Greenspan-IMF (see Greenspan, 1999) which emphasizes the point of reserves accumulation in emerging countries. The Guidotti-Greenspan-IMF rule states that "countries should hold liquid reserves equal to their foreign liabilities coming due within a year" (Rodrik, 2006). The rule simply uses 'the ratio of short-term external debt to reserves' as the adequacy measure of reserves for a country.

## **1.2 Literature Review**

Box et al. (1970) in their book entitled "Times Series Analysis: Forecasting and Control", developed Autoregressive Integrated Moving Average (ARIMA) time series forecasting models, hence, these models are often referred to as the Box-Jenkins models. Although many other authors, such as Wold (1954) and Quenouille et al., (1957), have studied and discussed time series models, Box and Jenkins promoted their use and revealed their methodical approach to seasonal data. Their approach, which is well-known as the Box-Jenkins approach, possesses many appealing features (Chatfield, 2000).

Vandaele (1983) pointed out that it can allow the manager who only has data on the past years' sales to forecast future sales without having to search for other time-series data such as customer's income, prices, etc.

Guha et al. (2011) in the paper "Gold Price Forecasting Using ARIMA Models", applied an ARIMA time series model to forecast the future gold price in India based on past data to mitigate the risk in the purchase of gold, hence giving a guideline for investors as to when to buy or sell the yellow metal. Guha further revealed that in India gold was, in earlier times, only purchased at the time of marriage or other rituals but then it later gained importance in the eyes of investors also, thus it became necessary to predict the price of gold with suitable methods. His study was based on secondary monthly data for the gold price which was collected from the Multi Commodity Exchange of India Ltd (MCX), ranging from November 2003 to January 2014. The Multi Commodity Exchange is a commodity future exchange based in India which started its operations from November 2003. The author performed all the relevant statistical analysis and comparing the fit statistics such Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Ljung's Box Q statistics). Guha et al., (2011) then decided to choose the ARIMA (1,1,1) as the best model for the collected data which could be used to forecast the future gold price. ARIMA (1,1,1) was chosen from six different model parameters as it provided the best model which satisfied all the criteria of fit statistics while the other five failed the fit statistics.

Khan (2013), in the paper "Forecasting the gold price: Box –Jenkins Approach", applied ARIMA models to develop the best model to predict future gold price or positions. Khan (2013) used sample data of gold prices (in USD per ounce) from 02 January 2003 to 01 March 2012. The data up until 02 January 2012 were used to build the model while the remaining data were used to check for the accuracy of the model. A unit root test was used to test for stationarity from the long-term daily gold positions data and the data was found to be non-stationary. The author applied the first differencing to the series and then repeated the stationarity test which led him to conclude that the data was therefore stationary at the first differencing. The E-views software was used to fit the ARIMA models that were identified using graphs, fit statistics, ACFs and PACFs of the residuals. After several iterations, the model selected for the data was an ARIMA (0,1,1). The author also used the following ways to measure accuracy from the model forecasts: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The author finally concluded that the ARIMA (0,1,1) model could be used for forecasting gold prices in the future.

Davis et al. (2014), in the paper "Modelling and Forecasting of Gold Prices on Financial Markets", used a monthly adjusted close price of gold from January 2003 to April 2012 (112 observed prices). The authors fitted an ARIMA model with the first 106 observed prices and the model was used for 6-step-ahead of forecasts. The forecast values were then compared to the original corresponding prices. The

actual values fell within the forecast limits which widened with increasing lead time (Deshmukh et al., 2014).

Mombeini et al. (2015), argue that “developing a precise and accurate model of the gold price is critical as gold is a difficult asset to manage because of its unique features”. In this study, the authors used an Artificial Neural Network (ANN) model for modelling the gold price and compared it with traditional statistical models of the Autoregressive Integrated Moving Average (ARIMA) type. The gold price data (220 monthly observations) from April 1990 to July 2008 from the study. The data comprised of 220 input vectors and their corresponding output vectors from the historical data of the gold price, which was divided into training and validation sets.

The authors first applied 200 observations (from April 1990 to November 2006) to formulate the model and the last 20 observations (from December 2006 to July 2008) were used to reflect the performance of the different constructed models. Based on the ARIMA model, the past observations of the gold price were used to formulate the model and obtain parameter estimates. The same was done for the ANN model. The authors have adopted three performance evaluation measures, namely, MAE, R-Square, and RMSE, to analyse the performances of various developed models. Their results showed that the ANN method was a powerful tool to model the gold price and could give better forecasting performance than the ARIMA method.

Petrica et al. (2016), investigated the limitations of using ARIMA models in financial and monetary economics using the behaviour of BET index and EUR/RON exchange rates. Many empirical studies have discovered two important features in the analysis of financial time series, that is, "fat-tails" where large losses or gains income has a higher probability than the normal distribution would suggest (Mandelbro et al., 1963), and "volatility clustering", where the occurrence of large (negative & positive) returns is not explainable by the arrival of new information, particularly on the market (Cont, 2005). The authors further revealed that these empirical properties cannot be captured ARIMA models, hence the limitation of these models. The basic version of the least-squares model (e.g., ARIMA) assumes that the expected value of all error terms when squared, is the same at any given point, that is, the variance of the error terms does not change over time. This phenomenon is known as homoscedasticity. This phenomenon needs advanced and sophisticated models to handle it accordingly.

### **1.2.1 Reserves Economic Aspects**

Financial crises are not new to less developed countries (emerging markets) but they have grown more gradually in the past centuries (Kannan et al., 2014). “The quick spread of the 1997 crisis in East Asia and of the 1982 crisis in Latin America has shown how changes in market perceptions can unexpectedly bring trouble to countries even when there has been no change in objective conditions” (Feldstein, 1999). Other countries that were also affected by devastating financial crises in the past centuries include Turkey in 1994 and 2001, Mexico in 1994/1995, Brazil in 1998 and Argentina in 2002 (Rodrik,

2006). The latest world global financial crises of the emerging countries, which occurred in 2008/2009, has left the whole world in a very painful and severe economic condition. Therefore, emerging countries which want to prevent the shocking effect of such crises must be prepared to save themselves rather than expecting to be helped by international countries.

Feldstein (1999), in the paper “Self-Protection For Emerging Market Economies”, has explained that emerging countries should protect themselves from the special effects of financial crises by not depending on the International Monetary Fund (IMF) or other international organisations nor expect that a “new global financial architecture” will create the world economy less dangerous. Furthermore, the country that has significant enough international liquidity, that is sufficient foreign exchange reserves, and a ready source of foreign currency loans is less likely to be the victim of a currency attack. Substantial liquidity also allows a country that is affected from within or without to protect or defend itself better and to make a more orderly adjustment. But the major challenge is to come up with a way to maximise liquidity at a reasonable cost.

Feldstein (1999) has also discussed three different main types of financial crises predominantly encountered in emerging markets: Balance Sheet Crisis; Bank-Run Currency Crisis; and Contagion Crisis and Irrational Speculation. The authors mentioned that a typically country may face a Balance Sheet Crisis when it has a market-determined “floating” exchange rate and a current account surplus, but still get attached because its short-term private and government liabilities denominated in foreign currencies exceed its foreign exchange reserves. Such a country is solvent because it can settle its future export earnings given enough time to do so; however it is illiquid in the sense that it is unable to meet its obligations immediately. Balance Sheet Crisis is also known as the balance of payments or exchange rate crisis, as described by Mudd et al., (2002) in the paper “Dealing with Financial Crises in Emerging Markets” as a crisis which mostly occurs when overall funds inflows to a country are more than offset by overall outflows. This crisis somehow leads the country to a situation where it has to borrow more abroad and/or to devalue its currency. The typical major effect of this kind of crisis is the exchange rate depreciation or a large amount of outflow of foreign exchange reserves from the central bank, which results in a major currency devaluation (Mudd et al., 2002). According to Feldstein (1999), the Balance Sheet Crisis risk exposure is based on the probability that most foreign creditors may worry that the country will not be able to meet its obligations because other foreign creditors will not continue to lend, so the country will be compelled to default or postpone its loans and devalue its currency in order to increase export earnings to pay back its debts rapidly. Otherwise, foreign creditors’ confidence in the country will be strengthened and they will stay to roll over the short-term debts or lengthen their maturity.

Feldstein (1999) discovered that the common feature of Bank-Run Currency Crisis is the widespread failure of domestic banks to meet their obligations. Most of these failures are the result of a currency



decline which raises the value of the bank's dollar liabilities relative to its domestic currency assets. The phenomenon of the currency decline also affects banks with dollar-denominated loans because the corporate borrowers have dollar liabilities and local-currency income. The main cause of the domestic banks' failure is that they accumulate too many bad loans and other investments so that their liabilities to depositors exceed the true value of their loan and other assets. For instance, the bank can only be safe when the depositors remain confident that they can withdraw their funds on demand. However, once depositors become worried that the bank might default (bank-run) from meeting its withdrawal demand requirements, they will abandon the bank which lacks the ability to meet further withdrawal obligations. The serious run on the currency which devalued it arises when enough depositors fear the bank's insolvency and start to withdraw their funds from all domestic banks and deposit them abroad where they feel they are safe. And if the government tries to resist, this may cause a run out of its foreign reserves and increase the risk of an exchange rate collapse (Feldstein, 1999). On the other hand, Mudd et al., (2002) defined Bank-Run Currency Crisis as a banking or financial system which has affected many emerging markets over the past centuries. The authors described this type of crisis as the crisis that occurs due to the internal issues in a country, such as when the commercial bank lenders within the country go bankrupt due to a dramatic drop in real estate prices which may cause a great loss to bank clients who no longer have a valuable asset in their possession, and hence default from their obligations with the bank. The Balance Sheet (balance of payment or exchange rate), and Bank-Run Currency (banking or financial system) crises are the two main varieties of financial crises that have been identified frequently in the history (Mudd et al., 2002). Finally, Feldstein (1999) described the contagion crisis and Irrational Speculation as the crisis which occurs because of the currency decline but is not due to current account imbalances, balance sheet mismatches or banking system failures. Despite the currency being able or allowed to float freely, it can sometimes be subjected to a sharp drop, just as an individual stock on the Stock Exchange can swing substantially in response to newly published information or due to market movements.

Mohanty et al. (2006), in the paper "Foreign exchange reserve accumulation in emerging markets: what are the domestic implications?", discussed some of the implications of the large-scale use of foreign exchange intervention or over-accumulation of foreign exchange reserves by emerging market economies to resist currency appreciation. The accumulation of foreign exchange reserves by emerging market economies has continued on an unusual scale for several years and many countries have adopted an accommodating monetary policy while intervening. The authors discovered that such intervention is good in that it resulted in a prolonged period of low-interest rates, kept the inflation rate under control and then eased one policy dilemma for central banks. Nevertheless, despite all the above advantages, large and prolonged foreign exchange reserves accumulation comes with some potential risks exposures other than near-term inflation. These potential risks involve the bank paying a huge amount for intervention (which led many Latin American countries to abandon such intervention in the early

1990s); unsustainable increases in credit and asset prices; and very liquid and perhaps manipulative banking systems. Another potential problem of the prolonged, and substantial accumulation of foreign exchange reserves is the monetary imbalances, that is, the implications for the balance sheets of the central bank. Greater shifts in the balance sheet variables can sometimes have serious macroeconomic effects which depend on how the related risk exposures are managed and on how the intervention is financed. Again, the balance sheet effects might reduce the effectiveness of sterilisation, with possible inflationary implications.

Mminele (2013) in the paper “Note on the foreign exchange market operations of the South African Reserve Bank (SARB)”, expounds the South African experiences regarding foreign exchange market intervention. The author clearly explained a historical perspective on the bank’s foreign currency intervention in the foreign exchange market, and the bank’s strategies in foreign exchange operations – also during the global financial crisis in the year 2008/2009. The institutional arrangement concerning the foreign exchange reserves accumulation was also discussed. Mminele (2013) classified the history of the bank’s operations in the foreign exchange market in three phases, that is, extensive use of the forward market to provide forward cover; intervention to support the depreciating currency and support market functioning; and lastly, reserves accumulation.

Mminele (2013) stated that the bank intervened in the forward market in September 1985 after the announcement of the foreign debt standstill, and legal action was taken against the South African economic and financial state. The country was no longer allowed to participate in the international capital markets and was also prohibited from borrowing from the International Monetary Fund (IMF) or any other official agencies. The bank then considered using the forward market to encourage and facilitate the use of foreign trade credits by domestic corporates as a process for it to, among other things, accumulate foreign exchange reserves.

The bank also used forward markets in the 1990s, predominantly to State-Owned Enterprises (SOEs). At this time, the country’s foreign debt amounted to US\$24 billion and the bank had US\$12 billion Net Open Foreign-currency Position (NOFP) and eventually increased it to over US\$25 billion in 1995. The bank provided forward cover to SOEs since the hedging of the foreign requirements of the SOEs was mainly via completely forward contracts. Fortunately, the bank managed to reduce its NOFP significantly from US\$25.8 billion in March 1995 to US\$8.5 billion in March 1996. This was done mostly by buying foreign exchange from the market and purchasing the proceeds of government bond issues abroad (Mminele, 2013).

The bank intervened in the foreign exchange market by increasing the forward book to US\$22 billion when the exchange rate of the rand devalued sharply in 1996. The bank successfully managed to reduce the NOFP by almost US\$10 billion in 1997. The bank intervened again in 1998 with the NOFP returning to the levels recorded ten years before, when the emerging markets crisis occurred. In 1998 the bank

officially declared the suspension of the trade of foreign exchange to the market and it started allowing the rand exchange rate to be determined by the factor of its demand and supply. The bank took the opportunity to increase the level of its country's official foreign exchange reserves and in 2003, after the NOFP was excluded, the bank started accumulating reserves by purchasing foreign exchange reserves on a spot basis. The foreign exchange reserves accumulation strategy has been successful over the past decade. Since it was launched in March 2003, the official reserves improved from a negative NOFP of US\$1.4 billion to an international liquidity position of US\$47.9 billion as at the end of December 2012 (Mmminele, 2013).

### **1.3 Volatility**

Volatility is a quantity used to measure the degree of price movement in the stock, futures, options or any other financial instruments (Hull, 2003). It can also be defined as the variation of an asset's returns from their average (Avellen, 2005). In practice, it is quite complex to analyse volatility; that is, it is not directly observable and it cannot be predicted over a long period. However, statistical models may be used to forecast it over a short period (Avellen, 2005). Volatility complexity has led most researchers to ask themselves whether it is possible to estimate this seeming complex quantity.

### **1.4 Problem Statement**

A series that changes a lot and swings widely, such as price series or an economic indicator, is said to be "volatile", that is, it has volatility (Kotze, 2005). The official gold and foreign exchange reserves positions series are not different from such series. It is important for the central bank to understand the risk that lies on the volatile data, especially for the analysis and management of reserves. This thesis will focus on the South African official gold and foreign exchange reserves positions, and it will carefully study the effect of the presence of volatility in the data and the implications caused by this presence when creating the model to predict future values on the reserves' position. The data that will be used here comprises of two variables, that is, South African official gold and foreign exchange reserves positions. The question of interest here is to find out how to deal with the volatility of the data to develop a model that is efficient to be used in the prediction process.

### **1.5 Aim and Objectives**

This thesis is focused on the analysis of the reserves and the importance of the reserve's management by the central bank in South Africa. The main aim of this thesis is to develop SARIMA or ARIMA and ARCH/GARCH models to forecast the expected future reserves positions and deal with the volatility, respectively from reserves positions data and forecast reserves volatility among all two variables, viz., official gold reserves position and foreign exchange reserves position. The discussion based on the model output with recommendations will be also provided at the end of the thesis.

## 1.6 Significance of the Study

The reserves are the most important assets of the country, especial in times of financial crises or disasters. Therefore, understanding their trajectory is very important, hence the significance of this study. This study will play a pivotal role in the analysis of reserves and help the central bank of the country to make fact-based decisions that are being driven by data concerning the accumulation or decumulation of reserves.

## 1.7 Thesis Structure

This thesis is structured as follows: Chapter 2 has the theory of exploratory data analysis in time-series, different statistical methods used to explore time-series data are described and these methods are normally used to give insights from the dataset being used. Chapter 3 comprises of homoscedastic models' theory (i.e., SARIMA); these models were then later used in Chapter 5 to forecast the expected future reserves positions. Such models simply assume that the mean and variance are constant over time, while in real-life data it has been discovered that the variance usually changes with time, hence the limitations of these models. Then Chapter 4 introduces the so-called conditional heteroscedastic models; these models allow the conditional variance to change with time, hence their flexibility. In this chapter different volatility models are discussed, such as Autoregressive Conditional Heteroscedasticity (ARCH) models (Engle, 1982), and their generalization; the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models (Bollerslev, 1986); Exponential Weighted Moving Average (EWMA), and their extension, that is, the Exponential GARCH models (Nelson, 1991); Integrated GARCH (IGARCH) models and Threshold GARCH (TARCH) models by Zakoian (1994) and Glosten et al., (1993). The Multivariate ARCH and GARCH models are discussed in the last section of this chapter, however, the application of these models is beyond the scope of this thesis.

In Chapters 5 and 6 we apply the homoscedastic and volatility (heteroscedastic) models respectively, from a time series dataset that has two variables, viz., South African gold reserves positions and foreign exchange reserves positions. The dataset used in this study and the source where it was obtained or extracted is also described in these Chapters. Chapter 5 further explores the data by breaking it down and plotting graphs to understand the different components (trends, seasonality, and random) of the data. The different formal statistical tests were conducted to check for stationarity in the data. Chapters 7 and 8 describe the Holt-Winters' method and their application respectively, which were then benchmarked with the ARIMA models developed earlier in Chapter 5. Lastly, Chapter 9 has the thesis conclusion and recommendations.

## Chapter 2:

### 2. Theory of Exploratory Data Analysis in Time Series

#### 2.1 Introduction

This chapter covers the theory of time series exploratory data analysis and different components of exploring time series data to make more sense of the data and the entire story behind it. This involves the methodology on how to break down time series data and different data transformation techniques used in time series data exploration.

#### 2.2 Exploratory Data Analysis and Data Transformations Theory

This section describes different techniques of exploring and transforming time series data from non-stationary series to a stationary one. Data transformation techniques used to get the data that can be modelled as a stationary time series without any complications are discussed. These transformation techniques are commonly used to eliminate trend, seasonal and cyclic components, and to accomplish approximate constancy of level and variability with time (Brockwell et al., 2002).

##### 2.2.1 Stationary and Non-Stationary Series

This sub-section reviews the basic methods of dealing with stationary and non-stationary time series. The series can be *deterministic* if future values can be exactly predicted from its past values (Chatfield, 2000). However, in most cases, the future values are not completely determined from the past values and such series are stochastic, or random. In time series analysis the model for a stochastic time series is usually called a *stochastic* process.

Stationarity is an essential property to define a time series process (Pfaff et al., 2008). A stationary time series assumes that the mean is constant with probabilistic properties that are constant over time (Box et al., 2008). More formally, suppose  $\{X_t\}$  is a time series process, then  $\{X_t\}$  is said to be covariance-stationary, or weakly stationary, if the following are independent of time:

$$E(X_t) = E[X_{t-k}] = \mu \quad \forall t, k \quad (2.1)$$

$$Var(X_t) = \gamma_0 < \infty \quad \forall t, \quad (2.2)$$

$$Cov(X_t, X_{t-k}) = \gamma_k \quad \forall t, k. \quad (2.3)$$

Otherwise, the series is non-stationary, that is, it exhibits some trends, cyclic, or any type of seasonality. The stationary time series allows most modelling methods including the popular Box-Jenkins methodology (Nash et al., 1970). Thus, stationarity of the data needs to be firstly achieved before modelling time series. The stationarity of the data is assessed by carefully examining its Auto-

Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots (Enders, 2010). If the ACF of the time series starts high and gradually dies down, then such a series may be non-stationary. However, if the ACF of the time series either cuts off or dies rapidly, then such series may be stationary (Hassan, 2014). The autocorrelation is the correlation between observations that are time  $k$  apart (Walton, 1997). If the ACF of the series at any given lag  $k$  is significantly different from zero, then the serial dependence among the observations must be included in the final model. The autocorrelation at lag  $k$  is usually denoted by  $\rho_k$  and its mathematical expression is

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{[\text{Var}(X_t) \cdot \text{Var}(X_{t+k})]}}, \quad \text{for } k = 0, \pm 1, \pm 2, \dots \quad (2.4)$$

and the ACF can be displayed in a graph (correlogram) of  $\rho_k$  against the given lag  $k$ . From the equation above it can be shown that for  $k = 0$  the expression for  $\rho_k$  is

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_t)}{\sqrt{[\text{Var}(X_t) \cdot \text{Var}(X_t)]}} = \frac{\text{Var}(X_t)}{\sqrt{[\text{Var}(X_t)]^2}} = 1, \text{ for } k = 0 \quad (2.5)$$

since for a stationary time series  $\text{Var}(X_t) = \text{Var}(X_{t+k}) \quad \forall t, k$ .

Similarly, the partial autocorrelation at lag  $k$  measures the correlation among observations that are  $k$  lags apart. However, the PACF at lag  $k$  removes, or “partials out”, all intervening lags (Hassan, 2014).

### 2.2.2 Variance Transformation

Typically, there are many methods developed to stabilize the variability in the data. When the variability of the data increases or decreases with the level, the Box-Cox or Log Transformation is one of the most popular technique used to stabilize the variability of the data (Brockwell et al., 2002). Suppose  $X_1, X_2, \dots, X_{n-1}, X_n$  are the original observations of the time series  $\{X_t\}$ . The Box-Cox transformation  $g_\alpha$  converts them to  $g_\alpha(X_1), g_\alpha(X_2), \dots, g_\alpha(X_{n-1}), g_\alpha(X_n)$ , where

$$g_\alpha(x) = \frac{x^\alpha - 1}{\alpha}, \alpha \neq 0, \\ = \log(x), \alpha = 0. \quad (2.6)$$

The choice of the quantity  $\alpha$  can be made visually by looking at the pattern of the time series plot, and the suitable choice of this quantity can often transform the variability of the data to nearly constant. For example, if the data is positive and its standard deviation increases linearly with level, the variability can be stabilized by selecting  $\alpha = 0$ .

### 2.2.3 Moving Average

A time series constructed by taking averages of several sequential values of another time series is called a moving average (Hyndman, 2009). More compactly, a two-sided moving average (MA) of order  $m = 2k + 1$  can be constructed by taking an average of series original  $\{X_t\}$ ,  $k$  periods around each point and inclusive of the point. The mathematical expression is given by:

$$\begin{aligned}
Y_t &= \frac{1}{m} \sum_{j=-k}^k x_{t+j} \\
&= \frac{1}{2k+1} \sum_{j=-k}^k x_{t+j}, \quad t = k+1, k+2, \dots, n-k,
\end{aligned} \tag{2.7}$$

where  $\{Y_t\} = \{y_{k+1}, y_{k+2}, \dots, y_{n-k}\}$  make up a new smoothed time series generated from averages of the original time series  $\{X_t\}$ . Similarly, a one-sided moving average of the time series  $\{X_t\}$  is given by

$$\begin{aligned}
Y_t &= \frac{1}{m} \sum_{j=-k}^k x_{t+j} \\
&= \frac{1}{k+1} \sum_{j=-k}^k x_{t+j}, \quad t = k+1, k+2, \dots, n-k,
\end{aligned} \tag{2.8}$$

(Hyndman, 2009).

It is important to explicitly note that in this context the moving average is different from the MA(q) component in the ARIMA description that will be explained in detail when we get to Chapter 3. The moving average here can be simply defined as a data smoothing technique, whereas the later definition of moving average MA(q) as part of the ARIMA context will be referring to error lags combinations.

#### 2.2.4 Data Decomposition

There are many methods for the elimination of the trend and seasonality from the data. Most common are the *differencing* and *classical decomposition* of the series into a trend component, a seasonal component and a random residual component.

Differencing means taking the difference between consecutive observations or between observations a year apart (Hyndman, 2001). The differencing method is preferable if the data or series has a deterministic linear trend (Vandaele, 1983). For instance, suppose  $\{X_t\}$  is the time series with seasonal components of period 12 and a deterministic linear trend, then the differenced series of  $\{X_t\}$  at lag 1, say  $\{Y_t\}$  can be used to remove the trend and is given by

$$Y_t = \nabla X_t = X_t - X_{t-1}, \tag{2.9}$$

where  $\nabla$  is called the backward difference operator. All the seasonal components in the series  $\{X_t\}$  can be eliminated by this transformation, but now using the differencing at lag 12, that is

$$Y_t = \nabla_{12}X_t = X_t - X_{t-12}, \quad (2.10)$$

Another compact way of writing the differencing equation is by using the so-called backshift operator  $B$ . Again, suppose  $\{X_t\}$  and  $\{Y_t\}$  are the original non-stationary time series and the first difference time series of  $\{X_t\}$ , respectively. Then using the backshift operator, we have  $BX_t = X_{t-1}$  such that

$$Y_t = \nabla X_t = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t, \quad (2.11)$$

where  $\nabla = (1 - B)$ . Making use of the backshift operator, equation (2.10) can be written as

$$Y_t = \nabla_{12}X_t = X_t - X_{t-12} = (1 - B^{12})X_t. \quad (2.12)$$

Differencing can be performed repeatedly until the trend has totally disappeared, thus the second differencing of the time series  $\{X_t\}$  which is used to eliminate the quadratic trend can be expressed as

$$\begin{aligned} Y_t &= \nabla(\nabla X_t) = \nabla^2 X_t = (1 - B)^2 X_t \\ &= (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned} \quad (2.13)$$

In general, the powers of the backshift operator and backward difference operator for trend elimination are defined as  $B^d X_t = X_{t-d}$  and  $\nabla^d = (1 - B)^d$ , respectively. Therefore, the general expression of the  $d$ -differencing for  $\{X_t\}$  can be written as

$$Y_t = \nabla^d X_t = (1 - B)^d X_t. \quad (2.14)$$

The backshift operator for seasonal models is denoted as  $B^s$  such that  $B^s X_t = X_{t-s}$ , where “s” indicates the seasonal period per year. Thus, the general equation for seasonal differencing may be written as

$$Y_t = \nabla_s X_t = X_t - X_{t-s} = (1 - B^s)X_t, \quad (2.15)$$

(Chatfield, 2000).

### 2.2.5 Components of Time Series

Generally, the time series has two different types of variations which can be observed from the data and hence eliminated where necessary. These are systematic variations (trends, seasonal, and cyclical) and random variations (irregular component, and inherent background noise in the series).

Classical decomposition of the time series  $\{X_t\}$  is based on the two models formed by the combination of the time series variations or components:



$$\text{Additive Model: } X_t = m_t + c_t + s_t + \varepsilon_t, \quad (2.16)$$

$$\text{Multiplicative Model: } X_t = m_t \times c_t \times s_t \times \varepsilon_t, \quad (2.17)$$

where  $\{X_t\}$  is the observation at time  $t$ ,  $m_t$  is a trend component,  $c_t$  is a cyclical component,  $s_t$  is a seasonal component and  $\varepsilon_t$  is a random noise component, which is stationary with mean zero (Brockwell et al., 2002).

The aim is to estimate the component  $m_t$ ,  $c_t$  and  $s_t$  and deduct them from the data to produce a sequence of errors or residuals (estimated noise) that can then be modelled as a stationary time series.

### **1. Trend Component**

This component is the long-term pattern of a time series. When a time series exhibits an increasing long-term pattern, then it has a positive trend. However, when a time series exhibits a gradual decreasing pattern then it has a negative trend. A time series that shows neither an increasing nor a decreasing long-term pattern is stationary in the mean (Chatfield, 2000).

### **2. Cyclical Component**

The cyclical component can be identified in a series as an upwards and downwards movement pattern around a given trend. Its period normally depends on the type of business or industry being examined.

### **3. Seasonal Component**

This type of variation occurs when a time series displays regular fluctuations at a fixed time interval every year, for example, weekly, monthly, quarterly or half-yearly. Similar patterns of behaviour can be detected throughout these intervals (Chatfield, 2000).

### **4. Random Component**

The random or irregular component is unpredictable and can be described as the random changes in a time series caused by non-recurring factors such as natural disasters (earthquakes, volcanos, or tsunamis, etc.). All-time series has this type of component which makes it a random variable. During the forecasting processes, the main purpose is to model all the systematic components to the point that the only component left unexplained is the random component.

## Chapter 3

### 3. Theory of Homoscedastic Models: S/ARIMA Models

#### 3.1 Introduction

In 1926, Yule introduced the class of models known as Autoregressive (AR) models and eleven years later, Slutsky (1937) came up with the Moving Average (MA) models. Wold (1938) joined both AR and MA schemes to form composite time series models known as ARMA models. Box et al. (1970) popularized the use of the ARMA models and described them as an important forecasting tool in time series data analysis (Chatfield, 2000). Although many other authors and/or researchers, such as Wold (1954) and Quenouille (1957) studied and discussed these types of time series models, Box and Jenkins spread their use, established how to extend their application to seasonal data, and made the methodology operational. Hence, these models are often referred to as Box-Jenkins Models. These models are flexible and able to determine simulated values by combining measures from historical data with associated temporal variations (Kao et al., 2000). This class of models comprises of two polynomials, that is, an Autoregressive scheme and Moving Average scheme, hence the acronym ARMA stands for ‘AutoRegressive Moving Average’.

An autoregressive moving-average (ARMA) model simply joins the concepts of AR and MA models with the purpose of keeping the minimum number of parameters (Matei, 2009). Before considering how the AR and MA schemes are combined to form an ARMA model, it is necessary to briefly discuss each of them separately.

#### 3.2 Autoregressive Models

The Autoregressive (AR) process is a stochastic process generally used in statistical calculations in which future values are estimated based on a weighted sum of the past values (Chatfield, 2000). An autoregressive model operates under the premise that the past observations influence the current ones. The most commonly used type of autoregressive model is the first-order autoregressive AR (1) model, meaning that the current value depends heavily on the immediately preceding value. Suppose that  $\{X_t\}$  is the time series with the sets of observations  $\{X_{t-1}, X_{t-2}, \dots, X_{T-1}, X_T\}$  collected over  $t \in [1, T]$  time interval. The AR (1) model can be expressed as

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad (3.1)$$

where  $X_{t-1}$  is the past observation of  $X_t$  associated with parameter  $\phi$ , and  $\varepsilon_t$  is the error term of the model.

An AR (2) model says that the present value depends on the previous two values and it can be expressed as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t. \quad (3.2)$$

In general, an AR(p) is the  $p^{th}$  – order model in which the current value simply depends on the past  $p^{th}$ - observations of the series and its mathematical expression is

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (3.3)$$

where  $\phi_1, \phi_2, \dots, \phi_p$  are the AutoRegressive parameters to be estimated and which describe the effect of the past values on  $X_t$ , subscript  $p$  is the order of the process,  $\varepsilon_t$  is the error term which is assumed to be a white noise with mean 0 and variance  $\sigma_\varepsilon^2$  and  $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$  are the past values of the process (Chatfield, 2000).

### 3.3 Moving Average Models

The moving average can be described as a stochastic process used in statistical calculations in which future values are estimated using the weighted sum of the past errors. Suppose  $\{X_t\}$  is an MA (1) process and then it can be expressed as

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}. \quad (3.4)$$

An MA (2) which is a second-order process is given by

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \quad (3.5)$$

Generally, an MA (q) process is of the  $q^{th}$  – order and its mathematical expression is

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3.6)$$

where  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}\}$  are the error terms at the time periods  $t-1, t-2, \dots, t-q$ , respectively and  $\theta_1, \theta_2, \dots, \theta_q$  are the Moving Average parameters, which describe the effect of the past errors on  $X_t$  (Chatfield, 2000).

### 3.4 Building an ARMA Model

An Autoregressive Moving Average (ARMA) model is the flexible type of time series models which can be used in modelling a large class of stationary time series as given the order of p, the number of AR terms, and/or q, the number of MA terms, are correctly identified (Makridakis et al., 1997). Since the ARMA model comprises of the p-AR and q-MA schemes or terms, the general time series process  $\{X_t\}$  can be modelled as a combination of past  $\{X_t\}$  observations and/or past  $\{\varepsilon_t\}$  error terms (Wold, 1938). The simplest form of ARMA models is the one which combines AR (1) and MA (1) to form a univariate ARMA (1,1) model given by

$$\begin{aligned} X_t - \phi_1 X_{t-1} &= \alpha_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ X_t &= \alpha_0 + \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \end{aligned} \quad (3.7)$$

where  $\alpha_0$  is constant. Thus, the general mathematical ARMA (p, q) model expression which combines both schemes can be written as

$$\begin{aligned} X_t &= \alpha_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \\ X_t &= \alpha_0 + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \end{aligned} \quad (3.8)$$

where  $\phi_i, i = 1, 2, \dots, p$  and  $\theta_j, j = 1, 2, \dots, q$  are the autoregressive and moving average parameters, respectively, and  $\{\varepsilon_t\}, t = 1, 2, \dots, T$  is a sequence of independent, identically normally distributed random variables, i.e., white noise. Using a backward shift operator B, ( $BX_t = X_{t-1}$ ), the ARMA model can be expressed more compactly as

$$\phi(B)X_t = \alpha_0 + \theta(B)\varepsilon_t, \quad (3.9)$$

where the root of the polynomials  $\phi(B)$  and  $\theta(B)$  are restricted to lie outside the unit circle to ensure stationarity and invertibility, respectively (Walton, 1997). Generally, there are four main steps required to use the ARMA (p, q) model above (Wold, 1938):

1. The original series  $\{X_t\}$  must be stationary about its mean and variance.
2. The suitable order of p and/or q need to be clearly specified.
3. The values of the parameters,  $\phi_i, i = 1, 2, \dots, p$  and/or  $\theta_j, j = 1, 2, \dots, q$  must be estimated using some non-linear optimization procedure that minimizes the sum of square errors or any other appropriate loss function.
4. The perfect way of modelling seasonal series must be anticipated, and the suitable order of such models should be identified.

Based on the first step above, in reality most time series are non-stationary and ARMA models cannot be useful when are applied directly to those series. One convenient way of dealing with non-stationary series is to apply ‘differencing’ to the original series and transform it to be stationary. The differencing is applied repeatedly until the series is stationary. Then the  $d^{th}$  – differences may be denoted as  $(1 - B)^d X_t$  using the backward shift operator, where  $d$  is the order of differencing used to convert the non-stationary series. If the series is differenced  $d$  times before an ARMA (p, q) model is fitted, then the model for the original undifferenced series is denoted as ARIMA (p, d, q), where the “I” in the acronym stands for “Integrated” and “d” denotes the number of differences taken to transform non-stationary series to stationary (Chatfield, 2000). Thus, the generalized mathematical equation of (3.9) gives the ARIMA model which can be written as

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t, \quad (3.10)$$

If the series  $\{X_t\}$  has a seasonality effect, then the ARIMA model in equation (3.10) can be generalized to the class of time series models known as SARIMA, where the “S” in the acronym stands for “Seasonal”. The backward shift operator for seasonal models is denoted as  $B^s$  such that  $B^s X_t = X_{t-s}$ , where “s” indicates the seasonal period per year. Thus, the seasonal differencing may be written as  $\nabla_s X_t = X_t - X_{t-s} = (1 - B^s)X_t$ . Therefore, SARIMA model with non-seasonal and seasonal terms of the order (p, d, q) and (P, D, Q) respectively, is written as  $SARIMA(p, d, q) \times (P, D, Q)_s$  which can generally be expressed as

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D X_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (3.11)$$

where  $\Phi$  and  $\Theta$  are the polynomials in  $B^s$  of order P and Q, respectively.

### 3.4.1 Box-Jenkins Methodology

Box and Jenkins proposed the approach to ARIMA models which is known as a “Box-Jenkins Methodology”. This approach became extremely popular in the 1970s among academics when it was shown through many practical studies that it could outperform many other econometric models that were well-known by that time (Makridakis et al., 1997). Box-Jenkins methodology includes identification of a suitable ARIMA model, fitting it to the relevant data and using it to predict future trends from the data (Hyndman, 2001). In general, these are the three stages of building a Box-Jenkins time series model. However, later studies and explanation of the Box-Jenkins procedure (e.g., Wheelwright, Makridakis and Hyndman) often added a preliminary stage of “data preparation” and can be listed more methodically as follow:

1. Data preparation
2. Model identification
3. Model parameter coefficients estimation
4. Diagnostic checking and model validation.
5. Forecasting using a fitted model.

If the selected model fails after checking the diagnostics, the entire procedure is repeated from stage 2 to till stage 5.

#### 3.4.1.1 Data Preparation

This is the first stage in the Box-Jenkins Methodology, and it includes data transformation processes. Usually, financial and/or economic data exhibits unusual behaviour or variation, therefore, application of data transformations can help to stabilize the variance in such data (Hyndman, 2001). Differencing is applied to data until there are no obvious patterns such as trends or seasonality left in the data.

Furthermore, working with a stationary series is more convenient in the Box-Jenkins Methodology than an original non-stationary series.

Box-Jenkins models are applicable to stationary series only, which has all properties listed in equation (2.1) – (2.3).

Stationarity in a time series can be assessed through its plots. If a series does exhibit some form of trend or any type of time series components (seasonality, cyclical, or random effect), then such series is non-stationary. It can also be detected through an autocorrelation plot (ACF) of a series, where a slow decaying in the ACF suggests non-stationarity in a series. If the autocorrelations of a series start high and decline gradually but never cut off, then such series is non-stationary, and the differencing or any relevant transformation should be then applied (Hassan, 2014). The Dickey-Fuller (DF) statistical test or Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test can be then used for stationarity checks.

#### 1. Dickey-Fuller Formal Statistical Test:

The null hypothesis for the Dickey-Fuller statistical test assumes that the series is non-stationary. The idea behind the DF test is to test whether the change in series  $\{X_t\}$  can be explained by lagged value and linear trend. This test was developed by Dickey et al., (1979) to test whether a unit root is present in an autoregressive model. The stationarity condition is that for an AR (1) model  $|\phi| < 1$ . When  $\phi = 1$ , this corresponds to a non-stationary random walk. The general idea is to find an AR model for the observations  $\{X_t\}$  as defined in section 2.2 and test whether  $\phi \geq 1$  or  $|\phi| \leq 1$ .

Consider the AR (1) model defined in equation (3.1), the unit root hypothesis of non-stationarity against the alternative hypothesis of stationarity can be given as:

$$H_0: \phi \geq 1 \text{ vs. } H_0: \phi < 1.$$

Dickey-Fuller derived the alternative hypotheses by subtracting  $X_{t-1}$  from both sides of equation (3.1) (Nielsen, 2006):

$$\begin{aligned} X_t - X_{t-1} &= \phi X_{t-1} - X_{t-1} + \varepsilon_t \\ \Delta X_t &= (\phi - 1)X_{t-1} + \varepsilon_t \\ \Delta X_t &= \delta X_{t-1} + \varepsilon_t. \end{aligned} \tag{3.12}$$

where  $\delta = \phi - 1$ . Therefore, the unit root hypotheses become:

$$H_0: \delta = 0 \text{ vs. } H_0: \delta < 0.$$

The null hypothesis is tested using the t-statistics which is given by this formula:

$$\hat{t}_n = \frac{1-\hat{\phi}}{\sqrt{\hat{\sigma}^2(\sum_{t=2}^n X_{t-1}^2)^{-1}}}, \quad df = n - 1, \quad (3.13)$$

where  $\hat{\phi}$  and  $\hat{\sigma}^2$  are estimators of  $\phi$  and  $\sigma^2$  (variance of  $\varepsilon_t$ ), respectively. The null hypothesis is accepted when  $\hat{t}_n$  is greater than the critical value and we can conclude that the series under consideration is non-stationary. On the other hand, the null hypothesis is rejected if  $\hat{t}_n$  is less than the critical value and we can conclude that the series is stationary. The p-value of the test can also be used to draw a conclusion about the null hypothesis; if the p-value is less 0.05 the null hypothesis is rejected, otherwise it is accepted (Mushtaq, 2011).

Dickey and Fuller extended the original DF test because this test is a normal test (usually at 5% significance level) and is not reliable when the error terms  $\varepsilon_t$  in equation (3.12) are autocorrelated. When the autocorrelations of  $\varepsilon_t$  are large, the significance of the test becomes more inaccurate (Jurgen et al., 2011). The ignoring of these autocorrelations will basically mean that the null hypothesis can be rejected at a low significance level (e.g., 5%), while the significance level lies at, for example, 20%. This is the reason Dickey and Fuller extended the original DF to Augmented Dickey-Fuller (ADF) test to overcome this issue. The ADF test is normally used to test larger and more complex sets of time series models by removing all autocorrelations in the time series and it uses the same procedure as the DF test (Dwyer et al., 2012).

## 2. Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Formal Statistical Test:

The null hypothesis for the KPSS statistical test assumes that the series is stationary around a deterministic trend. This formal statistical test was proposed by Kwiatkowski among others in 1992. The null hypothesis is rejected at 5% level of significance if the p-value is less than 0.05, which means that further differencing is required. In the case where KPSS test p-value is greater than 0.05, the null hypothesis is accepted, and this implies that the series is stationary.

### 3.4.1.2 Model Identification

This is where the adequate model is identified from the ARIMA model's family. The autocorrelation and partial autocorrelation functions are the most useful tools for the model identification step.

Box-Jenkins models can be extended to comprise the seasonal autoregressive and seasonal moving average terms. The autocorrelation function can also be used to detect the seasonality order(s), which can be revealed by the spikes at lag 1s, 2s, 3s, etc. To transform a series with seasonality to be stationary, the  $s^{th}$  differences may need to be applied from the original series or raw data.

Once the issue of stationarity and seasonality has been resolved, then the order of the AR(p) and MA(q), and/or AR(P) and MA(Q) may be identified. These can be identified using the series autocorrelation and partial autocorrelation functions. Specifically, the ACF plot is used to determine the order of the

Moving Average MA(q) component, whereas PACF plot can help in determining the order of the Autoregressive AR(p) component. In general, for an AR (1) process, the sample autocorrelation plot should show an exponentially declining appearance. However, for higher-order processes, that is AR(p), to identify such order, the sample partial autocorrelation plot should be analysed with the sample autocorrelation plot. If an AR(p) process is appropriate, the typical behaviour of the sample autocorrelation function should decay gradually, and the sample partial autocorrelation function should cut off quickly after p lags. For the MA(q) process, the sample autocorrelation and partial autocorrelation functions should cut off quickly after q lags and decay gradually, respectively (Dobre et al., 2008).

For ARIMA models without any seasonal component, ACF and PACF can be interpreted as follows:

1. If ARIMA (p, d, 0) model is appropriate then the ACF plot of transformed/differenced data shows an exponential decaying, and PACF plot shows a significant spike at only lag p and none beyond that.
2. If ARIMA (0, d, q) model is appropriate then the PACF plot of transformed/differenced data shows an exponential decaying, and ACF plot shows a significant spike at only lag q and nothing beyond q (Hyndman et al., 2014, p.76).

For SARIMA models the seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF (Hyndman et al., 2014, p.85):

1. For ARIMA (0,0,0) (0,0,1)<sub>12</sub>, the ACF plot will have a spike at only lag 12 and no other significant spike from other lags, and PACF will show exponential decaying in the seasonal lags, such as 12;24;36; ...
2. For ARIMA (0,0,0) (1,0,0)<sub>12</sub> model, there will be exponential decaying in the seasonal lags of the ACF and a single significant spike at lag 12 only in the PACF (Hyndman et al., 2014, p.85).

### 3.4.1.3 Model Selection and Parameter Estimation

The model which best fits the data is usually selected using diverse model selection criteria, such as Information Criterion (Akaike, 1973). The following statistical model selection criteria are normally used to select the orders of (p, q) for an ARMA or ARIMA process identified in stage 2 above:

- Fit all ARMA (p, q) or ARIMA (p, d, q) models with  $0 < p < p_{max}$  and  $0 < q < q_{max}$ , for chosen values of maximal orders.
- Let  $\hat{\sigma}^2(p, q)$  be the Maximum Likelihood Estimation (MLE) of  $c = Var(\varepsilon_t)$ , the variance of ARMA or ARIMA innovations under Gaussian or Normal assumptions.
- Select (p, q) to minimize  $\sigma^2$  and the following:
  - Akaike Information Criterion



$$AIC(p, q) = \log[\hat{\sigma}^2(p, q)] + 2 \frac{p+q}{n},$$

- Bayes Information Criterion

$$BIC(p, q) = \log[\hat{\sigma}^2(p, q)] + \log(n) \frac{p+q}{n},$$

- Hannan-Quinn Criterion

$$HQC(p, q) = \log[\hat{\sigma}^2(p, q)] + 2\log(\log(n)) \frac{p+q}{n}.$$

The smaller magnitude of these criteria above indicates the best model fitted.

Model estimation is simply the efficient use of the data to draw the conclusion about the parameters conditional on the adequacy of the fitted model (Box et al., 2008). The maximum likelihood estimation is usually the ideal method used to estimate the model parameters (Dobre et al., 2008). In this stage the numerous graphs based on the transformed and differenced data are used to try to find the possible ARMA or ARIMA model which might provide a good fit to the data, and the values of the model coefficients which provide the best fit to the data are estimated using sophisticated computational algorithms (Hyndman, 2001).

#### 3.4.1.4 Diagnostic Checking and Model Validation

Once the model has been identified, then model diagnostic checking step follows, where the fitted model is checked in relation to the data to verify its accuracy and/or inadequacies and therefore carry out model improvement where necessary. The autocorrelation function of the residuals and the cumulative periodogram of the residuals are the main two model diagnostics techniques usually used in time series analysis (Box et al., 2008). For a good model with errors always show a consistent behaviour for that model (Brockwell et al., 2002). The residuals are given as

$$e_t = X_t - \hat{X}_t, \quad t = 1, 2, \dots, T, \quad (3.14)$$

where  $\{\hat{X}_t\}$  is the linear mean-square predictor of  $\{X_t\}$ . The rescaled one-step residuals prediction errors can be given as

$$\hat{Z}_t = \frac{e_t}{\sqrt{a_{t-1}}}, \quad (3.15)$$

where  $a_{t-1} = \frac{E(e_t)^2}{\sigma^2}$  and  $\sigma^2$  is the white noise variance of the fitted model. If the data is accurately predicted by the fitted ARMA (p, q) or ARIMA (p, d, q) model with the white noise sequence  $\{\varepsilon_t\}$ , then for large samples, the properties of  $\{\hat{Z}_t\}$  should imitate those of  $\{\varepsilon_t\}$ . To check the goodness of fit for the fitted model we therefore examine the rescaled residual series  $\{\hat{Z}_t\}$  and check whether it resembles a realization of the white noise sequence  $\{\varepsilon_t\}$ . Sometimes the histogram for the rescaled residuals is used for model appropriateness check, and in such instances the mean of the residuals is expected to be closer to zero if the fitted model is appropriate. Furthermore, if we assume that the

rescaled residuals data are Gaussian, then the shape of its histogram should be like that of a normal density with a mean zero and variance of one (Brockwell et al., 2002). The histogram plot of the residuals versus time  $t$  should resemble the realization of a white noise sequence  $\{\varepsilon_t\}$  if the fitted model is appropriate. However, if the fitted model is inappropriate it should show some trends, cycles and/or non-constant variance (Walton, 1997). If at least 95% of the rescaled residuals lie inside the bound  $\pm 1.96$  or if there are rescaled residuals closer to this bound, then the fitted model can be regarded as Gaussian. The inspection of the corresponding qq-plots can be done to check for the distribution of the residuals with either the normal distribution or t-distribution. The Jarque-Bera and Box-Ljung statistic tests for normality and correlation in the residuals are also computed respectively. Furthermore, for the appropriate fitted model, the sample ACF and PACF of the observed residuals should lie inside the bound  $\pm \frac{1.96}{\sqrt{n}}$ , that is, roughly 95% of the time and these bounds are shown in the ACF and/or PACF graphs.

#### 3.4.1.5 Forecasting Using the Fitted Model

This is the final stage of the Box-Jenkins Methodology and it entails what the whole procedure is intended to accomplish. Usually, it is straightforward to compute forecasts once the model is identified, selected, estimated and checked. Suppose we have the observations  $X_1, X_2, \dots, X_{n-1}, X_n$  of a series  $X_t$  that is assumed to be appropriately modelled as an ARMA(p, q) or ARIMA(p, d, q) process, then we can forecast future values of  $X_{n+h}$  by calculating the linear combination  $P_n(X_{n+h})$  of  $X_1, X_2, \dots, X_{n-1}, X_n$  that reduces the mean squared error  $E[X_{n+h} - P_n(X_{n+h})]^2$ . Of course, this can be easily done by a computer.

The underlying strategy of Box-Jenkins Methodology has now grown to a point that is applicable to a wide range of statistical modelling situations (Hyndman, 2001). It entails a good framework which makes it easy for many data analysts to visualize their dataset and come up with a good model which can be helpful in answering their question(s) of interest about the data.

### 3.5 ARMA and/or ARIMA Models Limitations

Despite the advantages of using ARMA and/or ARIMA models, they are not without their limitations. Most studies in financial time series data analysis reported that this type of data has the three following stylized facts: *fat-tail*, that is, large losses or gains are coming at a higher probability than the normal distribution would suggest; *volatility clustering*, that is, large changes tend to be followed by large changes of either sign, and small changes tend to be followed by small changes (Mandelbro et al., 1963); and *leverage effect*, that is, the tendency for changes in stock prices to be negatively correlated with changes in stock volatility (Black, 1976). Therefore, these empirical properties cannot be captured by ordinal ARMA and/or ARIMA models, since these models simply assume that the expected value of all error terms when squared, is the same at any given point; that is, the variance of the error terms does not change over time (Petrica et al., 2016). Then more advanced and sophisticated models to deal with

the data that have these stylized facts are needed. Some of these facts can be captured by making use of volatility models such as ARCH and GARCH models for volatility clustering and use of heavy-tailed distributions governing the innovations for fat-tail. In the next chapter, we discuss volatility models.

## Chapter 4

### 4. Volatility Models

#### 4.1 Introduction

In Chapter 3, the time series modelling considered assumed time-independent or constant variance of the errors  $\{\epsilon_t\}$  (a phenomenon known as homoscedasticity). However, statistically speaking, in real-life data, such as financial data, this assumption does not hold since variance usually changes with time (a phenomenon known as heteroscedasticity). This is because the financial market may react differently to political disorders, economic crises, natural disasters or wars. Furthermore, most empirical studies have proven that financial data suffers from the following main three effects:

- serial dependence is present in the data (leverage effect),
- volatility changes over time (volatility clustering),
- distribution of the data is heavy-tailed, asymmetric, hence not Gaussian (fat-tail) (Posedel, 2018, p.244).

In 1982, Robert Engle introduced the so-called Autoregressive Conditional Heteroscedasticity (ARCH) models to model the time-varying volatility often observed in economical or financial time series data. His doctoral student Tim Bollerslev then introduced a useful generalization of these models in 1986, known as Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. These models can easily accommodate the dynamics of conditional heteroscedasticity by treating it as a variance to be modelled. The ARCH models simply cater for variation in the variance by relating the error variance to the past errors, and GARCH models cater for it by including the previous conditional variances as well. The goal of these models in financial applications is to provide a volatility measure, like standard deviation, that can be used in financial decisions concerning risk analysis, portfolio selection, and derivative pricing. They also play a significant role when the objective of the study is to analyse and forecast volatility.

#### 4.2 An ARCH(s) Model Overview

Let  $c_t$  be a random variable that has a mean and a variance conditional on the information  $\mathbb{T}_{t-1}$  ( $\sigma$ -field generated by  $c_{t-i}, i \geq 1$ ). Then the random variable  $c_t$  is an ARCH(s) process if it is stationary and if it satisfies for all  $t$  and some strictly positive-valued process  $\sigma_t$ , the two equations which follow:

$$c_t = \sigma_t \epsilon_t = \sqrt{\sigma_t^2} \epsilon_t, \quad (4.1)$$

where  $\{\epsilon_t\}$  is a sequence of independent, identically distributed random variables with a mean zero and unit variance. However,  $\{\epsilon_t\}$  can be other white noise with no need to be Gaussian as in the ordinal least squares regression. This implies that  $c_t|\mathbb{T}_{t-1} \sim G(0, \sigma_t^2)$ , where  $G$  is the distribution (characteristically assumed to be a normal or leptokurtic one). The second equation is the ARCH model of order  $p$  defined by the conditional variance:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 c_{t-1}^2 + \dots + \alpha_s c_{t-s}^2, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^s \alpha_i c_{t-i}^2,\end{aligned}\tag{4.2}$$

where  $\alpha_0 > 0$  and  $\alpha_i \geq 0, i = 1, 2, \dots, s-1, s$ , to ensure that the conditional variance is positive.

The ARCH( $s$ ) model of  $c_t$  has the following useful properties:

1. The conditional mean and unconditional mean of  $c_t$  are both equal to zero, i.e.,  $E(c_t|\mathbb{T}_{t-1}) = E(c_t) = 0$ .
2. The conditional variance of  $c_t$  has a nontrivial positive-valued parametric function of  $\mathbb{T}_{t-1}$  and it is given by  $\sigma_t^2 = E(c_t^2|\mathbb{T}_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i c_{t-i}^2$
3. The unconditional variance of  $c_t$  is given by  $\sigma_t^2 = Var(c_t | c_{t-1}, c_{t-2}, \dots) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i}$ .
4. The unconditional distribution of  $c_t$  is leptokurtic.

The proofs for the above ARCH model properties will be shown later, and for simplicity to show or prove these properties, we will use an ARCH (1) which will be introduced shortly; the special case of the general ARCH( $s$ ) model where  $s=1$ .

If the condition  $\alpha_i \geq 0$  in equation (4.2), to ensure that the conditional variance  $\sigma_t^2$  is positive for all  $t$  is relaxed, then the ARCH( $s$ ) model can be rewritten in the following natural way to achieve positivity of the conditional variance:

$$c_t^2 = \sigma_t^2 \epsilon_t^2 = \alpha_0 + \alpha_1 c_{t-1}^2 + \dots + \alpha_s c_{t-s}^2 + P_t,\tag{4.3}$$

where  $P_t = \sigma_t^2(\epsilon_t^2 - 1) = c_t^2 - \sigma_t^2$  (Talke, 2003).

#### 4.2.1 Building an ARCH( $s$ ) Model

Developing an ARCH model is relatively easy; once the statistical test has been conducted for ARCH Effect and found to be significant, one can use the PACF of  $c_t^2$  to determine the ARCH order. Before we go to a procedure on how to use the PACF of  $c_t^2$ , we first recall the ARCH model given by equation (4.2) above and note that for any given sample,  $c_t^2$  is an unbiased estimate of  $\sigma_t^2$  (Tsay, 2005). Therefore,

from equation (4.3) it can be shown that  $P_t$  is an uncorrelated series with mean 0. One can easily notice that the ARCH( $s$ ) model in equation (4.3) is in the form of an AR( $s$ ) for  $c_t^2$  (see equation (3.3) in Chapter 3 for an AR( $s$ ) model format), except that  $\{P_t\}$  is not an independent and identically distributed series. Since  $\{P_t\}$  are not identically distributed, the least-squares estimates of the prior model are consistent, but not efficient. The PACF of  $c_t^2$  may not be effective with a significantly small sample size.

#### 4.2.1.1 Fat – Tailed Distributions

This thesis considers three different probability distributions, or we simply assume that the error term  $\epsilon_t$  follows these distributions. These are the standard normal distribution, the student- $t$  (STD), the standardized student- $t$  distribution (SSTD), and the generalized error distribution (GED). The log-likelihood functions of these distributions are used for parameters estimation. The reason for using the log-likelihood functions, instead of the original likelihood function, is because maximizing the conditional likelihood function is equivalent to maximizing its logarithm. The main purpose for using different distributions (other than the normal distribution) is because the third and fourth moment of the normal distribution could be too restrictive (Nasstrom, 2003, p.25). The student- $t$  distribution has one parameter for modelling the fourth moment and this parameter is also estimated using the maximum likelihood estimation (Nasstrom, 2003, p.25). Again, these distributions (i.e., heavy-tailed distributions) helps us to cater for fat – tail, which cannot be captured by the normal distribution, hence their significance in modelling financial data.

#### 4.2.1.2 Parameter Estimation for an ARCH( $s$ ) Model

The parameters for an ARCH( $s$ ) model are usually estimated using the maximum likelihood function. Typically, there are three common likelihood functions used for ARCH model estimation. Under the assumption of normality, the likelihood function of an ARCH( $s$ ) model is given by

$$\begin{aligned} f(c_1, c_2, \dots, c_T | \alpha) &= f(c_T | \mathcal{T}_{T-1}) f(c_{T-1} | \mathcal{T}_{T-2}) \dots f(c_{s+1} | \mathcal{T}_s) f(c_1, \dots, c_s | \alpha) \\ &= \prod_{t=s+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right) \times f(c_1, \dots, c_s | \alpha), \end{aligned} \quad (4.4)$$

where  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_s)'$  and  $f(c_1, \dots, c_s | \alpha)$  is the joint probability density function (pdf) of  $c_1, c_2, \dots, c_s$ . The function  $f(c_1, \dots, c_s | \alpha)$  is usually deleted from the above likelihood function when the sample size is large enough. This is done simply because this causes the likelihood function to be more complicated to evaluate. Therefore, this results in using the conditional likelihood function

$$f(c_{s+1}, c_2, \dots, c_T | \alpha, c_1, c_2, \dots, c_s) = \prod_{t=s+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right), \quad (4.5)$$

where  $\sigma_t^2$  can be evaluated recursively. Since maximizing the conditional likelihood function is equivalent to maximizing its logarithm, for simplicity we will then use the logarithm of the conditional likelihood to find the maximum likelihood estimates (MLEs) of an ARCH( $s$ ) under normality and it is given by

$$l(c_{s+1}, c_2, \dots, c_T | \alpha, c_1, c_2, \dots, c_s) = \sum_{t=s+1}^T \left( -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{1}{2} \frac{c_t^2}{\sigma_t^2} \right) \quad (4.6)$$

By eliminating the term  $(\ln(2\pi))$  that does not involve any parameters, the log likelihood can be simplified to

$$l(c_{s+1}, c_2, \dots, c_T | \alpha, c_1, c_2, \dots, c_s) = -\frac{1}{2} \sum_{t=s+1}^T \left( \ln(\sigma_t^2) + \frac{c_t^2}{\sigma_t^2} \right), \quad (4.7)$$

where  $\sigma_t^2 = \alpha_0 + \alpha_1 c_{t-1}^2 + \dots + \alpha_s c_{t-s}^2$  can be evaluated recursively.

Under the Standardized Student- $t$  distribution assumption, it is assumed that  $\epsilon_t$  follows a heavy tailed distribution. Suppose  $y_m$  has a Student- $t$  distribution with  $m$  degrees of freedom. Then  $\text{Var}(y_m) = \frac{m}{(m-2)}$  for  $m > 2$ , and therefore  $\epsilon_t = \frac{y_m}{\sqrt{\frac{m}{(m-2)}}}$ . Thus, the pdf of  $\epsilon_t$  is given by

$$f(\epsilon_t | m) = \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left[\frac{m}{2}\right] (\sqrt{(m-2)\pi})} \frac{1}{\sigma_t} \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right)^{-\frac{(m+2)}{2}}, m > 2. \quad (4.8)$$

The conditional likelihood function of  $m$  and  $\alpha$  is obtained using the equation  $c_t = \sigma_t \epsilon_t$  and can be written as

$$\begin{aligned} f(c_{s+1}, c_2, \dots, c_T | \alpha; m, c_1, c_2, \dots, c_s) \\ = \prod_{t=s+1}^T \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left[\frac{m}{2}\right] (\sqrt{(m-2)\pi})} \frac{1}{\sigma_t} \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right)^{-\frac{(m+2)}{2}}, m > 2. \end{aligned} \quad (4.9)$$

Now, again we use the log-likelihood function to find the maximum likelihood estimates (MLEs) as we did under the normality assumption. Then the likelihood function from equation (4.9) becomes

$$l(c_{s+1}, c_2, \dots, c_T | \alpha; m, c_1, c_2, \dots, c_s) = - \sum_{t=s+1}^T \left[ \frac{m+1}{2} \ln \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right]. \quad (4.10)$$

The log-likelihood function to estimate  $m$  jointly with other parameters is given by

$$\begin{aligned} l(c_{s+1}, c_2, \dots, c_T | \alpha; m, (c_1, c_2, \dots, c_s)) \\ &= (T-s) \left\{ \ln \Gamma \left[ \frac{(m+1)}{2} \right] - \ln \left[ \Gamma \left( \frac{m}{2} \right) \right] - \frac{1}{2} \ln[(m-2)\pi] \right\} \\ &\quad - \sum_{t=s+1}^T \left[ \frac{m+1}{2} \ln \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right] \\ &= (T-s) \left\{ \ln \Gamma \left[ \frac{(m+1)}{2} \right] - \ln \left[ \Gamma \left( \frac{m}{2} \right) \right] - \frac{1}{2} \ln[(m-2)\pi] \right\} \\ &\quad + l(c_{s+1}, c_2, \dots, c_T | \alpha; m, c_1, c_2, \dots, c_s). \end{aligned} \quad (4.11)$$

Lastly, we may assume that  $\epsilon_t$  has a generalized error distribution (GED) and its pdf can be given by

$$f(\epsilon_t) = \frac{m \cdot \exp \left( -\frac{1}{2} \left| \frac{\epsilon_t}{\omega} \right|^m \right)}{\omega 2^{(1-\frac{1}{m})} \Gamma \left( \frac{1}{m} \right)}, \quad -\infty < \epsilon_t < \infty \text{ and } 0 < m \leq \infty, \quad (4.12)$$

where  $\omega = \left[ 2^{(-\frac{2}{m})} \frac{\Gamma(\frac{1}{m})}{\Gamma(\frac{3}{m})} \right]^{\frac{1}{2}}$  and  $\Gamma(\cdot)$  is the usual gamma function ( $\Gamma(y) = \int_0^\infty (x^{y-1} e^{-x}) dx$ ).

Then the likelihood function  $f(c_{s+1}, c_2, \dots, c_T | \alpha; m, c_1, c_2, \dots, c_s)$  and log-likelihood function  $l(c_{s+1}, c_2, \dots, c_T | \alpha; m, c_1, c_2, \dots, c_s)$  can be easily obtained by following the same procedure as under the assumption of normality and Student- $t$  distribution. For the tail-thickness parameter  $m = 2$  and the pdf equals to the one for a standard normal distribution and it has heavy tails when  $m < 2$  (Bollerslev et al., 1994).

#### 4.2.1.3 An ARCH(s) Model Diagnostic

The standardized residuals for an ARCH model are given by

$$\hat{c}_t = \frac{c_t}{\sigma_t}, \quad (4.13)$$

which forms a sequence of independent and identical distributed random variables. Therefore, the adequacy and accuracy of the fitted ARCH model can be checked by investigating the series  $\{\hat{c}_t\}$ . Certainly, the Ljung-Box statistical tests of  $\hat{c}_t$  and  $\hat{c}_t^2$  are usually used to check for the mean equation



adequacy and the volatility equation validation, respectively. The model distribution validation can be checked using the skewness, kurtosis, and quantile-to-quantile plot (qq-plot) of  $\{\hat{c}_t\}$  (Tsay, 2005).

#### 4.2.1.4 Forecasting with an ARCH(s) Model

Consider the series  $c_1, c_2, \dots, c_{T_0}$  and let  $c_{T_0}(h)$  for  $h = T_0 + 1, T_0 + 2, \dots$ , denote the  $h$ -steps ahead forecast from the origin  $T_0$ . If we let  $f(c)$  be a function of the observed series, then  $c_{T_0}(h)$  can be considered as a minimum mean square error predictor that minimizes  $E[c_{T_0+1} - f(c)]^2$ . Since  $E[c_{T_0}(h)] = 0$  is not necessary the forecast a series  $c_t$ , we use  $c_t^2$  instead of the standard series  $c_t$  (Talke, 2003).

Now for an ARCH(s) model, the 1-step ahead forecast of  $\sigma_{T_0+1}^2$  from the origin  $T_0$  is

$$\sigma_{T_0}^2(1) = \alpha_0 + \alpha_1 c_{T_0}^2 + \dots + \alpha_s c_{T_0+1-s}^2. \quad (4.14)$$

The 2-step ahead forecast is

$$\sigma_{T_0}^2(2) = \alpha_0 + \alpha_1 c_{T_0}^2 + \alpha_2 c_{T_0}^2 + \dots + \alpha_s c_{T_0+2-s}^2. \quad (4.15)$$

In general, the  $h$ -step ahead forecast for  $\sigma_{T_0+h}^2$  is

$$\begin{aligned} \sigma_{T_0}^2(h) &= \alpha_0 + \alpha_1 c_{T_0}^2 + \alpha_2 c_{T_0}^2 + \dots + \alpha_s c_{T_0+h-s}^2, \\ &= \alpha_0 + \sum_{i=1}^s \alpha_i c_{T_0}^2 (h - i), \end{aligned} \quad (4.16)$$

where  $c_{T_0}^2(h - i) = c_{T_0+h-s}^2$  if  $(h - i) \leq 0$  (Tsay, 2005).

#### 4.2.2 An ARCH (1) Model Overview

An ARCH (1) model is the special case of the general ARCH(s) model introduced in section 4.2 where  $s=1$ . Without the loss of generality, suppose  $c_t$  is a random variable that has a mean and a variance conditionally on the information  $\mathbb{T}_{t-1}$  ( $\sigma$ -field generated by  $c_{t-1}$ ). As in ARCH(s) except that  $s=1$ ,  $c_t$  is an ARCH (1) process if it is stationary and if it satisfies for all  $t$  and some strictly positive-valued process  $\sigma$ , the following two equations:

$$c_t = \sigma_t \epsilon_t = \sqrt{\sigma_t^2} \epsilon_t, \quad (4.17)$$

where  $\epsilon_t \sim N(1,0)$  and,

$$\sigma_t^2 = \alpha_0 + \alpha_1 c_{t-1}^2, \quad (4.18)$$

where  $\alpha_0 > 0$  and  $0 \leq \alpha_1 < 1$ , to ensure that the conditional variance is positive.

All the properties presented in section 4.2 for the general ARCH(s) models also hold for the ARCH (1) models; however, formulae become more complex for higher-order ARCH models. So below we use an ARCH (1) model to prove them.

From equation (4.18), an ARCH (1) model can also be presented by,

$$c_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 c_{t-1}^2}, \quad (4.19)$$

then to show the

1. Conditional mean:

$$E(c_t | \mathcal{T}_{t-1}) = E\left(\epsilon_t \sqrt{\alpha_0 + \alpha_1 c_{t-1}^2} \mid \mathcal{T}_{t-1}\right) = E(\epsilon_t | \mathcal{T}_{t-1}) \sqrt{\alpha_0 + \alpha_1 c_{t-1}^2} = 0.$$

Unconditional mean:

$$E(c_t) = E\left(\epsilon_t \sqrt{\alpha_0 + \alpha_1 c_{t-1}^2}\right) = E(\epsilon_t) E\left(\sqrt{\alpha_0 + \alpha_1 c_{t-1}^2}\right) = 0.$$

2. Conditional Variance:

$$\begin{aligned} \text{Var}(c_t | \mathcal{T}_{t-1}) &= E(c_t^2 | \mathcal{T}_{t-1}) - [E(c_t | \mathcal{T}_{t-1})]^2 \\ &= E(\epsilon_t^2 | \mathcal{T}_{t-1}) (\alpha_0 + \alpha_1 c_{t-1}^2) - 0 \\ &= \alpha_0 + \alpha_1 c_{t-1}^2 \end{aligned}$$

Unconditional Variance (proof using the law of total variance):

$$\begin{aligned} \text{Var}(c_t) &= E[\text{Var}(c_t | \mathcal{T}_{t-1})] + \text{Var}[E(c_t | \mathcal{T}_{t-1})] \\ &= E(\alpha_0 + \alpha_1 c_{t-1}^2) + \text{Var}(0) \\ &= \alpha_0 + \alpha_1 E(c_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \text{Var}(c_{t-1}) \end{aligned}$$

Since  $c_t$  is stationary,  $\text{Var}(c_t) = \text{Var}(c_{t-1})$ .

Therefore,  $\text{Var}(c_t) = \frac{\alpha_0}{1-\alpha_1}$ .

3. Leptokurtic unconditional distribution:

$$\text{Kurt}(c_t) = \frac{E(c_t^4)}{[E(c_t^2)]^2} = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3, \text{ such that}$$

$$E(c_t^4) = E[E(c_t^4 | \mathcal{T}_{t-1})] = E[\alpha_t^4 E(\epsilon_t^4 | \mathcal{T}_{t-1})]$$

$$\begin{aligned}
&= E(\epsilon_t^4)E(\alpha_0 + \alpha_1 c_{t-1}^2)^2 \\
&= 3[\alpha_0 + 2\alpha_0\alpha_1 E(c_{t-1}^2) + \alpha_1^2 E(c_{t-1}^4)]
\end{aligned}$$

$$\text{So } E(c_t^4) = 3 \frac{\alpha_0^2(1+\alpha_1)}{(1+\alpha_1)(1-3\alpha_1^2)} \text{ and } E(c_t^2) = \frac{\alpha_0}{1-\alpha_1}$$

$$\text{Hence } Kurt(c_t) = \frac{E(c_t^4)}{[Var(c_t)]^2} = \frac{E(c_t^4)}{[E(c_t^2)]^2} = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Therefore, for an ARCH (1) process, we must ensure  $0 \leq \alpha_1 < \frac{1}{3}$  for the  $E(c_t^4)$  to exist and the kurtosis of  $c_t$  will be always greater than 3. This shows that the excess kurtosis of  $c_t$  is positive, that is, its curvature is high in the middle of the distribution and the tails of  $c_t$  are fatter or heavier than those of a normal distribution. This means that the shock  $c_t$  of a conditional Gaussian ARCH (1) model is more likely than a Gaussian white noise series to yield “outliers” (Tsay, 2005).

#### 4.2.2.1 Parameter Estimation for an ARCH (1) Model

Now, we use ARCH (1) model to derive parameters since derivation using ARCH (1) is easy because there are few parameters to be estimated than in the generalized ARCH (s) model discussed in subsection 4.2.1.2. Under the assumption of normality, the maximum likelihood estimation method can be used to estimate the parameters  $\alpha_0$  and  $\alpha_1$  for an ARCH (1) model and its likelihood function is given by

$$\begin{aligned}
f(c_1, c_2, \dots, c_T | \alpha) &= f(c_T | \mathcal{T}_{T-1}) f(c_{T-1} | \mathcal{T}_{T-2}) \dots f(c_2 | \mathcal{T}_1) f(c_1 | \alpha) \\
&= \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right) \times f(c_1 | \alpha),
\end{aligned} \tag{4.20}$$

where  $\alpha = (\alpha_0, \alpha_1)'$  and  $f(c_1 | \alpha)$  is the joint probability density function (pdf) of  $c_1, c_2$ . The function  $f(c_1 | \alpha)$  is usually deleted from the above likelihood function when the sample size is large enough. This is done simply because it causes the likelihood function to be more complicated to evaluate. Therefore, this results in using the conditional likelihood function

$$f(c_2, \dots, c_T | \alpha; c_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right), \tag{4.21}$$

where  $\sigma_t^2$  can be evaluated recursively. Since maximizing the conditional function is equivalent to maximizing its logarithm, for simplicity we will then use the logarithm of the conditional likelihood to find the maximum likelihood estimates (MLEs) of an ARCH (1) under normality assumption and it is given by

$$l(\alpha|c_1) = \ln f(c_2, \dots, c_T | \alpha; c_1) = \sum_{t=2}^T \left( -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{1}{2} \frac{c_t^2}{\sigma_t^2} \right) \quad (4.22)$$

By eliminating the term  $(\ln(2\pi))$  that does not involve any parameters, the log-likelihood can be simplified to

$$l(\alpha|c_1) = \ln f(c_2, \dots, c_T | \alpha; c_1) = -\frac{1}{2} \sum_{t=2}^T \left( \ln \sigma_t^2 + \frac{c_t^2}{\sigma_t^2} \right), \quad (4.23)$$

where  $\sigma_t^2 = \alpha_0 + \alpha_1 c_{t-1}^2$  can be evaluated recursively.

Under the standardized Student- $t$  distribution assumption, it is normal to assume that  $\epsilon_t$  follows a heavy-tailed distribution. Suppose  $y_m$  has a Student- $t$  distribution with  $m$  degrees of freedom. Then  $\text{Var}(y_m) = \frac{m}{(m-2)}$  for  $m > 2$ , and therefore  $\epsilon_t = \frac{y_m}{\sqrt{\frac{m}{(m-2)}}}$ . Thus, the pdf of  $\epsilon_t$  is given by

$$f(\epsilon_t|m) = \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left(\frac{m}{2}\right)(\sqrt{(m-2)}\pi)^{\frac{1}{2}}\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \quad (4.24)$$

The conditional likelihood of  $c_t^2$  is obtained using the equation  $c_t = \sigma_t \epsilon_t$  and can be written as

$$f(c_2, \dots, c_T | \alpha; m, c_1) = \prod_{t=2}^T \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left(\frac{m}{2}\right)(\sqrt{(m-2)}\pi)^{\frac{1}{2}}\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \quad (4.25)$$

Now, again we use the log-likelihood function to find the maximum likelihood estimates (MLEs) as we did under the normality assumption. Then the likelihood function becomes

$$l(\alpha|c_1) = \ln f(c_2, \dots, c_T | \alpha; m, c_1) = -\sum_{t=2}^T \left[ \frac{m+1}{2} \ln \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right]. \quad (4.26)$$

The log-likelihood function to estimate  $m$  jointly with other parameters is given by

$$\begin{aligned} & \ln f(c_2, \dots, c_T | \alpha; m, c_1) \\ &= (T-1) \left\{ \ln \Gamma\left[\frac{(m+1)}{2}\right] - \ln \left[ \Gamma\left(\frac{m}{2}\right) \right] - \frac{1}{2} \ln[(m-2)\pi] \right\} \\ & \quad - \sum_{t=2}^T \left[ \frac{m+1}{2} \ln \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right] \\ &= (T-1) \left\{ \ln \Gamma\left[\frac{(m+1)}{2}\right] - \ln \left[ \Gamma\left(\frac{m}{2}\right) \right] - \frac{1}{2} \ln[(m-2)\pi] \right\} \\ & \quad + \ln f(c_2, \dots, c_T | \alpha; c_1). \end{aligned} \quad (4.27)$$

Lastly, we may assume that  $\epsilon_t$  has a generalized error distribution (GED) and its pdf can be given by

$$f(y) = \frac{m \cdot \exp\left(-\frac{1}{2} \left| \frac{y}{\omega} \right|^m\right)}{\omega 2^{(1-\frac{1}{m})} \Gamma\left(\frac{1}{m}\right)}, -\infty < y < \infty \text{ and } 0 < m \leq \infty, \quad (4.28)$$

where  $\omega = \left[ 2^{(-\frac{2}{m})} \frac{\Gamma(\frac{1}{m})}{\Gamma(\frac{3}{m})} \right]^{\frac{1}{2}}$  and  $\Gamma(\cdot)$  is the usual gamma function ( $\Gamma(y) = \int_0^\infty (x^{y-1} e^{-x}) dx$ ).

Then the likelihood function  $f(c_2, \dots, c_T | \alpha; m, c_1)$  and log-likelihood function  $l(c_2, \dots, c_T | \alpha; m, c_1)$  can be easily obtained by following the same procedure as we did under normality and Student- $t$  distribution assumptions. For the tail-thickness parameter  $m = 2$  then the pdf equals the one for the standard normal distribution and it has heavy tails when  $m < 2$  (Bollerslev et al., 1994).

#### 4.2.2.2 Forecasting with the ARCH (1) Model

Consider the series  $c_1, c_2, \dots, c_{T_0}$  and let  $c_{T_0}(h)$  for  $h = T_0 + 1, T_0 + 2, \dots$ , denote the  $h$ -steps ahead forecast from the origin  $T_0$ . If we let  $f(c)$  be a function of the observed series, then  $c_{T_0}(h)$  can be considered as a minimum mean square error predictor that minimizes  $E[c_{T_0+1} - f(c)]^2$ . Now for an ARCH( $s$ ) model, the 1-step ahead forecast  $\sigma_{T_0+1}^2$  from the origin  $T_0$  is

$$\sigma_{T_0}^2(1) = \alpha_0 + \alpha_1 c_{T_0}^2 \quad (4.29)$$

The 2-step ahead forecast  $\sigma_{T_0+2}^2$  is

$$\begin{aligned} \sigma_{T_0}^2(2) &= \alpha_0 + \alpha_1 \sigma_{T_0}^2(1). \\ &= \alpha_0 + \alpha_1 (\alpha_0 + \alpha_1 c_{T_0}^2) \\ &= \alpha_0 + \alpha_1 \alpha_0 + \alpha_1^2 c_{T_0}^2 \end{aligned} \quad (4.30)$$

In general, the  $h$ -step ahead forecast  $\sigma_{T_0+h}^2$  is

$$\begin{aligned} \sigma_{T_0}^2(h) &= \alpha_0 + \alpha_1 c_{T_0}^2 (\sigma_{T_0}^2(h-1)) \\ &= \alpha_0 \left( \sum_{i=0}^{h-1} \alpha_1^i \right) + \alpha_1^h c_{T_0}^2 \end{aligned} \quad (4.31)$$

(Tsay, 2005).

#### 4.2.3 ARCH Models Limitations

Apart from the flexibility and properties of the ARCH models discussed in the previous subsection, the model has some weaknesses too.

- The model assumes that positive and negative shocks have the same effects on volatility since it depends on the square of the previous shocks. However, it is known that the price of a financial asset reacts differently to positive and negative shocks.
- The ARCH model is rather restrictive. For example,  $\alpha_1^2$  of an ARCH (1) model must be in the interval  $\left[0, \frac{1}{3}\right]$  if the series has a finite fourth moment. The constraint tends to be more complex for higher-order ARCH models. In fact, it limits the ability of an ARCH model with normally distributed innovations to capture excess kurtosis.
- The ARCH models only provide a mechanical way to describe the behaviour of the conditional variance. They do not describe or introduce any new source demonstrating variations of a financial time series.
- ARCH models are likely to over-estimate the volatility because they react slowly to largely isolated shocks to the stock returns series. (Tsay, 2005, p.106).

### 4.3 Exponentially Weighted Moving Average (EWMA)

Before introducing the Generalized ARCH model, we first discuss one of the volatility models, the EWMA. The EWMA forecasts are used to model the variances (Guo, 2012). This class of models can be written as

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) c_{t-1}^2, \quad (4.32)$$

where  $\lambda$  is a constant ( $0 < \lambda < 1$ ),  $\sigma_t^2$  is the squared volatility (the variance rate) on day  $t$  and the volatility of a market variable on day  $t$ , estimated at the end of day  $t-1$  is given by  $\sigma_t$  (Guo, 2012). The equation (4.32) is known as a volatility formula and it has the flexibility to adapt to market changes because of its functionality form (Nasstrom, 2003). The estimate,  $\sigma_t$ , of the volatility for day  $t$  is calculated from  $\sigma_{t-1}$  and  $c_{t-1}$ , and large values of  $c_{t-1}^2$  will directly affect the volatility estimate (Hull, 2003). The EWMA approach to estimate the volatility works by substituting for  $\sigma_{t-1}^2$  in equation (4.32) and keep doing so for  $n$ -steps. Thus,

substituting for  $\sigma_{t-1}^2$  to get

$$\begin{aligned} \sigma_t^2 &= \lambda[\lambda \sigma_{t-2}^2 + (1 - \lambda) c_{t-2}^2] + (1 - \lambda) c_{t-1}^2, \\ \sigma_t^2 &= (1 - \lambda)[c_{t-1}^2 + \lambda c_{t-2}^2] + \lambda^2 \sigma_{t-2}^2, \end{aligned} \quad (4.33)$$

substituting for  $\sigma_{t-2}^2$  gives

$$\sigma_t^2 = (1 - \lambda)[c_{t-1}^2 + \lambda^2 c_{t-2}^2 + \lambda^3 c_{t-3}^2] + \lambda^3 \sigma_{t-3}^2, \quad (4.34)$$

continuing in the same manner, the general equation is

$$\sigma_t^2 = (1 - \lambda) \sum_{i=1}^n \lambda^{i-1} c_{t-i}^2 + \lambda^n \sigma_{t-n}^2. \quad (4.35)$$

For large  $n$ , the term  $\lambda^n \sigma_{t-n}^2$  is significantly small and if is ignored, equation (4.35) can be simplified to

$$\sigma_t^2 = \sum_{i=1}^n \alpha_i c_{t-i}^2 \quad (4.36)$$

where  $\alpha_i = (1 - \lambda)\lambda^{i-1}$  are the weights for  $c_i^2$ 's and decrease exponentially at rate  $\lambda$  as they move back through time and each weight is  $\lambda$  times the previous weight (Guo, 2012).

#### 4.4 The Generalized ARCH (GARCH) Model

In the GARCH models, the squared volatility  $\sigma_t^2$  depends on the previous squared volatilities as well as previous squared values of the model, therefore in that sense they are generalized ARCH models. Thus, the ARCH model has been replaced by GARCH models which were developed by Bollerslev (1986) and Taylor (1986) independently of each other (Terasvirta, 2006, p.4).

##### 4.4.1 A GARCH ( $r, s$ ) Model

Let  $c_t$  be a random variable that has a mean and a variance conditionally on the information  $\mathbb{T}_{t-1}$  ( $\sigma$ -field generated by  $c_{t-i}, i \geq 1$ ). Then the random variable  $c_t$  is a GARCH ( $r, s$ ) process if it is stationary and if it satisfies for all  $t$  and some strictly positive-valued process  $\sigma$ , the following two equations:

$$c_t = \sigma_t \epsilon_t = \sqrt{\sigma_t^2} \epsilon_t, \quad (4.37)$$

where  $\epsilon_t \sim \text{IIDN}(1,0)$  and,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i c_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (4.38)$$

with  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, 2, \dots, r-1, r$ ,  $\beta_j \geq 0, j = 1, 2, \dots, s-1, s$  to ensure that the conditional variance is positive. Note that the condition for  $\{\epsilon_t\}$  to be stationary is:  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_j) < 1$ , where  $\alpha_i = 0$  for  $i > r$  and  $\beta_j = 0$  for  $j > s$ . If  $s = 0$  then the GARCH ( $r, s$ ) model equals to ARCH( $s$ ) model. The constraint on  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_j)$  also ensures that the unconditional variance of  $c_t$  is finite, while its conditional variance  $\sigma_t^2$  changes over time (Matei, 2009, p. 52). The series  $\{\epsilon_t\}$  can have other white noise distributions. There is no need for it to be Gaussian as in the ordinal least squares regression.

#### 4.4.1.1 Basic Properties of an GARCH Model

Similarly, GARCH models have their properties as do ARCH models. Below we show and/or prove some of the most important GARCH model properties.

- Uniqueness and Stationarity:

The GARCH  $(r, s)$  model in equation (4.38) has a unique and stationary solution if

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1, \quad (4.39)$$

(Straumann, 2005).

- Zero Mean

Let the set of information at time  $t - 1$  be

$$\mathbb{T}_{t-1} = \sigma\{c_t, -\infty < i \leq t - 1\}.$$

The mean of  $c_t$  is zero in any model where  $\sigma_t$  is quantifiable concerning  $\mathbb{T}_{t-1}$  (which is the case in the GARCH (1,1) model).

$$\begin{aligned} E(c_t) &= E(\sigma_t \epsilon_t) \\ &= E[E(\sigma_t \epsilon_t | \mathbb{T}_{t-1})] \\ &= E[\sigma_t E(\epsilon_t | \mathbb{T}_{t-1})] \\ &= E[\sigma_t E(\epsilon_t)] \\ &= E[\sigma_t \cdot 0] \\ &= 0. \end{aligned} \quad (4.40)$$

- Lack of Serial Correlation

In the same manner as the zero-mean above, we can simply show that  $c_t$  is uncorrelated to  $c_{t+h}$  for  $h > 0$ :

$$\begin{aligned} E(c_t c_{t+h}) &= E(c_t \sigma_{t+h} \epsilon_{t+h}) \\ &= E[E(c_t \sigma_{t+h} \epsilon_{t+h} | \mathbb{T}_{t+h-1})] \\ &= E[c_t \sigma_{t+h} E(\epsilon_{t+h} | \mathbb{T}_{t+h-1})] \\ &= 0. \end{aligned} \quad (4.41)$$

- Unconditional Variance

We need an alternative representation of  $c_t^2$  which will enable us to compute  $E(c_t^2)$ . Now we define the sequence

$$\varepsilon_t = c_t^2 - \sigma_t^2 \quad (4.42)$$



$$= \sigma_t^2(\epsilon_t^2 - 1).$$

We can further show that  $\epsilon_t$  has a mean zero, hence it can be treated as a “white noise”. Now

$$\begin{aligned} c_t^2 &= \sigma_t^2 + \epsilon_t \\ &= \alpha_0 + \sum_{i=1}^r \alpha_i c_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 + \epsilon_t \\ &= \alpha_0 + \sum_{i=1}^r \alpha_i c_{t-i}^2 + \sum_{j=1}^s \beta_j c_{t-j}^2 - \sum_{j=1}^s \beta_j \epsilon_{t-j} + \epsilon_t. \end{aligned} \quad (4.43)$$

If we let  $N = \max(r, s)$ ,  $\alpha_i$  and  $\beta_j$  for  $i > r$  and  $j > s$  respectively, then equation above can be written as

$$c_t^2 = \alpha_0 + \sum_{i=1}^N (\alpha_i + \beta_i) c_{t-i}^2 - \sum_{j=1}^s \beta_j \epsilon_{t-j} + \epsilon_t. \quad (4.44)$$

Equation (4.44) above shows that  $c_t^2$  is an ARMA model or process with martingale difference innovations.

By stationarity assumption (i.e.,  $E(c_t^2) = E(c_{t+h}^2)$ ), the unconditional variance can be obtained as

$$\begin{aligned} E(c_t^2) &= \alpha_0 + \sum_{i=1}^N (\alpha_i + \beta_i) E(c_{t-i}^2) - \sum_{j=1}^s \beta_j E(\epsilon_{t-j}) + E(\epsilon_t) \\ &= \alpha_0 + \sum_{i=1}^N (\alpha_i + \beta_i) E(c_{t-i}^2) \\ &= \alpha_0 + \sum_{i=1}^N (\alpha_i + \beta_i) E(c_t^2) \end{aligned}$$

$$\left[1 - \sum_{i=1}^N (\alpha_i + \beta_i)\right] E(c_t^2) = \alpha_0$$

which gives

$$E(c_t^2) = \frac{\alpha_0}{[1 - \sum_{i=1}^N (\alpha_i + \beta_i)]}. \quad (4.45)$$

This result also shows the importance of property number one above.

- Heavy-tails of  $c_t^2$

GARCH (1,1) model is the special case of the GARCH ( $r, s$ ) model where  $r=s=1$  and this model can easily be heavy-tailed. Therefore, for simplicity, we use a GARCH (1,1) model to show or prove the property for  $c_t^2$  being heavy-tailed. The proof is shown under GARCH (1,1) model in sub-section 4.4.2 (Fryzlewicz, 2007, p.5-7).

#### 4.4.1.2 Parameter Estimation for the GARCH ( $r, s$ ) Model

Maximum likelihood estimation has become a very popular method for GARCH model parameter estimation (Princ et al., 2012, p.82). This method involves maximizing the likelihood function that is formulated under normality assumption of error and suitable regularity conditions; maximum likelihood estimates are consistent and asymptotically normal distributed (Princ et al., 2012, p.82). Thus, the maximum likelihood estimations (MLEs) method can be used to estimate the parameters of the GARCH ( $r, s$ ) model, where those parameters are  $\alpha_0 > 0, \alpha_i, i = 1, 2, \dots, r-1, r, \beta_j, j = 1, 2, \dots, s-1, s$ . Under the assumption of normality, the likelihood function can be written as

$$\begin{aligned} f(c_1, c_2, \dots, c_T, \sigma_1^2, \sigma_2^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}) &= f(c_T | \mathbb{T}_{T-1}) f(c_{T-1}, \sigma_{T-1}^2 | \mathbb{T}_{T-2}) \dots f(c_2, \sigma_1^2 | \mathbb{T}_1) f(c_1, \sigma_1^2 | \boldsymbol{\varphi}) \\ &= \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right) \times f(c_1, \sigma_1^2 | \boldsymbol{\varphi}), \end{aligned} \quad (4.46)$$

where  $\boldsymbol{\varphi} = (\alpha_0, \alpha_1, \dots, \alpha_r; \beta_1, \beta_2, \dots, \beta_s)'$  and the set of information at time  $t-1$   $\mathbb{T}_{t-1} = \{c_1, c_2, \dots, c_{t-1}; \sigma_1^2, \sigma_2^2, \dots, \sigma_{t-1}^2\}$ . It is too complicated to evaluate the exact pdf  $f(c_1, \sigma_1^2 | \boldsymbol{\varphi})$  and therefore to estimate  $\boldsymbol{\varphi}$ , it is usually convenient to drop it from the likelihood function and then use the conditional likelihood of the form

$$\begin{aligned} f(c_{s+1}, \dots, c_T, \sigma_{r+1}^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}; c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) \\ &= f(c_T, \sigma_T^2 | \mathbb{T}_{T-1}) \dots f(c_{s+1}, \sigma_{r+1}^2 | \boldsymbol{\varphi}; c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) \\ &= \prod_{t=\max(r,s)}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right). \end{aligned} \quad (4.47)$$

As in the ARCH( $s$ ) model, the maximum likelihood is equivalent to maximizing its logarithm. Thus, the following conditional log-likelihood can be used for parameter estimation

$$\begin{aligned} l(\boldsymbol{\varphi} | c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) &= \ln f(c_{s+1}, \dots, c_T, \sigma_{r+1}^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}; c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) \\ &= \ln f(c_T, \sigma_T^2 | \mathbb{T}_{T-1}) \dots \ln f(c_{s+1}, \sigma_{r+1}^2 | \boldsymbol{\varphi}; c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) \\ &= -\frac{1}{2} \sum_{t=\max(r,s)}^T \left( \ln \sigma_t^2 + \ln 2\pi + \frac{c_t^2}{\sigma_t^2} \right). \end{aligned} \quad (4.48)$$

Then removing the term  $\ln 2\pi$  since it does not involve any parameter, the conditional log-likelihood becomes

$$l(\boldsymbol{\varphi}|c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) = -\frac{1}{2} \sum_{t=\max(r,s)}^T (\ln \sigma_t^2 + \frac{c_t^2}{\sigma_t^2}), \quad (4.49)$$

(Talke, 2003, p.30).

The parameter  $\boldsymbol{\varphi}$  is solved recursively using either the Newton-Raphson and/or Fisher Scoring method.

Under the standardized Student- $t$  distribution assumption, we assume that  $\epsilon_t$  follows a heavy-tailed distribution. Suppose  $y_m$  has a Student- $t$  distribution with  $m$  degrees of freedom. Then  $Var(y_m) = \frac{m}{(m-2)}$  for  $m > 2$ , and therefore  $\epsilon_t = \frac{y_m}{\sqrt{\frac{m}{(m-2)}}}$ . Thus, the pdf of  $\epsilon_t$  is given by

$$f(\epsilon_t|\boldsymbol{\varphi}, m) = \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left[\frac{m}{2}\right] (\sqrt{(m-2)\pi})} \frac{1}{\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \quad (4.50)$$

The conditional likelihood of  $c_t^2$  is obtained using the equation  $c_t = \sigma_t \epsilon_t$  and can be written as

$$\begin{aligned} f(c_1, c_2, \dots, c_T, \sigma_1^2, \sigma_2^2, \dots, \sigma_T^2|\boldsymbol{\varphi}, m) \\ = \prod_{t=2}^T \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left[\frac{m}{2}\right] (\sqrt{(m-2)\pi})} \frac{1}{\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \end{aligned} \quad (4.51)$$

Now, again we use the log-likelihood function to find the maximum likelihood estimates (MLEs) as we did under the normality assumption. Then the likelihood function becomes

$$\begin{aligned} l(\boldsymbol{\varphi}, m|c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) &= l(\boldsymbol{\varphi}, m) = \ln f(c_t|\mathbb{T}_{T-1}) \\ l(\boldsymbol{\varphi}, m|c_1, \dots, c_s, \sigma_1^2, \dots, \sigma_r^2) &= l(\boldsymbol{\varphi}, m) \\ &= - \sum_{t=\max(r,s)}^T \left[ \frac{m+1}{2} \ln \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right) + \frac{1}{2} \ln(\sigma_t^2) \right] \end{aligned} \quad (4.52)$$

Lastly, we may assume that  $\epsilon_t$  has a generalized error distribution (GED) and its pdf can be given by

$$f(c_1|\boldsymbol{\varphi}, m) = \frac{m \cdot \exp\left(-\frac{1}{2} \left|\frac{c_t}{\omega \cdot \sigma_t}\right|^m\right)}{\omega \cdot \sigma_t 2^{(1-\frac{1}{m})} \Gamma\left(\frac{1}{m}\right)}, 0 < m \leq \infty, \quad (4.53)$$

where  $\omega = [2^{(-\frac{2}{m})} \frac{\Gamma(\frac{1}{m})}{\Gamma(\frac{3}{m})}]^{\frac{1}{2}}$  and  $\Gamma(\cdot)$  is the usual gamma function ( $\Gamma(y) = \int_0^\infty (x^{y-1} e^{-x}) dx$ ).

Then log-likelihood function

$$l(\boldsymbol{\varphi}|m) = \ln(m) - \left(-\frac{1}{2} \left| \frac{c_t}{\omega \cdot \sigma_t} \right|^m\right) - \ln \left[ \omega 2^{\left(1-\frac{1}{m}\right)} \right] - \ln \Gamma \left( \frac{1}{m} \right) \quad (4.54)$$

For the tail-thickness parameter  $m = 2$  and the pdf equals to the one for a standard normal distribution and it has heavy tail distribution when  $m < 2$  (Bollerslev et al., 1994).

#### 4.4.1.3 Forecasting with the GARCH (r, s) Model

The capacity of forecasting with GARCH models has been broadly discussed by Poon et al., (2003). The forecast is conducted by taking a conditional expectation. For GARCH (r, s) model, the forecast one-step ahead from the origin  $T_0$  is given by

$$c_{T_0+1}^2 = \alpha_0 + \sum_{i=1, j=1}^{\max(r,s)} (\alpha_i + \beta_j) E[c_{T_0+1-i}^2 | \mathcal{T}_{T_0}] - \sum_{j=1}^s \beta_j E[v_{T_0+1-j} | \mathcal{T}_t], \quad (4.55)$$

where  $c_{T_0}^2, \dots, c_{T_0+1-\max(r,s)}^2$  and  $\sigma_{T_0}^2, \dots, \sigma_{T_0+1-r}^2$  are assumed to be known as origin time  $T_0$  (Talke, 2003).

In general, the  $h$ -step ahead forecast from the time origin  $T_0$  can be written as

$$c_{T_0+h}^2 = \alpha_0 + \sum_{i=1, j=1}^{\max(r,s)} (\alpha_i + \beta_j) E[c_{T_0+h-i}^2 | \mathcal{T}_{T_0}] - \sum_{j=1}^s \beta_j E[v_{T_0+h-j} | \mathcal{T}_t], \quad (4.56)$$

where  $E[c_{T_0+h-i}^2 | \mathcal{T}_{T_0}]$  is specified recursively by equation (4.56) for  $i < h$ . If  $i \geq h$  then  $E[c_{T_0+h-i}^2 | \mathcal{T}_{T_0}] = c_{T_0+h-i}^2$  and  $E[v_{T_0+h-i} | \mathcal{T}_{T_0}] = v_{T_0+h-i}$ , and  $E[c_{T_0+h-i}^2 | \mathcal{T}_{T_0}] = 0$  for  $i < h$  (Talke, 2003, p.30 & 31).

#### 4.4.2 A GARCH (1,1) Model

It is easy to describe the general GARCH model and its properties in detail using the reduced model where the orders of  $r$  and  $s$  are specified. Suppose  $r=s=1$  from the GARCH (r, s) in equation (4.38) then a random variable  $c_t$  has a GARCH (1,1) model if

$$\begin{aligned} \epsilon_t | \mathcal{T}_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 c_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \quad (4.57)$$

where  $\alpha_0 > 0$ ,  $\alpha_1, \beta_1 \geq 0$  and  $\alpha_1 + \beta_1 < 1$  are imposed to ensure that the conditional variance  $\sigma_t^2$  is positive (Nasstrom, 2003, p.23). The structure of equation (4.57) shows that a huge value of  $c_{t-1}^2$  or  $\sigma_{t-1}^2$  leads to a huge value for a squared volatility  $\sigma_t^2$ . Thus, the huge value of  $c_{t-1}^2$  tends to be followed by another huge value of  $c_t^2$ . This is known as volatility clustering (Tsay, 2005, p.114).

#### 4.4.2.1 Properties of a GARCH (1,1) Model

- The moments of  $c_t$  are one of its interesting properties. To show them, consider the GARCH (1,1) model in equation (4.57), then a necessary and sufficient condition for existence of the  $2k^{th}$  moment is

$$\mu(\alpha_1, \beta_1, k) = \sum_{j=0}^k \binom{k}{j} d_j \alpha_1^{k-j} < 1 \quad (4.58)$$

where  $\alpha_0 = 1$ ,  $d_j = \prod_{i=1}^j (2i - 1)$ .

Therefore the  $2k^{th}$  moment can be described by the recursive formula

$$E(c_t^{2k}) = d_k \cdot \left[ \sum_{l=0}^{k-1} d_l^{-1} E(c_t^{2l}) \alpha_0^{k-l} \binom{k}{k-l} \mu(\alpha_1, \beta_1, l) \right] \cdot [1 - \mu(\alpha_1, \beta_1, k)]^{-1}, \quad (4.59)$$

Because of the symmetry properties of the Gaussian distribution, it follows that if the  $2k^{th}$  moment does exist, then  $2(k-1)^{th}$  moment is equals to zero (Nasstrom, 2003, p.23).

- It can also be proved that if  $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ , then

$$\frac{E(c_t^4)}{[E(c_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3. \quad (4.60)$$

Thus, like ARCH models, a GARCH (1,1) model has a heavy-tail distribution (i.e., its tail is heavier than that of a normal distribution) (Tsay, 2005, p.115).

To show that we begin by assuming the condition below

$$E(\alpha_1 c_t^2 + \beta_1)^{\frac{d}{2}} > 1 \quad (4.61)$$

for some  $d > 0$ . This condition is satisfied if  $\epsilon_t \sim N(0,1)$  (but of course not only in this case).

Then we have

$$\begin{aligned} \sigma_{t+1}^2 &= \alpha_0 + \alpha_1 c_t^2 + \beta_1 c_t^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) c_t^2, \end{aligned} \quad (4.62)$$

and using the independence of  $\epsilon_t$  and  $\mathbb{T}_{t-1}$  we get

$$\begin{aligned} E(\sigma_{t+1}^d) &= E[\alpha_0 + (\alpha_1 \epsilon_t^2 + \beta_1) \sigma_t^2]^{\frac{d}{2}} \\ &\geq E[(\alpha_1 \epsilon_t^2 + \beta_1) \sigma_t^2]^{\frac{d}{2}} \\ &= E(\alpha_1 \epsilon_t^2 + \beta_1)^{\frac{d}{2}} E(\sigma_t^d). \end{aligned} \quad (4.63)$$

If  $E(\sigma_t^d)$  were finite, then by stationarity, it would be equivalent to  $E(\sigma_{t+1}^d)$ . Then we would simplify it to

$$1 \geq E(\alpha_1 \epsilon_t^2 + \beta_1)^{\frac{d}{2}}, \quad (4.64)$$

which would contradict with the assumption in equation (4.46). Therefore,  $E(\sigma_t^d)$  is infinite, which implies that  $E(c_t^2)$  is infinite. Thus,  $c_t$  does not have all finite moments, hence it is heavy-tailed (Fryzlewicz, 2007, p.5-7).

- The GARCH (1,1) model offers a simple parametric function that can be used to define the volatility evolution (Tsay, 2005, p.115).

#### 4.4.2.2 The Kurtosis of GARCH (1,1) Model

The kurtosis of volatility is one of the important features when assessing the variability of the estimated volatility. This section simply shows the derivation for the excess kurtosis of a GARCH (1,1) model. Even though the derivation is done in GARCH (1,1) the same idea applies to all other GARCH models; this was done only for simplicity. Consider the GARCH (1,1) model given in equation (4.57) where  $\{\epsilon_t\}$  is an independent and identical distributed sequence satisfying the following

$$E(\epsilon_t) = 0, \quad (4.65)$$

$$Var(\epsilon_t) = 1, \quad (4.66)$$

$$E(\epsilon_t^4) = K_\epsilon + 3, \quad (4.67)$$

where  $K_\epsilon$  is the excess kurtosis of the innovation  $\epsilon_t$ . Based on the model assumption, we know the following properties:

- $Var(c_t) = E(c_t^2) = E(\sigma_t^2) = \frac{\alpha_0}{[1-(\alpha_1+\beta_1)]}$ , (see equation 4.45 for a general proof of this property).
- $E(c_t^4) = (K_\epsilon + 3)E(\sigma_t^4)$  given that  $E(\sigma_t^4)$  does exist.

Now by taking the square of the GARCH (1,1) volatility model in equation (4.57), we get

$$\begin{aligned} [\sigma_t^2]^2 &= [\alpha_0 + \alpha_1 c_{t-1}^2 + \beta_1 \sigma_{t-1}^2]^2 \\ \sigma_t^4 &= \alpha_0^2 + \alpha_1^2 c_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_0 \alpha_1 c_{t-1}^2 + 2\alpha_0 \beta_1 \sigma_{t-1}^2 + 2\alpha_1 \beta_1 \sigma_{t-1}^2 c_{t-1}^2. \end{aligned} \quad (4.68)$$

The expectation of the equation (4.68) using the above two properties we obtain

$$\sigma_t^4 = \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(K_\epsilon + 2) - (\alpha_1 + \beta_1)^2]}, \quad (4.69)$$

where  $0 \leq (\alpha_1 + \beta_1) < 1$  and  $[1 - \alpha_1^2(K_\epsilon + 2) - (\alpha_1 + \beta_1)^2] > 0$ . Therefore, if the excess kurtosis of  $c_t$  exists, it is then

$$\begin{aligned} K_c &= \frac{E(c_t^4)}{[E(c_t^2)]^2} - 3 \\ &= \frac{(K_\epsilon + 3)[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2} - 3. \end{aligned} \quad (4.70)$$

We can write this equation in a more compact way by considering the following two cases:

1. When  $\epsilon_t$  is normally distributed. In this case  $K_\epsilon = 0$ , with some algebra, it can be shown that the excess kurtosis of  $c_t$  is written as

$$K_c^{(g)} = \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2}, \quad (4.71)$$

where the superscript  $(g)$  denotes Gaussian distribution for  $\epsilon_t$ . This equation has two significant consequences:

- i. the kurtosis of  $c_t$  exists only if  $[1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2] > 0$ ,
  - ii. the corresponding GARCH (1,1) model does not have heavy tails if  $\alpha_1 = 0$  which implies that  $K_c^{(g)} = 0$ .
2. When  $\epsilon_t$  is not normally distributed. Using the previous result, we have

$$\begin{aligned} K_c &= \frac{K_\epsilon - K_\epsilon(\alpha_1 + \beta_1) + 6\alpha_1^2 + 3K_\epsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon[1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2] + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon + K_c^{(g)} + \frac{5}{6}K_\epsilon K_c^{(g)}}{1 - \frac{1}{6}K_\epsilon K_c^{(g)}}. \end{aligned} \quad (4.72)$$

This result holds for all GARCH models provided that their respective kurtosis does exist, and it was originally obtained by George C. Tiao; refer to Bai et al., (2003) (Tsay, 2005, p. 145-146).

Remarks:

- i. One can notice that if  $\beta_1 = 0$  from GARCH (1,1) model in equation (4.57), then the model reduces to an ARCH (1) model. In this case, if  $\epsilon_t$  is normally distributed, the kurtosis of  $c_t$  is given by

$$K_c^{(g)} = \frac{6\alpha_1^2}{(1 - 3\alpha_1^2)}, \quad (4.73)$$

and when  $\epsilon_t$  is not Gaussian distribution then

$$\begin{aligned} K_c &= \frac{(K_\epsilon + 3)(1 - \alpha_1^2)}{1 - (K_\epsilon + 3)\alpha_1^2} - 3 \\ &= \frac{K_\epsilon + 2K_\epsilon\alpha_1^2 + 6\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon(1 - 3\alpha_1^2) + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon + K_c^{(g)} + \frac{5}{6}K_\epsilon K_c^{(g)}}{1 - \frac{1}{6}K_\epsilon K_c^{(g)}}. \end{aligned} \quad (4.74)$$

- ii. From the results for the GARCH (1,1) model, the coefficient  $\alpha_1$  plays a pivotal role in determining the tail behaviour of  $c_t$ . Now it can be seen that if  $\alpha_1 = 0$ , then  $K_c^{(g)} = 0$  and  $K_c = K_\epsilon$ . In this regard, the tail behaviour of  $c_t$  imitates that of the standardized noise  $\epsilon_t$ . However, if  $\alpha_1 > 0$ , then  $K_c^{(g)} > 0$  and it implies that  $c_t$  has heavy tails.
- iii. When  $\epsilon_t$  has a standardized student- $t$  distribution with  $m$  degrees of freedom, then we get

$$E(\epsilon_t^4) = \frac{6}{(m - 4)} + 3, \text{ if } m > 4. \quad (4.75)$$

Thus, the excess kurtosis of  $\epsilon_t$  is given by

$$K_\epsilon = \frac{6}{(m - 4)}, \text{ for } m > 4. \quad (4.76)$$

and the excess kurtosis of  $c_t$  becomes



$$K_c = \frac{[6 + (m + 1)K_c^{(g)}]}{[m - 4 - K_c^{(g)}]}, \quad (4.77)$$

given that  $\frac{1-2\alpha_1^2(m-1)}{(m-4)-(\alpha_1+\beta_1)^2} > 0$ .

(Tsay, 2005, p.146-147).

#### 4.4.2.3 Parameter Estimation with GARCH (1,1) Model

The maximum likelihood estimations (MLEs) method can be used to estimate the parameters for GARCH (1, 1) model, where those parameters are  $\alpha_0$ ,  $\alpha_1$ , and  $\beta_1$ . The log-likelihood function for normally distributed standardized errors  $\{\epsilon_t\}$  is

$$\begin{aligned} f(c_1, c_2, \dots, c_T, \sigma_1^2, \sigma_2^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}) &= f(c_T | \mathbb{T}_{T-1}) f(c_{T-1}, \sigma_{T-1}^2 | \mathbb{T}_{T-2}) \dots f(c_2, \sigma_1^2 | \mathbb{T}_1) f(c_1, \sigma_1^2 | \boldsymbol{\varphi}) \\ &= \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right) \times f(c_1, \sigma_1^2 | \boldsymbol{\varphi}), \end{aligned} \quad (4.78)$$

where  $\boldsymbol{\varphi} = (\alpha_0, \alpha_1; \beta_1)'$ . It is too complicated to evaluate the exact pdf  $f(c_1, \sigma_1^2 | \boldsymbol{\varphi})$  and therefore to estimate  $\boldsymbol{\varphi}$ , it is usually convenient to drop it from the likelihood function and then use the conditional likelihood of the form

$$\begin{aligned} f(c_2, \dots, c_T, \sigma_2^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}; c_1, \sigma_1^2) &= f(c_T, \sigma_T^2 | \mathbb{T}_{T-1}) \dots f(c_2, \sigma_2^2 | \boldsymbol{\varphi}; c_1, \sigma_1^2) \\ &= \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{c_t^2}{2\sigma_t^2}\right). \end{aligned} \quad (4.79)$$

Since the maximum likelihood is equivalent to maximizing its logarithm, then conditional log-likelihood can be used for parameter estimation

$$\begin{aligned} l(\boldsymbol{\varphi} | c_1, \sigma_1^2) &= \ln f(c_2, \dots, c_T, \sigma_2^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}; c_1, \sigma_1^2) \\ &= \ln f(c_T, \sigma_T^2 | \mathbb{T}_{T-1}) \dots \ln f(c_2, \sigma_2^2 | \boldsymbol{\varphi}; c_1, \sigma_1^2) \\ &= -\frac{1}{2} \sum_{t=2}^T \left( \ln \sigma_t^2 + \ln 2\pi + \frac{c_t^2}{\sigma_t^2} \right). \end{aligned} \quad (4.80)$$

Then removing the term  $\ln 2\pi$  since it does not involve any parameter, the conditional log-likelihood becomes

$$l(\boldsymbol{\varphi} | c_1, \sigma_1^2) = -\frac{1}{2} \sum_{t=2}^T \left( \ln \sigma_t^2 + \frac{c_t^2}{\sigma_t^2} \right), \quad (4.81)$$

(Talke, 2003, p.23; Princ et al., 2012, p.82). The parameter  $\boldsymbol{\varphi}$  is solved recursively using either the Newton-Raphson and/or Fisher Scoring method.

Under the standardized Student- $t$  distribution assumption, we assume that  $\epsilon_t$  follows a heavy-tailed distribution. Suppose  $y_m$  has a Student- $t$  distribution with  $m$  degrees of freedom. Then  $Var(y_m) = \frac{m}{(m-2)}$  for  $m > 2$ , and therefore  $\epsilon_t = \frac{y_m}{\sqrt{\frac{m}{(m-2)}}}$ . Thus, the pdf of  $\epsilon_t$  is given by

$$f(\epsilon_t | \boldsymbol{\varphi}, m) = \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left(\frac{m}{2}\right)(\sqrt{(m-2)}\pi)^{\frac{1}{2}}\sigma_t} \frac{1}{\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \quad (4.82)$$

The conditional likelihood of  $c_t^2$  is obtained using the equation  $c_t = \sigma_t \epsilon_t$  and can be written as

$$\begin{aligned} f(c_1, c_2, \dots, c_T, \sigma_1^2, \sigma_2^2, \dots, \sigma_T^2 | \boldsymbol{\varphi}, m) \\ = \prod_{t=2}^T \frac{\Gamma\left[\frac{(m+1)}{2}\right]}{\Gamma\left(\frac{m}{2}\right)(\sqrt{(m-2)}\pi)^{\frac{1}{2}}\sigma_t} \frac{1}{\sigma_t} \left(1 + \frac{c_t^2}{(m-2)\sigma_t^2}\right)^{-\frac{(m+2)}{2}}, m > 2. \end{aligned} \quad (4.83)$$

Now, again we use the log-likelihood function to find the maximum likelihood estimates (MLEs) as we did under the normality assumption. Then the likelihood function becomes

$$\begin{aligned} l(\boldsymbol{\varphi}, m | c_1, \sigma_1^2) &= l(\boldsymbol{\varphi}, m) = \ln f(c_t | \mathbb{T}_{T-1}) \\ l(\boldsymbol{\varphi}, m | c_1, \sigma_1^2) &= l(\boldsymbol{\varphi}, m) = - \sum_{t=2}^T \left[ \frac{m+1}{2} \ln \left( 1 + \frac{c_t^2}{(m-2)\sigma_t^2} \right) + \frac{1}{2} \ln(\sigma_t^2) \right]. \end{aligned} \quad (4.84)$$

Lastly, we may assume that  $\epsilon_t$  has a generalized error distribution (GED) and its pdf can be given by

$$f(c_1 | \boldsymbol{\varphi}, m) = \frac{m \cdot \exp\left(-\frac{1}{2} \left| \frac{c_t}{\omega \cdot \sigma_t} \right|^m\right)}{\omega \cdot \sigma_t 2^{(1-\frac{1}{m})} \Gamma\left(\frac{1}{m}\right)}, 0 < m \leq \infty, \quad (4.85)$$

where  $\omega = \left[ 2^{\left(-\frac{2}{m}\right)} \frac{\Gamma\left(\frac{1}{m}\right)}{\Gamma\left(\frac{3}{m}\right)} \right]^{\frac{1}{2}}$  and  $\Gamma(\cdot)$  is the usual gamma function ( $\Gamma(y) = \int_0^\infty (x^{y-1} e^{-x}) dx$ ).

Then log likelihood function

$$l(\boldsymbol{\varphi} | m) = \ln(m) - \left( -\frac{1}{2} \left| \frac{c_t}{\omega \cdot \sigma_t} \right|^m \right) - \ln \left[ \omega 2^{(1-\frac{1}{m})} \right] - \ln \Gamma\left(\frac{1}{m}\right) \quad (4.86)$$

As in general GARCH ( $r, s$ ), for the tail-thickness parameter  $m = 2$  and the pdf equals to the one for a standard normal distribution and it has heavy tails when  $m < 2$  (Bollerslev et al., 1994).

#### 4.4.2.4 Forecasting with GARCH (1,1) Model

Since the GARCH models can be represented as ARMA models (see equation 4.44), the forecast of a GARCH model can just as well be obtained using the same procedure as in the ARMA model. Now we

can use the GARCH (1,1) model in equation (4.57) and the 1-step ahead forecast from the origin  $T_0$  can be

$$\sigma_{T_0+1}^2 = \alpha_0 + \alpha_1 c_{T_0}^2 + \beta_1 \sigma_{T_0}^2, \quad (4.87)$$

where  $c_{T_0}^2$  and  $\sigma_{T_0}^2$  are known at the time origin  $T_0$ .

for several steps ahead, we can use  $c_t^2 = \sigma_t^2 \epsilon_t^2$  and GARCH (1,1) equation can be rewritten as

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 + \alpha_1 \sigma_t^2 (\epsilon_{t+1}^2 - 1). \quad (4.88)$$

When  $t = T_0 + 1$ , the equation (4.88) becomes

$$\sigma_{T_0+2}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{T_0+1}^2 + \alpha_1 \sigma_{T_0+1}^2 (\epsilon_{T_0+1}^2 - 1). \quad (4.89)$$

Therefore, the 2-step ahead from the origin  $T_0$  is given by

$$\sigma_{T_0+2}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{T_0+1}^2, \quad (4.90)$$

because  $E(\epsilon_{T_0+1}^2 - 1 | \mathcal{T}_{T_0}) = 0$ .

Thus, the general equation for  $h$ -steps ahead from the origin  $T_0$  is

$$\sigma_{T_0+h}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{T_0}^2 (h - 1), \quad (4.91)$$

where  $h > 1$  (Tsay, 2005, p.115).

#### 4.4.3 GARCH Models Limitations

The GARCH model successfully captures volatility clustering and thick-tailed returns, however, it is unable to accommodate several other stylized facts such as to explain the *leverage effects* which are usually observed in the financial time series (Bollerslev et al., 1994, p.2 969). The leverage effects were first discovered by Black (1976) and they represent the tendency of variation in the price of stocks to be negatively correlated with changes in the stock volatility. Briefly, the effect of a shock upon the volatility is asymmetric, which means that the impacts of *good news* (positive lagged residual) and of *bad news* (negative residual) are different. Therefore, like everything else, the GARCH model is not perfect, hence it needs improvement as well. So, researchers have developed different alternative models as extensions of the GARCH model, including TARCH, IGARCH, EGARCH, OGARCH, M-GARCH, APARCH and PC-GARCH, to name a few (Matei, 2009, p.52).

#### 4.5 Extensions of the GARCH Models

The main objectives of volatility analysis are to explain its cause, while the structure of the time series is valuable for forecasting (Engle, 2001, p.16). Therefore, a variety of asymmetric GARCH models have been developed, such as the ones mentioned above in sub-section 4.4.2, that is, Exponential GARCH of Nelson (1991), Threshold ARCH model attributed to Rabemananjara et al., (1993), and

Jaganathan and Runkle (1993), Integrated GARCH modes, TARCH, OGARCH, M-GARCH, APARCH, PC-GARCH and GJ-GARCH, etc. (Engle, 2001, p.16).

#### 4.5.1 The Exponential GARCH (EGARCH) Model

The Exponential GARCH (EGARCH) model was first introduced by Nelson (1991). The model can account for an asymmetric response to a shock. The main advantage of the EGARCH model is that the squared volatility is being modelled as a logarithm  $\log(\sigma_t^2)$  which allows it to be positive even if the parameters are negative, hence there is no need to impose the constraints on the parameters to avoid them being non-negative (Matei, 2009, p.52). In GARCH model, the conditional variance is a function for only the magnitudes of the lagged residuals and not their signs, whereas in the EGARCH model the conditional variance depends on both the size and the sign of lagged residuals (Bollerslev et al., 1994, p.2969). The EGARCH  $(r, s)$  models may be defined as

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^r h_i(\epsilon_{t-i}) + \sum_{j=1}^s \beta_j \ln \sigma_{t-j}^2, \quad (4.92)$$

where  $h_i(\epsilon_{t-i}) = \alpha_i \epsilon_{t-i} + \psi_i(|\epsilon_{t-i}| - E|\epsilon_{t-i}|)$ ,  $i = 1, \dots, r$ . It can be seen from equation (4.92) that there is no need for a parameter restriction to ensure positivity of the conditional variance  $\sigma_t^2$ . The parameters  $\alpha_i$ ,  $i = 1, \dots, r$ , make an asymmetric response to shocks possible. (Terasvirta, 2006, p.20).

#### 4.5.2 The Integrated GARCH Model

The Integrated GARCH (IGARCH) model was created specifically to model data that has persistent changes in volatility. It is a special case of a GARCH model when the sum of estimated parameters (except  $\alpha_0$ ) equals to one, that is

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j = 1. \quad (4.93)$$

The IGARCH (1,1) model can be written as

$$c_t = \sigma_t \epsilon_t, \quad (4.94)$$

where  $0 < \beta_1 < 1$  and  $\{\epsilon_t\}$  is defined as for the GARCH models. When  $\alpha_0 = 0$  this model is similar

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) c_{t-1}^2, \quad (4.95)$$

to the Exponential Weighted Moving Average (EWMA)-procedure presented in equation (4.32).

(Nasstrom, 2003, p.24).

#### 4.5.3 The Threshold GARCH Model

The Threshold GARCH (TARCH) model is one of the asymmetric models which was developed by Zakoian (1994) and Glosten et al., (1993). Threshold GARCH model works as the Exponential GARCH

model, in the sense that it also accounts for the asymmetric effects of the news. However, in the TARCH model, the leverage effect is expressed in a quadratic form whereas in the EGARCH model it is expressed in the exponential form. Thus, a TARCH ( $s, p$ ) model for modelling a conditional variance is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \delta_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 + \sum_{k=1}^q \gamma_k \delta_{t-k}^2 I_{t-k}^- \quad (4.96)$$

The leverage effects exist only if  $\gamma_k \neq 0$ , and the news response is asymmetric. When  $\gamma_k = 0$  for all values of  $k$ , then the TARCH model takes the form of a standard GARCH model. (Matei, 2009, p.53).

The TARCH (1,1) model can be written as

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \delta_{t-1}^2 + \gamma_1 \delta_{t-1}^2 I_{t-1}^-, \\ &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + (\alpha_1 + \gamma_1 I_{t-1}^-) \delta_{t-1}^2, \end{aligned} \quad (4.97)$$

where  $I_{t-1}^- = 1$  if  $\delta_{t-1} < 0$ , and  $I_{t-1}^- = 0$  if  $\delta_{t-1} > 0$ . The *good news* is represented by  $\delta_{t-1} > 0$  and *bad news* by  $\delta_{t-1} < 0$ . It is obvious that good and bad news have different outcomes on the conditional variance  $\sigma_t^2$ . (Výrost, 2003, p.18).

#### 4.5.4 The Asymmetric Power ARCH

An APARCH model is one of the more general GARCH models which was implemented by Laurent and Peters (2002) in the GARCH toolbox. The structure of a model is written as

$$\sigma_t^k = \alpha_0 + \sum_{i=1}^r \alpha_i (|c_{t-i}^2| - \gamma c_{t-1})^k + \sum_{j=1}^s \beta_j \sigma_t^2 \quad (4.98)$$

which usually involves many other GARCH models as special cases. (Nasstrom, 2003, p.25).

#### 4.6 Testing for an ARCH Effect

The ARCH effect is known as the conditional heteroscedasticity and its test is equivalent to check for serial correlations of squared series; such correlation is referred to as the volatility clustering (Chen et al., 2002, p.1). To get an understanding of the dependence structure of time series data, it is useful to calculate sample autocorrelations and check the significance of the relevant statistics such as the Ljung-Box test ( $Q_{LB}$ ) of Ljung and Box (1970) or ( $Q_{BP}$ ) test of Box and Pierce (1978). Thus, for demonstration purpose, let  $c_t$  be a series of errors or residuals. Then the Ljung-Box statistic is used to test if  $\{c_t\}$  series is a white noise sequence, where the hypotheses to be tested are

- $H_0$  : The first  $l$  – lags of the autocorrelation function of  $\{c_t\}$  series are zero,  
 $H_1$ : The first  $l$  – lags of the autocorrelation function of  $\{c_t\}$  series are not zero.

The Ljung-Box statistic is represented by

$$Q_{LB}(l) = T(T+2) \sum_{k=1}^l \left( \frac{\hat{\rho}_k^2(c_t)}{T-k} \right), \quad (4.99)$$

where  $T$  is the sample size,  $l$  is the total number of lags, and  $\hat{\rho}_k^2(c_t)$  is the squared sample autocorrelation of residuals series  $c_t$  at lag  $k$ . The sample autocorrelation  $\hat{\rho}_k(c_t)$  is given by

$$\hat{\rho}_k(c_t) = \frac{\sum_{t=k+1}^T (c_t - \hat{\mu}_T)(c_{t-k} - \hat{\mu}_T)}{\sum_{t=1}^T (c_t - \hat{\mu}_T)^2} \quad (4.100)$$

where  $\hat{\mu}_T = \frac{1}{T} \sum_{t=1}^T c_t$ .

If  $\{c_t\}$  is an i.i.d sequence and  $c_t$  has a finite fourth moment, the  $Q_{LB}(l)$  statistic has the asymptotic null distribution  $\chi^2(l)$ . (Tsay, 2005, p.106; Chen et al., 2002, p.2).

Now to test for an ARCH Effect or conditional heteroscedasticity, the Lagrange Multiplier LM test of Engle (1982) is used. This test is like the Ljung-Box ( $Q_{LB}$ ) test, except that it is based on sample autocorrelations of the squared series  $\{c_t^2\}$ . The Lagrange Multiplier (LM) test statistic is given by

$$Q_{LM}(l) = T(T+2) \sum_{k=1}^l \left( \frac{\hat{\rho}_k^2(c_t^2)}{T-k} \right), \quad (4.101)$$

where  $T$  is the sample size,  $l$  is the total number of lags, and  $\hat{\rho}_k^2(c_t^2)$  is the squared sample autocorrelation of residuals series  $c_t$  at lag  $k$ . The sample autocorrelation  $\hat{\rho}_k(c_t^2)$  is given by

$$\hat{\rho}_k(c_t^2) = \frac{\sum_{t=k+1}^T (c_t^2 - \hat{\sigma}_T)(c_{t-k}^2 - \hat{\sigma}_T)}{\sum_{t=1}^T (c_t^2 - \hat{\sigma}_T)^2} \quad (4.102)$$

where  $\hat{\sigma}_T = \frac{1}{T} \sum_{t=1}^T c_t^2$ .

Under the null hypothesis, the Lagrange Multiplier test statistic is asymptotically distributed as a chi-square distribution  $\chi^2(l)$  with  $l$  degrees of freedom. Thus, the null hypothesis is rejected if  $Q_{LM}(l) > \chi^2(\alpha)$ , where  $\chi^2(\alpha)$  is a 100(1 -  $\alpha$ ) percentile of a chi-square distribution with  $l$  degrees of freedom.

(Chen et al., 2002, p.2; Tsay, 2005, p.106; Wang, et. al., 2005, p.58-59).

#### 4.7 An ARCH Model Selection Criteria

Model section criteria are one of the important processes in data analysis, because all the inference, evaluation of data and good forecasting results depend heavily on the true model specification. Typically, there are four most common methods used for model selection, however, in this thesis we will focus mainly on two of them. The first one is Akaike's information criterion. The Akaike Information Criterion, in the case of a least-squares estimation under normal distribution, can be given as

$$AIC = -2 \log(\hat{\sigma}^2) + 2k, \quad (4.103)$$

where  $\hat{\sigma}^2$  is the estimated model error variance and  $k$  is the number of free parameters in the model.

If the model includes ARCH type errors, the following modification was recommended by Bieren (2006)

$$AIC = -2 \log(\hat{\sigma}^2) + 2(k) - 1 - \log(2\pi). \quad (4.104)$$

or

$$AIC = -2 \log(\mathcal{L}(\hat{\phi}|c)) + 2k, \quad (4.105)$$

$$AIC = -2 \log(\mathcal{L}(\hat{\phi}|c)) + 2k, \quad (4.106)$$

where  $\mathcal{L}(\hat{\phi}|c)$  is the value of the likelihood function in  $\hat{\phi}$ , the vector of the estimated parameters, given the observed data  $c$  and  $k$  is the number of parameters. (Chuffart, 2015, p.295).

The AIC is suitable for a small sample, and with a large sample can be important in selecting the right model, hence it is not consistent. Therefore, many improvements have been made to overcome this inconsistency in this criterion. Then, the second criterion is Bayesian Information Criteria (BIC) which was developed to accommodate model selection in large samples and it has the form:

$$BIC = -2 \log(\mathcal{L}(\hat{\phi}|c)) + k \log(T), \quad (4.107)$$

where  $N$  is the sample size. In contrast to AIC, BIC has weak performance in selecting a model from a small sample. (Chuffart, 2015, p.295).

These model selection criteria need several different models to be fitted, then the values of AIC or BIC are compared among the fitted models. The model with the least value of these information criteria is the best (Box et al., 2008, p.211-212).

## 4.8 Chapter Summary

In this chapter different volatility models were discussed as well as their parameter estimation procedures. The derivation of Maximum Likelihood Estimation for ARCH and/ GARCH models is provided with different distributions of the innovations. Forecasting methodologies for these models were also explained.

## Chapter 5

### 5. Application of SARIMA Models to Reserves Time Series Data

A time series is a gathering of observations made sequentially over a certain period. The typical examples in which time series are developed are economics and finance, environmental data, meteorology and hydrology, medicine, demographics, quality control and engineering (Reinert, 2010). The main objective of time series analysis is to build models for relating the behaviour of an individual or multiple time series and come up with a methodology for specifying, estimating and validating or assessing an appropriate model for specific data (DuFour, 2008).

S/ARIMA models are the most used flexible class of forecasting models used in time series analysis that utilize historical data to make predictions as described in Chapter 3. In this chapter, we apply these types of models to analyse and forecast the time series for South African official gold reserves, and foreign exchange reserves positions. From Chapter 3, fitting S/ARIMA models involves five steps: data preparation, model identification, model parameter coefficients estimation, fitted model diagnostics, and forecasting using the fitted model. We will follow these steps in fitting S/ARIMA models to gold and foreign exchange reserves positions series.

#### 5.1 Data Description

The time-series datasets used in this study comprises of the South African Official Gold Reserves Positions, and Foreign Exchange (FX) Reserves Positions. The unit measurement of these reserves positions is ZAR million (R000). These reserves positions (observations) were collected every month-end from 31 January 1994 to 31 May 2017. This is secondary monthly data obtained from South African Reserve Bank (SARB) website: (<https://www.resbank.co.za>), accessed on 12 April 2017.

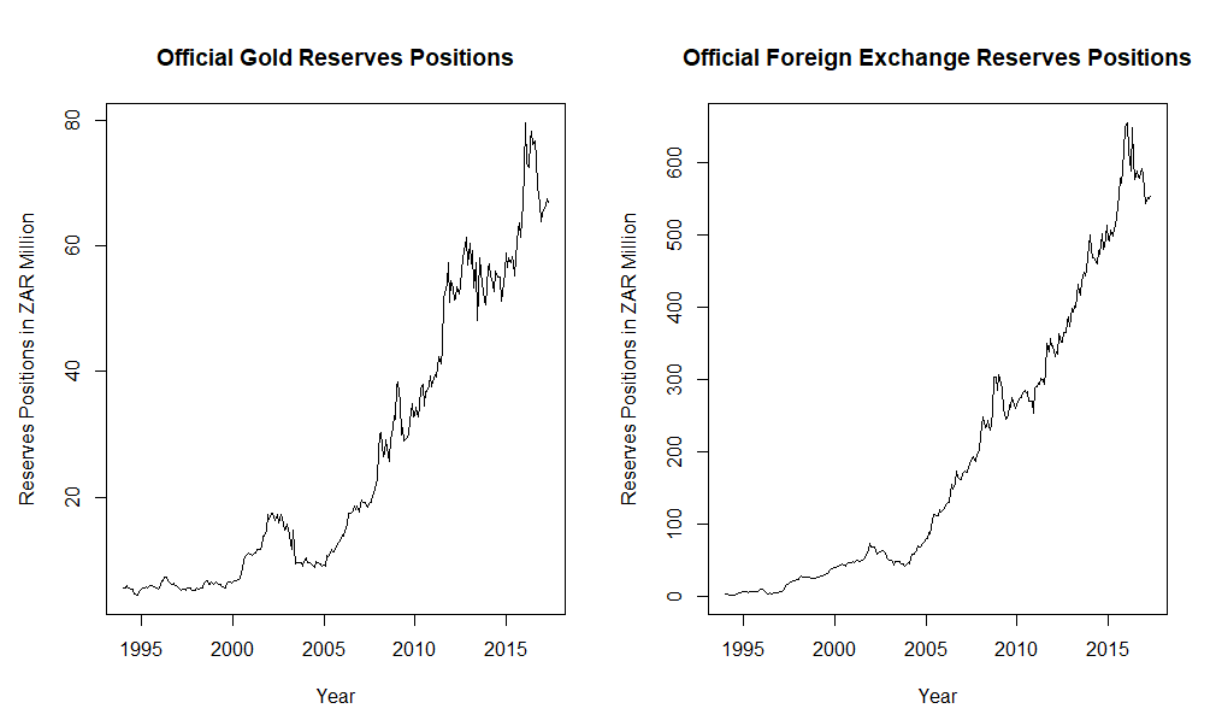
In every time series data analysis, a good starting point is to plot the series to have an insight of the data trends and visually examine it for any outliers, volatility, or irregularities, and thereby identify whether the series is stationary or not. A series can be considered as stationary when its mean, variance, and autocovariance are time-invariant, that is independent of time. Fitting some models like ARIMA model requires this assumption to hold for a series. This makes sense in a statistical context since ARIMA uses previous lags of series to model its behaviour and modelling stable series with consistent properties reduces the uncertainty. In this study, we will use both KPSS and ADF formal statistical tests in R-studio to test for stationarity of the data after taking differences into account, where applicable.

Autocorrelation plots or autocorrelation function (ACF) can also play a pivotal role in visualizing whether the series is stationary since it displays the correlation between observations at different lags. Usually, if the ACF plots suggest that the series is correlated, then generally, there are some trend or seasonal components, hence statistical properties are not constant over time and therefore the series is



non-stationary. In addition, ACF plots can also be used in suggesting the order of differencing if the series is non-stationary.

The plots for these reserves' positions are displayed in Figure 5.1. It can be seen from these plots that positions for both reserves are generally rising in the long-run.



(a) Gold Reserves Positions

(b) Foreign Exchange Reserves Positions

Figure 5.1 : Gold and foreign exchange reserves positions time series plots.

## 5.2 Gold Reserves Data Analysis

Table 5.1 below has descriptive statistics. The table shows that the standard deviation is R21.84 million; this indicates a high level of volatility in the gold reserves positions. The average monthly gold reserves positions is R26.02 million. The minimum is R4.38 million and the maximum is R79.66 million. This huge gap between minimum and maximum supports the high variability of positions in the gold reserves. To check the normality assumption, kurtosis and skewness can be used. Under the normality distribution, the kurtosis and skewness should be equal to 3 and 0, respectively. When the data is asymmetric it normally yields a positive or negative skew, otherwise the data is symmetrically distributed. The coefficient of kurtosis that is less than or greater than 3, indicates the flatness and peakedness, respectively. Table 5.1 shows a positive skew of 0.79, which implies that the distribution of the gold reserves position has a long right tail and a deviation from normality. The coefficient of kurtosis is -0.80, showing that the distribution is not normal but flattened.

Table 5.1: Descriptive statistics for gold reserves positions.

<b>Descriptive Statistic</b>	<b>Values</b>
No. of Obs.	281,00
Min.	4,38
Max.	79,66
Mean	26,02
Median	16,44
Range	75,28
Std. Dev.	21,84
Std. Error	1,30
Kurtosis	- 0,80
Skewness	0,79
Sum	7 311,65

### 5.2.1 Data Preparation

In this sub-section we explore gold reserves positions data to identify if there is a presence of some time series components such as trend, seasonality, and some random variation which caused variability of the data. We will use two formal statistical tests to check for stationarity and if the results from both tests come out showing that the data is non-stationary, we will, therefore, use the appropriate transformation techniques to transform the data to be stationary.

Figure 5.2 below shows the gold reserves positions plot; the data is clearly non-stationary, with an upward trend and some variation (which looks like seasonality with period 12) across, so variance stabilization and detrending might be required to transform the data. The ACF plot also shows significant autocorrelations with many lags, and its values start high and gradually die down slowly but never cuts off, which usually symbolizes non-stationarity in the series.

However, there is a necessity for a formal statistical test for stationarity to verify our visual inspection. We used two statistical tests defined in sub-section 3.4.1.1 to check for stationarity into the data, namely, Augmented Dickey-Fuller and KPSS tests. Table 5.2 shows the results from two statistical tests for stationarity.

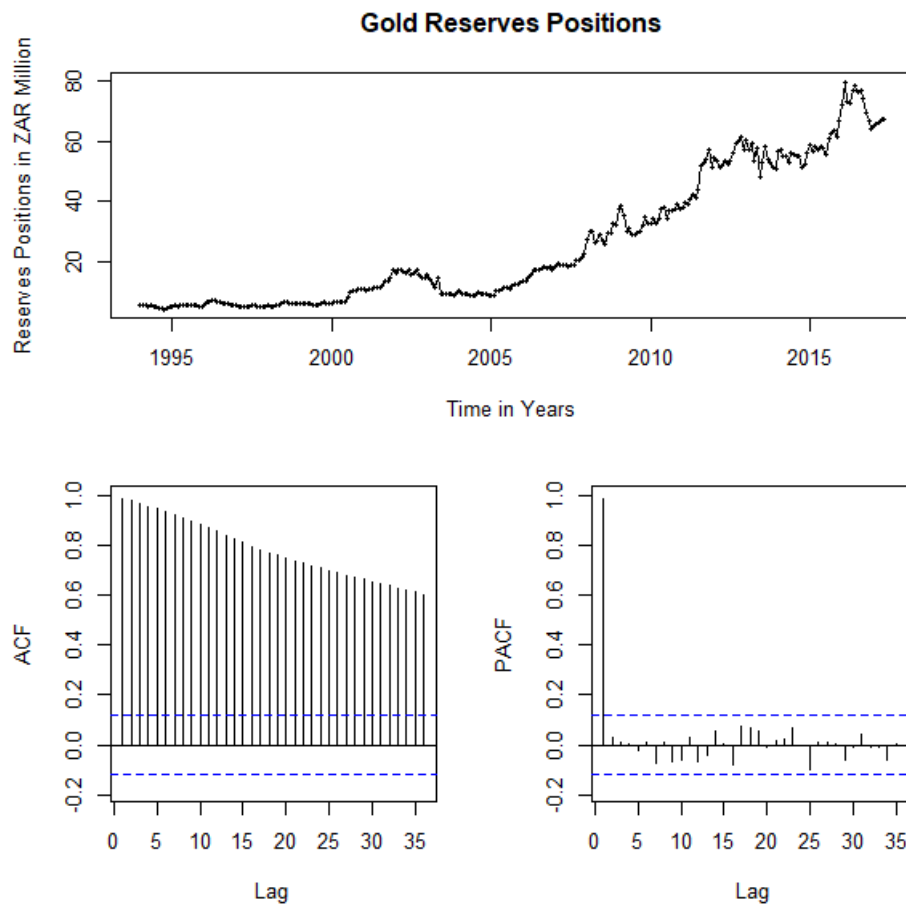


Figure 5.2: Gold reserves time series, ACF and PACF plots.

Table 5.2: Results of the ADF and KPSS tests for the stationarity of the original series of gold reserves positions.

Augmented Dickey-Fuller Test	
Data	Gold Reserves Positions
Dickey-Fuller	-2.180
Lag Order	6.000
P-value	0.500
KPSS Test	
Data	Gold Reserves Positions
KPSS Level	4.242
Lag Parameter	5.000
P-value	0.010

The results from two statistical tests in Table 5.2 show that the p-value of Augmented Dickey-Fuller test is greater than 0.05 (p-value = 0.500 > 0.05). Therefore, we fail to reject the null hypothesis of non-stationarity of the series. The p-value of KPSS test is less than 0.05 (p-value = 0.01 < 0.05) which implies that we have enough evidence to reject the null hypothesis of stationarity of the series. Now based on the above results we can conclude that gold reserves data is non-stationary at 5% level of significance, hence we need to apply data transformation.

The original series was log transformed by applying Box-Cox technique with  $\lambda=0$  which is equivalent to log transformation. The log transformed series was differenced to remove the linear trend ( $d=1$ ) and seasonality period 12 ( $D=1$ ). Figure 5.3 displays the time series plot, ACF and PACF plots of the resultant series. The series is now looking stationary since the plot of the data shows an oscillating pattern around 0 with no visible trend, which may suggest stationarity in the series. This is also being supported by the sample ACF and PACF that lie within the bounds  $\pm 1.96/\sqrt{n}$  roughly 95% of the time, which suggests that this series has no serial correlation. However, it is always necessary to conduct a formal statistical test to check for stationarity to verify visual inspection before jumping to a conclusion. We use two statistical tests to test for stationarity into the data, namely Augmented Dickey-Fuller and KPSS tests. Table 5.3 shows the results from two statistical tests for stationarity.

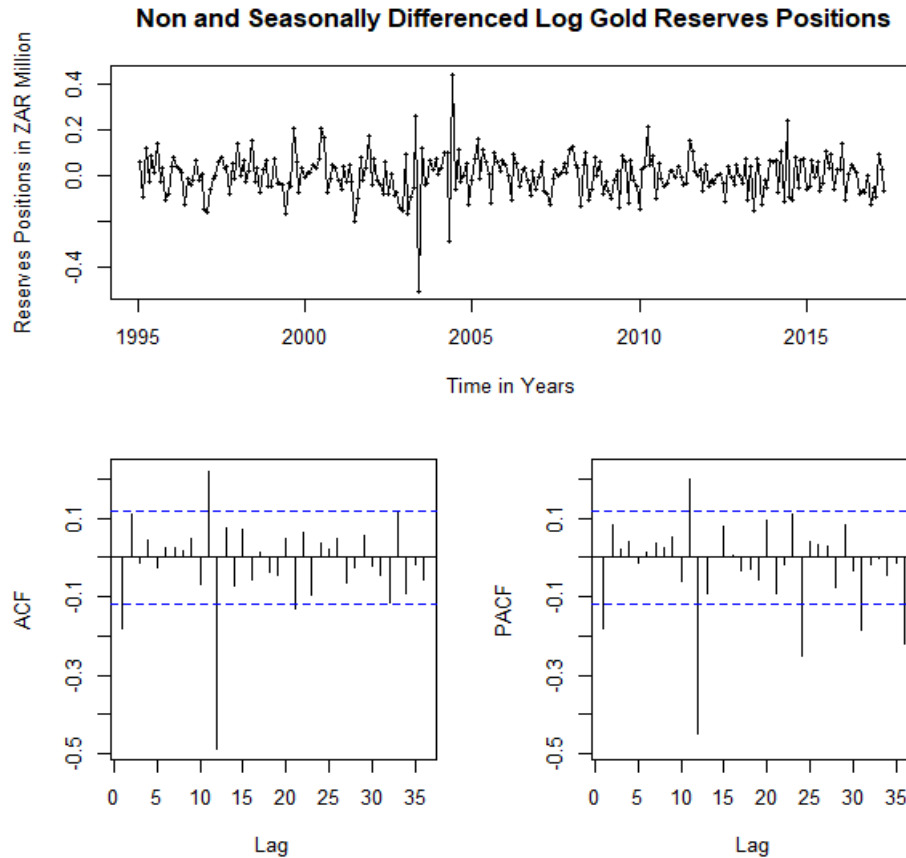


Figure 5.3: The time series, ACF and PACF plots of detrended ( $d=1$ ) and period 12 de-seasonalized ( $D=1$ ) log transformed gold reserves positions.

The results from the statistical tests in Table 5.3 show that the p-value of the Augmented Dickey-Fuller test is less than 0.05 (p-value = 0.01 < 0.05). Therefore, the null hypothesis of non-stationarity of the series is rejected. The p-value of KPSS test is greater than 0.05 (p-value = 0.1 > 0.05) which implies that we do not have enough evidence to reject the null hypothesis of stationarity of the series. Now based on these results we can conclude that differenced log gold reserves data is stationary at 5% level of significance.

Table 5.3: Results of the ADF and KPSS tests for the stationarity of the non- and seasonally differenced log gold reserves positions.

<b>Augmented Dickey-Fuller Test</b>	
Data	Gold Reserves Positions
Dickey-Fuller	-6.668
Lag Order	6.000
P-value	0.010
<b>KPSS Test</b>	
Data	Gold Reserves Positions
KPSS Level	0.077
Lag Parameter	5.000
P-value	0.100

### 5.2.2 Model Identification

Our aim is to find an appropriate SARIMA model to predict future positions of the gold reserves based on the ACF and PACF plots shown in Figure 5.3. The significant spike at lag 1 in the ACF suggests a non-seasonal MA (1) component, and the significant spike at lag 12 and 24 in the ACF suggests a seasonal MA (2) component. Consequently, we begin by fitting an SARIMA (0,1,1)(0,1,2)<sub>12</sub> model [or ARIMA (0,1,1)(0,1,2)<sub>12</sub> model], with non-seasonal MA(1), seasonal MA(2) components and first non-seasonal and seasonal differences ( $d=1$  and  $D=1$ ) and then vary both values of  $(p, q)$  by adding or subtracting 1 ( $\pm 1$ ) from this model to identify the exact possible values of AR and MA components. Looking at the sample PACF plot from Figure 5.3, we could have chosen the SARIMA (1,1,0) (2,1,0)<sub>12</sub> model as our initial model. To verify our initial guessed model, we can use AUTO.ARIMA() function (Hyndman and Khandakar: JSS, 2008) from forecast r package to generate an appropriate model. The AUTO.ARIMA() function selects a number of differences  $d$  and  $D$  via unit root test and  $(p, q)$  by minimizing AIC (Hyndman et al., 2014, pp.78). This function gives SARIMA (0,1,1) (0,1,2)<sub>12</sub>, which is aligned to our initial identified model.

### 5.2.3 Model Diagnostics and Parameter Coefficients Estimation

Table 5.4 shows a set of different possible models fitted to the data. To select an appropriate model from the set we need to start by examining ACF plots from their residuals to identify if there is no correlation present among the model residuals. If the order of parameters for a specific model and structure are correctly specified, then we expect to have no significant autocorrelations among the residuals. However, the examination of residuals alone is not enough to conclude that the fitted model is appropriate or not. Hence, we must use other criteria to ensure the selection of the right model.

Normally, the criteria used in selecting the appropriate model fitted include but are not limited to AIC, RMSE, log-likelihood, and sometimes significance of parameters. The smaller the magnitude of AIC and RMSE and the larger log-likelihood the better the model fitted.

Table 5.4: Goodness of fit statistics of the different models fitted to the series of gold reserves positions.

Model	Significant of Parameter, SARIMA	AIC	RMSE	Log-Likelihood	Ljung-Box Test p-values
1. SARIMA (0,1,1) (0,1,2) <sub>12</sub>	Only $\hat{\theta}_1$ sig.	-669	1.840	339	0.324
2. SARIMA (0,1,2) (0,1,2) <sub>12</sub>	Only $\hat{\theta}_1$ sig.	-669	1.837	340	0.459
3. SARIMA (1,1,2) (0,1,2) <sub>12</sub>	Only $\hat{\theta}_1$ sig.	-668	1.841	340	0.446
4. SARIMA (1,1,1) (0,1,2) <sub>12</sub>	Only $\hat{\theta}_1$ sig.	-669	1.834	339	0.385
<b>5. SARIMA (1,1,0) (0,1,2)<sub>12</sub></b>	<b>All sig.</b>	<b>-670</b>	<b>1.838</b>	<b>339</b>	<b>0.353</b>
6. SARIMA (0,1,0) (0,1,2) <sub>12</sub>	All sig.	-670	1.852	338	0.239

From Table 5.4 above all models are quite close which makes it hard to select the appropriate one, though the 5<sup>th</sup> and 6<sup>th</sup> models are looking better in terms of AIC and significant of parameters. Since residuals autocorrelation is okay for all models (all uncorrelated), it is hard to infer from them. Therefore, the choice is between only the 5<sup>th</sup> and 6<sup>th</sup> models in the table. However, looking at the two models' RMSE and log-likelihood it is quite clear that the best model is SARMA (1,1,0) (0,1,2)<sub>12</sub>, since it has the smallest RMSE (1.838) and largest log-likelihood (339) compared to SARIMA (0,1,0) (0,1,2)<sub>12</sub> model with RMSE of 1.852 and 338 log-likelihood.

We tried other models with non-seasonal AR terms as well, but none that gave a smaller AIC value. Consequently, we chose the SARIMA (1,1,0) (0,1,2)<sub>12</sub> model. Its residuals are plotted in Figure 5.4. All the spikes are now within the significance limits except one spike at lag 11, so by being generous, we can infer that residuals appear to be white noise. The Ljung-Box test in Table 5.5 also shows that the residuals for this model have no remaining autocorrelations ( $Q^*(df = 24) = 22.838$ , & p-value = 0.353 > 0.05). The summary of the fitted model with its parameter coefficients estimates is shown in Table 5.6 with all coefficient parameters being significant.

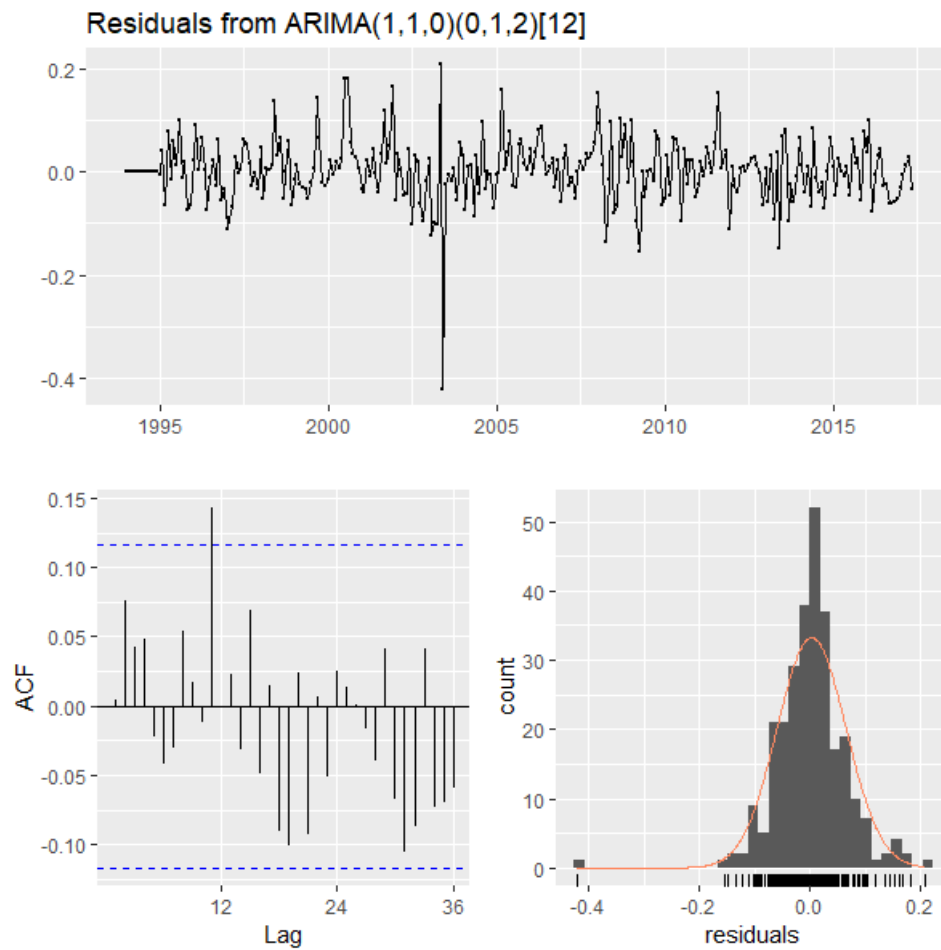


Figure 5.4: Residuals from SARIMA (1,1,0) (0,1,2)<sub>12</sub> model.

Table 5.5: Ljung-Box test results for residuals from selected fitted SARIMA (1,1,0) (0,1,2)<sub>12</sub> model.

Ljung-Box Test		
Q*	Df	p-value
22.838	24	0,3523



Table 5.6: Parameter estimates of the selected SARIMA (1,1,0) (0,1,2)<sub>12</sub> model for the gold reserves series.

Parameter Estimates	Coefficient	Std Error	Confidence Interval	
			2.50%	97.50%
$\hat{\phi}_1$	-0.091	0.062	-0.212	-0.021
$\hat{\theta}_1$	-0.964	0.089	-1.139	-0.789
$\hat{\theta}_2$	-0.036	0.062	-0.157	-0.085

Then the selected model to forecast gold reserves positions can be mathematically expressed as:

$$\hat{X}_t = -0.091\hat{X}_{t-1} + 0.964\varepsilon_{t-12} + 0.036\varepsilon_{t-24}. \quad (5.1)$$

#### 5.2.4 Forecasting Gold Reserves Positions Using SARIMA Models

The forecasting process using the selected model is simple in R-studio. First, the forecast horizon ( $h$  periods ahead) is specified and forecasts can then be run. The 24 months forecasts using the fitted model are shown in Figure 5.5. The large and rapidly increasing prediction intervals show that the gold reserves positions forecasts have increasing prediction error over time. There seems to be a slight upward trend, hence the gold reserves positions are expected to increase in the next 24 months.

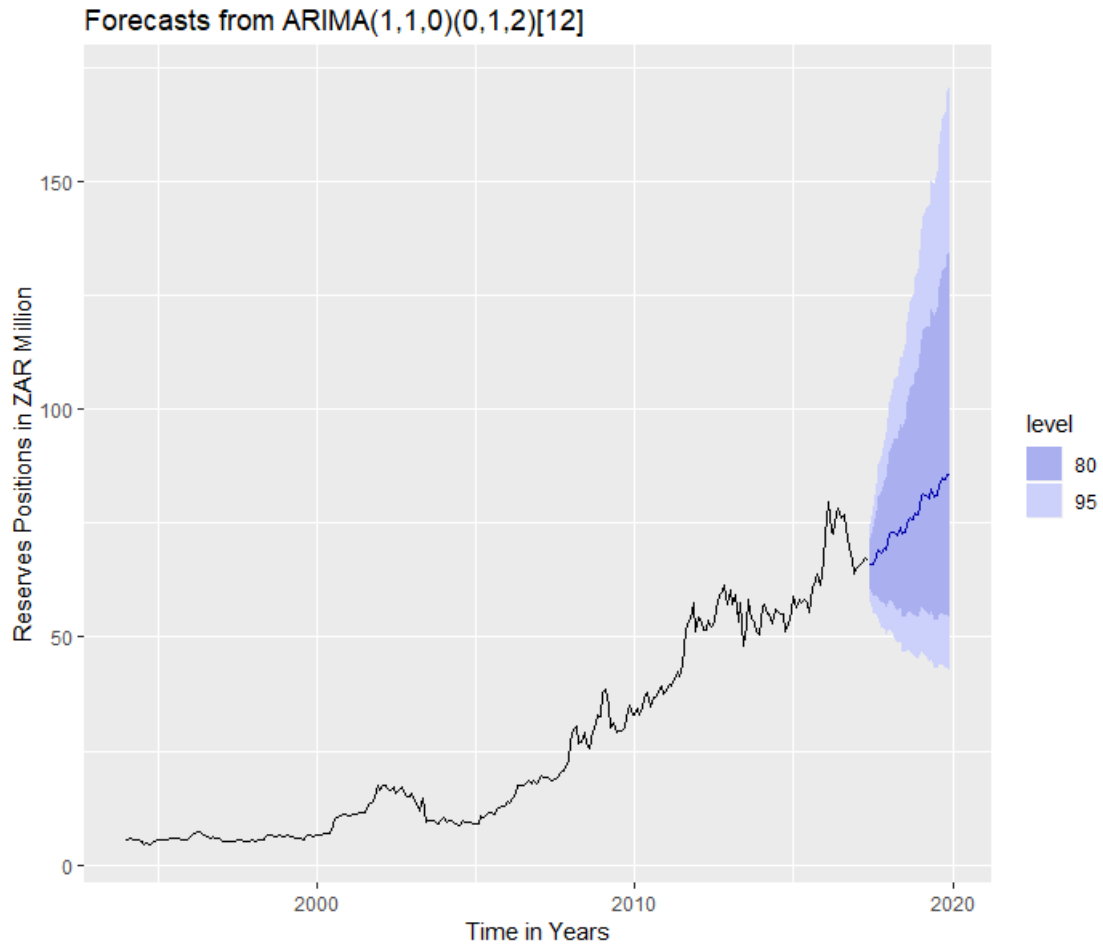


Figure 5.5: Gold reserves positions forecasts from SARIMA (1,1,0) (0,1,2)<sub>12</sub> model.

The gold reserves positions forecasts obtained using the SARIMA (1,1,0) (0,12)<sub>12</sub> model are shown in Appendix A.

### 5.3 Foreign Exchange Reserves Data Analysis

In this section, we identify estimates and fit the SARIMA model to the foreign exchange reserves positions.

The result in Table 5.7 shows that the standard deviation is R190.27 million, this implies a high level of volatility in the foreign exchange reserves positions. The average monthly foreign exchange reserves positions is R193.70 million. The minimum is R1.62 million, and the maximum is R654.98 million; this wider gap between minimum and maximum supports the high variability of positions in the foreign exchange reserves. The result from Table 5.7 shows a positive skew of 0.81, which implies that the distribution of the foreign exchange reserves positions has a long right tail and a deviation from normality. The coefficient of kurtosis is -0.62, implies that the distribution is not normal but flattened.

Table 5.7: Descriptive statistics for foreign exchange reserves positions

Descriptive Statistic	Values
No. of Obs.	281,000
Min.	1,618
Max.	654,981
Mean	193,704
Median	113,575
Range	653,363
Std. Dev.	190,274
Std. Error	11,351
Kurtosis	- 0,623
Skewness	0,806
Sum	54 430,763

### 5.3.1 Data Preparation

Here we explore foreign exchange reserves positions data, to identify the presence of some time series components, such as trend, seasonality, and some random variation caused variability of the data. Two formal statistical tests are used to test for stationarity and if the results from both tests show that the data is non-stationary, the data transformation techniques will be performed to convert the data.

Figure 5.6 below shows the foreign exchange reserves positions plot; the data is clearly non-stationary, with an upward trend and some variation or fluctuation which might suggest seasonality of period 12 since it is a monthly data, so variance stabilization and detrending or differencing might be required to transform the data. The ACF plot also shows significant autocorrelations with many lags, and its values start high and gradually die down slowly but never cut off. Therefore, we conduct a formal statistical test for stationarity to verify our visual inspection. Table 5.8 shows the results from Augmented Dickey-Fuller and KPSS statistical tests for stationarity.

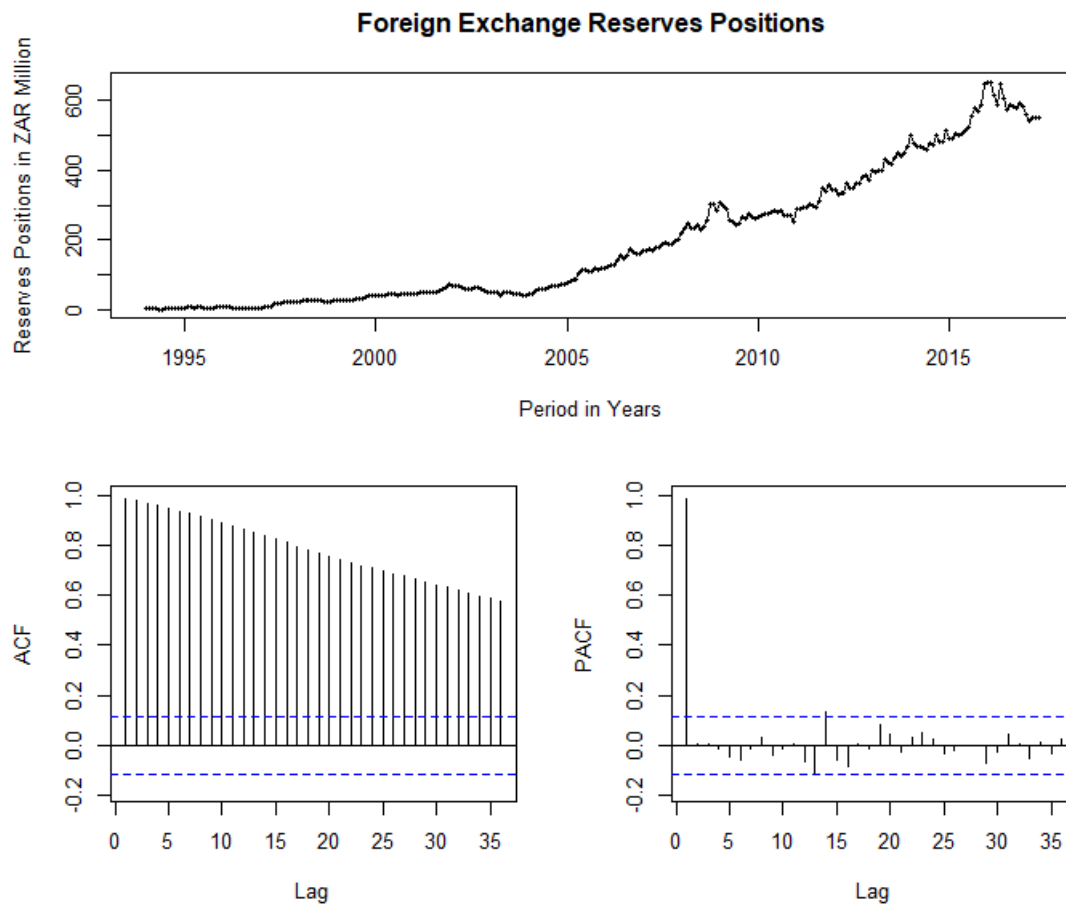


Figure 5.6: Foreign exchange reserves time series, ACF and PACF plots.

Table 5.8: Results of the ADF and KPSS tests for the stationarity of the foreign exchange reserves positions.

Augmented Dickey-Fuller Test	
Data	FX Reserves Positions
Dickey-Fuller	-1.792
Lag Order	6.000
P-value	0.664
KPSS Test	
Data	FX Reserves Positions
KPSS Level	4.339
Lag Parameter	5.000
P-value	0.010

The results from two statistical tests in Table 5.8 above show that the p-value of Augmented Dickey-Fuller test is greater than 0.05 (p-value = 0.664 > 0.05), therefore we fail to reject the null hypothesis of non-stationarity of the series. The p-value from KPSS test is less than 0.05 (p-value = 0.01 < 0.05) which implies that we have enough evidence to reject the null hypothesis of stationarity of the series. Now based on the above results we can conclude that foreign exchange reserves data is non-stationary at 5% level of significance.

Figure 5.7 shows transformed data with the Box-Cox variance transformation technique and non-seasonal and seasonal differencing. We have removed variability from the series by applying the Box-Cox technique using `Box-Cox.lambda()` R function to automate the value of lambda (lambda = 0.4216916). The log transformation (where lambda = 0) technique was firstly applied to the series, however, it has failed to completely remove the variability from the series. The presence of the trend and seasonal components from the data were removed by non-seasonal and seasonal differencing of order 1 (d=1 and D=1), respectively. The data is now looking stationary since the plot of the data shows an oscillating pattern around 0 with no visible trend, which may suggest stationarity in this series. This is also being supported by the sample ACF and PACF that lie within the bounds  $\pm 1.96/\sqrt{n}$  roughly 95% of the time, which suggests that this series has no serial correlation. Now we conduct a formal statistical test to check for stationarity to verify the visual inspection. Table 5.9 shows the results from Augmented Dickey-Fuller and KPSS statistical tests for stationarity.

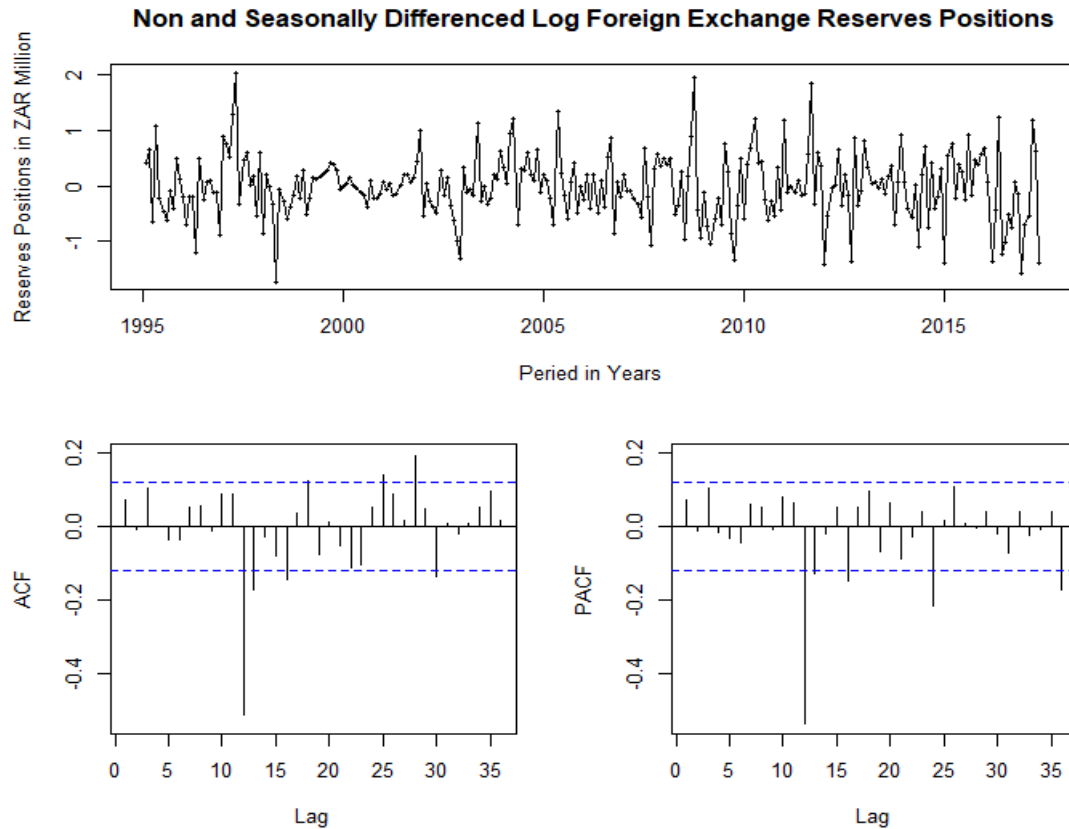


Figure 5.7: The time series, ACF and PACF plots of detrended ( $d=1$ ) and period 12 de-seasonalized ( $D=1$ ) log transformed foreign exchange reserves positions.

The results from two statistical tests in Table 5.9 below show that a p-value of Augmented Dickey-Fuller test is less than 0.05 ( $p\text{-value} = 0.01 < 0.05$ ), therefore we reject the null hypothesis of non-stationarity of the series. The p-value from KPSS test is greater than 0.05 ( $p\text{-value} = 0.10 > 0.05$ ) which implies that we do not have enough evidence to reject the null hypothesis of stationarity of the series. Now based on these results we can conclude that differenced log foreign exchange reserves series is stationary at 5% level of significance.

Table 5.9: Results of the ADF and KPSS tests for the stationarity of the non- and seasonally differenced log foreign exchange reserves positions.

<b>Augmented Dickey-Fuller Test</b>	
Data	FX Reserves Positions
Dickey-Fuller	-5.682
Lag Order	6.000
P-value	0.010
<b>KPSS Test</b>	
Data	FX Reserves Positions
KPSS Level	0.068
Lag Parameter	5.000
P-value	0.100

### 5.3.2 Model Identification

Our aim is now to find an appropriate SARIMA model based on the ACF and PACF shown in Figure 5.7. The significant spike at lags 12 and 24 from the PACF suggests a seasonal AR (2) component(s). Therefore, our initial model is SARIMA (0,1,0) (2,1,0)<sub>12</sub> model, indicating non-seasonal and seasonal differencing of order 1 ( $d=1$  &  $D=1$ ), and seasonal AR (2) components. Similarly, based on the ACF interpretation, we could have started with an SARIMA (0,1,0) (0,1,1)<sub>12</sub> model. Now we can vary both values of (p, q) by adding or subtracting 1 ( $\pm 1$ ) from this model to identify the exact possible values of AR and MA components. The AUTO.ARIMA () function has suggested the model SARIMA (0,1,0) (0,0,2)<sub>12</sub>, which is completely different from our initial guessed model, hence we will fit this model as well.

### 5.3.3 Model Diagnostics and Parameter Coefficients Estimation

Table 5.10 shows a set of different possible models fitted into the data and to select an appropriate model from the set we need to look at all the selection criteria; that is, the model with a smallest AIC and RMSE, largest log-likelihood, significant parameter coefficients estimates, and uncorrelated residuals.

Table 5.10: Goodness of fit statistics of the different models fitted to the series of foreign exchange reserves positions.

Model	Significant Parameters, SARIMA	AIC	RMSE	Log-Likelihood	Ljung-Box Test p-values for Residuals
<b>SARIMA (0,1,0) (2,1,0)<sub>12</sub></b>	<b>All sig.</b>	<b>361</b>	<b>11.394</b>	<b>-165</b>	<b>0.087</b>
SARIMA (1,1,0) (2,1,0) <sub>12</sub>	Only $\hat{\phi}_1$ not sig.	362	11.852	-177	0.083
SARIMA (0,1,1) (2,1,0) <sub>12</sub>	Only $\hat{\theta}_1$ not sig.	363	11.843	-177	0.083
SARIMA (1,1,1) (2,1,0) <sub>12</sub>	Only $\hat{\theta}_1$ not sig.	363	11.872	-177	0.072
SARIMA (0,1,0) (0,0,2) <sub>12</sub>	Only $\hat{\theta}_2$ not sig.	337	11.349	-175	0.137

From Table 5.10 above the 1<sup>st</sup> and 5<sup>th</sup> models look more appropriate in terms of AIC and RMSE. However, looking at the two models' significance of parameter coefficients and their log-likelihood it is quite clear that the best model is SARIMA (0,1,0) (2,1,0)<sub>12</sub> since all its parameter coefficients are significant at 5% level of significance and it has the largest log-likelihood. Therefore, we select this model as our best model to be fitted to foreign exchange reserves positions data.

The selected model residuals are plotted in Figure 5.8. All the spikes are now within the significance limits; therefore, residuals appear to be white noise. The Ljung-Box test in Table 5.11 also shows that the residuals for this model are not autocorrelated ( $Q^*(df = 22) = 31.483$ , &  $p\text{-value} = 0.087 > 0.05$ ). The summary of the fitted model with its parameter coefficients estimates is shown in Table 5.12.



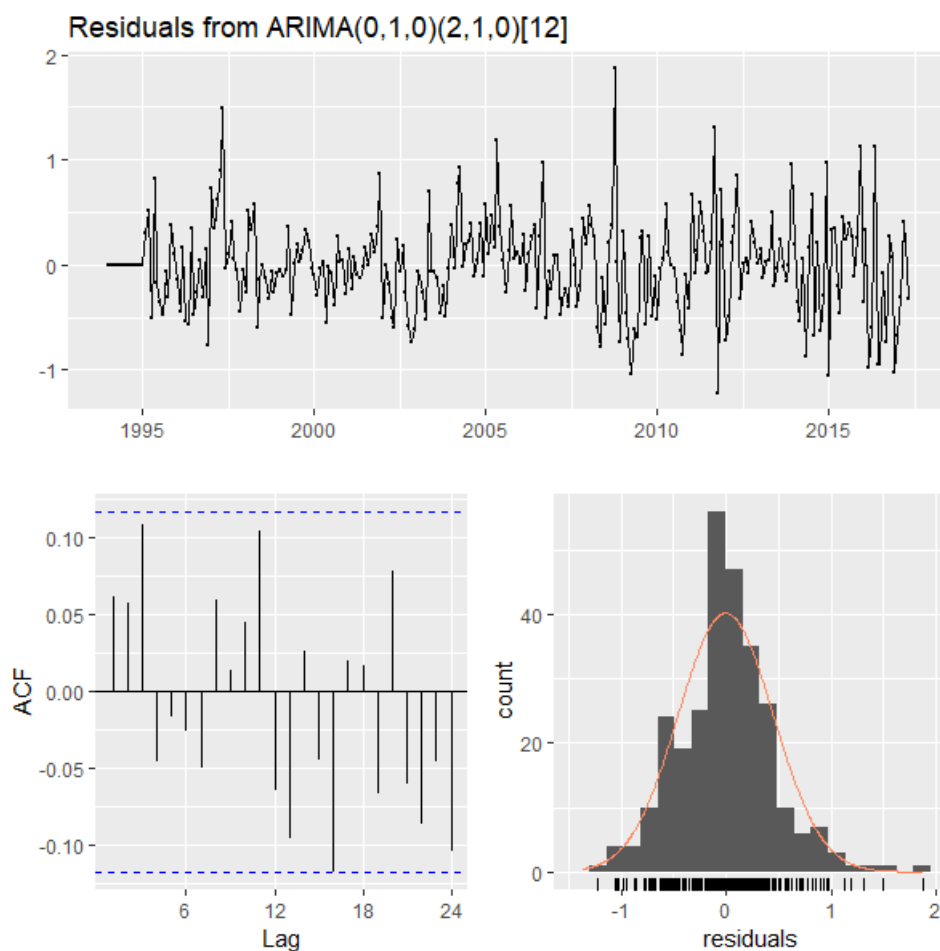


Figure 5.8: Residuals from SARIMA (0,1,0) (2,1,0)<sub>12</sub> model.

Table 5.11: Ljung-Box test results for residuals from selected fitted SARIMA (0,1,0) (2,1,0)<sub>12</sub> model.

Ljung-Box Test		
Q*	Df	p-value
31,483	22	0,087

Table 5.12: Parameter estimates of the selected SARIMA (0,1,0) (2,1,0)<sub>12</sub> model for the foreign exchange reserves series.

Parameter Estimates	Coefficient	Std Error	Confidence Interval	
			2.50%	97.50%
$\hat{\phi}_1$	-0.760	0.051	-0.878	-0.643
$\hat{\phi}_2$	-0.355	0.061	-0.474	-0.235

The selected model to forecast foreign exchange reserves positions can be mathematically expressed as:

$$\hat{X}_t = -0.760\hat{X}_{t-12} - 0.355\hat{X}_{t-24} \quad (5.2)$$

### 5.3.4 Forecasting Foreign Exchange Reserves Positions using SARIMA models

Since the fitted Seasonal ARIMA model has satisfied all required conditions it can now be used for forecasting. Forecasts for foreign exchange reserves positions for the next 24 months are shown in Figure 5.9. The large and rapidly increasing prediction intervals show that the foreign exchange reserves positions have increasing forecast error margins in time, thus forecast accuracy decreases with an increasing forecasting horizon.

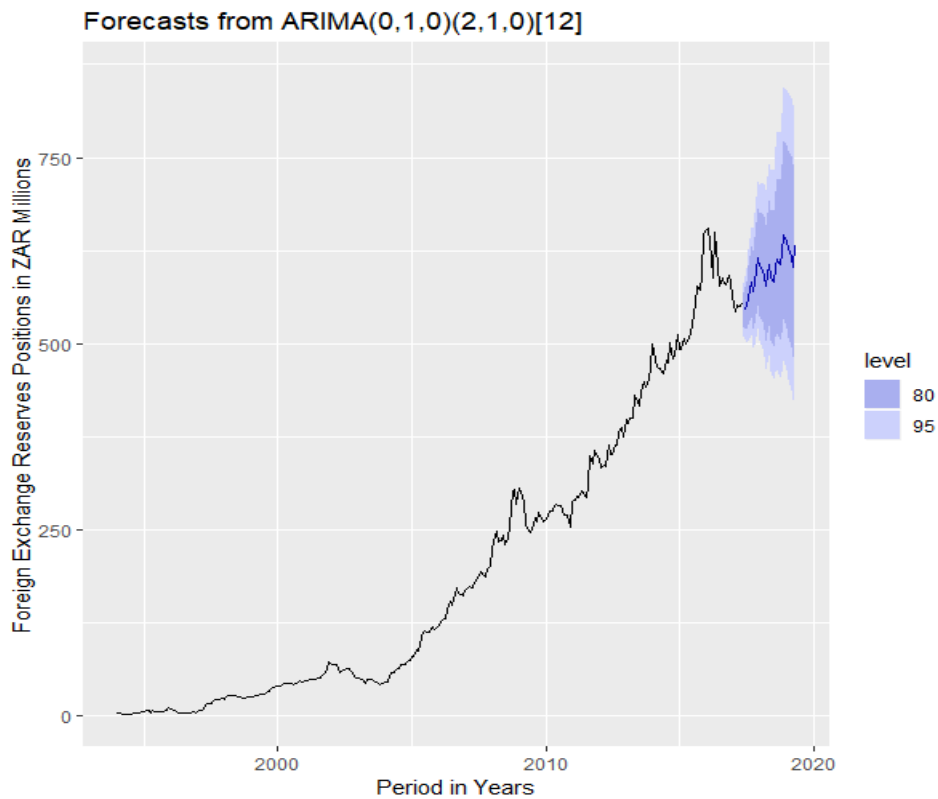


Figure 5.9: Forecasts from SARIMA (0,1,0) (2,1,0)<sub>12</sub> model.

The foreign exchange reserves positions forecasts obtained using the SARIMA (0,1,0) (2,1,0)<sub>12</sub> model are shown in Appendix A.

## 5.4 Chapter Summary

In this chapter, two datasets were analysed, that is, data for gold reserves positions, and foreign exchange (FX) reserves positions. The method of Box-Jenkins was applied to these datasets to forecast the future reserves positions up to end of May 2019. Now, since we have seen that our data has some

variation, we test whether there is a necessity to apply volatility models to cater for variability from the series or the fitted SARIMA models have dealt with these variations completely.

Table 5.13 shows serial correlation test results of the residuals from both models, SARIMA (1,1,0) (0,1,2)<sub>12</sub> and SARIMA (0,1,0) (2,1,0)<sub>12</sub> for gold reserves series and foreign exchange reserves series, respectively. The result ( $\chi^2(23) = 22.838$ , p-value=0.470) from Table 5.13 shows that there is no serial correlation from the residuals of the SARIMA (1,1,0) (0,1,2)<sub>12</sub> model (gold reserves series model). Hence the mean model fitted for gold reserves series has removed the correlation completely from this series. On the other hand, the result ( $\chi^2(24) = 31.483$ , p-value=0.140) from Table 5.13 also shows that there is no serial correlation from the residuals of the SARIMA (0,1,0) (2,1,0)<sub>12</sub> model (foreign exchange reserves series model). Again, this implies that the fitted model has removed the correlation completely from the foreign exchange series.

Table 5.13: The gold and foreign exchange reserves serial correlation test.

Data	X-Squared	df	p-value
Gold Reserves Series (Mean Model Residuals)	22,838	23	0,470
Foreign Exchange Reserves Series (Mean Model Residuals)	31,483	24	0,140

Table 5.14 shows the results from the statistical test for serial correlation in a variance (heteroscedastic or ARCH Effect test). The test shows that squared residuals for SARIMA (1,1,0) (0,1,2)<sub>12</sub> model are uncorrelated ( $\chi^2(23) = 22.455$ , p-value=0.493), hence the gold reserves series does not have ARCH Effect. However, the squared residuals for SARIMA (0,1,0) (2,1,0)<sub>12</sub> model seem to be correlated ( $\chi^2(24) = 36.787$ , p-value=0.046), therefore the foreign exchange reserves series has an ARCH Effect.

Table 5.14: The gold and foreign exchange reserves ARCH Effect (heteroscedasticity) test from the SARIMA model residuals.

Data	X-Squared	df	p-value
Gold Reserves Series (Mean Model Squared Residuals)	22,455	23	0,493
Foreign Exchange Reserves Series (Mean Model Squared Residuals)	36,787	24	0,046

Since we found that there is no ARCH Effect from the gold reserves series, there is no need to fit the volatility (ARCH and/or GARCH) models. Therefore, the initial fitted SARIMA (1,1,0) (0,1,2)<sub>12</sub> model is the best model for making predictions up to 24 months for the monthly gold reserves positions in South Africa. Hence the Central Reserve Bank of South Africa can use this model to make predictions. We believe that this model will help the central bank to make fact-based decisions and stabilize the whole economy for the country.

The foreign exchange reserves series has been found to have an ARCH Effect, which means the initial fitted SARIMA (0,1,0) (2,1,0)<sub>12</sub> model has failed to deal with the volatility from this series. Therefore, there is a necessity to fit the volatility (ARCH and/or GARCH) models for this series to cater for variability.

In conclusion the gold reserves series or data can be modelled adequately by SARIMA model whilst the SARIMA model is not adequate for the foreign exchange reserves series or data thus this is carried over to chapter 6.

## Chapter 6

### 6. Volatility Models Application to Foreign Exchange Data

#### 6.1 Introduction

This chapter focuses on the application of ARCH and GARCH models to the data sets introduced in Chapter 5. The chapter demonstrates how the heteroscedasticity models' theory in Chapter 4 is applied to real data. These models were fitted using the R-Studio statistical software.

#### 6.2 Model Selection Criteria

The choice for the best model can be based on the information criteria such as AIC and BIC as mentioned in Chapter 4, the smaller the AIC and/or BIC values, the better the model fitted. The  $R^2$  and RMSE can also be used as the best model selection criteria, the larger and smaller the values of  $R^2$  and RMSE, respectively, the better the model. In this study, we will base our best model selection on the AIC and BIC. However, the AIC and BIC only are not enough to infer that the fitted model is the best; hence, below are some other conditions the model must satisfy:

1. All the parameters estimates should be significant,
2. The model fitting iterative procedures should converge, and
3. The sum of the parameters  $\alpha_i$  and  $\beta_j$ , for  $i = 1, 2, \dots, p$  and  $j = 1, 2, \dots, q$  in the model should be less or equal to one,
4. The model selected should be parsimonious, that is, it must have as few parameters as possible.

#### 6.3 Analysis of Foreign Exchange Reserves

From Chapter 5, the selected mean model for foreign exchange reserves series was SARIMA (0,1,0) (2,1,0)<sub>12</sub> with  $p = 0$  and  $q = 0$ . Hence, we shall attempt to fit the SARIMA (0,1,0) (2,1,0)<sub>12</sub> – GARCH ( $s, r$ ) model or the standard GARCH ( $s, r$ ) model, since both parameters from the mean model are equals to zero (i.e.,  $p = q = 0$ ). Hence, in this chapter you will notice that we often use GARCH ( $s, r$ ) than SARIMA ( $p, q$ ) – GARCH ( $s, r$ ). To get the most robust GARCH model we may need to fit different GARCH models to the data with the different distributions described in Chapter 4. Usually, the standard order of  $s$  and  $r$  is less than or equal to 2, which implies that the values of  $s$  and  $r$  can vary between 0, 1, and 2. However, this is not always the case as sometimes the values of  $s$  and/ or  $r$  can be greater than 2 since this is informed by the number of lags which remove the ARCH Effect and autocorrelation in the residuals after fitting the GARCH model.

Table 6.1: GARCH – type models fitted with normal distribution.

Model	AIC	BIC	Log-Likelihood	Sig. Parameters at 5% L.o.S
GARCH (1,0)	1.219	1.271	-167.251	All sig.
GARCH (2,0)	1.220	1.284	-166.371	$\hat{\alpha}_2$ not sig.
<b>GARCH (1,1)</b>	<b>1.187</b>	<b>1.252</b>	<b>-161.820</b>	<b>All sig.</b>
GARCH (1,2)	1.184	1.262	-160.410	$\hat{\beta}_1$ not sig.
GARCH (2,1)	1.198	1.276	-162.317	$\hat{\alpha}_2$ not sig.
GARCH (2,2)	1.189	1.279	-159.994	$\hat{\alpha}_2$ and $\hat{\beta}_1$ not sig.

Table 6.1 shows the results from different models fitted with the normal  $\epsilon_t$  distribution; the ARCH (1) or GARCH (1,0) model seems to be more suitable than the ARCH (2) or GARCH (2,0) model since all parameters for ARCH (1) are significant and it has the smallest AIC and BIC values when compared to the other models. However, we cannot infer that this is the best model to modelled the data when we compare it with the general GARCH ( $s, r$ ) model where  $s, r \geq 1$ . Therefore, we have to discard the ARCH (1) model and consider other GARCH ( $s, r$ ) models. As described in Chapter 4, GARCH models are better than ARCH models. From Table 6.1, it is clearly seen that the GARCH (1,1) model is the model of best fit to the data, since it is the only model with all significant parameters, and it is the only one that has the smallest AIC and BIC values, and highest log-likelihood.

Table 6.2 : MLE parameter estimates for fitted GARCH (1,1) model, conditional on the normal distribution.

Parameter Estimates	Coefficients' Estimates	Std. Error	t-value	p-value
$\hat{\omega}$	0.024	0.016	2.053	0.040
$\hat{\alpha}_1$	0.229	0.074	3.094	0.002
$\hat{\beta}_1$	0.684	0.084	8.192	< 0.001
Skew	1.307	0.133	9.847	< 0.001

Table 6.2 shows the results of the fitted model with all parameter estimates being significant at the 5% level of significance. There is now a need to perform the diagnostic tests of the fitted model to check if the standardized residuals are stationary and normally distributed.

Table 6.3: Normality test statistics with p – values for standardized residuals tests from GARCH (1,1) model under the standard normal distribution.

Statistical Test	Standardised Residuals	Test Statistic	Test Statistic Value	p-value
Jarque-Bera Test	R	Chi <sup>2</sup>	54.306	< 0.001
Shapiro-Wilk Test	R	W	0.965	< 0.001
Ljung-Box Test	R	Q (10)	7.580	0.670
Ljung-Box Test	R	Q (15)	12.071	0.674
Ljung-Box Test	R	Q (20)	18.012	0.587
Ljung-Box Test	R <sup>2</sup>	Q (10)	7.606	0.667
Ljung-Box Test	R <sup>2</sup>	Q (15)	14.772	0.468
Ljung-Box Test	R <sup>2</sup>	Q (20)	16.088	0.711
LM Arch Test	R	TR <sup>2</sup>	10.836	0.543

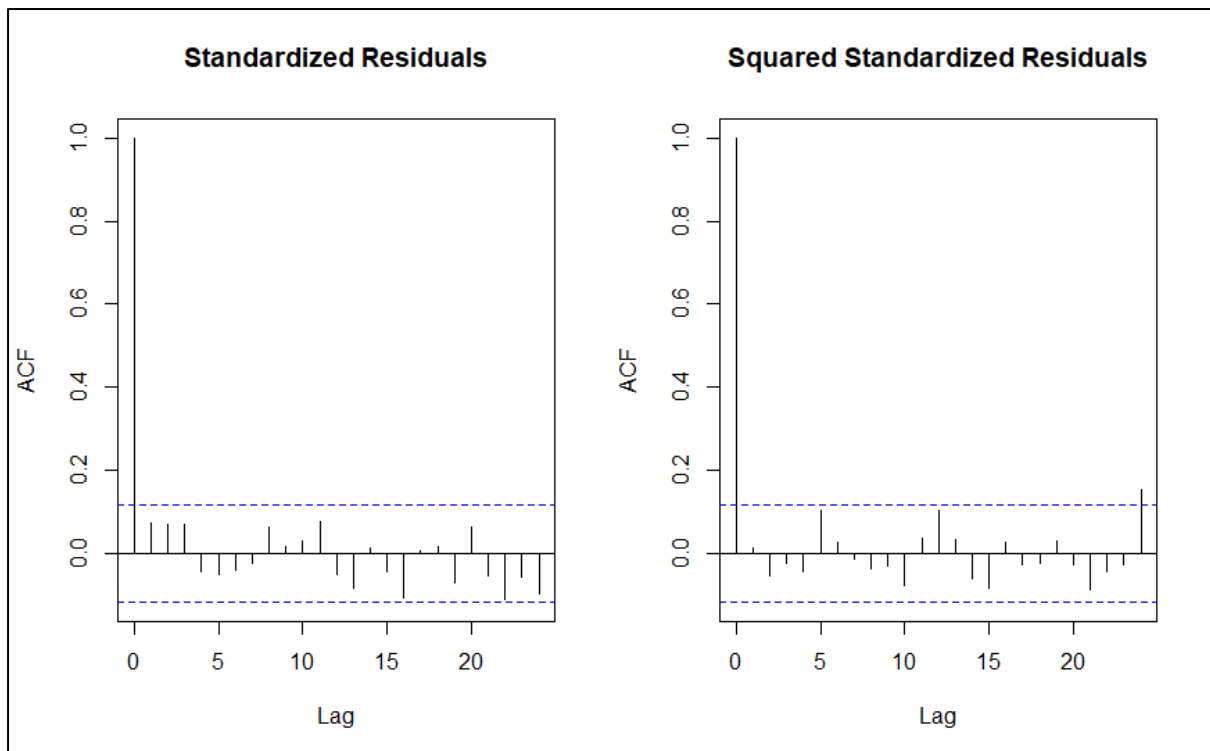


Figure 6.1: Standardized and squared standardized residuals ACF for GARCH (1,1) model fitted with standard normal distribution.

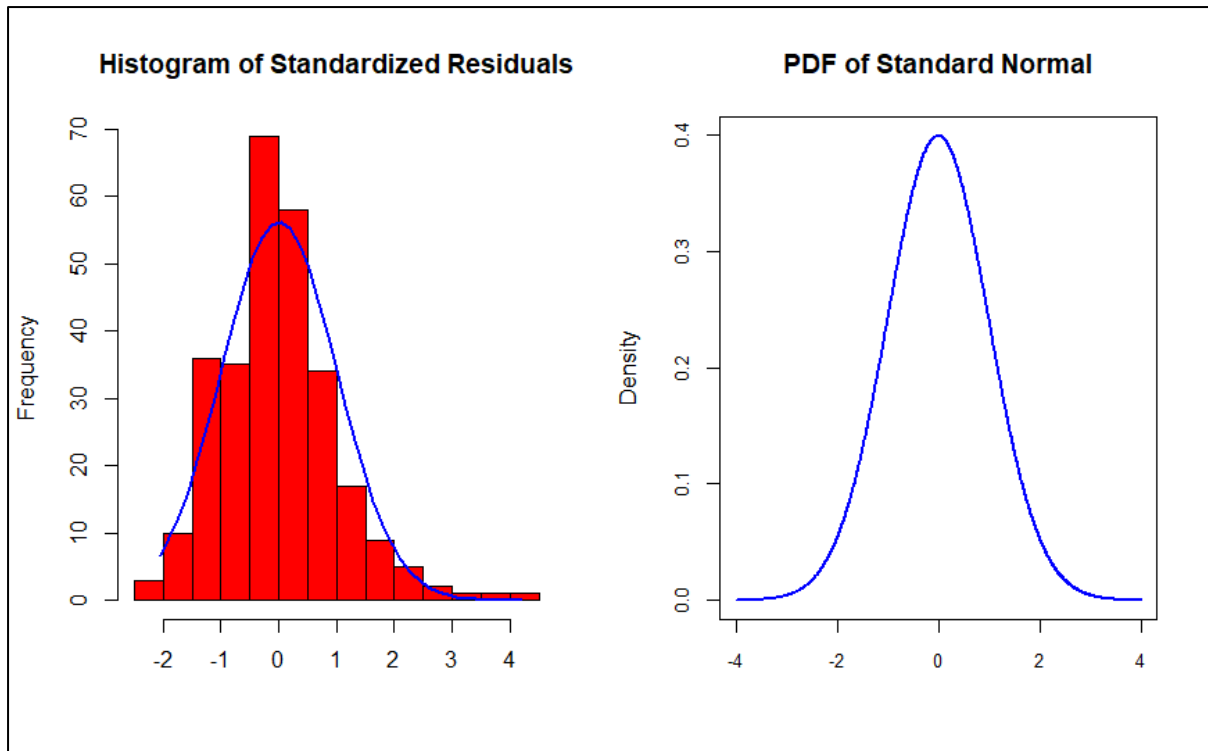


Figure 6.2: Standardized residuals frequency distribution vs. standard normal distribution density.

The results in Table 6.3 show that there is no evidence of autocorrelated residuals, and no ARCH Effect present in the series. This is supported by the Ljung-Box and LM Arch tests, where all the test statistics of standardized (R) and squared standardized residuals ( $R^2$ ) are not significant at the 5% level of significance. Figure 6.1 shows the autocorrelation function (ACF) of residuals and squared residuals used to assess if there is neglected dependence in the conditional mean and variance, respectively. Overall, there is weak evidence of auto-correlation in the standardized residuals and their squared values, so that, from this standpoint, the GARCH (1,1) model with a normal distribution seems to be well specified to model the foreign exchange reserves positions series.

However, the Shapiro – Wilk test ( $W=0.965$ ,  $p - \text{value} < 0.001$ ) and Jarque – Bera test ( $\text{Chi}^2=54.306$ ,  $p - \text{value} < 0.001$ ) are both significant, thus their null hypotheses of normally distributed residuals are rejected. This means that the standardized residuals do not come from the normal (Gaussian) distribution, hence the indication of significant departure from normality. This is supported by Figure 6.2, which compares the frequency distribution of the standardized residuals with the standard normal distribution. The histogram of the standardized residuals has a large peak at the centre of the distribution and a fatter right tail relative to the normal, although this is difficult to see in the graph. Since normality assumption is rejected, we must consider models with different distributions such as the student-t (std), skewed student-t (sstd), and the generalized error (ged) governing the innovations.



Table 6.4: GARCH (1,1) models fitted with student –  $t$  (std), skewed student –  $t$  (sstd) and generalized error (ged) distribution.

Model	Distribution	AIC	BIC	Log-Likelihood	Sig. Parameters at 5% L.o.S
1. GARCH (1,1)	std	1.178	1.243	-160.490	$\hat{\omega}$ and $\hat{\beta}_1$ not sig.
2. GARCH (1,1)	ged	1.740	1.252	-158.953	$\hat{\omega}$ and $\hat{\alpha}_2$ not sig.
<b>3. GARCH (1,1)</b>	<b>sstd</b>	<b>1.159</b>	<b>1.224</b>	<b>-157.861</b>	<b>All sig.</b>

Tables 6.4 shows maximum likelihood estimation parameter of GARCH (1,1) model fitted with all the three distribution, viz., student- $t$  distribution (std), skewed student- $t$  distribution (sstd), and generalized error distribution (ged). since it has the smallest AIC and BIC, respectively. When we compared the three GARCH (1,1) models fitted with different distributions, it is clearly that the best model is GARCH (1,1) fitted with skewed student –  $t$  distribution (sstd). This model has smallest AIC, BIC, and largest log – likelihood, respectively. Even its parameters are all significant at 5% level of significance.

Furthermore, the QQ – plots in figure 6.3, shows that the GARCH (1,1) model fitted with skewed student- $t$  distribution is the only model with points that fall approximately along the 45-degree reference line of the QQ – plot (i.e., data points are scattered around the straight line) when compared to the others.

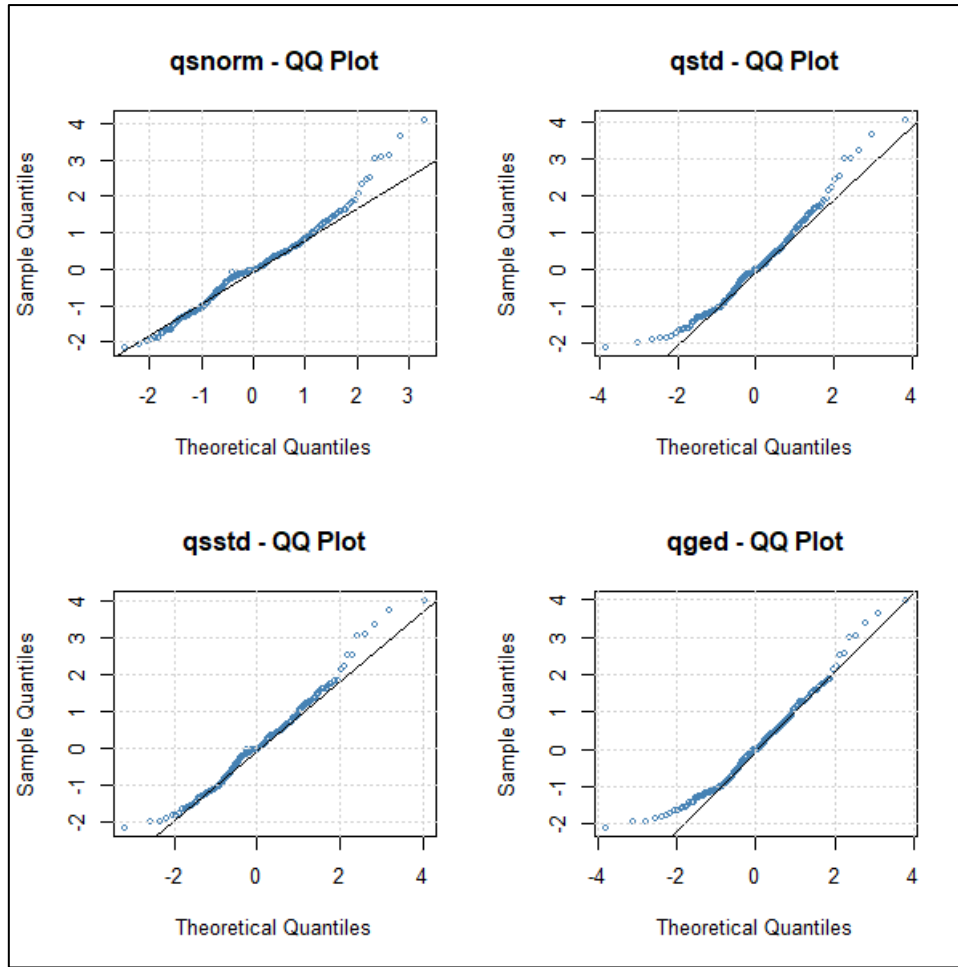


Figure 6.3: QQ-Plots for GARCH (1,1) model fitted with different distributions governing the innovations.

The parameter estimates of the GARCH (1,1) model fitted with the skewed student –  $t$  distribution is shown in Table 6.5 and all the parameters are significant at the 5% level of significance.

Table 6.5: MLE parameter estimates of GARCH (1,1) model fitted with skewed student –  $t$  distribution (sstd).

Parameter Estimates	Coefficients' Estimates	Std. Error	t-value	p-value
$\hat{\omega}$	0.019	0.012	1.596	0.0110
$\hat{\alpha}_1$	0.280	0.105	2.666	0.008
$\hat{\beta}_1$	0.677	0.087	7.824	0.000
Skew	1.170	0.114	10.234	0.000
Shape	7.043	3.543	1.988	0.047

The residuals diagnostics under the skewed student- $t$  distribution are presented in Table 6.6. The Ljung-Box test for standardized residuals and squared standardized residuals for all lags reveals that there is no evidence that residuals are autocorrelated and conditionally heteroscedastic (ARCH Effect), respectively at the 5% level of significance. The Shapiro-Wilk and Jarque-Bera tests are not significant at the 5% level of significance, which confirms that the standardized residuals are not from the normal distribution as oppose to the results from Table 6.3.

Table 6.6: Standardized residuals tests for GARCH (1,1) model under skewed student- $t$  distribution.

Statistical Test	Standardized Residuals	Test Statistic	Test Statistic Value	p-value
Jarque-Bera Test	R	Chi <sup>2</sup>	62.484	0.634
Shapiro-Wilk Test	R	W	0.962	0.596
Ljung-Box Test	R	Q (10)	7.409	0.686
Ljung-Box Test	R	Q (15)	11.458	0.720
Ljung-Box Test	R	Q (20)	17.282	0.635
Ljung-Box Test	R <sup>2</sup>	Q (10)	7.573	0.670
Ljung-Box Test	R <sup>2</sup>	Q (15)	14.346	0.499
Ljung-Box Test	R <sup>2</sup>	Q (20)	15.649	0.738
LM Arch Test	R	TR <sup>2</sup>	11.182	0.513

The ACF plots for standardized residuals and squared standardized residuals are shown in Figure 6.4. The plots show that the ACF of standardized residuals are not autocorrelated since the sample ACF lies within the bounds  $\pm 1.96/\sqrt{n}$  roughly 95% of the time, which indicates the good fit of the model. The squared standardized residuals also show no autocorrelation present, and this means that the conditional heteroscedasticity (ARCH Effect) has been completely removed from the series.

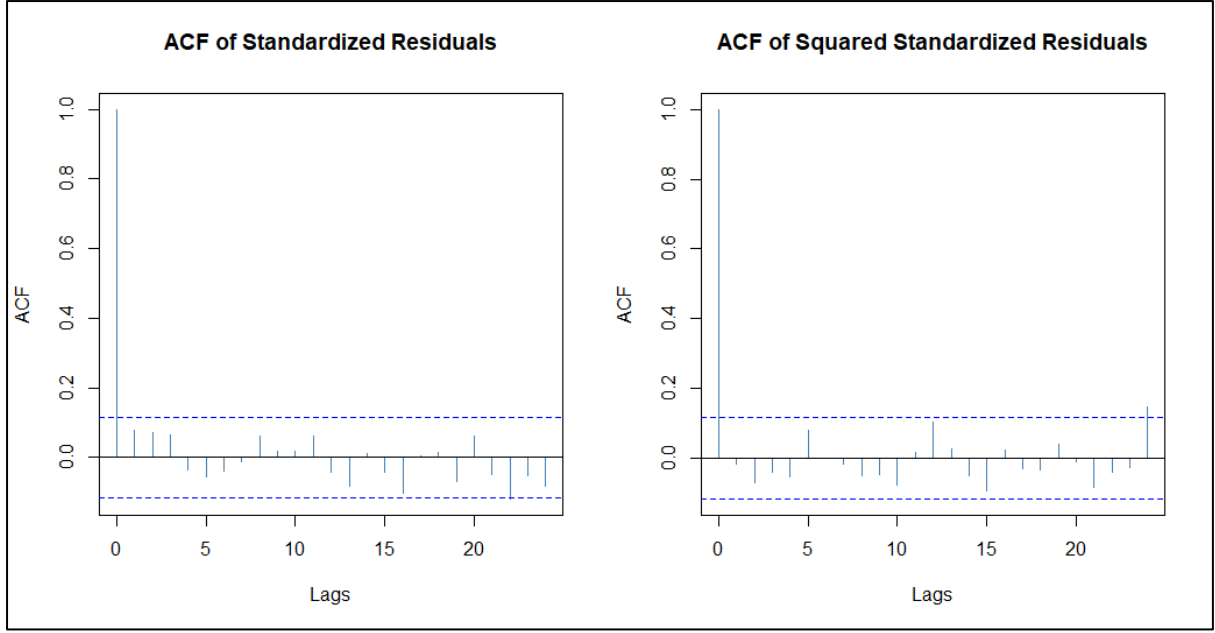


Figure 6.4: Standardized residuals ACF for GARCH (1,1) model with skewed student- $t$  distribution.

The equation of the fitted model can be written as:

$$\begin{aligned}\hat{\sigma}_t^2 &= \hat{\omega} + \hat{\alpha}_1 \epsilon_{t-1}^2 + \hat{\beta}_1 \sigma_{t-1}^2, \\ \hat{\sigma}_t^2 &= 0.019 + 0.280 \epsilon_{t-1}^2 + 0.677 \sigma_{t-1}^2.\end{aligned}\tag{6.1}$$

Figure 6.5 shows the forecasts obtained using the GARCH (1,1) model fitted with the skewed student- $t$  distribution governing the innovations. The plots show that the mean for foreign exchange reserves is expected to remain constant over the 24 months while the variance is expected to increase exponentially. Appendix B shows the 24 months' forecasts using the fitted model.

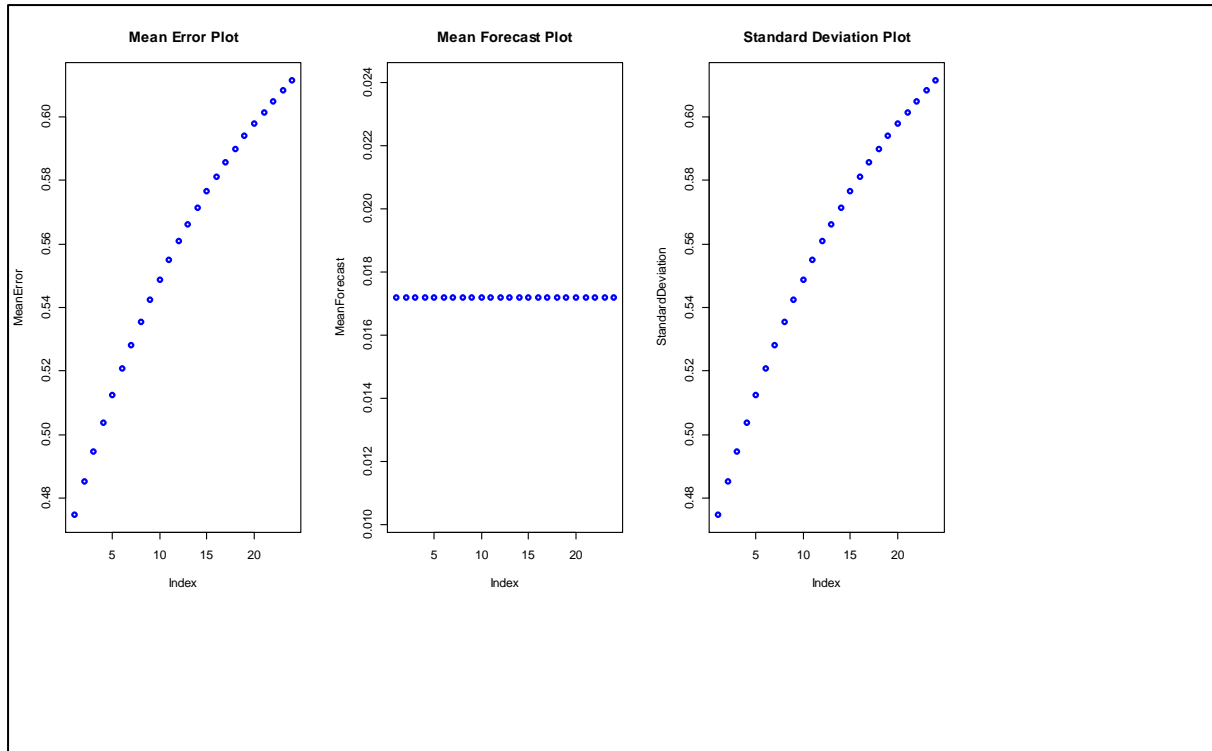


Figure 6.5: The forecasts from GARCH (1,1) model fitted with skewed student- $t$  distribution governing the innovations.

## 6.4 Chapter Summary

In this chapter, volatility models were fitted to foreign exchange (FX) reserves series or data. The necessity to fit volatility models to this type of series was found in chapter 5 where we developed the mean model SARIMA (0,1,0) (2,1,0)<sub>12</sub>. Therefore, used its standardized residuals and squared standardized residuals to test for serial correlation and ARCH Effect from the series. The test results showed that foreign exchange reserves data was uncorrelated but exhibited an ARCH Effect, thus SARIMA models could not adequately modelled this type of data. Hence the reason we proceed to fit the volatility models to modelled the data, so that the variability is being captured. The best volatility model for foreign exchange reserves data was found to be a GARCH (1,1) model fitted with skewed student- $t$  distribution governing the innovations (SARIMA (0,1,0) (2,1,0)<sub>12</sub> – GARCH (1,1) – SSTD).

## Chapter 7

### 7. Holt-Winters' Method in Time Series Analysis

#### 7.1 Introduction

It is always a brilliant idea when forecasting to compare the developed models with simple benchmark models to see which model outperforms other models. In this chapter, we discuss Holt-Winters method as compared to S/ARIMA forecasting models discussed and developed in Chapter 3 and Chapter 5, respectively. This method is used extensively in most time series analysis for forecasting purposes and was developed by both Holt (1957) and Winters (1960). They have extended Holt's original Exponential Smoothing method so as to capture both trend and seasonality from the data (Hyndman et al., 2018).

#### 7.2 Holt-Winters' Linear Trend Method

This method extends simple exponential smoothing to allow for the forecasting of data with a trend. This method involves a forecast equation and two smoothing equations, that is, the level and the trend equations:

Forecast Equation:

$$\hat{X}_{t+h|t} = l_t + hb_t, \quad (7.1)$$

Level Equation:

$$l_t = \alpha X_t + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (7.2)$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (7.3)$$

where  $l_t$  and  $b_t$  are the estimates of the level and the trend of the series at time  $t$ , respectively. The  $\alpha$  and  $\beta^*$  are the smoothing parameters for the level ( $0 \leq \alpha \leq 1$ ), and trend ( $0 \leq \beta^* \leq 1$ ), respectively (Hyndman et al., 2014).

Since the level equation shows that  $l_t$  is a weighted average of observation  $X_t$  and 1-step forecast ahead for time  $t$ , and trend equation shows that  $b_t$  is the weighted average of the estimated trend at time  $t$  based on  $l_t - l_{t-1}$  and  $b_{t-1}$ , it implies that the forecast function is no longer flat but trending (Hyndman et al., 2017).

Therefore, the forecasts generated by Holt-Winters' linear trend method display a constant trend (i.e., whether going down or up continually) forever into the future. Now this method tends to overestimate, or underestimate forecast values particularly when forecasting for longer forecast time horizons. This

phenomenon led Gardner and McKenzie (1985) to introduce an additional parameter from the above equations that "dampens" the trend to a flat line sometime in the future.

### 7.3 Holt-Winters' Damped Trend Method

In combination with the Holt-Winters' linear trend method's parameters ( $\alpha$  &  $\beta^*$ ), this method involves a damping parameter  $\varphi$  ( $0 < \varphi < 1$ ):

Forecast Equation:

$$\hat{X}_{t+h|t} = l_t + (\varphi^1 + \varphi^2 + \dots + \varphi^h)b_t, \quad (7.4)$$

Level Equation:

$$l_t = \alpha X_t + (1 - \alpha)(l_{t-1} + \varphi b_{t-1}), \quad (7.5)$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\varphi b_{t-1}. \quad (7.6)$$

This method is equivalent to the Holt-Winters' linear method if the damped parameter is equal to one ( $\varphi = 1$ ). Usually, the value of the damped parameter is restricted to a minimum of 0.8 and a maximum of 0.98, because damping has a very strong effect especial for values that are less than 0.8, and values close to 1 will eventually mean that a damped model is not able to be distinguished from a non-damped model (Gardner et al., 1985).

### 7.4 Holt-Winters' Seasonal Method

The Holt-Winters' Seasonal method is also made up of the forecast equation and three smoothing equations; the first two are the same as in the previous methods (i.e., for level and trend components) with the additional third one being a seasonal component. These components again have their respective smoothing parameters, that is,  $\alpha$  for the level,  $\beta^*$  for the trend, and  $\gamma$  for the seasonal (Holt, 1957). In this study,  $k$  is used to denote the period of the seasonality, for example, 4 for quarterly data, 12 for monthly data, and 52 for weekly data.

This method has two variations that vary in nature of the seasonal component:

#### 1. Additive Seasonality Method:

This method is only suitable when the seasonal variations are roughly constant through the series. Here the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation, the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero. The equations form of this method can be expressed as:

Forecast Equation:

$$\hat{X}_{t+h|t} = l_t + hb_t + s_{t+h-k(m+1)}, \quad (7.7)$$

Level Equation:

$$l_t = \alpha(X_t - s_{t-k}) + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (7.8)$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (7.9)$$

Seasonal Equation:

$$s_t = \gamma(X_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-k}, \quad (7.10)$$

where  $\gamma$  is the smoothing parameter for the seasonal ( $0 < \gamma < 1$ ) and  $m$  is the integer part of  $(h-1)/k$ , which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample (Holt,1957; Winters,1960).

## 2. Multiplicative Seasonality Method:

This method is preferred when the seasonal variations are changing proportionally to the level of the series. Here the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately  $k$  (Holt,1957; Winters,1960). The equations form of this method can be expressed as:

Forecast Equation:

$$\hat{X}_{t+h|t} = (l_t + hb_t)s_{t+h-k(m+1)}, \quad (7.11)$$

Level Equation:

$$l_t = \alpha \frac{X_t}{s_{t-k}} + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (7.12)$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (7.13)$$

Seasonal Equation:

$$s_t = \gamma \frac{X_t}{(l_{t-1} - b_{t-1})} + (1 - \gamma)s_{t-k}. \quad (7.14)$$



## 7.5 Holt-Winters' Damped Seasonal Method

Damping can be also included in both additive and multiplicative seasonal Holt-Winters methods. The Holt-Winters method with a damped trend and multiplicative seasonality often produces robust and accurate forecasts for seasonal data (Gardner et al., 1985). The Holt-Winters Seasonal equations with a damping parameter can be given as:

Forecast Equation:

$$\hat{X}_{t+h|t} = [l_t + (\varphi^1 + \varphi^2 + \dots + \varphi^h)b_t]s_{t+h-k(m+1)}, \quad (7.15)$$

Level Equation:

$$l_t = \alpha \frac{X_t}{s_{t-k}} + (1 - \alpha)(l_{t-1} + \varphi b_{t-1}), \quad (7.16)$$

Trend Equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\varphi b_{t-1}, \quad (7.17)$$

Seasonal Equation:

$$s_t = \gamma \frac{X_t}{(l_{t-1} - \varphi b_{t-1})} + (1 - \gamma)s_{t-k}. \quad (7.18)$$

## 7.6 Chapter Summary

In this chapter we briefly described the theory of Holt-Winters' methods in time series analysis. The four types of Holt-Winters' methods were defined, viz., linear trend method, damped trend method, seasonal method, and damped seasonal method, with their respective representative equations.

## Chapter 8

### 8. Application of Holt-Winters Method in Time Series Data

#### Analysis

##### 8.1 Introduction

Holt-Winters' exponential smoothing estimates the level, slope and seasonal component at the current time point, using their respective smoothing parameters. The parameters  $\alpha$ ,  $\beta^*$ ,  $\gamma$ , and  $\phi$  all have values between 0 and 1. The values close to 0 means that relatively little weight is placed on the most recent observations when making forecasts of future values. In this chapter we fit Holt – Winters' method to gold reserves series or data as compared to SARIMA models previously developed in Chapter 5. Since SARIMA models were not found to be the most robust models for foreign exchange reserves data, we cannot fit Holt – Winters method to this type of data, as the best model has been already found to be SARIMA (0,1,0) (2,1,0)<sub>12</sub> – GARCH (1,1) – SSTD in Chapter 6.

##### 8.2 Gold Reserves Data Analysis

We have seen from the SARIMA model that gold reserves monthly positions data has an upward trend and seasonality of period 12, therefore Holt-Winters seasonal method can be used to forecast its future positions. We apply Holt-Winters method with both additive and multiplicative seasonality to forecast monthly positions of the South African gold reserves. Figure 8.1 shows the data from 1994 and the forecasts for 30 June 2017–31 May 2019. The monthly forecasts positions for South African gold reserves obtained using Holt-Winters method with additive and multiplicative seasonality are also shown in Figure 8.1 and both shows an upward trend:

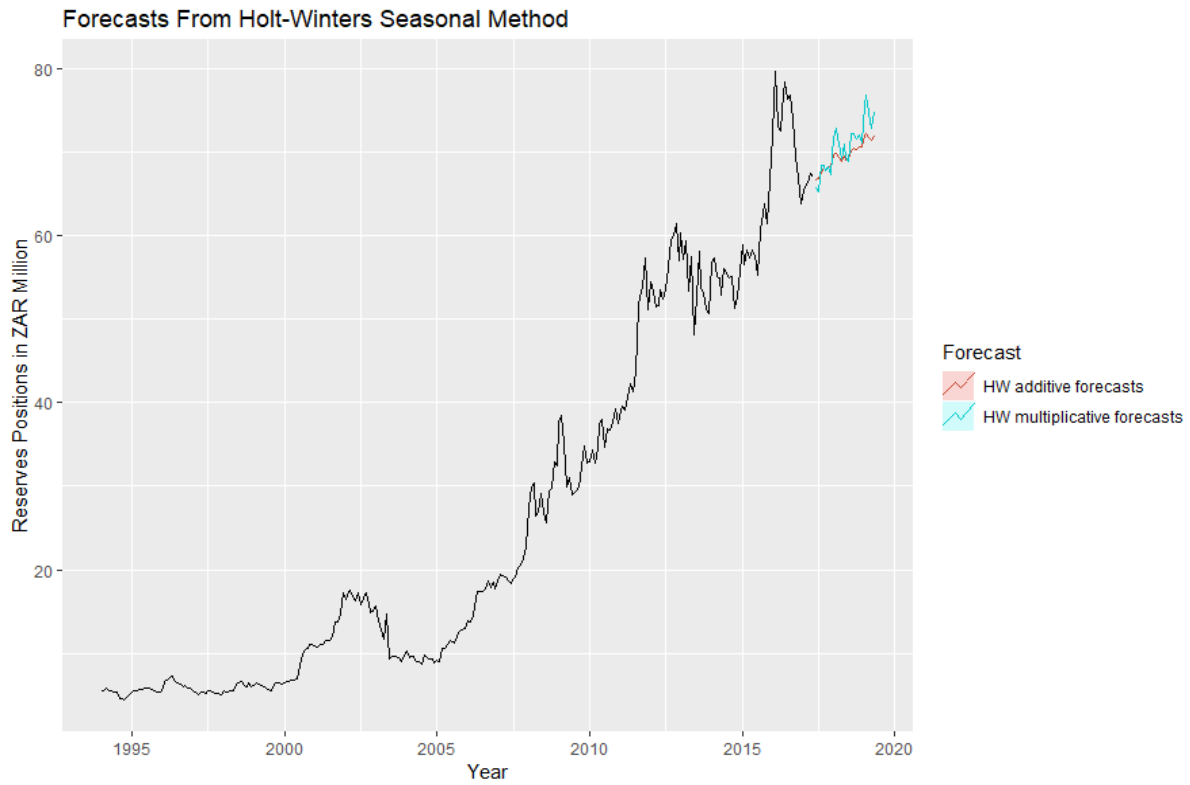


Figure 8.1: Forecasting official gold reserves positions in South Africa using the Holt-Winters' method with both additive and multiplicative seasonality.

Table 8.1: Applying Holt-Winters method with additive seasonality for forecasting official gold reserves positions in South Africa, notice that the additive seasonal component sums to approximately zero. The smoothing parameters and initial estimates for the components were estimated by minimizing RMSE ( $\hat{\alpha} = 0.8619$ ,  $\hat{\beta}^* = 0.0001$ ,  $\hat{\gamma} = 0.0002$ , and  $AIC=1967.574$  and  $RMSE=1.861433$ ).

Table 8.1: HW additive seasonal smoothing components (Gold Reserves Data).

Smoothing Components		
$l_t$	$b_t$	$s_t$
5,1332	0,2012	-0,2945
		-0,0205
		-0,1712
		0,1615
		-0,0191
		-0,7478
		-0,6275
		0,0032
		-0,3919
		0,2199
		0,9106
		0,9774
Sum	0,000	

Table 8.2: Applying Holt-Winters method with multiplicative seasonality for forecasting official gold reserves positions in South Africa, notice that the multiplicative seasonal component sums to approximately  $k=12$ . The smoothing parameters and initial estimates for the components have been estimated by minimizing RMSE ( $\hat{\alpha} = 0.9131$ ,  $\hat{\beta}^* = 0.0047$ ,  $\hat{\gamma} = 0.0001$ , and AIC=1753.022 and RMSE=1.846398).

Table 8.2: HW multiplicative seasonal smoothing components (Gold Reserves Data).

$l_t$	$b_t$	$s_t$
4,1133	0,4747	0,9756
		0,9941
		0,9908
		1,0048
		1,0083
		0,9662
		0,9786
		1,0046
		0,984
		1,0128
		1,0465
		1,0336
	<b>Sum</b>	<b>12,000</b>

Because both methods presented in Tables 8.1 and 8.2 above consist of the same number of parameters to estimate, we can compare the RMSE from both models. In this case, the method with multiplicative seasonality fits the data better, since it has the smallest RMSE (1.846398) compared to the additive seasonality method (RMSE=1.861433). This was to be expected, as it reflected from the two sets of forecasts (Figure 8.1); the forecasts generated by the method with the multiplicative seasonality display larger and increasing seasonal variation as the level of the forecasts increases compared to the forecasts generated by the method with additive seasonality.

Now we can try to incorporate the damping parameter and observe if the model from the Holt-Winters method with a multiplicative seasonality does improve the forecasts by including the damping parameter, since empirical studies have shown that Holt-Winters' method with a damped trend and multiplicative seasonality often produces robust and accurate forecasts for seasonal data (Gardner et al., 1985).

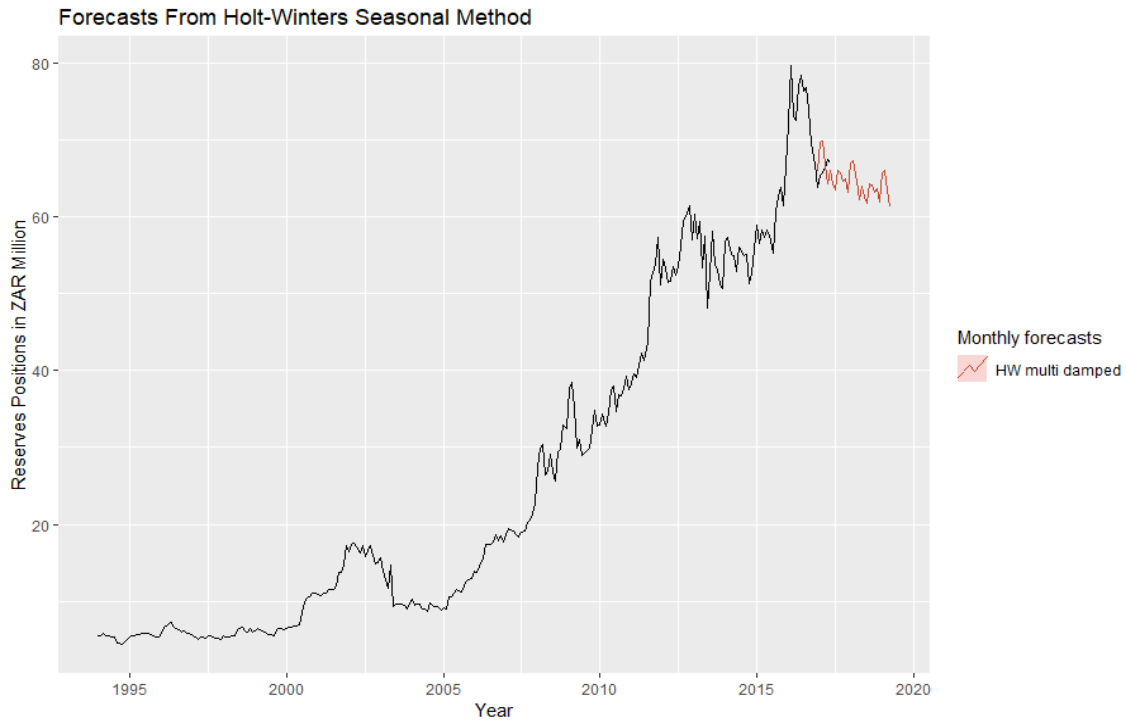


Figure 8.2: Forecasts of monthly gold reserves positions from Holt-Winters' Damped method with multiplicative seasonality.

Clearly, the model has improved and identified the monthly seasonal pattern and the slightly decreasing trend at the end of the data.

Table 8.3: Applying Holt-Winters damped method with multiplicative seasonality for forecasting official gold reserves positions in South Africa, again, we notice that the multiplicative seasonal component sums to approximately  $k=12$  since the data has a seasonality of period 12. The smoothing parameters and initial estimates for the components have been estimated by minimizing RMSE ( $\hat{\alpha} = 0.8277$ ,  $\hat{\beta}^* = 0.0406$ ,  $\hat{\gamma} = 0.0001$ , and  $AIC=1692.634$  and  $RMSE=1.83982$ ).

Table 8.3: HW damped multiplicative seasonal smoothing components (Gold Reserves Data).

<b>Smoothing Components</b>		
$l_t$	$b_t$	$s_t$
4,4702	0,4817	0,9694
		0,9926
		0,9917
		1,0086
		1,0149
		0,9715
		0,9803
		1,0061
		0,9849
		1,0103
		1,0399
		1,0297
	<b>Sum</b>	<b>12,000</b>

The RMSE (1.83982) from Holt-Winters damped method with multiplicative seasonality shows that the model has improved compared to the previous Holt-Winters method without a damping parameter (RMSE = 1.846398).

### 8.3 Chapter Summary

In this chapter, we have used the Holt-Winters seasonal method to find forecasts for gold reserves positions. The Holt-Winters damped method with multiplicative seasonality yielded good forecasts for this data compared to Holt-Winters damped method with additive seasonality. This was to be expected, as the time plot from data shows that the seasonal variation in the series increases with time.

## Chapter 9

### 9. Conclusion and Recommendations

In this study, reserves positions for gold and foreign exchange were studied and their 24 months (from May 2017– May 2019) predictions were obtained with the use of Seasonal ARIMA models. The Holt-Winters seasonal method was also used to forecast future gold reserves positions as the benchmarking method to early developed SARIMA model. The study also used SARIMA – GARCH models to forecast volatility for foreign exchange reserves positions since this data exhibited variation in the variance (ARCH Effect). The best SARIMA model to forecast gold reserves positions was found to be SARIMA (1,1,0) (0,1,2)<sub>12</sub>. The SARIMA (0,1,0) (2,1,0)<sub>12</sub> – GARCH (1,1) – SSTD model was found to be the best model to forecast volatility from the foreign exchange reserves data.

The best Holt-Winters method to forecast gold reserves positions was found to be the Holt-Winters damped method with multiplicative seasonality. Looking at the model accuracy, it is quite clear that in this case the SARIMA (1,1,0) (0,1,2)<sub>12</sub> model (RMSE=1.838338) outperformed the Holt-Winters damped method with multiplicative seasonality (RMSE=1.83982) in forecasting gold reserves positions.

Tables 9.1 represents the observed values (for previous 5 months) versus forecasts for gold reserves positions from the SARIMA model and Holt-Winters damped method with multiplicative seasonality. The forecasts from SARIMA model seem to be closer to the observed values. This confirms that SARIMA model have produced better forecasts than the Holt-Winters seasonal models.

Table 9.1: Observed versus forecasts for gold reserves positions.

Months	Observed (Million)	Forecasts	
		SARIMA (1, 1, 0) (0, 1, 2) <sub>12</sub>	HW Method With Multi Seasonality
Jan-17	R 65,303	R 67,448	R 67,332
Feb-17	R 65,769	R 68,200	R 66,548
Mar-17	R 66,338	R 67,800	R 62,946
Apr-17	R 67,457	R 66,990	R 59,872
May-17	R 66,958	R 68,928	R 60,918

More studies need to be conducted in the analysis of reserves to ensure that the central reserve bank is given advices or guidance on how to manage reserves of the country and when to accumulate reserves and/or reduce the level of reserves accumulation. The focus of this study was on creating forecasts for reserves, hence, to inform South African Reserve Bank of the reserves trajectory to be expected in future and be able to make fact-based decisions concerning reserves management. In the future studies, it



might be of the interest to look closely at the models for multivariate analysis in the same context of reserves positions analysis as an extension of the current univariate models.

## 10. References

1. Amini, H., Cont, R., & Minca, A. (2016). Resilience to contagion in financial networks. *Mathematical finance*, 26(2), 329-365.
2. Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.
3. Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.
4. Avellen, L. Med. (2005). *Essays on Multiple Exchange Rate Systems* (Doctoral dissertation).
5. Bai, X., Russell, J. R., & Tiao, G. C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of econometrics*, 114(2), 349-360.
6. Black, F. (1976). The pricing of commodity contracts. *Journal of financial economics*, 3(1-2), 167-179.
7. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
8. Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. *Handbook of econometrics*, 4, 2959-3038.
9. Box, G. E., Jenkins, G. M., & Reinsel, G. (1970). Time series analysis: forecasting and control Holden-day San Francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*.
10. Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). Forecasting. *Time Series Analysis*, 137-191.
11. Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). *Introduction to time series and forecasting* (Vol. 2). New York: Springer.
12. Chatfield, C. (2000). *Time-series forecasting*. Chapman and Hall/CRC.
13. Chen, Y. T., & Kuan, C. M. (2002). Time irreversibility and EGARCH effects in US stock index returns. *Journal of Applied Econometrics*, 17(5), 565-578.
14. Chuffart, T. (2015). Selection criteria in regime switching conditional volatility models. *Econometrics*, 3(2), 289-316.
15. Cont, R. (2005). Long range dependence in financial markets. In *Fractals in engineering* (pp. 159-179). Springer, London.
16. Davis, R., Dedu, V.K., & Bonye, F. (2014). Modelling and Forecasting of Gold Prices on Financial Markets. *American International Journal of Contemporary Research Vol. 4, No.3; March 2014*.
17. Deshmukh, R., & Pawar, D. (2014) prediction of gold prices using time series analysis.
18. Dickey, D. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.

19. Dickey, D. A., & Fuller, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.
20. Dobre, I., & Alexandru, A. A. (2008). Modelling the unemployment rate using the Box-Jenkins procedure.
21. Dufour, J. M. (2008). Introduction to time series analysis. In *McGill University Research Paper* (p. 1).
22. Enders, C. K. (2010). Applied missing data analysis. *Guilford press*.
23. Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of economic perspectives*, 15(4), 157-168.
24. Engle, R. F. (1982). A general approach to Lagrange multiplier model diagnostics. *Journal of Econometrics*, 20(1), 83-104.
25. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
26. Engle, R. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1), 122-150.
27. Feldstein, M. (1999). A self – help guide for emerging markets. *FOREIGN AFFAIRS-NEW YORK*-, 78, 93-109
28. Fischer, S. (2001). Exchange rate regimes: is the bipolar view correct? *Journal of economic perspectives*, 15(2), 3-24.
29. Fryzlewicz, P. (2007). Lecture notes: Financial time series, arch and garch models. *University of Bristol*.
30. Gardner Jr, E. S., & McKenzie, E. D. (1985). Forecasting trends in time series. *Management Science*, 31(10), 1237-1246.
31. Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess on stocks. *Journal of Finance*, 48, 1779-1801.
32. Greenspan, A. (1999, April). Currency reserves and debt. In *Remarks before the World Bank Conference on Recent Trend in Reserve Management, Washington, DC*.
33. Guha, B. &. (2011). Gold Price Forecasting Using ARIMA Model. *Journal of Advanced Management Science Vol. 4, No. 2*.
34. Guha, B., & Bandyopadhyay, G. (2011). Gold Price Forecasting Using ARIMA Model. *Journal of Advanced Management Science Vol. 4, No. 2, March 2016*.
35. Guo, H. (2012, September). Estimating Volatilities by the GARCH and the EWMA model of PetroChina and TCL in the Stock Exchange Market of China. In *International Scientific Conference Managing and Modelling of Financial Risks*.
36. Hassan, J. (2014). ARIMA and regression models for the prediction of daily and monthly clearness index. *Renewable Energy*, 68, 421-427.

37. Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages. (*O.N.R. Memorandum No. 52*). Carnegie Institute of Technology, Pittsburgh USA.
38. <https://www.resbank.co.za/AboutUs/PublicAwareness/Documents/FactSheet5.pdf> Accessed on [30 February. 17].
39. <https://www.resbank.co.za> Accessed on [15 June. 17]
40. Hull, J. C. (2003). *Options futures and other derivatives*. Pearson Education India.
41. Hyndman, R. J. (2001). ARIMA processes. *Data jobs: Data science knowledge*.
42. Hyndman, R. J. (2007). *Automatic time series forecasting*. In Book of Abstracts.
43. Hyndman, R. J. (2014). *Forecasting: Principles & Practice*. NY.
44. Hyndman, R. J., & Athanasopoulos, G. (2017). Forecasting: Principles and practice. otexts; 2014. *The book is freely available as an online book at [www. otexts. org/fpp](http://www.otexts.org/fpp). Alternatively, a print version is available.*
45. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
46. Hyndman, R. J., and G. Athanasopoulos. "Forecasting: Principles & Practice. NY." (2014).
47. Hyndman. (2009). Density forecasting for long-term peak electricity demand. *IEEE Transaction on Power System*, 25(2), 1142-1153.
48. IMF, A. (2014). World Economic Outlook (WEO).
49. ScottKannan, P., Scott, A., & Terrones, M. E. (2014). From recession to recovery: how soon and how strong. *Financial crises: causes, consequences, and policy responses*, 239.
50. Kao, J. J., & Huang, S. S. (2000). Forecasts using neural network versus Box-Jenkins methodology for ambient air quality monitoring data. *Journal of the Air & Waste Management Association*, 50(2), 219-226.
51. Khan, M. M. A. (2013). Forecasting of gold prices (Box Jenkins approach). *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 662-670.
52. Kotze. (2005).
53. Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
54. Makridakis, S., & Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*, 16(3), 147-163.
55. Mandelbro, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, Vol. 36, No. 4, October 1963.
56. Massarrat Ali Khan, M. (2013). Forecasting of Gold Prices (Box Jenkins Approach). *International Journal of Emerging Technology and Advanced Engineering Vol. 3, Issue 3, March 2013*.
57. Matei, M. (2009). Assessing volatility forecasting models: why GARCH models take the lead. *Romanian Journal of Economic Forecasting*, 12(4), 42-65.

58. McLeod, A. I. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of time series analysis*, 4(4), 269-273.
59. Mminele, D. (2013). Note on the foreign exchange market operations of the South African Reserve Bank. *BIS Paper*, (73x).
60. Mohanty, M. S., & Turner, P. (2006). Foreign exchange reserve accumulation in emerging markets: what are the domestic implication?. *BIS Quarterly Review*, September.
61. Mombeini, H., & Yazdani-Chamzini, A. (2015). Modelling Gold Price via Artificial Neural Network. *Journal of Economics, Business, and Management*, Vol. 3, No. 7, July 2015.
62. Mudd, S. G. (2002). Dealing with financial crises in emerging markets. *Thunderbird International Business Review*, 44(3), 399-430.
63. Mushtaq, R. (2011). Augmented Dickey-Fuller test.
64. Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *Journal of hydrology*, 10(3), 282-290.
65. Näsström, J. (2003). Volatility modelling of asset prices using Garch models.
66. Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
67. Nielsen, B. (2006). Order determination in general vector autoregressions. In *Time series and related topics* (pp. 93-112). Institute of Mathematical Statistics.
68. Patrica, A. C., Stancu, S., & Tindeche, A. (2016). Limitation of ARIMA models in financial and monetary economics. *Theoretical & Applied Economics*, 23(4).
69. Pfaff, B., & Stigler, M. (2008). vars: VAR Modelling. *R package version*, 1-3.
70. Posedel, P. (2018). Asymptotic analysis for a simple explicit estimator in Barndorff-Nielsen and Shephard stochastic volatility models. *Local Organizing Committee*.
71. Princ, P. E. T. E. R., & Školuda, V. Á. C. L. A. V. (2012). A Note on GARCH (1, 1) Estimation via Different Estimation Methods. *Methods*, 81-85.
72. Quenouille, M. H., & Quenouille, M. H. (1957). *The analysis of multiple time-series* (No. 519.232 Q84). London: Griffin.
73. Rabemananjara, R. &. (1993). Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics*, 8(1), 31-49.
74. Rabemananjara, R., & Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility. *Journal of applied econometrics*, 8(1), 31-49.
75. Reddy, Y. V. (2002). India's foreign exchange reserves: policy, status, and issues. *Economic and Political Weekly*, 1906-1914.
76. Reinert, G. (2010). Time series.
77. Rodrik, D. (2006). The social cost of foreign exchange reserves. *International Economic Journal*, 20(3), 253-266.

78. Steiner, A. (2013). How central banks prepare for financial crises—An empirical analysis of the effects of crises and globalization on international reserves. *Journal of International Money and Finance*, 33, 208-234.
79. Straumann. (2005).
80. Talke, I. S.. (2003). Modelling volatility in financial time series data. *Master's thesis, Mathematics, Statistics and Information Technology*. University of KwaZulu-Natal, Pietermaritzburg.
81. Taylor, S.J. B. (1986/99). *A historical analysis of monetary policy rules*. In *Monetary policy rules* (pp. 319-348). *Modelling financial time series*, University of Chicago Press 2<sup>nd</sup> ed., Chichester: Wiley.
82. Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. *Handbook of economic forecasting*, 1, 413-457.
83. Tsay, R. S. (2005). *Analysis of financial time series*. 2<sup>nd</sup> ed., Hoboken, New Jersey: Wiley.
84. Vandaele, W. (1983). *Applied time series and Box-Jenkins models* (No. 04; HB3730, V3.).
85. Výrost, V. G. T. (2003). Application of GARCH Models in forecasting the volatility of the Slovak share index (SAX). *Biatec*, 11, 2.
86. Walton Jr, T. L. (1997/2005). Short term storm surge forecasting. *Journal of Coastal Research*, 421-429.
87. Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., & Ma, J. (2005). Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes.
88. Weisweiler, R. (2017). *Foreign Exchange*. Routledge.
89. Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3), 324-342.
90. Wold, H. (1938). *A study in the analysis of stationary time series* (Doctoral dissertation, Almqvist & Wiksell).
91. Wold, H. (1954). Causality and econometrics. *Econometrica: Journal of the Econometric Society*, 162 – 177.
92. Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931-955

## 11. Appendices

Appendix A: Official gold and foreign exchange reserves positions forecasts from Seasonal ARIMA models and Holt-Winters' method.

Month	Official Gold Reserves Positions Forecasts		FX Reserves Positions Forecasts	
	<i>SARIMA</i> (1, 1, 0)(0, 1, 2) <sub>12</sub>	HW Method With Multi Seasonality	<i>SARIMA</i> (0, 1, 0)(2, 1, 0) <sub>12</sub>	
	Forecast	Forecast	Forecast	
Jun-17	R 65,807	R 65,407	R 545,068	
Jul-17	R 66,084	R 64,764	R 550,289	
Aug-17	R 67,755	R 67,600	R 563,270	
Sep-17	R 68,980	R 67,142	R 582,218	
Oct-17	R 68,499	R 65,969	R 570,899	
Nov-17	R 69,425	R 65,989	R 582,624	
Dec-17	R 69,185	R 64,419	R 615,014	
Jan-18	R 72,546	R 68,382	R 604,029	
Feb-18	R 73,084	R 69,031	R 599,485	
Mar-18	R 72,978	R 67,046	R 592,505	
Apr-18	R 72,260	R 65,335	R 578,127	
May-18	R 74,050	R 66,717	R 605,897	
Jun-18	R 72,452	R 64,988	R 589,183	
Jul-18	R 72,859	R 64,386	R 583,041	
Aug-18	R 74,756	R 67,240	R 601,765	
Sep-18	R 76,269	R 66,818	R 612,565	
Oct-18	R 75,891	R 65,678	R 606,545	
Nov-18	R 77,065	R 65,723	R 621,673	
Dec-18	R 76,927	R 64,183	R 646,880	
Jan-19	R 80,740	R 68,154	R 636,776	
Feb-19	R 81,351	R 68,821	R 627,750	
Mar-19	R 81,194	R 66,861	R 616,343	
Apr-19	R 80,315	R 65,170	R 602,593	
May-19	R 82,406	R 66,564	R 631,844	

Appendix B: Mean and standard deviation forecasts using GARCH(1,1) model with skewed student-t distribution.

<b>Month</b>	<b>Mean Forecast</b>	<b>Mean Error</b>	<b>Std. Deviation</b>
Jun-17	0,0172	0,4747	0,4747
Jul-17	0,0172	0,4851	0,4851
Aug-17	0,0172	0,4948	0,4948
Sep-17	0,0172	0,5039	0,5039
Oct-17	0,0172	0,5126	0,5126
Nov-17	0,0172	0,5207	0,5207
Dec-17	0,0172	0,5283	0,5283
Jan-18	0,0172	0,5356	0,5356
Feb-18	0,0172	0,5424	0,5424
Mar-18	0,0172	0,5489	0,5489
Apr-18	0,0172	0,555	0,555
May-18	0,0172	0,5608	0,5608
Jun-18	0,0172	0,5663	0,5663
Jul-18	0,0172	0,5715	0,5715
Aug-18	0,0172	0,5765	0,5765
Sep-18	0,0172	0,5812	0,5812
Oct-18	0,0172	0,5857	0,5857
Nov-18	0,0172	0,5899	0,5899
Dec-18	0,0172	0,5940	0,5940
Jan-19	0,0172	0,5978	0,5978
Feb-19	0,0172	0,6015	0,6015
Mar-19	0,0172	0,6050	0,6050
Apr-19	0,0172	0,6083	0,6083
May-19	0,0172	0,6115	0,6115



### 11.1 Rstudio Code Used:

```
install.packages('zoo')  
  
install.packages('forecast')  
  
install.packages('tseries')  
  
install.packages('ggplot2')  
  
install.packages("rugarch")  
  
install.packages("fGarch")  
  
install.packages('expsmooth')  
  
install.packages("quantmod")  
  
install.packages("DescTools")
```

```
library("TTR")  
  
library('zoo')  
  
library('forecast')  
  
require('rugarch')  
  
library('ggplot2')  
  
library('tseries')  
  
library('rugarch')  
  
library('fGarch')  
  
library('expsmooth')  
  
library(quantmod)  
  
library(DescTools)
```

```
install.packages("data.table")  
  
library(data.table) #used to install the package setnames
```

```
setnames(Dataset, old=c("Time", "Gold Reserves Position", "Foreign Exchange Reserves Positions", "Gross Reserves Position"), new=c("Time", "Gold Reserves", "FX Reserves", "Gross Reserves"))
```

```
attach(Dataset)
```

```
X<-ts(Dataset$`Gold Reserves Position`*1000/1000000, frequency = 12, start = c(1994,1))
```

```
Y<-ts(Dataset$`Foreign Exchange Reserves Positions`*1000/1000000, frequency = 12, start = c(1994,1))
```

```
Z<-ts(Dataset$`Gross Reserves Position`*1000/1000000, frequency = 12, start = c(1994,1))
```

```
#####
```

```
# GARCH MODELS APPLICATION:
```

```
#####
```

```
#####GOLDRESERVEDATA#####
```

```
# Serial Correlation & Arch Effect Test:
```

```
X.d12.cox=BoxCox(X, lambda =0)
```

```
autoplot(X.d12.cox)
```

```
X.d12 = diff(X.d12.cox, lag=1)
```

```
X.d112=diff(X.d12, lag=12)
```

```
hist(X.d112, main="", breaks=20, freq=FALSE, col="grey")
```

```
fit.X = Arima(X, order = c(1,1,0), seasonal=list(order=c(0,1,2),period=12), lambda =0, include.mean=FALSE, include.constant=FALSE, method = "CSS-ML")
```

```
Box.test(fit.X$residuals, lag = 24, type = "Ljung-Box")
```

```
Box.test(fit.X$residuals^2, lag = 24, type = "Ljung-Box")
```

```

# Fitting ARMA+GARCH Models:

goldgarch=garchFit(formula = ~arma(1,1)+garch(1,0),data =fit.X$residuals, include.mean =
TRUE, trace=F, include.skew = NULL, include.shape = NULL, cond.dist = "norm")

MeanAD(residuals(fxgarch), FUN = mean, na.rm = FALSE)

summary(goldgarch)

plot(goldgarch)

predgold = predict(goldgarch, n.ahead = 24)

Gold_Spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1,0)),
mean.model = list(armaOrder = c(1,1)), distribution.model = "norm")

Gold_fit <- ugarchfit(spec = Gold_Spec, data = X )

Predgold<- ugarchboot(Gold_fit, n.ahead=10, method= c("partial", "full")[1] )

$$$$$$$$$$$$$$$$$$$$$$$$$$$$FOREIGNRESERVESDATA$$$$$$$$$$$$$$$$$$$$$$$$$$$$

# Serial Correlation & Arch Effect Test:

Y.d12.cox=BoxCox(Y, lambda = "auto")

autoplot(Y.d12.cox)

Y.d12 = diff(Y.d12.cox, lag=1)

Y.d112=diff(Y.d12, lag=12)

fit.Y = Arima(Y, order = c(0,1,0), seasonal=list(order=c(2,1,0),period=12), lambda = "auto",
include.mean=FALSE, include.constant=FALSE, method = "CSS-ML")

Box.test(fit.Y$residuals, lag = 24, type = "Ljung-Box")

Box.test(fit.Y$residuals^2, lag = 24, type = "Ljung-Box")

fxgarch=garchFit(formula = ~garch(1,1),data =fit.Y$residuals, include.mean = TRUE, trace=F,
include.skew = NULL, include.shape = NULL, cond.dist = "sstd")

MAD(residuals(fxgarch), na.rm = FALSE)

MeanAD(residuals(fxgarch), FUN = mean, na.rm = FALSE)

summary(fxgarch)

round(fxgarch@fit$matcoef, 3)

```

```
plot(fxgarch)

predfx = predict(fxgarch, n.ahead = 24)

predfx1= round(predfx, 4)

plot(predfx1$meanForecast, main="Mean Forecast Plot", ylab="MeanForecast", col="blue",
lwd=2)

plot(predfx1$meanError, main="Mean Error Plot", ylab="MeanError",col="blue", lwd=2)

plot(predfx1$standardDeviation,          main="Standard          Deviation          Plot",
ylab="StandardDeviation",col="blue", lwd=2)
```