UNIVERSITY OF KWAZULU-NATAL

Modelling of Volatility in the South African Mining Sector: Application of ARCH and GARCH Models

By

Rugiranka Tony Gaston

A Thesis Submitted in fulfilment of
the academic requirements for the Degree of
Master of Science
in
Statistics

under the supervision of
Prof S. Ramroop
and
Prof H.G. Mwambi

School of Mathematics, Statistics and Computer Science
Pietermaritzburg Campus
South Africa
2016

Declaration

The research work described in this thesis was carried out under the supervision of Prof Shaun Ramroop and the co-supervision Prof H.G. 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 it is duly acknowledged in the text.

Mr Rugiranka Tony Gaston	Date
Prof Shaun Ramroop	Date
Prof Henry G. Mwambi	Date

Acknowledgements

Thanks to the Almighty God, for His grace and unconditional love that helped me to complete this thesis. I would also like to thank my supervisors, Prof Shaun Ramroop and Prof Henry Mwambi, for their help and faith in me while working on this dissertation. This work would not have been done without their valuable input and advice. Thanks to Dr Rose Kuhn, Michael Bowes and Michele Bowes for their unlimited support and their unconditional love as parents. Thanks go to all my friends and colleagues who supported me and remained close to me during this time, especially Dr Moise Majyambere, Innocent Hakizimana Abubakar, Dr Jean Baptiste Twayigira, Eric Bariraje, Lambert Nizeyimana, Emmanuel Nshimiyimana, Chantal Ingabire, Jean Chrysostome Kanamugire, Seth Hakizimana, Olivier Niyitegeka, Celeste Johns, Thando Sibiya, Jean Claude Mugisha, Jean Claude Manariyo, Andile Gumede, Edna Manda, Donna Ongoma, Patrick Ndayizigamiye, Victor Bangamwabo, Alexis Munyengabe and others who have made huge contributions to my studies. I would like to thank all UKZN Pietermaritzburg library staff members, Department of Student Housing and School of Mathematics, Statistics and Computer Science for providing an enjoyable and inspiring working environment. Last but not least, thanks to my parents and family members for keeping me in their prayers and lives regardless of me being away. I would like to extend my deepest gratitude to Dr Paul François Muzindutsi who has always been a good brother, showing support, care and love.

Abstract

The primary objectives of this study were to model the volatility in the Johannesburg Stock Exchange (JSE) using the two mining companies' data. The companies are Impala Platinum Holdings Limited and Harmony Gold Limited and data set used was the daily closing price for the period of 3 January 1995 to 3 July 2014. The data vectors consist of log transformed returns of the daily closing price. The methods used to model this volatility were the Autoregressive Conditional Heteroscedasticity (ARCH), Generalized ARCH (GARCH), Exponential GARCH (EGARCH), Asymmetric Power ARCH (APARCH) and Glosten, Jagannathan, Runkle GARCH (GJR-GARCH) models for univariate models, Dynamic Conditional Correlation (DCC) GARCH model for multivariate and stochastic volatility (SV) models. Before the univariate GARCH models used, the presence of ARCH effect should be checked and this was found in both mining companies. The autocorrelation in the mean was present as well, which could then be removed by modeling the AR (Autoregresive) model, and hence using the residual to run GARCH models. The AR1+GJR-GARCH(1,2) model with Skewed Student-t distribution (sstd) error was the best fit model for Harmony Gold Ltd, while the AR2+EGARCH (1,2) model with Student-t distribution (std) was found to be the best fit model for Impala Platinum Holdings Ltd. The results reveal that there is a high volatility persistence in both mining companies with higher volatility in the Harmony Gold Limited mining company. The results from the stochastic volatility models were in agreement with those obtained from the univariate GARCH models. The Dynamic Conditional Correlation (DCC) GARCH model was also employed between the two daily data mining companies within the same period as the one from GARCH type models and stochastic volatility. The overall finding indicated that the correlation between the two mining companies was varied over time.

Dedication

This thesis is dedicated to my brothers, Jean Claude Kubwimana and Gaspard Maniraho, for their guidance, support and love during the time we were together. I will always wish them the best wherever there are.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.0.1 Background

South Africa is one of the world leaders in the mining Industry. The country is well-known for its richness of mineral resources, accounting for a significant proportion of world production and reserves. The economy of South Africa is built on diamond and gold mining, with gold accounting for more than one-third of exports (Sorensen 2011). South Africa is also a primary mine of coal and manufacturer of manganese and chrome(Yager 2004). The discovery of gold on Witwatersrand led to the founding of the Johannesburg Stock Exchange (JSE) in 1887 (Muzindutsi 2011). The launch of the JSE facilitated the process of raising capital for the development of the gold mines (Jefferis & Okeahalam 1999). The JSE is heavily influenced by the treasures of the mining sector (Jefferis & Okeahalam 1999) and as a result, South African companies are central participants in the worldwide industry.

The South African mining sector has been volatile in recent years and this volatility is as a consequence of both endogenous and exogenous factors. One example of these variables triggering the volatility in the mining sector is the high labour costs, which are motivated by the high number of strikes in South Africa (Prno & Slocombe 2012). This high level of volatility in the mining sectors is associated with heightened risks for investors interested in this sector, since the volatility can be considered as a fundamental factor in the global financial market. This volatility is related to the risk that can be taken in order to obtain a reward (Neokosmidis et al. 2009). Risk and rewards are correlated with each other, and it is necessary to generate with a method that can minimize the risk and maximize the reward, resulting in a positive return. Risk is determined by the variance and the square root of the variance is known as volatility. This can be shown by Black & Scholes' (1973) model which is employed to judge the value of options in financial derivatives. The higher the volatility the higher the risk since volatility changes over time. This implies that unidentified volatility may lead to economic losses for investors which eventually affect the job security in the mining sector. Hence, this study focuses on the identifying and modelling the volatility in the South African mining sector, through the application of appropriate statistical models.

This method of modelling and forecasting volatility of financial time series is powerful in financial decisions, such as portfolio selection, option pricing, monetary policy making and risk management. Since modelling volatility of time series is one of the primary factors to determine the option price for stock and stock indices, this process of modelling volatility may also improve the efficiency of interval forecast (Liu 2009). Modelling and forecasting volatility is also useful to identify the best model and to improve the accuracy of volatility forecasting. Volatility is characterised by the following three factors:

Volatility Clustering

This phenomenon of the volatility clustering indicated by Mandelbrot (1963) alludes to the fact that small price changes were followed by small prices changed of either sign or large price changes were followed by large price changes of either sign. This was later confirmed by Fama (1965), who stated that the uncertainty of stock prices vary with time where the variance is used as a measure. This volatility clustering is also considered as the period where there is a wide swing in prices for an extended time period followed by a relatively calm period. This means that the variance of financial time series varies over time (Engle 2001, Gujarati 2004).

Volatility Persistence

The concepts of volatility persistence and volatility clustering are closer. The volatility clustering implies volatility persistence (Engle et al. 2001). The volatility persistence can also defined, by considering the expected value of variance of returns l period in the future to be defined as

$$\sigma_{t+l}/t \equiv E_t[(r_{t+l} - \epsilon_{t+l})^2].$$

As a result, the forecast of future volatility depends upon information in today's information set such as today's returns. In this case, the volatility persistence can be stated as the uncertainty today's return, which has a large effect on the forecast variance several periods in the future (Engle et al. 2001).

Leverage effect

Leverage effect is a phenomenon often observed in return series and Black (1976) explains that this phenomenon occurs when the negative shock for stock prices has a greater impact than positive shock with a change in volatility. The leverage effect implies that the relationship between returns and volatility is asymmetric (Kirchgässner et al. 2012). In this case, when bad news occurs in the market, there is a decrease in stock price.

There is much literature concerning the leverage effect, and the reader is referred to Andersen & Bollerslev (1997), Lux & Marchesi (2000), Friedmann & Sanddorf-Köhle (2002), Iori (2002), Siklos & Skoczylas (2002), Jacobsen & Dannenburg (2003), Cont (2007), Bentes et al. (2008), Kang & Yoon (2009), McMillan & Ruiz (2009), Thupayagale (2010), Tseng & Li (2011), etc.
Modelling and forecasting in finance and economic time series was based on the conditional first moments, and dependency on higher moments was treated as a nuisance (Bollerslev et al. 1994). By using the ARMA model, there is an assumption of stationarity, and this assumption implies that the time series exhibits a constant variance, which means that the variance remains constant over time. However, many financial time series are often covariance non-stationary. The importance of modelling risk has recently increased, which has led to the development

of models to allow for time-varying variance and covariance (Bollerslev et al. 1994). In general, there are two main classes of volatility models. First are the observation-driven models, which are defined as the conditional variance as a function of past observations of the returns and these are ARCH-type models. Second are the parameter-driven models, where the conditional variance is specified as an unobserved component that conforms to some underlying latent stochastic process (Talke 2003). The parameter-driven models are known as Stochastic Volatility models (Taylor 1986).

The history of ARCH models is well known owing to the work of Engle (1982), as he introduced these models. The ARCH model was introduced to model the change in volatility over time. This means that the variance of distribution change over time and the large change period tends to be followed by a large period, while small change periods are assumed to be followed by a small change period of either positive or negative sign. Consider the AR (q) or the ARMA (q,0,0) model which is the same as the ARCH (q) model, is given by the following equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \omega_t,$$

where σ_t^2 is the measure of volatility, parameter $\alpha_i \geq 0$ for $i = 0, ..., p$ and $\alpha_1 + ... + \alpha_q < 1$, ω_t is white noise (error term) and ϵ_{t-i}^2 is the initial condition variance. Thus, this model postulates that volatility in the current period is related to its value in the previous period plus a white noise error term ω_t. This means that if α_1 was positive, it suggests that volatility was high in the previous period, then it will continue to be high in the current period and this is an indication of volatility clustering. If α_1 is zero, there is no volatility clustering, and this statistical significance test of estimated α_1 can be found by using t-test or F-test (Gujarati 2004).

However, while the ARCH model is the best model for modelling and forecasting volatility. It has some disadvantages such as the requirement of many parameters to fit the data. The GARCH model, which is the Generalized ARCH model was developed independently by Bollerslev (1986), Taylor (1986) to overcome the shortcoming of ARCH models. The general equation of GARCH (p,q) model is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{1.1}$$

Therefore, the extensions of GARCH model were introduced since the GARCH model was not able to detect the leverage effect, which is the asymmetric impact on volatility (Black 1976, Gouriéroux 1997, Tsay 2010, Daníelsson 2011). The existing model of volatility is characterized by symmetric and asymmetric models. In the symmetric models, the conditional variance depends only on the magnitude, and not on the sign of the underlying asset, while in the asymmetric models the shocks of the same magnitude, positive or negative, have different effects on future volatility (Ahmed & Suliman 2011). More explanation and application are given in subsequent chapters of the thesis. Thus, some extensions of GARCH models are the following:

1. EGARCH model (Exponential GARCH model), which was the first model to take into account the leverage effect proposed by Nelson (1991).

2. GJR-GARCH model (Glosten, Jagannathan, and Runkle GARCH model) which was developed by Glosten et al. (1993) to capture the leverage effects.

3. PGARCH model (Power GARCH model) which was proposed by Taylor (1986) then later by Schwert (1989) to capture any power of the absolute returns.

4. APARCH model (Asymmetric Power GARCH model) which was introduced by Ding et al. (1993) and which allows both power effect and leverage effects.

5. TGARCH model (Threshold GARCH model) which was introduced by Zakoian (1994) to capture leverage effects.

6. IGARCH model (Integrated GARCH model) which was introduced by Engle & Bollerslev (1986).

7. AGARCH model (Asymmetric GARCH model) which was introduced by Engle (1990) to allow for asymmetric effect of negative and positive innovations.

After developing ARCH, GARCH models and their extensions, modelling and forecasting the volatility of financial and economics time series has become the pivotal area of research. Forecasting volatility can be considered as a measurement for financial decision, such as portfolio selection, option pricing, risk management and monetary policy making (Liu 2009, Bekaert et al. 2013). This method of modelling and forecasting volatility can help us to identify the best model and to improve the accuracy of volatility forecasting.

1.1 Literature Review

In recent years, there has been a vast amount of literature on modelling and forecasting stock market volatility in both developed and developing countries around the world. After Engle (1982) developed the ARCH model and Bollerslev (1986) developed its extension, the Generalised ARCH model's many extensions were introduced and a large number of researchers were interested in using ARCH, GARCH models, and extensions of GARCH models to model and forecast volatility in financial environments and economic. Finding of research studies which used ARCH, GARCH models and their extensions to model and forecast stock market volatility are as follows:

Peters (2001) used GARCH, EGARCH, GJR and APARCH models with three distributions (normal, student-t and skewed student-t), to model and forecast volatility of two major European stock indices (FTSE100 and DAX30). The results suggested that the asymmetric models perform better than symmetric models, and within asymmetric models, EGARCH model performs better than GJR and APARCH models. The results also suggested that the non-normal distribution provides better in-sample results than a normal distribution.

Talke (2003) used ARCH, GARCH models and stochastic volatility models to model volatility in financial time series and to model the volatility of the exchange rate of the South African Rand against three foreign currencies (the US Dollar, the Swiss Franc and the UK Pound) all on a daily, weekly and monthly basis. In general, the GARCH models were found to be superior to the ARCH models for all sampling frequencies with an exception of the weekly UK Pound against SA Rand and monthly Swiss Franc against SA Rand data. Furthermore, the ARCH and GARCH models compared to stochastic volatility models were found to be the best in describing the stylized facts of asset returns for the daily and the weekly US Dollar against SA Rand and the stochastic volatility models were found to better in describing the stylized asset returns for the monthly US Dollar against the SA Rand. Finally, in times of high volatility the stochastic volatility models performed better than the ARCH and GARCH models.

Alberg et al. (2008) used GARCH, EGARCH and GJR-GARCH models to investigate the performance of volatility forecasting for standard Poor's 100 stock index series with three different types of distributions (normal, student-t and skewed generalized error distribution (SGED)). Their empirical results showed that SGED and student-t distributions performed slightly better than a normal distribution, and GJR-GARCH model was the best to achieve the most accurate volatility forecasts.

Liu & Hung (2010) used GARCH, EGARCH and GJR-GARCH models to investigate the performance of volatility forecasting for standard Poor's 100 stock index series with three different types of distributions (normal, student-t and skewed generalized error distribution(SGED)). Their empirical results showed that the GJR-GARCH model with SGED and student-t distributions performed better to generate the accurate volatility forecasts.

Su (2010) used GARCH and EGARCH models to estimate the financial volatility of daily returns in China. They analyzed whether the long-term volatility was more extensive during the crisis period than before the crisis, and they also compared the movements of the return volatility of the Chinese stock market to the other stock markets before and throughout the crisis period. The empirical results suggested that the EGARCH model fit the sample data better than the GARCH model in modelling the volatility of Chinese Stock returns. The result also showed that the long-term volatility was more volatile during the crisis period.

Dralle (2011) used ARCH and GARCH models compared to stochastic volatility model, in his topic of modelling the volatility of financial time series data. The data set used in fitting the ARCH, GARCH and Stochastic volatility models was daily closing prices of four gold mining companies registered on the JSE. The ARCH, GARCH models were fitted with normal and student-t distributions errors, and the model with student-t distribution was the best fit to all four data set. Hence the results from ARCH, GARCH and Stochastic volatility models were the same, in showing the high persistence volatility in all four gold mining companies. However, Stochastic volatility model is restricted to the form of an AR(1) (Autoregressive(1)) owing to its complication of fitting higher order modes.

Goudarzi & Ramanarayanan (2011) used GARCH, EGARCH and GJR-GARCH models in modelling asymmetric volatility in the Indian stock market. The BSE500 stock index was used, and their results showed that the TGARCH (1,1) model can be a good model of the asymmetric conditional volatility process for daily returns series of BSE500.

Abdalla & Winker (2012), used GARCH, GARCH-M, TGARCH and P-GARCH models to model and estimate stock return volatility in two African markets. First was a Sudanese stock market (Khartoum Stock Exchange, KSE) and the second was an Egyptian stock market (Cairo and Alexandria Stock EXCHANGE, CASE). The result from GARCH (1,1) indicates that the conditional volatility of KSE was an explosive process, while it was quite persistent for the CASE index return series. The results from GARCH-M (1,1) showed that there is statistically significant in both KSE and CASE. Finally, using asymmetrical models, the EGARCH and TGARCH models, the results show that there is significant evidence for the existence of the leverage effect in KSE and CASE returns index series. Thus, the PGARCH model provided the same results except for CASE.

Hajizadeh et al. (2012) used GARCH, EGARCH and GJR-GARCH models in a hybrid modelling approach for forecasting the volatility of S &P 500 index returns. The two hybrid models based on EGARCH and Artificial Neural Networks were considered, and their results showed that the EGARCH (3,3) model was the best compared to the second model.

Jiang (2012) used GARCH model, EGARCH model and GJR-GARCH model (named by Glosten, Jagannathan, and Runkle (1993)) to analyze and predict the different stock markets, under the normal and t-distribution errors. They selected five global stock market indices: NASDAQ's daily index, Standard Poor's 500 daily index, FTSE100 daily indices, HANG SENG daily index and NIKKEI daily index. To predict the conditional variance, the best model was chosen based on the model with the smallest Root Mean Square Error (RMSE). The results of the NASDAQ stock return, show that GJR-GARCH (2,2) model under the normal distribution is a better model to forecast the future conditional variance since is the one with smallest RMSE. For SP500 stock return, GJR-GARCH (2,1) model under the normal distribution, is the best model to predict the future conditional variance. For the FTSE100 stock return, GJR-GARCH (2,2) model under the normal distribution, is the best model to forecast the future conditional variance since it has the smallest RMSE. For NIKKEI stock return, GJR-GARCH (1,1) model under normal distribution, is a better model to forecast the future conditional variance. For HANG SENG stock return, GARCH (1,1) model under the normal distribution, is the best model than others to forecast the future conditional variance.

Mandimika & Chinzara (2012) in their study, analyzed the behaviour of volatility on the South African stock market. Three models across three error distributions were compared, and only two models were used since GARCH-in-Mean is symmetric. The asymmetric models used are TARCH-M and EGARCH-M models, and the TARCH-M model under the assumption of Generalized Error Distribution (GED) was the most appropriate model for the conditional volatility of the South African stock market. In their findings volatility was also largely persistent and asymmetric. The risk at both the aggregate and the disaggregate level was generally not a priced factor in the South African stock market and finally, volatility generally increases over time and its trend structurally breaks during financial crises and major global shocks.

A time series analysis of the Stock NSE 20-share index from January 1998 to March 2007 was used Achia et al. (2013). They tested volatility for the period of election in 2000-2003 and the period prior to and after electoral season 1998-2001 and 2004-2006. They found that the volatility had a positive serial correlation in the market as they well expected. They also tested whether the efficient market hypothesis (EMH) held in the case of the NSE20 share index, checking ARMA (1,1,1), GARCH (1,1) and the random walk process under EMH. They found that the hypothesis was not satisfied, ARMA (1,1,1) and GARCH (1,1) model are fitted to the data and the random walk does not hold under EMH in the data.

In the study by Mzamane (2013), univariate and multivariate GARCH models were proposed to extend the volatility model in the JSE index. The daily log-returns of the JSE index were used for the time frame between 6 June 1995 to 30 June 2012. Mzamane intended to investigate the volatility in the market using GARCH, GJR-GARCH, EGARCH and APARCH models. The findings shows that the GJR-GARCH model was suitable for the detection of volatility in the JSE index, where the volatility in the residuals and leverage effect was present in JSE index returns. In addition, the multivariate GARCH models were used to explore the dynamics of the correlation between the JSE index, FTSE-100 and NASDAQ-100 index on the basis of weekly returns from 6 June 1995 to 30 June 2012. The findings revealed that the correlation was dynamic and persistent.

1.2 Objectives

The primary objectives of this study are to review the statistical properties of the GARCH models and to use these models to study volatility in the South African mining sector. The secondary objectives of the study are to:

a) Determine the best fitting ARCH-type model and stochastic volatility models for data from the South African Mining Sector with daily data.

b) Compare the ARCH models and their extensions along with the stochastic volatility model.

c) Identify the volatility of JSE Companies in mining sector using ARCH-type models, the DCC-GARCH and stochastic volatility models.

1.2.1 Chapter Summary

This chapter introduced to the study, how South African Mining companies are compared to other countries in the world. It explained how South African companies are key players in global industries. It also explained various types of volatility, which is the main aim of the study. In this chapter the theory and the history of ARCH, GARCH, and its extensions was discussed, which have been used in the application of the data analysis. It also included a brief discussion of previous studies, how they modelled volatility using ARCH, GARCH models and their extensions with different distributions. The previous studies about modelling volatility using GARCH models provided insight into procedures to follow. Finally, this chapter explained the objectives of the study, which have been evaluated at the end.

Chapter 2

Exploratory Data Analysis

This Chapter comprises two sections, where the first section is the data description, and the second section is the data exploration. In section two the test and different plots for each company are presented. The various software programmes could be used such as: SAS, EVIEWS, SPLUS, and R softwares just to name a few. In this Chapter R and EVIEWS softwares were suggested as the first option to be used (Talke 2003, Tsay 2005). However R was the main focus in comparison to EVIEWS because:

- EVIEWS does not have provision for studentilized-t and skewed t options, while R has all these options.
- EVIEWS is only available through purchase while R is easily downloadable as it is freeware.
- EVIEWS does not include JGR-GARCH and IGARCH models, while R comes with many optional packages.
- Graphics for R are simple and precise unlike those from EVIEWS, which do not clearly show details.
- It is easy to extend with user-defined functions in R unlike in EVIEWS.

For more details about R software see Cryer & Chan (2008), Shumway & Stoffer (2010) .

2.1 Data Description

In this section, two data sets were employed to model volatility in the JSE Companies, using ARCH, GARCH, multivariate and Stochastic volatility models. The data used are daily closing price for each company and the selected mining companies are registered on the JSE market company. The companies selected are Harmony Gold Mining Limited and Impala Platinum Holdings Limited. The data for the JSE closing price were obtained with permission from the McGregor BFA website (http:// research.mcgregorbfa.com/default.aspx). The return series is preferential to using the actual price series since the return series is manageable and has more attractive statistical properties compared to the price series. The simple return is defined as

$$\hat{R}_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{2.1}$$

where P_t and P_{t-1} are the exchange rate, closing prices in the period t and $t-1$ respectively (Ruppert 2011). The log-returns has been used in the data analysis, since the log-returns has more tractable statistical properties and the data tend to be multiplicative (Tsay 2010). Thus the equation of log-returns given by

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}), \tag{2.2}$$

where R_t is the daily returns for period t, P_t and P_{t-1} are all shares companies for days t and $t-1$ respectively, while ln is the natural logarithm.

2.2 Data Exploration

2.2.1 Harmony Gold Mining Limited

The data for Harmony Gold Mining Limited used in this study consist of a daily closing price from 3 January 1995 to 3 July 2014. This yields a total of 4870 time series observations. The log-returns were used to analyze 4869 time series observations because one observation was lost due to differencing the daily closing price series.

Table 2.1: Descriptive Statistic for Harmony Gold Mining Daily Returns

Results in	R
Descriptive Statistics	Value
nobs	4869
Minimum	-0.2287
Maximum	0.1728
Mean	0.0000
Median	0.0000
Skewness	-0.3008
Stdev	0.0356
Kurtosis	3.7513

Table 2.2: Harmony Gold Mining Daily Log Returns Test for Normality

Test	Statistic	P-value
Jarque-Bera test	4339.4467	$< 2.2e^{-16}$
Shapiro-Wilk	0.9532	$< 2.2e^{-16}$

The descriptive statistics are given in Table 2.1. The result is clearly seen that the standard deviation is 3.6%, this implies a high level of volatility in the market. The average daily return is 0%. The minimum is -0.2287, and maximum is 0.1728, and this wide gap between minimum and maximum support the high variability of prices change in the JSE mining sector of the Harmony Gold Mining company. To check the normality assumption, the Jarque-Bera (J-B) test, Shapiro-Wilk test, skew and kurtosis can be considered. Under the null hypothesis of normal distribution, the J-B, and Shapiro-Wilk are 0, skewness is 0, and kurtosis is 3. In the series when there is a positive or a negative skewness distribution, this is an indication of asymmetry in returns data, and when the coefficient of kurtosis is less than or greater than 3, this suggests flatness and peakedness respectively. Table 2.1 shows a negative skew -0.3008, and this implies that the distribution has a long left tail and a deviation from normality.

The coefficient of kurtosis 3.7513, implies that the empirical distribution is not normal but peaked. This is supported by highly significant J-B and Shapiro-Wilk statistics in Table 2.2, and also Figure 2.4, which reveals that there is access of kurtosis. Figure 2.1 reveals that there is a period of large price movements in the data and small price movements. This is an indication of volatility clustering in the Harmony Gold Mining company.

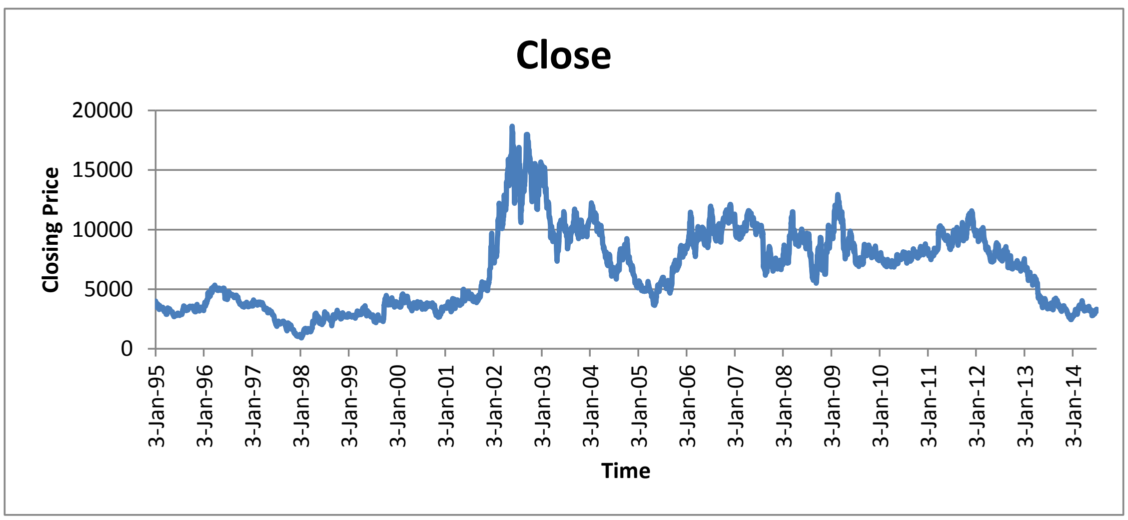


Figure 2.1: Harmony Gold Mining Daily Closing Prices in Rand

Figure 2.2 reveals that the data does not come from the normal distribution since the points do not fall approximately along the 45-degree reference line. As a result it can be concluded that the JSE mining sector of Harmony Gold Mining Limited does not conform to the normal distribution. Figure 2.3 it is clearly seen that there is evidence of volatility clustering in the Harmony Gold Mining Limited company. Moreover, Figure 2.5 reveals that there is some evidence of high volatility clustering in the years of 1998 and 2012. There is also some period of medium, and stable volatility clustering in the Harmony Gold Mining company.

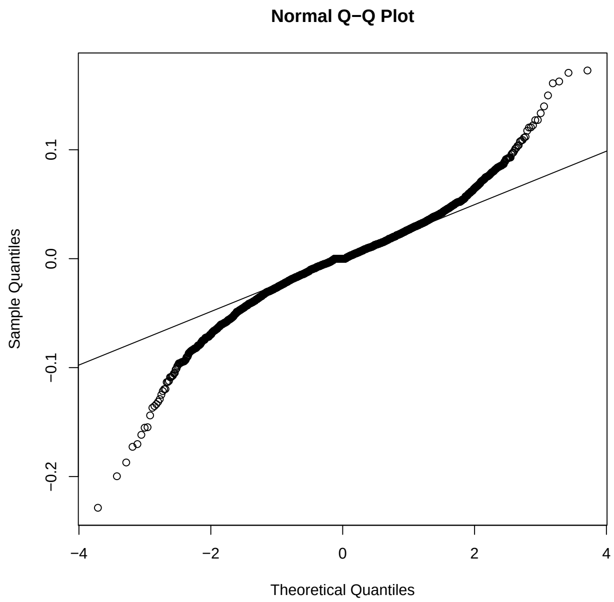


Figure 2.2: Q-Q Plot of Daily Returns for Harmony Gold Mining

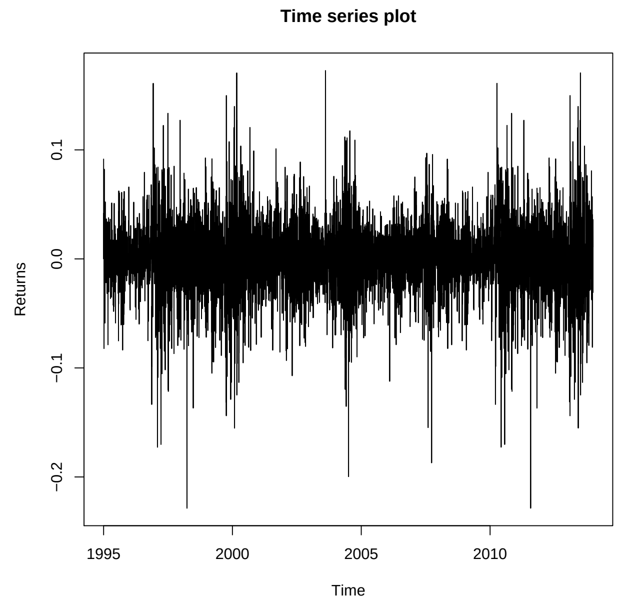


Figure 2.3: Harmony Gold Mining Daily Returns

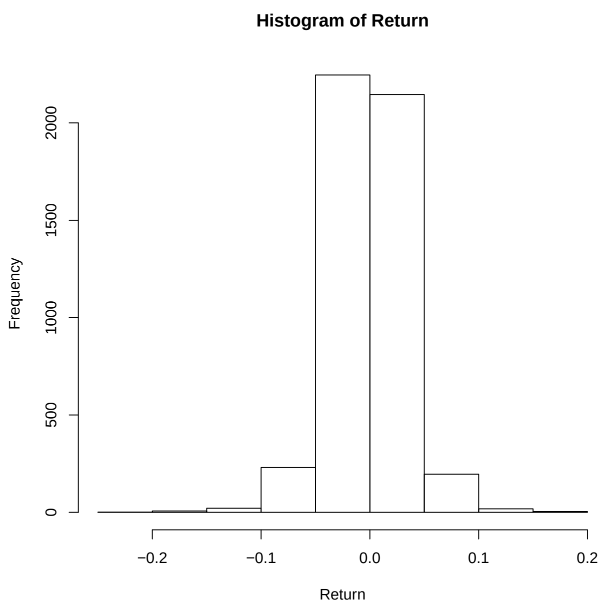


Figure 2.4: Histogram of Daily Returns for Harmony Gold Mining

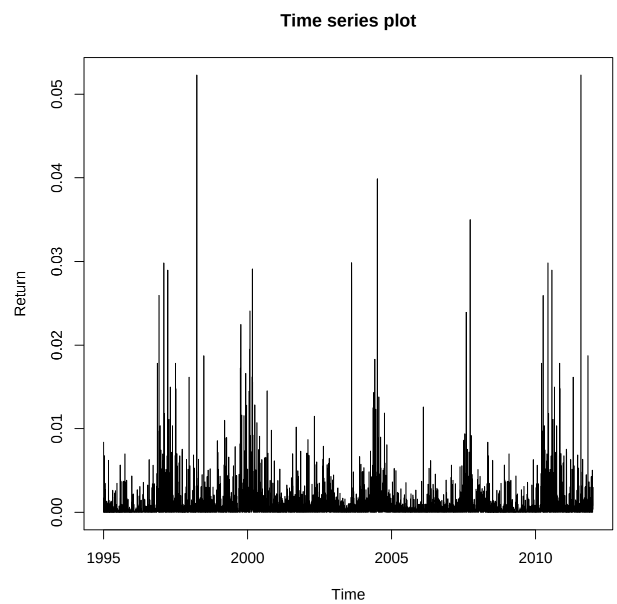


Figure 2.5: Harmony Gold Mining Daily Squared Returns

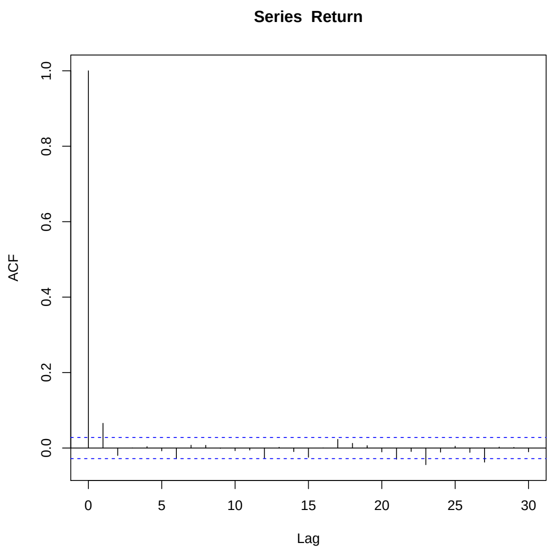


Figure 2.6: ACF for Harmony Gold Mining Daily Returns

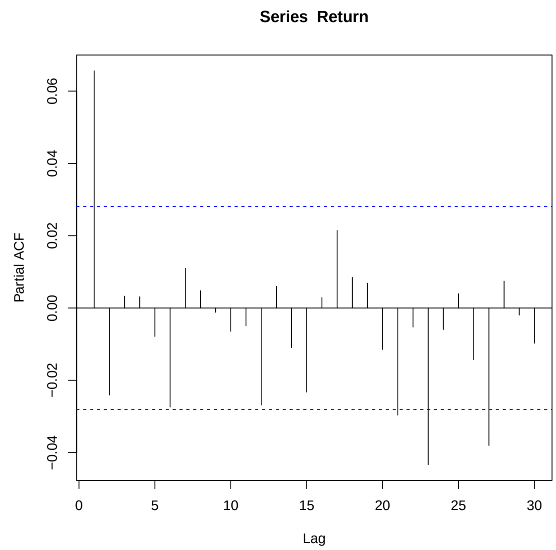


Figure 2.7: PACF for Harmony Gold Mining Daily Returns

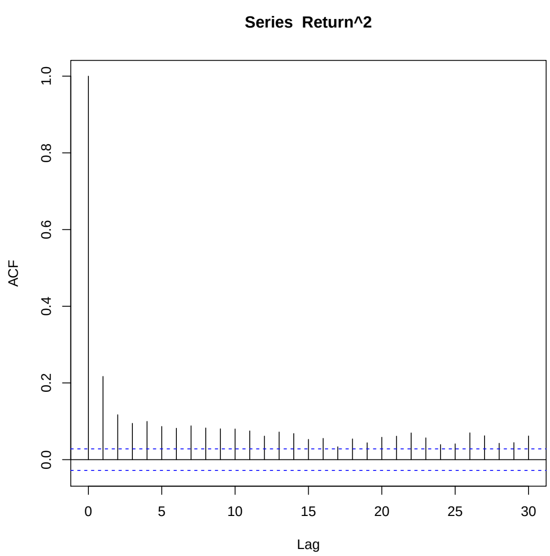


Figure 2.8: ACF for Harmony Gold Mining Daily Squared Returns

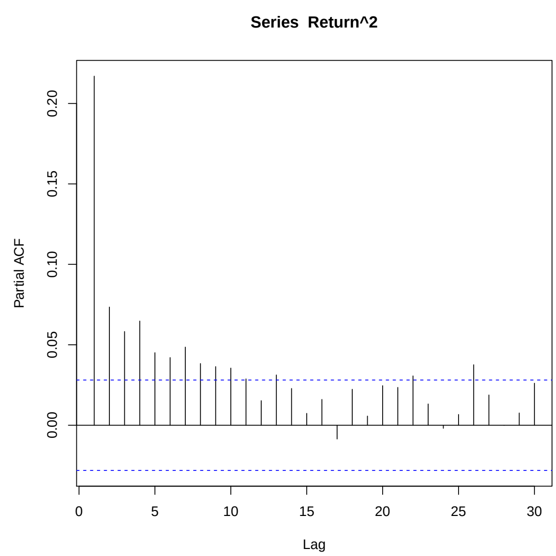


Figure 2.9: PACF for Harmony Gold Mining Daily Squared Returns

Figure 2.6 and 2.7 respectively reveal that the autocorrelation (ACF) of the returns exhibit a minor serial correlation at lags 1, 21 and 27, while the partial autocorrelation (PACF) has a significant spike at lags 1, 21 and 27. Figure 2.8 and 2.9 respectively, show that both ACF and PACF have a significant spike which is the indication of presence of ARCH effect in the data.

Table 2.3: Box-Ljung Q-Statistic for Autocorrelation

Lags	Critical-value	Statistic	p-value
5	11.0710	23.2080	0.0003
10	18.3070	27.9224	0.0019
15	24.9960	34.8455	0.0026
20	31.4100	38.9232	0.0068
25	37.6520	53.8398	0.0007
30	43.7730	61.8758	0.0005

Table 2.4: Engle's ARCH Test for Heteroscedasticity

Lags	Critical-value	Statistic	p-value
5	11.0710	299.9200	$< 2.2e^{-16}$
10	18.3070	338.5100	$< 2.2e^{-16}$
15	24.9960	350.5500	$< 2.2e^{-16}$
20	31.4100	356.9400	$< 2.2e^{-16}$
25	37.6520	364.8400	$< 2.2e^{-16}$
30	43.7730	375.6700	$< 2.2e^{-16}$

Table 2.3 shows a high test statistic for all lags compared to its critical-values, and the corresponding p-values are less than 0.005 significance level. Thus the null hypothesis rejected, which says that there is no autocorrelation and fail to reject the alternative hypothesis which says that there is autocorrelation in the returns data.

Table 2.4 shows higher test statistics compared to their corresponding critical values, and the p-values to all lags are less than 0.005 significance level. Thus the null hypothesis was rejected and it was concluded that there is heteroscedasticity in the returns data.

2.2.2 Impala Platinum Holdings Limited

The data used for Impala Platinum Holdings Limited was a daily closing price from 3 January 1995 to 3 July 2014. The total of 4870 observations during the period of sampling was considered. Table 2.5 shows a higher level of volatility in the market as it is indicated by the higher value of standard deviation. There is also a wide gap between minimum and maximum and this support the high variability of change in price for the JSE mining sector of Impala Platinum Holdings Limited company. Table 2.5 shows descriptive statistics where a positive skew 0.2301, implies that the distribution has a long right tail and a deviation from normality. The coefficient of kurtosis 4.4943 implies that the empirical distribution is not normal but peaked. This is supported by highly significant J-B and Shapiro-Wilk statistics in Table 2.6, and also in Figure 2.10, reveals that there is an excess of kurtosis.

Table 2.5: Descriptive Statistic for Impala Platinum Holdings Limited Daily Returns

Results in	R
Descriptive Statistics	Value
nobs	4869
Minimum	-0.1502
Maximum	0.2412
Mean	-0.0005
Median	0.0000
Skewness	0.2301
Stdev	0.0272
Kurtosis	4.4943

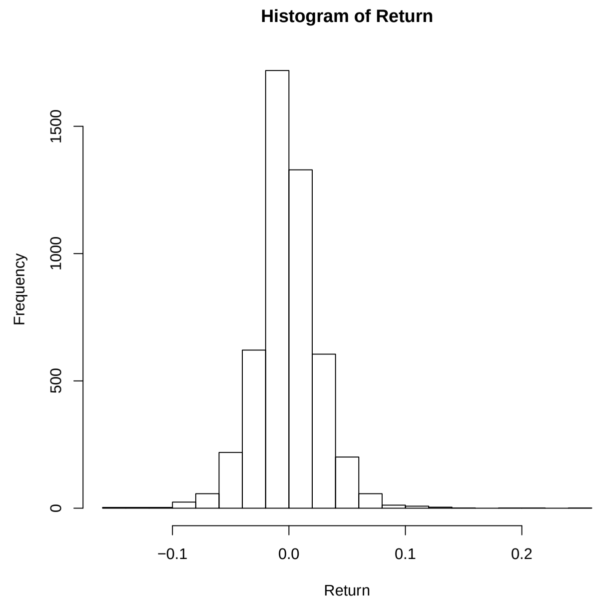


Figure 2.10: Histogram of Daily Returns for Impala Platinum Holdings Limited

Table 2.6: Impala Platinum Holdings Limited Daily Log Returns Test for Normality

Test	Statistic	P-value
Jarque-Bera test	3076.9831	$<2.2e^{-16}$
Shapiro-Wilk	0.9636	$<2.2e^{-16}$

Table 2.7 shows a higher test statistic for all lags to compare to its critical values, and the corresponding p-values are less than 0.005 significance level. Thus the null hypothesis rejected, which says that there is no autocorrelation and accept the alternative hypothesis which says that there is autocorrelation in the returns data. Finally Table 2.8 shows higher test statistics to compare to their corresponding critical values, and the p-values for all lags are less than 0.005 significance level. The null hypothesis is rejected and thus conclude that there is heteroscedasticity in the returns data.

Table 2.7: Box-Ljung Q-Statistic for Autocorrelation

Lags	Critical-value	Statistic	p-value
5	11.0700	77.0143	$<3.553e^{-15}$
10	18.3070	85.8087	$< 3.619e^{-14}$
15	24.9960	98.3439	$<2.687e^{-14}$
20	31.4100	104.9814	$<1.601e^{-13}$
25	37.6520	114.0874	$<2.409e^{-13}$
30	43.7730	115.1738	$<6.492e^{-12}$

Table 2.8: Engle's ARCH Test for Heteroscedasticity

Lags	Critical-value	Statistic	p-value
5	11.0700	266.6038	$<2.2e^{-16}$
10	18.3070	294.1367	$< 2.2e^{-16}$
15	24.9960	311.8893	$< 2.2e^{-16}$
20	31.4100	330.5144	$< 2.2e^{-16}$
25	37.6520	335.4408	$< 2.2e^{-16}$
30	43.7730	348.5383	$< 2.2e^{-16}$

Figure 2.13 also confirms that the data does not come from the normal distribution since the points do not fall approximately along the 45-degree reference line. In Figure 2.11 it can be seen that there is evidence of volatility clustering in the Impala Platinum Holdings Limited company. Moreover, Figure 2.12, reveals that there is some evidence of high volatility clustering, also some periods of medium, and stable volatility clustering in the Impala Platinum Holdings Limited company.

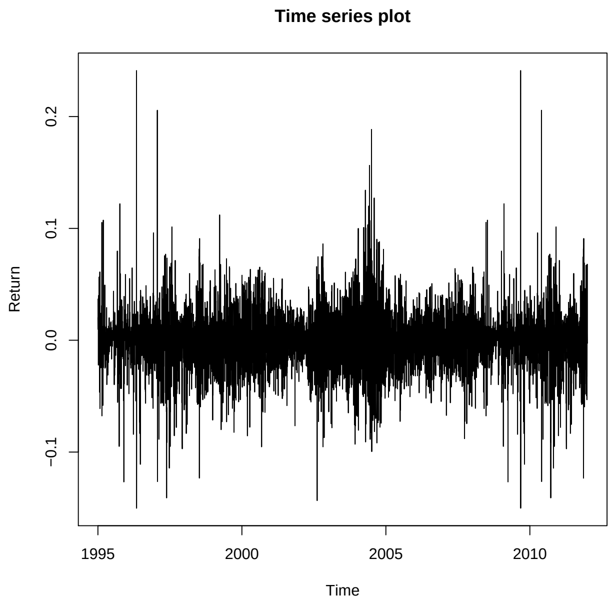


Figure 2.11: Impala Platinum Holdings Limited Daily Returns

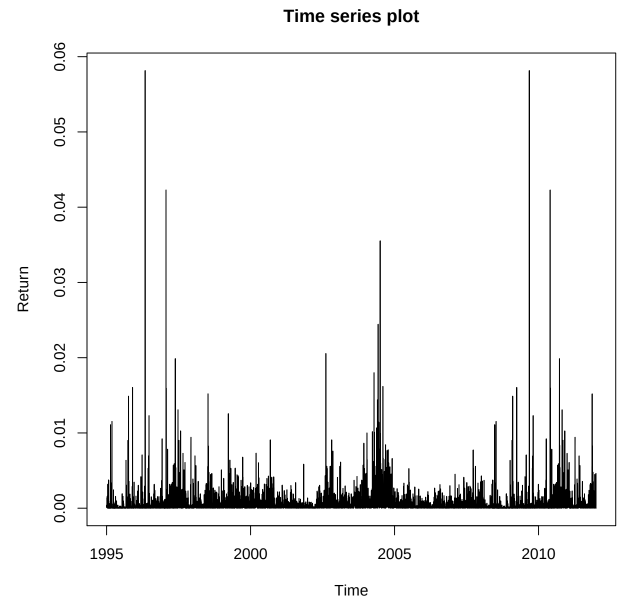


Figure 2.12: Impala Platinum Holdings Limited Daily Squared Returns

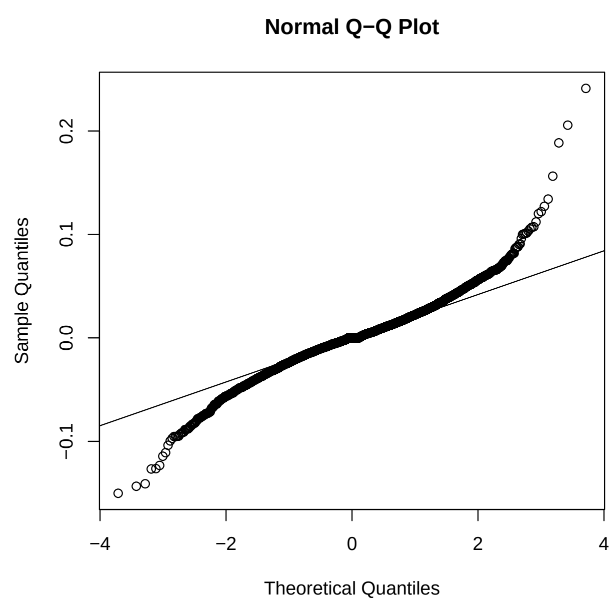


Figure 2.13: Q-Q Plot of Daily Returns for Impala Platinum Holdings Limited

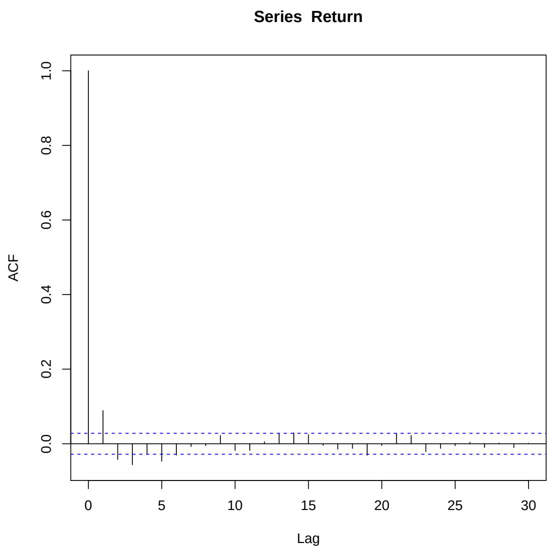


Figure 2.14: ACF for Impala Platinum Holdings Limited Daily Returns

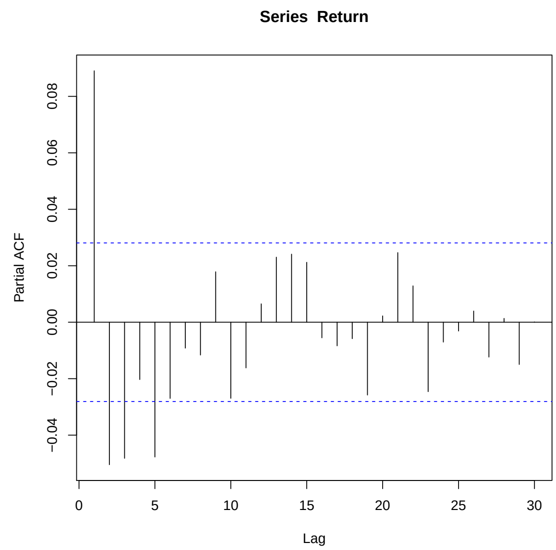


Figure 2.15: PACF for Impala Platinum Holdings Limited Daily Returns

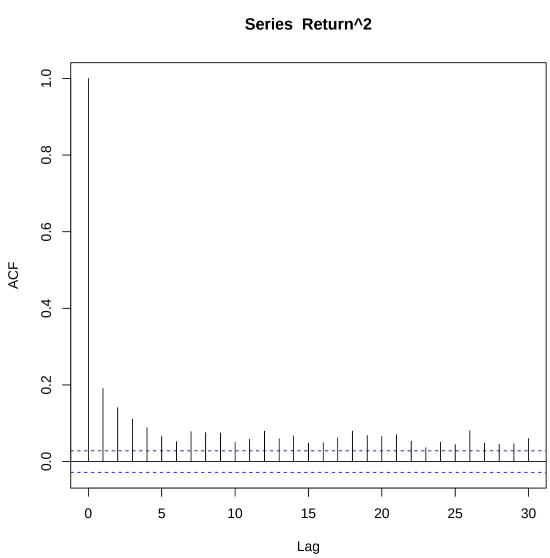


Figure 2.16: ACF for Impala Platinum Holdings Limited Daily Squared Returns

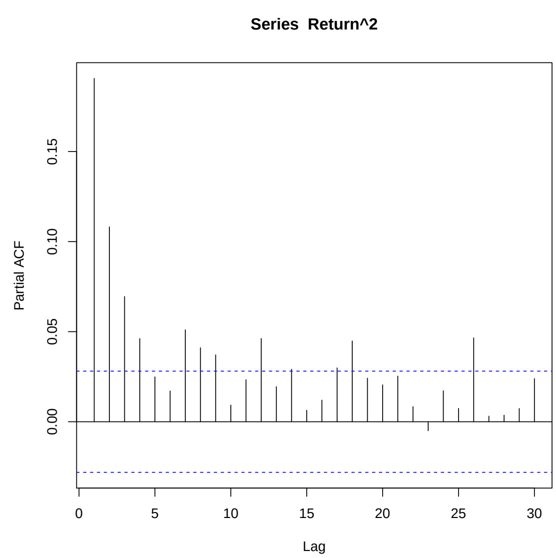


Figure 2.17: PACF for Impala Platinum Holdings Limited Daily Squared Returns

Figure 2.14, and 2.15 respectively reveal that the ACF of the returns exhibits a minor serial correlation at lags 1, 2, 3 and 5, while the PACF has a significant spike at lags 1, 2, 3 and 5. Figure 2.16, and 2.17 respectively, the ACF and PACF both show significant spikes which is an indication of presence of ARCH effect in the data.

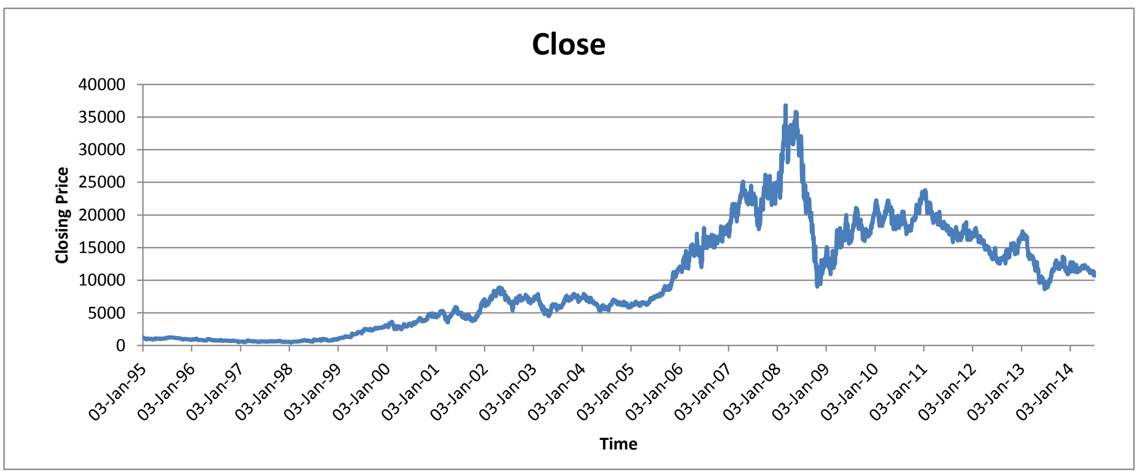


Figure 2.18: Impala Platinum Holdings Limited Daily Closing Prices in Rand

Figure 2.18 reveals that there is a period of large price movements in the data and small price movements. This is an indication of volatility clustering in the Impala Platinum Holdings Limited company.

2.2.3 Chapter Summary

In this chapter different daily data of the two JSE listed mining companies were analyzed over the same time period. The two companies considered are Harmony Gold Mining Limited and Impala Platinum Holdings Limited, with the data being from the period of 3 January 1995 to 3 July 2014. The presence of ARCH effect and autocorrelation was found in both the companies' data. Thus the use of the ARCH and GARCH models was applicable to JSE mining companies. It was clearly seen that the data from the two companies did not conform to the normal distribution, and the presence of volatility clustering was found in both companies. The application of GARCH is presented in Chapter four, where the best model for the data has been considered with the matching error distributions, since the normal distribution has failed. The best model also should not have the presence of ARCH effect nor autocorrelation. These results were obtained from R software using the fGARCH, and FinTS package.

Chapter 3

ARCH and GARCH models

3.1 The ARCH model

ARCH (Autoregressive Conditional Heteroscedasticity) is the first model introduced by Engle (1982) to model change in volatility (Shumway & Stoffer 2006). The AR comes from the fact that the models are autoregressive models in square return. The conditional component comes from the fact that in these models, the next period's volatility is conditional on the information in this period. In other words, heteroscedasticity means non-constant volatility. This is different from the ARMA (Autoregressive Moving Average), where the error terms are assumed to be independent.

3.1.1 The ARCH (1) model

The basic and very useful model in financial time series with time-varying volatility. The ARCH (1) is defined as the Autoregressive Conditional Heteroscedasticity of order one and was introduced by Engle (1982). Let us consider y_t as the return of an asset at time t that is given by

$$y_t = \mu_t + \epsilon_t,$$

where μ_t is expressed as an ARMA model with some explanatory variables given by

$$\mu_t = \phi_0 + \sum_{i=1}^{r} \beta_i y_{it} + \sum_{i=1}^{p} \gamma_i y_{t-i} - \sum_{i=1}^{q} \delta_i \epsilon_{t-i}$$

(Tsay 2005). Thus, more explanation about ARMA and any related models can be found in (Tsay 2005, Box et al. 2008).

A process $\epsilon_1, \epsilon_2, ..., \epsilon_t$ is called ARCH (1) if it can be written as

$$\epsilon_t = \sigma_t y_t, \tag{3.1}$$

where the random variables y_t are independent and identically distributed with mean zero and variance one and σ_t^2 satisfies the following constraints

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2, \tag{3.2}$$

where α_0 and $\alpha_1 \geq 0$ (Engle 1982).

Under the normality assumption of ϵ_t, the process can express conditional on Φ_{t-1} as

$$\epsilon|\Phi_{t-1} = \epsilon_t|\epsilon_{t-1} \sim N(0, \sigma_t^2).$$

From equation 3.1 and 3.2 it is clear that a large past squared mean-corrected return or shock implies a large conditional variance σ_t^2 for the mean-corrected return ϵ_t. Hence, ϵ_t tends to consider a large value in absolute value (Tsay 2005). This means that from the two equations, large shocks tend to be followed by another large shock. This characteristic is identical to volatility clusterings observed in asset returns (Bollerslev et al. 1992). Thus the stochastic process ϵ_t is a martingale difference sequence and, therefore, the white noise process. Consider ϵ_t be an ARCH (1) process with $Var[\epsilon_t] = \sigma_t^2 < \infty$, it follows that ϵ_t is a white noise process. Thus the conditional mean of ϵ_t equal to zero can be shown in the following equations

$$\begin{aligned} E[\epsilon_t|\Phi_{t-1}] &= E[\sigma_t y_t|\Phi_{t-1}] \\ &= E[\sigma_t(y_t|\Phi_{t-1})] \\ &= \sigma_t(0) \\ &= 0. \end{aligned}$$

This implies that $E[\epsilon_t] = 0$ and

$$\begin{aligned}
Cov[\epsilon_t, \epsilon_{t-k}] &= E[\epsilon_t \epsilon_{t-k}] - E[\epsilon_t]E[\epsilon_{t-k}] \\
&= E[\epsilon_t \epsilon_{t-k}] \\
&= E[E[\epsilon_t \epsilon_{t-k}|\Phi_{t-1}]] \\
&= E[\epsilon_{t-k} E[\epsilon_t|\Phi_{t-1}]] \\
&= 0.
\end{aligned}$$

Since ϵ_t is a martingale difference sequence, it is an uncorrelated sequence process. Suppose that the process ϵ_t is a second-order stationary ARCH (1) process with $Var[\epsilon_t]$. The unconditional variance of ϵ_t it follows that,

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1},$$

and this can be shown in the following equation. Consider the definition of variance of ϵ_t, $Var[\epsilon_t] = \sigma^2 < \infty$. This follows that,

$$\begin{aligned}
Var[\epsilon_t] &= E[\epsilon_t^2] - (E[\epsilon_t])^2 \\
&= E[\epsilon_t^2] \\
&= E[E[\epsilon_t^2|\Phi_{t-1}]] \\
&= E[\sigma_t^2] \\
&= E[\alpha_0 + \alpha_1 \epsilon_{t-1}^2] \\
&= \alpha_0 + \alpha_1 E[\epsilon_{t-1}^2] \\
&= \alpha_0 + \alpha_1 E[\epsilon_t^2]. \qquad (3.3)
\end{aligned}$$

Since ϵ_t portrays a second-order stationarity, that is $E[\epsilon_t^2] = E[\epsilon_{t-1}^2]$, therefore we have

$$Var[\epsilon_t] = \alpha_0 + \alpha_1 Var[\epsilon_t]$$

which implies that

$$Var[\epsilon_t] = \sigma^2 = \frac{\alpha_0}{1 - \alpha_1}$$

when $\alpha_1 < 1$ (Tsay 2005, Shumway & Stoffer 2006).

The variance of ϵ_t is required to be positive, therefore we require $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$. If the innovation y_t is symmetrically distributed around zero, then all odd moments of ϵ_t are equal to zero. Under the assumption of normal distribution the existence of higher even moments can be derived.

The tail behaviour of the distribution of asset returns tends to exhibit occasional extreme value and to study that behaviour we need the fourth moment of ϵ_t to be defined. The requirement conditional of variance for ϵ_t to be positive is that $0 < \alpha_1 < 1$ (Gouriéroux 1997, Tsay 2005).

Suppose that the process ϵ_t is an ARCH (1) process, $y_t \sim N(0,1)$ and $E[\epsilon_t^4] = k < \infty$. Thus we get

$$E[\epsilon_t^4] = \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} \tag{3.4}$$

where $3\alpha_0^2 < 1$.

The unconditional distribution of ϵ_t is leptokurtic and the fourth moment of ϵ_t must be finite. Let us assume that the series ϵ_t is fourth-order stationary, thus

$$[\epsilon_t^4] = E[\epsilon_{t-1}^4].$$

Therefore

$$\begin{aligned} \epsilon_t^4 &= E[E[\epsilon_t^4|\Phi_{t-1}] \\ &= E[E[\sigma_t^4 y_t^4|\Phi_{t-1}] \\ &= 3E[\sigma_t^4]. \end{aligned}$$

Since $y_t \sim N(0,1), E[y_t^4] = 3$, thus

$$\begin{aligned} E[\epsilon_t^4] &= 3E[(\alpha_0 + \alpha_1\epsilon_{t-1}^2)^2 \\ &= 3E[\alpha_0^2 + 2\alpha_0\alpha_1\epsilon_{t-1}^2 + \alpha_1^2\epsilon_{t-1}^4] \\ &= 3\alpha_0^2 + 6\alpha_0\alpha_1 E[\epsilon_{t-1}^2] + 3\alpha_1^2 E[\epsilon_{t-1}^4] \\ &= 3\alpha_0^2 + 6\alpha_0\alpha_1 E[\epsilon_{t-1}^2] + 3\alpha_1^2 E[\epsilon_t^4]. \end{aligned}$$

Consider $E[\epsilon_t^4]$ as a subject, we find that

$$(1-3\alpha_1^2)E[\epsilon_t^4] = 3\alpha_0^2 + 6\alpha_0\alpha_1 E[\epsilon_{t-1}^2]$$

therefore

$$E[\epsilon_t]^4 = \frac{3\alpha_0^2 + 6\alpha_0\alpha_1 E[\epsilon_{t-1}^2]}{1 - 3\alpha_1^2}$$

and by substituting $E[\epsilon_t^2]$ by $Var[\epsilon_t] = \frac{\alpha_0}{1-\alpha_1}$ in equation 3.4 we obtained that

$$\begin{aligned} E[\epsilon_t^4] &= \frac{3\alpha_0^2 - 3\alpha_0^2\alpha_1 + 6\alpha_0^2\alpha_1}{(1-3\alpha_1^2)(1-\alpha_1)} \\ &= \frac{3\alpha_0^2 + 3\alpha_0^2\alpha_1}{(1-3\alpha_1^2)(1-\alpha_1)} \\ &= \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} \end{aligned}$$

(Box et al. 2008). Therefore the kurtosis of ϵ_t is given by

$$\begin{aligned} Kurt[\epsilon_t] &= \frac{E[\epsilon_t^4]}{(E[\epsilon_t^2])^2} \\ &= \frac{\dfrac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)}}{\dfrac{\alpha_0^2}{(1-\alpha_1)^2}} \\ &= \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3. \end{aligned}$$

Thus, for the ARCH (1) process it is required that $0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$ for the fourth-order moment and the conditional kurtosis to exist. Furthermore, the excess kurtosis of ϵ_t is heavier than that of normal distribution (Talke 2003, Tsay 2005, Shumway & Stoffer 2006, Tsay 2010). Most of the cases, when using the ARCH and GARCH models are necessary to consider modelling the squared residuals ϵ_t^2 and this becomes applicable when forecasting using ARCH model. Thus modelling the squared residuals using ARCH (1) model gets the following equation

$$\epsilon_t^2 = \sigma_t^2 \tau_t^2, \tag{3.5}$$

and since $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$, we get

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \omega_t,$$

where $\omega_t = \sigma_t^2(\tau_t^2 - 1)$ (Talke 2003).

3.1.2 Estimation of ARCH (1) model

The parameters of ARCH (1) model, which are α_0 and α_1, need to be estimated. Under the assumption of normality the method of maximum likelihood estimator is used to estimate those parameters of ARCH (1) model. Consider the number of observations $\epsilon_1, \epsilon_2, ..., \epsilon_T$, then the likelihood can be written as a product of conditional:

$$f(\epsilon_1, \epsilon_2, ..., \epsilon_T|\tau) = f(\epsilon_T|y_{t-1})f(\epsilon_{T-1}|y_{t-2})...f(\epsilon_2|y_1)f(\epsilon_1|\tau),$$

$$f(\epsilon_1, \epsilon_2, ..., \epsilon_T|\tau) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2})f(\epsilon_1|\tau), \tag{3.6}$$

where the conditional maximum likelihood estimates $\tau = (\hat{\alpha}_0, \hat{\alpha}_1)$. However, it is too complicated to get the probability density function of ϵ_1, which is $f(\epsilon_1|\tau)$. Thus it is usually not used when the sample size is sufficiently large (Tsay 2010). In this case the conditional likelihood function is used, we have the following equations

$$f(\epsilon_2, ..., \epsilon_T|\tau;\epsilon_1) = f(\epsilon_T|y_{t-1})f(\epsilon_{T-1}|y_{t-2})...f(\epsilon_2|y_1)f(\epsilon_2|\tau;\epsilon_1)$$

$$f(\epsilon_2, ..., \epsilon_T|\tau;\epsilon_1) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2}), \tag{3.7}$$

Under normality the logarithm of likelihood is used, along the maximizing the conditional likelihood function is equivalent to maximizing the logarithm of the conditional likelihood. Thus the function becomes

$$l(\epsilon_2, ..., \epsilon_T|\tau, \epsilon_1) = \sum_{t=2}^{T} -\frac{1}{2}\ln(2\Pi) - \frac{1}{2}\ln(2\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2},$$

since the term $\ln(2\Pi)$ does not include any parameter to be estimated, and the above equation can be written as

$$l(\epsilon_2, ..., \epsilon_T|\tau, \epsilon_1) = -\sum_{t=2}^{T}(\frac{1}{2}\ln(\sigma_t^2) + \frac{\epsilon_t^2}{2\sigma_t^2}), \tag{3.8}$$

where $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2$ is evaluated recursively (Talke 2003, Tsay 2010). Note that the maximization of the equation 3.8, with respect to τ, where $\tau = (\hat{\alpha}_0, \hat{\alpha}_1)$ is a non-linear optimization problem. The non-linear optimization approach is used to maximize the conditional log-likelihood with respect to τ. Thus the Newton-Raphson and its alternative Fisher's Scoring method is used to solve the non-linear equations.

The Newton-Raphson Method

This is the most important method to solve a non-linear equation. This method starts with an initial guess for the solution, the second order is obtained by approximating the function to be maximized in the neighborhood of the initial guess by the second-degree polynomial. Newton-Raphson also finds the location of maximum value of the polynomial. Therefore, the third guess is obtained by finding the location of its maximum. This method always generates a sequence of guesses. Thus this method determines the value of $\hat{\alpha}$ at which the function $L(\alpha)$ is maximized (Agresti 2002). Let

$$U' = (\frac{\partial l(\alpha)}{\partial \alpha_0}, \frac{\partial l(\alpha)}{\partial \alpha_1}, ...) \tag{3.9}$$

and we let H be the Hessian matrix with entries

$$h_i j = \frac{\partial^2 l(\alpha)}{\partial \alpha_i \partial \alpha_j} \tag{3.10}$$

Consider $U^{(t)}$ and $H^{(t)}$ to be U and H calculated at $\alpha^{(t)}$. Assume $\alpha^{(t)}$ to be the guess for $\hat{\alpha}$ at step t, where t=1,2,.... Each step approximates $l(\alpha)$ near $\alpha^{(t)}$ by the terms up to second order of its Taylor series expansion is given by

$$l(\alpha) \approx \alpha^{(t)} + U^{(t)\prime}(\alpha - \alpha^{(t)}) + \frac{1}{2}(\alpha - \alpha^{(t)})' H^{(t)}(\alpha - \alpha^{(t)}). \tag{3.11}$$

Thus, the next guess is obtained by solving the following equation

$$\frac{\partial l(\alpha)}{\partial \alpha} \approx U^{(t)} + H^{(t)}(\alpha - \alpha^{(t)}) = 0. \tag{3.12}$$

Assuming that $H^{(t)}$ is nonsingular, the next guess can be expressed as

$$\alpha^{(t+1)} = \alpha^{(t)} - [(H^{(t)})^{-1} U^{(t)}] \tag{3.13}$$

This iteration can proceed until change in $l(\alpha)^{(t)}$ between two successive steps becomes sufficiently small. Thus, the Maximum Likelihood (ML) estimator becomes the limit of $\alpha^{(t)}$ as

$t \to \infty$ (Agresti 2002). Therefore, the Hessian matrix equation is given by

$$H = \begin{pmatrix} \frac{\partial^2 l(\partial)}{\partial\alpha_0\partial\alpha_0} & \frac{\partial^2 l(\partial)}{\partial\alpha_0\partial\alpha_1} \\ \\ \frac{\partial^2 l(\partial)}{\partial\alpha_1\partial\alpha_0} & \frac{\partial^2 l(\partial)}{\partial\alpha_1\partial\alpha_1}. \\ \\ \end{pmatrix} \tag{3.14}$$

The Fisher Scoring Method

This is an alternative to the Newton-Raphson method. This method is used to solve likelihood equations and is similar to Newton-Raphson method. However, this method used expected value in the equation 3.12 instead of using the Hessian matrix (Agresti 2002). Let $I^{(t)}$ be the expected information matrix with element

$$-E\left(\frac{\partial^2 l(\alpha)}{\partial\alpha_i\partial\alpha_j}\right) \tag{3.15}$$

and the equation 3.15 is evaluated at $\alpha^{(t)}$. The equation of Fisher Scoring method is given by

$$\alpha^{(t+1)} = \alpha^{(t)} + (I^{(t)})^{-1}U^{(t)}, \tag{3.16}$$

(Agresti 2002). The maximization of likelihood is the same as maximizing the log-likelihood. Therefore we can maximize the equation 3.8 for the ARCH (1) model using Newton-Raphson and Fisher Scoring methods, using the idea of equation 3.9 and equation 3.10. This can be shown in the following equations:

$$\frac{\partial l(\alpha)}{\partial\alpha_0} = -\frac{1}{2}\sum_{t=2}^{T}\left(\frac{1}{\alpha_0+\alpha_1\epsilon_{t-1}^2} - \frac{\epsilon_t^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2}\right)$$

$$\frac{\partial l(\alpha)}{\partial\alpha_1} = -\frac{1}{2}\sum_{t=2}^{T}\left(\frac{\epsilon_{t-1}^2}{\alpha_0+\alpha_1\epsilon_{t-1}^2} - \frac{\epsilon_t^2\epsilon_{t-1}^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2}\right)$$

$$\frac{\partial^2 l(\alpha)}{\partial\alpha_0} = -\frac{1}{2}\sum_{t=2}^{T}\left(\frac{-1}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2} + 2\frac{\epsilon_t^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^3}\right)$$

$$\frac{\partial^2 l(\alpha)}{\partial\alpha_0\partial\alpha_1} = -\frac{1}{2}\sum_{t=2}^{T}\left(-\frac{\epsilon_{t-1}^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2} + 2\frac{\epsilon_t^2\epsilon_{t-1}^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^3}\right)$$

$$\frac{\partial^2 l(\alpha)}{\partial\alpha_1\partial\alpha_0} = -\frac{1}{2}\sum_{t=2}^{T}\left(-\frac{\epsilon_{t-1}^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2} + 2\frac{\epsilon_t^2\epsilon_{t-1}^2}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^3}\right)$$

$$\frac{\partial^2 l(\alpha)}{\partial^2\alpha_1} = -\frac{1}{2}\sum_{t=2}^{T}\left(-\frac{\epsilon_{t-1}^4}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^2} + 2\frac{\epsilon_t^2\epsilon_{t-1}^4}{(\alpha_0+\alpha_1\epsilon_{t-1}^2)^3}\right)$$

Using the above approach the equations 3.13 for Newton-Raphson method can be written as

$$\alpha^{(t+1)} = \alpha^{(t)} - \begin{pmatrix} \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_0\partial\alpha_0} & \frac{\partial^2 l(\partial^2 l(\alpha^{(t)}))}{\partial\alpha_0\partial\alpha_1} \\ \\ \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1} & \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1} \\ \\ \end{pmatrix}^{-1} \tag{3.17}$$

and 3.16

$$\begin{pmatrix} \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_0\partial\alpha_0} & \frac{\partial^2 l(\partial^2 l(\alpha^{(t)}))}{\partial\alpha_0\partial\alpha_1} \\ \\ \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1} & \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1} \\ \\ \end{pmatrix}$$

Using the above approach the equations 3.13 for Newton-Raphson method can be written as

$$\alpha^{(t+1)} = \alpha^{(t)} - \begin{pmatrix} \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_0\partial\alpha_0} & \frac{\partial^2 l(\partial^2 l(\alpha^{(t)}))}{\partial\alpha_0\partial\alpha_1} \\ \\ \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_0} & \frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1} \\ \\ \end{pmatrix}^{-1} \times \begin{pmatrix} \frac{\partial l(\alpha^{(t)})}{\partial\alpha_0} \\ \\ \frac{\partial l(\alpha^{(t)})}{\partial\alpha_1} \end{pmatrix}, \tag{3.18}$$

and equation 3.16 for the Fisher Scoring can be written as

$$\alpha^{(t+1)} = \alpha^{(t)} + \begin{pmatrix} E\left[\frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_0\partial\alpha_0}\right] & E\left[\frac{\partial^2 l(\partial^2 l(\alpha^{(t)}))}{\partial\alpha_0\partial\alpha_1}\right] \\ \\ E\left[\frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_0}\right] & E\left[\frac{\partial^2 l(\alpha^{(t)})}{\partial\alpha_1\partial\alpha_1}\right] \\ \end{pmatrix}^{-1} \times \begin{pmatrix} \frac{\partial l(\alpha^{(t)})}{\partial\alpha_0} \\ \\ \frac{\partial l(\alpha^{(t)})}{\partial\alpha_1} \end{pmatrix} \tag{3.19}$$

(Dralle 2011, Franke et al. 2011)

3.1.3 Forecasting using ARCH (1) model

In time series forecasting is one of main reasons for modelling volatility. Let us consider the series of $\alpha_1, \alpha_2, ..., \alpha_T$ at forecast origin k, the k-step ahead forecast for $T = 1, 2, ...,$ where $\epsilon_T(k)$ is the minimum mean square error predictor, that is $\epsilon_T(k)$ minimizes the function f(y) (Talke 2003). Thus

$$E[\epsilon_{T+k} - f(y)]^2,$$

where f(y) is a function of observations and is given by

$$\epsilon_{T+k} = E[\epsilon_{T+k}|\epsilon_1, \epsilon_2, ..., \epsilon_T]$$

(Shumway & Stoffer 2000, Tsay 2005). Thus taking a mean from equation 3.1 of ARCH (1) model, gives us the following equation

$$E[\epsilon_{k+h}|\epsilon_1, \epsilon_2, ..., \epsilon_k] = 0. \tag{3.20}$$

This forecasts for ϵ_k series is not helpful. Therefore we can consider forecasting for the squared returns ϵ_k, which is given by the AR (1) model from equation 3.2 (Shephard 1996, Tsay 2005). This result in:

$$\epsilon_k(h)^2\epsilon = E[\epsilon^2_{k+h}|\epsilon^2_1, \epsilon^2_2, ..., \epsilon^2_k].$$

Hence the 1-step ahead forecast is given by $\epsilon_k(1) = \hat{\alpha}_0 + \hat{\alpha}\alpha\alpha_1\epsilon_k^2$, where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the conditional maximum likelihood estimator since $\hat{\alpha}_0$ and α_1 are not known. Similarly the 2-step ahead forecast is given by

$$\epsilon_k(2)^2 E[\epsilon_{k+2|\epsilon_k}] = E[\sigma_{k+2}|\epsilon_k]$$

3.1.4 The ARCH (q) model

This is an extension of ARCH (1) model, where the conditional variance σ_t^2 depends on the last squared mean-corrected returns. Therefore, the model can be expressed as

$$\epsilon_t = \sigma_t y_t \tag{3.21}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + ... + \alpha_q\epsilon_{t-q}^2, \tag{3.22}$$

where y_t is the sequence of independent and identically distributed (iid) random variables with mean zero and variance one. Thus $\alpha_0 > 0$, and $\alpha_i > 0$ for all $i > 0$, where $i = 0, 1, 2, ..., q$ are unknown parameters. The conditional variance to be finite, the coefficients α_i must satisfy some regularity conditions for the unconditional variances of ϵ_t (Tsay 2010).

3.1.5 Estimation of ARCH (q) model

Under the normality assumption, the unknown parameters of ARCH (q) model are $\alpha_0, \alpha_1, ..., \alpha_q$. Those parameters are estimated using maximum likelihood approach, and the likelihood function of an ARCH (q) model is given by

$$f(\epsilon_1, \epsilon_2, ..., \epsilon_T|\tau) = f(\epsilon_T|y_{t-1})f(\epsilon_{T-1}|y_{t-2})...f(\epsilon_{q+1}|y_q)f(\epsilon_1, \epsilon_2, ..., \epsilon_q|\tau) \tag{3.23}$$

$$= \prod_{t=q+1}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2}) f(\epsilon_1|\tau) f(\epsilon_1, \epsilon_2 ..., \epsilon_q|\tau), \tag{3.24}$$

where $\tau = (\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_q)$ and $f(\epsilon_{q+1}|y_q)f(\epsilon_1, \epsilon_2, ..., \epsilon_q|\tau)$ is the joint probability density function of $\alpha_0, \alpha_1, ..., \alpha_q$. The equation 3.24 becomes

$$f(\epsilon_{q+1},...,\epsilon_T|\tau,\epsilon_1,...,\epsilon_q) = \prod_{t=q+1}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2}), \tag{3.25}$$

since the exact form of $f(\epsilon_1, \epsilon_2, ..., \epsilon_q|\tau)$ is complicated. Therefore, σ_t^2 can be evaluated recursively, when using the condition likelihood. Under normality the logarithm of likelihood is used, and when maximizing the conditional likelihood function is equivalent to maximizing the logarithm of the conditional likelihood. Thus the function becomes

$$l(\epsilon_{q+1},...,\epsilon_T|\tau,\epsilon_1,...,\epsilon_q) = \sum_{t=q+1}^{T} -\frac{1}{2}\ln(2\Pi) - \frac{1}{2}\ln(2\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2}.$$

Since the term $\ln(2\Pi)$ does not include any parameter to be estimated, the above equation can be written as

$$l(\epsilon_{q+1},...,\epsilon_T|\tau,\epsilon_1,...,\epsilon_q) = -\sum_{t=q+1}^{T} (\frac{1}{2}\ln(\sigma_t^2) + \frac{\epsilon_t^2}{2\sigma_t^2}),$$

where $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1} + ... + \alpha_q\epsilon_{t-q}$ is evaluated recursively (Tsay 2005).

3.1.6 Forecasting using ARCH (q) model

Forecasting using an ARCH (q) model, we use the same procedure as an ARCH (1) model. Let us consider the series of $\alpha_1, \alpha_2, ..., \alpha_T$ at forecast origin k, the k-step ahead forecast for $T = 1, 2, ...,$ where $\epsilon_T(k)$ is the minimum mean square error predictor, that is $\epsilon_T(k)$ minimizes the function f(y) (Talke 2003). Thus $E[\epsilon_{T+k} - f(y)]^2$, where f(y) is a function of observations, y (Shumway & Stoffer 2000, Tsay 2005). Thus the forecasts for ϵ_k series is not helpful since $E[y_T(k)] = 0$. Therefore we can consider forecasting using the AR (q) model for the squared mean-corrected returns, y_t^2. The 1-step ahead forecast for the squared returns is given by

$$y_T^2(1) = \sigma_T^2(1) = E[y_{T+1}^2|Y_T] = \hat{\alpha}_0 + \sum_{i=1}^{q} \alpha_i y_{T+1-i}^2,$$

where $i = 1, 2, ...q$ and α_i are substituted by their conditional maximum likelihood estimate $(\hat{\alpha}_0, \hat{\alpha}_1, ...\hat{\alpha}_q)$. Hence forecasting an ARCH (q) model at the the k-step ahead forecast for σ_{T+k}^2 is given by

$$\sigma_T^2(k) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \sigma_T^2(k-i),$$

for $k = 1, 2, ...,$ where $\sigma_T^2(k-i) = y_{T+k-i}^2$ if $i \geq k$ (Shephard 1996, Talke 2003, Tsay 2005).

Weakness of ARCH models

In the previous subsection, we discussed the advantages of ARCH models. There are also disadvantages when using ARCH models, as follows:

1. The ARCH model does not take into consideration the difference between positive and negative shock because it depends on the square of the previous shocks. This implies that positive and negative shocks are assumed to have the same effects on volatility.

2. The ARCH model is restrictive. This can be shown when the parameter α_1 of ARCH (1) model lies within this interval $[0, \frac{1}{3}]$, for the fourth moment of series to exist. Thus, this constraint becomes more complicated for higher order ARCH models.

3. The ARCH model provides a way of describing the behaviour of the conditional variance rather than showing us the causes of such behaviour.

4. The ARCH model can over predict volatility. This happens because of the slow response of the ARCH model to the largely isolated shock in the returns series (Tsay 2005).

3.1.7 The GARCH model

In this section, the GARCH (1,1) model, its estimation parameters and forecasting approach have been discussed. Then GARCH (p,q) has been discussed including its parameters estimation and forecasting approach. The GARCH (p,q) is the Generalized ARCH model with an order (p,q). This model was introduced by Bollerslev (1986), since the ARCH (q) models use many parameters to provide an adequate description of the data and it is difficult to estimate a model with such a large number of parameters. Thus the general equation of GARCH (p,q) is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3.26}$$

(Gouriéroux 1997, Tsay 2005, Daníelsson 2011).

3.1.8 The GARCH (1,1) model

A GARCH (1,1) model is a Generalized ARCH (1,1) model and is very common in most sectors of finance and economics. Therefore, this method can be also applicable in mining time series data. The GARCH (1,1) model was introduced by Bollerslev (1986) after a realization that the ARCH model is simple and requires many parameters to fit the data (Tsay 2010). Thus the GARCH (1,1) model is given by

$$y_t = \mu_t + \epsilon_t, \tag{3.27}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{3.28}$$

where y_t is Gaussian white noise and the constant mean μ_t . This means that the random variable y_t is independent and identically distributed with mean zero and variance one. The parameters from equation 3.29 must be restricted, for the σ_t^2 to be positive, and is shown as $\alpha_0 > 0$, $\alpha_1 \geq 0$, and $\beta_1 \geq 0$ (Tsay 2010). The equation 3.29 clearly shows that large past mean-corrected squared return ϵ_{t-1}^2 or past conditional variances σ_{t-1}^2 result in the large value of σ_t^2. This means that a large ϵ_{t-1}^2 tends to be followed by another large ϵ_t^2. Thus this indicates the behaviour of volatility clustering which is in financial time series (Tsay 2010). Let $\tau_t = \epsilon_t^2 - \sigma_t^2$, then it is clear that

$$\sigma_t^2 = \epsilon_t^2 - \tau_t, \tag{3.29}$$

and this can be also expressed as

$$\sigma_{t-1}^2 = \epsilon_{t-1}^2 - \tau_{t-1}. \tag{3.30}$$

This shows that the process of squared errors exhibits an ARMA (1,1) process with uncorrelated τ_t (Box et al. 2008, Tsay 2010). By plugging equation 3.31 into equation 3.30 and use equation

3.29, the GARCH (1,1) model can be derived by

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \epsilon_t^2 - \tau_t \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 (\epsilon_{t-1}^2 - \tau_{t-1}) \\
\sigma_t^2 + \tau_t &= \epsilon_t^2 \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \epsilon_{t-1}^2 - \beta_1 \tau_{t-1} + \tau_t \\
\epsilon_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 - \beta_1 \tau_{t-1} + \tau_t
\end{aligned}$$

where $(\alpha_1 + \beta_1) \leq 1$ for the unconditional variance of ϵ_t to be finite. The unconditional variance is infinite when $(\alpha_1 + \beta_1) = 1$ and undefined when $(\alpha_1 + \beta_1) > 1$. The Gaussian white noise τ_t is martingale difference series since $E[\tau_t] = 0$ and $cov[\tau_t, \tau_{t-l}] = 0$, where $l \geq 1$. Thus τ_t is a serially uncorrelated, however τ_t is not an independent and identical distributed (iid) sequence in general form. When we consider the second order of ϵ_t^2 and for stationarity proven by Bollerslev (1986), where the $var[\tau_t]$ is assumed to be finite. Therefore ϵ_t^2 is an ARMA (1,1) process following the second order stationary. Thus $E[\epsilon_t^2]$ is given as

$$E[\epsilon_t^2] = \alpha_0 + (\alpha_1 + \beta_1) E[\epsilon_{t-1}^2] = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \tag{3.31}$$

(Gouriéroux 1997, Daníelsson 2011).

The kurtosis for a GARCH (1,1) model is proven here, by following Tsay (2005). In this case the error τ_t is Gaussian white noise and $E[\tau_t] = 3$ if we assume the fourth-order to be stationary, hence

$$\begin{aligned}
E[\epsilon_t^4] &= E[E[\epsilon_t^4 | \epsilon_{t-1}]] \\
&= E[E[\sigma^4 \epsilon_t^4 | \epsilon_{t-1}]] \\
&= E[\sigma^4 E[\epsilon_t^4 | \epsilon_{t-1}]] \\
&= 3E[\sigma^4].
\end{aligned}$$

By considering

$$E[\sigma^4] = \alpha_0^2 + \alpha_1^2 E[\epsilon_{t-1}^4] + \beta_1^2 E[\sigma_{t-1}^4] + 2\alpha_0\alpha_1 E[\epsilon_{t-1}^2] + 2\alpha_0\beta_1 E[\sigma_{t-1}^2] + 2\alpha_1\beta_1 E[\sigma_{t-1}^2\epsilon_{t-1}^2]$$

and $E[\sigma_{t-1}^2\epsilon_{t-1}^2]$ can be shown as follows

$$\begin{aligned} E[\sigma_{t-1}^2\epsilon_{t-1}^2] &= E[E[E[\sigma_{t-1}^2\epsilon_{t-1}^2|\epsilon_{t-2}]]] \\ &= E[\sigma_{t-1}^4 E[\epsilon_{t-1}^2|\epsilon_{t-2}]] \\ &= E[\sigma_{t-1}^4] \end{aligned}$$

since $\epsilon_{t-1}^2 = \sigma_{t-1}^2\epsilon_{t-1}^2$. Knowing that expression

$$\begin{aligned} E[\epsilon_{t-1}^2] &= E[\sigma_{t-1}^2] \\ &= \frac{\alpha_0}{1-(\alpha_1+\beta_1)} \end{aligned} \tag{3.32}$$

is stationary, then by rearranging expression for $E[\sigma^4]$ we have the following expression

$$E(\sigma_t^4)[1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2] = \alpha_0^2 + \frac{2\alpha_0^2\alpha_1}{1-(\alpha_1+\beta_1)} + \frac{2\alpha_0^2\beta_1}{1-(\alpha_1+\beta_1)}.$$

Thus

$$\begin{aligned} E(\sigma_t^4) &= \frac{\alpha_0^2(1+\alpha_1+\beta_1)}{[1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2][(1-2\alpha_1^2)-(\alpha_1+\beta_1)^2]} \\ &= \frac{\alpha_0^2(1+\alpha_1+\beta_1)}{[1-(\alpha_1+\beta_1)][(1-2\alpha_1^2)-(\alpha_1+\beta_1)^2]} \end{aligned}$$

from equation 3.33 shows that $E[\epsilon_t^4] = E[\sigma_t^4]$, we have the following expression

$$\begin{aligned} E[\epsilon_t^4] &= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{[1-\alpha_1-\beta_1][1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2]} \\ &= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{[1-(\alpha_1+\beta_1)][(1-2\alpha_1^2)-(\alpha_1+\beta_1)^2]}. \end{aligned}$$

Therefore, the kurtosis of a GARCH (1,1) model is expressed as

$$kurtosis = \frac{E[\epsilon_t^4]}{(\sigma_2[\epsilon_t])^2} = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2} > 3 \tag{3.33}$$

Thus, the requirement condition for kurtosis to exist is that $1-2\alpha_1^2-(\alpha_1+\beta_1)^2 > 0$ and this shows that GARCH model has fatter tails than the normal distribution (Gouriéroux 1997, Talke 2003, Tsay 2005).

3.1.9 Estimation of GARCH (1,1) model

The parameters estimation of the GARCH (1,1) model, can be solved in the same way as for the ARCH (1) model. However, the conditional variance of the GARCH (1,1) model is dependent on the past variance, and the initial value of the past conditional variance which is σ_1^2 is required. To estimate the parameters of the GARCH (1,1) model, Bollerslev (1986) suggested that the unconditional variance for ϵ_t should be used as an initial value for this variance, and the equation can be written as

$$\sigma_1^2 = E[\epsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \tag{3.34}$$

(Talke 2003). Under the assumption of normality, the logarithm of likelihood is used, when we maximize the conditional likelihood function. By assuming that ϵ_t and σ_1^2 are both known, then the estimated parameters are $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$. Thus, the parameters estimation procedure is carried out as for the ARCH (1) model, then conditional maximum likelihood estimate can be shown as

$$l(\epsilon_2, ..., \epsilon_T | \tau, \epsilon_1, \sigma_1^2) = \sum_{t=2}^{T} -\frac{1}{2}\ln(2\Pi) - \frac{1}{2}\ln(2\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2}. \tag{3.35}$$

Since the term $\ln(2\Pi)$ does not include any parameter to be estimated, the above equation can be written as

$$l(\epsilon_2, ..., \epsilon_T | \tau, \epsilon_1) = -\sum_{t=2}^{T} (\frac{1}{2}\ln(\sigma_t^2) + \frac{\epsilon_t^2}{2\sigma_t^2}), \tag{3.36}$$

where $\tau = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$. The method of Newton-Raphson and its alternative Fisher's Scoring method used to solve the non-linear equation is the same as those for ARCH (1) model, except that σ_t^2 has a slightly different formulation (Tsay 2005, Daníelsson 2011).

3.1.10 Forecasting using GARCH (1,1) model

Forecasting with the GARCH (1,1) model can be obtained in the same way as for the ARCH (1) model. Consider equation 3.29 of the GARCH (1,1) model, and assume that the forecast origin is k. The one-step ahead volatility forecast is given by

$$\sigma^2_{T+1} = \alpha_0 + \alpha_1\epsilon^2_T + \beta_1\sigma^2_T,$$

where ϵ^2_T and σ^2_T are both known at time index T. Thus the above equation can be written as

$$\sigma^2_T(1) = \alpha_0 + \alpha_1\epsilon^2_T + \beta_1\sigma^2_T. \tag{3.37}$$

For multi-step ahead forecast, the equation

$$\epsilon^2_t = \sigma^2_t\tau^2_t \tag{3.38}$$

can be used and equation 3.29 is rewritten to get the following equation

$$\sigma^2_{T+1} = \alpha_0 + (\alpha_1 + \beta_1)\sigma^2_T + \alpha_1\sigma^2_T(\epsilon^2_T - 1) \tag{3.39}$$

The two-step ahead volatility forecast is the conditional mean of σ^2_{T+2}, which is $E[\sigma^2_{T+2}|\sigma_{T+1}, \epsilon_T]$ and is given by

$$\sigma^2_T(2) = \alpha_0 + \alpha_1\epsilon^2_T + \beta_1\sigma^2_T. \tag{3.40}$$

In general the k-step ahead volatility forecast can be given by

$$\sigma^2_T(k) = \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{k-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1}\sigma^2_T(1) \tag{3.41}$$

where $\alpha_1 + \beta_1 < 1$. Thus the equation 3.41 shows that as the forecast horizon goes to infinity, then multi-step ahead volatility forecast for a GARCH (1,1) model converges to the unconditional variance of ϵ_t (Daníelsson 2011).

3.1.11 The GARCH (p,q) model

The GARCH (p,q) model is an extension of the GARCH (1,1) model from (1,1) parameter to (p,q) parameters, and was introduced by Bollerslev (1986). The GARCH (p,q) model can be expressed as

$$\epsilon_t = \sigma_t y_t \tag{3.42}$$

$$\sigma^2_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i\epsilon^2_{t-i} + \sum_{j=1}^{T}\beta_j\sigma^2_{t-j}, \tag{3.43}$$

where α_0 is Gaussian white noise, and y_t is a sequence of independent and identical distributed random variables with mean zero and variance one. Thus for the variance to be positive the following conditions must be satisfied: $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\sum_{i=1}^{max(p,q)}(\alpha_i + \beta_j) < 0$ and we can consider that $\alpha_i = 0$ for $i > p$ and $\beta_j = 0$ for $j > q$ (Bollerslev 1986, Tsay 2010). Let $\tau_t = \epsilon_t^2 - \sigma_t^2$, then is clear that $\sigma_t^2 = \epsilon_t^2 - \tau_t$, and this can be expressed as

$$\sigma_{t-i}^2 = \epsilon_{t-i}^2 - \tau_{t-i}. \tag{3.44}$$

By substituting equation 3.45 into equation 3.44, we get the following equations

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \\
\epsilon_t^2 - \tau_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_1 (\epsilon_{t-j}^2 - \tau_{t-1}) \\
\epsilon_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}^2 - \beta_j \tau_{t-1} + \tau_t
\end{aligned}$$

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_j)\epsilon_{t-1}^2 + \tau_t - \sum_{j=1}^{q} \beta_j \tau_{t-j}, \tag{3.45}$$

where $\alpha_i = 0$, $\beta_j = 0$ for $i > p$ and $j > q$. The equation 3.45 shows that the GARCH (p,q) model can be expressed as an ARMA model for the square series ϵ_t^2. Therefore, the unconditional mean of an ARMA model is given by

$$E[\epsilon_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{max(p,q)}(\alpha_1 + \beta_1)}, \tag{3.46}$$

where

$$\sum_{i=1}^{max(p,q)} (\alpha_1 + \beta_1) < 1$$

(Gouriéroux 1997, Francq & Zakoian 2010, Tsay 2010).

3.1.12 Estimation of GARCH (p,q) model

The method of maximum likelihood estimator can be used to estimate the parameters for the GARCH (p,q) model. Under the assumption of normality, the estimates of the GARCH (p,q) model can be obtained in the same way as the ARCH (q) model, and those parameters to be estimated are $\alpha_0, \alpha_1, ..., \alpha_p$ and $\beta_1, \beta_2, ..., \beta_q$. We then maximize the conditional likelihood function and assume that $\alpha_1, \alpha_2, ..., \alpha_p$ and $\sigma_1^2, \sigma_2^2, ..., \sigma_q^2$ are both known. Thus, the likelihood function can be expressed as

$$\begin{aligned} & f(\epsilon_1, ..., \epsilon_T, \sigma_1^2, ..., \sigma_T^2|\tau) \\ = \quad & f(\epsilon_T, \sigma_T^2|y_{t-1})f(\epsilon_{T-1}, \sigma_{T-1}^2|y_{t-2})...f(\epsilon_2\sigma_2^2|y_1)f(\epsilon_1, \sigma_1^2|\tau) \\ = \quad & \prod_{t=2}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2})f(\epsilon_1, \sigma_1^2|\tau) \end{aligned} \tag{3.47}$$

where $\tau = (\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_p, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_q)$, and $y_{t-1} = \epsilon_1, \sigma_1^2, ..., \epsilon_{t-1}, \sigma_{t-1}$ is the set of information at time $t-1$. Thus equation 3.48 becomes

$$= \prod_{t=2}^{T} \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp(\frac{-\epsilon_t^2}{2\sigma_t^2}), \tag{3.48}$$

since the exact form of $f(\epsilon_1, \sigma_1^2|\tau)$ is complicated, then σ_t^2 can be evaluated recursively. Maximizing the conditional likelihood function is equivalent to maximizing the logarithm of the conditional likelihood. Hence we can use the conditional log-likelihood to get the following equation:

$$lf(\epsilon_1, ..., \epsilon_T, \sigma_1^2, ..., \sigma_T^2|\tau) = \sum_{t=q+1}^{T} -\frac{1}{2}\ln(2\Pi) - \frac{1}{2}\ln(2\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2}, \tag{3.49}$$

since the term $\ln(2\Pi)$ does not include any parameter to be estimated, the above equation can be written as

$$l(\epsilon_1, ..., \epsilon_T, \sigma_1^2, ..., \sigma_T^2|\tau)) = -\sum_{t=q+1}^{T} (\frac{1}{2}\ln(\sigma_t^2) + \frac{\epsilon_t^2}{2\sigma_t^2}), \tag{3.50}$$

where $p = max(p, q)$ (Gouriéroux 1997, Talke 2003). The method of Newton-Raphson and its alternative Fisher's Scoring method used to solve the non-linear equation is the same as for the ARCH (1) model, except that σ_t^2 has a slightly different formulation (Daníelsson 2011).

Parameter estimation with non-Gaussian error distribution

The assumption of normal distribution is mostly used to fit the data with the GARCH model, however in the case of financial time series, when the assumption of normality is violated for real data, the appropriate distributions to use are: student's t distribution; the double exponential distribution; and the generalized error distribution (Andersen et al. 2009).

GARCH model with a Student's t Distribution

A Student's t distribution is used to fit the GARCH model when the assumption of normality is violated for a real data. This idea of fitting a GARCH model with a Student's t distribution was proposed by Bollerslev (1986). Consider a random variable y_ν that has a student's distribution with ν degrees of freedom, such that ϵ_t follows a heavy-tailed Student's t distribution. If $\frac{\nu}{\nu-2}$ for $\nu > 2$, and $\epsilon_t = \frac{y_t}{\sqrt{(\nu-2)}}$, and probability density function of ϵ_t is given by

$$f(\epsilon_t|\nu) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{(\nu-2)\Pi}} \left(1 + \frac{\epsilon_t^2}{\nu-2}\right)^{-(\nu+1/2)}, \tag{3.51}$$

for $\nu > 2$ and $\Gamma(y)$ is the gamma function, which is know as $\Gamma(y) = \int\limits_a^b y^{(k-1)} e^{-y}\, dy$.

From equation $\epsilon_t = \sigma_t y_t$, conditional likelihood function of ϵ_t can be written as

$$f(\epsilon_{k+1}, ..., \epsilon_T|\tau, K_t) = \prod_{t=k+1}^{T} \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{(\nu-2)\Pi}} \frac{1}{\sigma_t} \left(1 + \frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}\right)^{-(\nu+1/2)}, \tag{3.52}$$

where $\nu > 2$ and $K_t = (\epsilon_1, \epsilon_2, ...\epsilon_k)$. The degrees of freedom of t distribution can be specified a priori or estimated jointly with other parameters. Then if it is pre-specified we use the value between 4 and 8. Thus, the conditional log-likelihood function obtained in the following equation when the degrees of freedom ν of student-t distribution is pre-specified

$$l(\epsilon_{k+1}, ..., \epsilon_T|\tau, K_t) = -\sum_{t=k+1}^{T} \left[(\frac{\nu+1}{2}) \ln \left(1 + \frac{\epsilon_t^2}{(\nu-2)\sigma_t^2} + \frac{1}{2}(\sigma_t^2)\right)\right] \tag{3.53}$$

when ν is jointly with other parameters, then the log-likelihood function can be estimated as

$$l(\epsilon_{k+1},..,\epsilon_T|\tau,\nu,K_t) = (T-k)\left[\ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - 0.5\ln[(\nu-2)\Pi]\right] \tag{3.54}$$
$$+\, l(\epsilon_{k+1},..,\epsilon_T|\tau,\nu,K_t),$$

where

$$l(\epsilon_{k+1},...,\epsilon_T|\tau,\nu,K_t) = -\sum_{t=k+1}^{T}\left[(\frac{\nu+1}{2})\ln\left(1+\frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}+\frac{1}{2}(\sigma_t^2)\right)\right]. \tag{3.55}$$

GARCH model with Skew-Student's t Distribution

When the empirical distribution of an asset returns data is skewed, then Student-t distribution becomes Skew-Student-t distribution (Andersen et al. 2009, Tsay 2010). Many researchers have been using the Skew-Student-t distribution, however we considered the approach by Fernández & Steel (1998), where they found out that skewness can be introduced into any continuous unimodal and symmetric univariate distribution. The equation 3.53 was used by Lambert et al. (2001), where the method of Fernández & Steel (1998) was applied to get the following equation

$$g(\epsilon_t|\eta,\nu) = \begin{cases} \frac{2}{\eta+\frac{1}{\eta}}\phi f[\eta(\phi\epsilon_t+\rho)|\nu] & \text{if } \epsilon_t < -\dfrac{\rho}{\phi}, \\ \frac{2}{\eta+\frac{1}{\eta}}\phi f[\frac{\eta(\phi\epsilon_t+\rho)}{\eta}|\nu] & \text{if } \epsilon_t \geq -\dfrac{\rho}{\phi}\ , \end{cases} \tag{3.56}$$

where $f(.)$ is the probability density function (pdf) of the standardized Student-t distribution from equation 3.53, $\eta(.)$ is the skewness parameters, and ν is the degrees of freedom. When $\eta(.)$ is squared it is a measure of the skewness and is given as the ratio of probability masses above and below the model of the distribution. The ρ, ϕ are parameters which are given as follows

$$\rho = \frac{\Gamma[(\nu-1)/2]\sqrt{\nu-2}}{\sqrt{\Pi}\Gamma(\nu/2)}(\eta-\frac{1}{\eta}), \tag{3.57}$$

$$\phi^2 = \left(\eta^2+\frac{1}{\eta^2}-1\right)-\rho^2. \tag{3.58}$$

GARCH model with Generalized Error Distribution (GED)

GED is a class of distribution close to normal distribution but with variation of kurtosis. This distribution is a powerful alternative when the assumption of conditional normality is violated (Emenike 2010). GED is also useful because it can be transformed from a normal distribution into a leptokurtic distribution (fat tail) or into a platykurtic distribution (thin tail). This distribution was proposed by Nelson (1991), to capture the fat or thin tail observed in the distribution of time series (Zivot & Wang 2007). Consider x_t as a random variable with mean zero and variance of one such that x_t has a GED distribution, then the probability density function (pdf) of x_t is given by

$$f(x_t) = \frac{\nu exp\left[-(1/2)|x_t/\nu|^{\nu}\right]}{\delta 2^{(\nu+1)/\nu}\Gamma(1/\nu)}, \tag{3.59}$$

where $\delta = \left[\frac{2^{-(2/\nu)}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2}$, and ν is a positive parameter conducting the thickness of the tail behaviour of the distribution. When $\nu = 2$ the equation 3.60 becomes the standard normal pdf; when $\nu < 2$, the density has thicker tails than the normal density; when $\nu > 2$, the density has thinner tails than the normal density; finally when $\nu = 1$, then the pdf of GED reduces to the pdf of double exponential distribution which is

$$f(x_t) = \frac{1}{\sqrt{2}} exp^{-\sqrt{2}|x_t|}$$

(Andersen et al. 2009).

3.1.13 Forecasting using GARCH (p,q) model

Forecasting volatility of return series with a GARCH (p,q) model is done in the same way as that of the GARCH (1,1) model, when we consider T as a starting date for forecasting with the GARCH (p,q) model. Thus the one-step ahead forecast for σ^2_{T+1} is given by

$$\sigma^2_T(1) = \hat{\alpha}_0 + \sum_{i=1}^{p}(\hat{\alpha}_i + \hat{\beta}_i)E[\epsilon^2_{T+1-i}|\tau_T] - \sum_{j=1}^{q}\hat{\beta}_j E[\upsilon_{T+1-j}|\tau_T], \tag{3.60}$$

where $\epsilon^2_T + ... + \epsilon^2_{T+1-p}$ are assumed to be known at time T. Therefore, in general the k-step ahead forecast for σ^2_{T+1} is given by

$$\sigma_T^2(k) = \hat{\alpha}_0 + \sum_{i=1}^{p}(\hat{\alpha}_i + \hat{\beta}_i)E[\epsilon_{T+k-i}^2|\tau_T] - \sum_{j=1}^{q}\hat{\beta}_j E[\upsilon_{T+k-j}|\tau_T] \tag{3.61}$$

where

$$E[\epsilon_{T+k-i}^2|\tau_T] = \epsilon_{T+k-i}^2,$$

for $i \geq k$

$$E[\upsilon_{T+k-i}|\tau_T] = 0$$

for $i < k$ and

$$E[\upsilon_{T+k-j}|\tau_T] = \upsilon_{T+k-j}$$

for $j \geq k$ (Talke 2003, Shumway & Stoffer 2006).

Weakness of GARCH models

In this chapter, we have noted that the advantages of the GARCH models which are used in modelling of volatility in the South African mining sector to overcome some of the weakness in the ARCH model. Although, the GARCH model has been shown to be the best to model the volatility on the South African mining sectors, there are also some disadvantages. These are as follows:

1. The non-negativity conditions may be violated by the estimated model.
2. The GARCH model cannot account for volatility clustering and leptokurtosis in a series.
3. The model does not allow for any direct feedback between the conditional variance and conditional mean (Brooks 2008).

3.1.14 Extension of GARCH model

Extensions the GARCH model have been proposed as a result of identified weaknesses. That is when the stock returns are some times negatively correlated with change in volatility. This means that volatility tends to rise following bad news and fall following good news. This process is called "Leverage effect", and since the GARCH model was not able to detect this leverage effect, extensions were introduced (Daníelsson 2011). There are many extensions of the GARCH model such as the Exponential GARCH model introduced for this function by Nelson (1991), the Power GARCH model proposed by Taylor (1986) and then later by Schwert (1989), the Asymmetric Power GARCH (APARCH) model proposed by Ding et al. (1993), the GJR-GARCH model named for Glosten et al. (1993), the GARCH-M (GARCH in the Mean) model introduced by Engle et al. (1987), the FIGARCH (Fractionally Integrated GARCH) model proposed by Baillie et al. (1996), the Threshold GARCH (TGARCH) model introduced by Zakoian (1994), the Quadratic GARCH (QGARCH) model proposed by Sentana (1995), etc. A few extension models of GARCH model have been discussed and used.

3.1.15 EGARCH model

The Exponential GARCH model was the first model introduced by Nelson (1991) to take into account the leverage effect and which specifies the conditional variance in the the logarithmic form. This means that the EGARCH model allows for asymmetric effect between positive and negative asset returns (Tsay 2010). Nelson also considered the weighted innovation as shown in the following equations

$$g(y_t) = \theta y_t + \gamma(|y_t| - E[|y_t|]), \tag{3.62}$$

where θ and γ are real constants. Thus y_t and $(|y_t| - E[|y_t|])$ are zero mean independent and identically distributed sequences with continuous distribution. Consequently $E[g(y_t)] = 0$ and asymmetry of $g(y_t)$ can be shown as

$$g(y_t) = \begin{cases} (\theta + \gamma)y_t - \gamma E(|y_t|) & \text{if } y_t \geq 0, \\ (\theta - \gamma)y_t - \gamma E[|y_t|] & \text{if } y_t < 0 \ , \end{cases} \tag{3.63}$$

by letting y_t be iid sequences such that $E[y - t] = 0$ and $var[y_t] = 1$ (Tsay 2010). Thus we said that the process is Exponential GARCH (p,q), if the following form of equations are satisfied:

$$\epsilon_t = \sigma_t y_t \tag{3.64}$$

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i g(y_{t-i}) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2, \tag{3.65}$$

where $g(y_{t-i}) = \theta y_{t-i} + \psi(|y_{t-i}| - E(|y_{t-i}|)$ and α_0, α_i, β_j and ψ

(Francq & Zakoian 2010). From equation 3.66 it is clearly seen that no parameter restrictions are needed to ensure positivity of σ^2. The parameters $\alpha_i, i = 1, ..., p$, give an asymmetric response to possible shocks (Andersen et al. 2009). The alternative formula for EGARCH (p,q) model can be written as

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{\epsilon_{t-i} + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \tag{3.66}$$

where σ_t^2 is the conditional variance since it is a one period ahead estimate for the variance calculated on any past information thought relevant, and α_0, α_i, β_j and γ_i are the parameters to be estimated. The advantage of the EGARCH model is that even if the parameters are negative, σ_t^2 is modelled (Su 2010). The parameter estimates from equation 3.67 can be defined as follows:

- α_i represents a magnitude effect or the symmetric effect of the model, which is the GARCH effect.

- β_j measures the persistence in conditional volatility irrespective of occurrences in the market. This means that if β is large, the volatility takes a long time to die out following a crisis in the market.

- γ_i this parameter measures the asymmetry or leverage effect. This is an important parameter, where the EGARCH model allows for testing of asymmetries and is expected to be positive in most empirical cases in order for the negative shock to increase in the future. When $\gamma = 0$, the model is symmetric, when $\gamma < 0$, the positive shock (good news) generates less volatility than negative shock (bad news), and when $\gamma > 0$, this means that positive innovations are more destabilizing than negative innovations.

Parameter estimation of EGARCH (p,q) is similar to that of ARCH and GARCH models, where the logarithm of likelihood is used under normality assumption to estimate the parameters in the model.

3.1.16 Forecasting using EGARCH model

Forecasting using the EGARCH model is the same as in ARCH and GARCH models that have been discussed in this Chapter. The EGARCH (1,1) model has been considered to demonstrate multi-step ahead forecasts. We assume that the parameters of the model are known and the innovations are standard Gaussian (Tsay 2010). Thus the model is shown as follow

$$\ln(\sigma_t^2) = (1-\alpha_1)\alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + g(\epsilon_{t-1}), \tag{3.67}$$

where

$$g(\epsilon_{t-1}) = \theta\epsilon_{t-1} + \gamma\left(|\epsilon_{t-1}| - \sqrt{\frac{2}{\Pi}}\right).$$

By applying the exponentials on both sides, the model becomes

$$\sigma_t^2 = \sigma_{t-1}^{(2\alpha_1)} exp[(1-\alpha_1)\alpha_0 exp[g(\epsilon_{t-1})]], \tag{3.68}$$

consider k as the forecast origin, then for the 1-step ahead forecast we have

$$\sigma_{k+1}^2 = \sigma_k^{(2\alpha_1)} exp[(1-\alpha_1)\alpha_0 exp[g(\epsilon_{t-1})]]. \tag{3.69}$$

Thus the 1-step ahead volatility forecast at the origin k is given by $\hat{\sigma}_k^2(1) = \sigma_{k+1}^2$, when we assume that all the quantities on the right-hand side of the above equation are known. Thus the 2-step ahead forecast has been shown as

$$\sigma_{k+2}^2 = \sigma_{k+1}^{(2\alpha_1)} exp[(1-\alpha_1)\alpha_0 exp[g(\epsilon_{t-1})]]. \tag{3.70}$$

Now by taking the expectation at the origin k, we got the following equation

$$
\begin{aligned}
E[exp[g(\epsilon)]] &= \int_{-\infty}^{\infty} exp\left[\theta\epsilon + \gamma\left(|\epsilon| - \sqrt{\frac{2}{\Pi}}\right)\right]\phi(\epsilon)d\epsilon \\
&= exp\left(-\gamma\sqrt{\frac{2}{\Pi}}\right)\left[\int_{-\infty}^{0} e^{(\theta+\gamma)\epsilon}\frac{1}{2\pi}e^{-\frac{\epsilon^2}{2}}d\epsilon + \int_{0}^{\infty} e^{(\theta-\gamma)\epsilon}\frac{1}{2\pi}e^{-\frac{\epsilon^2}{2}}d\epsilon\right] \\
&= exp\left(-\gamma\sqrt{\frac{2}{\Pi}}\right)\left[e^{\frac{(\theta+\gamma)^2}{2}}\psi(\theta+\gamma) + e^{\frac{(\theta-\gamma)^2}{2}}\psi(\gamma-\theta)\right],
\end{aligned}
\tag{3.71}
$$

where $\phi(\epsilon)$ is the pdf, and $\psi(x)$ the cumulative density function of the standard normal distribution. Therefore the 2-step-ahead volatility forecast is given by

$$
\begin{aligned}
\hat{\sigma}_k^2(2) = \hat{\sigma}_k^{(2\alpha_1)} exp\left[(1-\alpha_1)\alpha_0 - \gamma\sqrt{\frac{2}{\Pi}}\right][exp\left[\frac{(\theta+\gamma)^2}{2}\right]\psi(\theta+\gamma) \\
+ exp\left[\frac{(\theta-\gamma)^2}{2}\right]\psi(\gamma-\theta)].
\end{aligned}
\tag{3.72}
$$

Thus the j-step ahead volatility forecast, where l=1,2,..., can be given by

$$
\hat{\sigma}_k^2(l) = \sigma_k^{\hat{2}\alpha_1}(l-1)exp(\mu)[exp\left[\frac{(\theta+\gamma)^2}{2}\right]\psi(\theta+\gamma) + exp\left[\frac{(\theta-\gamma)^2}{2}\right]\psi(\gamma-\theta)],
$$

where $\mu = (1-\alpha_1)\alpha_0 - \gamma\sqrt{\frac{2}{\Pi}}$, and $\psi(\theta+\gamma)$ or $\psi(\gamma-\theta)$ are the parameters to be estimated (Tsay 2010).

3.1.17 The GJR-GARCH model

The GJR-GARCH model was named after Glosten et al. (1993). This model is helpful to detect the leverage effect in the data, and is also similar to the Threshold ARCH model (TGARCH) introduced by Zakoian (1994). The general conditional variance of the GJR-GARCH (p,q) model is given by

$$
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}(\alpha_i + \gamma_i M_{t-i})\epsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2, \tag{3.73}
$$

where M_{t-i} is the dummy variable, and

$$M_{t-i} = \begin{cases} 1 & \text{if } \epsilon_{t-1} < 1, \\ \\ 0 & \text{if } \epsilon_{t-1} \geq 1 \ . \end{cases} \tag{3.74}$$

The conditional for non-negative parameters is the same as those of the GARCH model ($\alpha_0 > 0, \alpha_1 > 0, \beta_1 \geq 0$ and $\alpha_1 + \gamma \geq 0$.) Provided that $\alpha_1 + \gamma \geq 0$, the model is still admissible, even if $\gamma < 0$. For the case of leverage effect, we would expect that $\gamma > 0$ (Brooks 2008, Tsay 2010). Note that the parameters' estimation of the model and forecasting volatility is similar to those of ARCH and GARCH models.

3.1.18 The Asymmetric Power ARCH (APARCH) model

The APARCH model was introduced by Ding et al. (1993), to take the leverage effect into account. The general formula of the APARCH (p,q) model is written as

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta, \tag{3.75}$$

where $\omega > 0$, $\delta \geq 0$, $\beta_j > 0$ for (j=1,..., q), $\alpha_i > 0$ for (i=1,.., p), and $|\gamma_i| < 1$ for (i=1,..., p). The conditionals for non-negative parameters are also the same as to those of GARCH model. The APARCH model is a special cases where the seven GARCH models are nested within this model and it is shown as follows:

- When $\delta = 2$, $\gamma_i = 0$ for (i=1,.., p) and $\beta_j = 0$ for (j=1,.., q), the APARCH model reduces to the ARCH model of Engle (1982).

- When $\delta = 2$, and $\gamma_i = 0$ for (i=1,.., p), the APARCH model reduces to the standard linear GARCH model of Bollerslev (1986).

- When $\delta = 2$, and $0 \leq \gamma_i \leq 1$, the APARCH model becomes the JGR-GARCH model of Glosten et al. (1993).

- When $\delta = 1$, and $0 \leq \gamma_i \leq 1$, the APARCH model becomes the TARCH model of Zakoian (1994).

- When $\delta = 1$, and $\gamma_i = 0$ for (i=1,.., p), the APARCH model becomes the Taylor-Schwert GARCH (TS-GARCH) model of Taylor (1986) and Schwert (1990).
- When $\delta \to 0$ and $\gamma_i = 0$ for (i=1,.., p), the APARCH model becomes the log-ARCH model of Geweke (1986) and Pantula (1986).
- When $\gamma_i = 0$ for (i=1,.., p) and $\beta_j = 0$ for (j=1,.., q), the APARCH model reduces to the Nonlinear ARCH (NARCH) model of Higgins & Bera (1992).

The estimation of parameters and volatility forecasting is similar to those of ARCH and GARCH models, which have been discussed in this chapter (Wurtz et al. 2006, Andersen et al. 2009, Tsay 2010).

3.1.19 Model Selection Criteria

We need to ascertain the most parsimonious model fitted to the data. If the two models have the same number of parameters, this mean that one model is nested within the other model. The maximum value of their likelihood function has to be compared. However, the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) models need to be used, when the models differ in number of parameters. The best model is the one with small AIC or BIC from the model that has been selected to fit the data (Box et al. 2008). The AIC and BIC are calculated in the following equations.

$$AIC(p) = \frac{-2}{N} \ln(ML) + \frac{2p}{N} \tag{3.76}$$

where ML is the maximum likelihood and N is the number of observations in the model. The one disadvantage of AIC and BIC is that using maximum likelihood requires a great deal of calculation and is thus time consuming (Box et al. 2008).

3.1.20 Testing for ARCH effect

In time series data when we model the volatility using ARCH and GARCH models it is very important to check the ARCH effect in the residuals before we fit the GARCH models. There are two tests that are used to test the ARCH effect. The first test is the Ljung-Box statistic test Q (m) for Autocorrelation (ACF), this test is applied to the ϵ_t^2 series, where the null hypothesis is that the first m lags of the ACF function of the ϵ_t^2 series are zero (McLeod & Li 1983, Tsay 2005). The Ljung-Box test is given by

$$Q(m) = N(N+2)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2}{N-i}, \tag{3.77}$$

where N is the sample size, m is the number of lags, and $\hat{\rho}_i^2$ is the estimate of the i^{th} ACF of the squared residuals.

$$\hat{\rho}_i = \frac{\sum_{i+1}^{N}(c-\hat{\mu})(\epsilon_{r-i}^2-\hat{\mu})}{\sum_{r=1}^{N}(\epsilon_{r-i}^2-\hat{\mu})^2}, \tag{3.78}$$

where $\hat{\mu}$ is the sample mean given by $\hat{\mu} = \frac{1}{N}\sum_{r=1}^{N}\epsilon_r^2$. Under the null hypothesis, Q(m) is asymptotically distributed as a chi-squared distribution with m degrees of freedom (McLeod & Li 1983, Box et al. 2008). The null hypothesis is that the first m lags at autocorrelation of the ϵ_t^2 series are zero (Tsay 2005). Thus we reject null the hypothesis if $Q(m) > \chi_m^2(\alpha)$, where $\chi_m^2(\alpha)$ is the $100(1-\alpha)$ percentile of a chi-squared distribution with m degrees of freedom (Tsay 2005). The second test which is more useful in testing the ARCH effect is the Lagrange Multiplier (LM) test of Engle (1983), and this test is equivalent to the normal F-statistic. Testing $\alpha_i = 0$, for i=1,2,3,....,m in the regression is given by

$$\epsilon_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + ... + \alpha_m\epsilon_{t-m}^2 + \epsilon_t, \tag{3.79}$$

where $t = m+1, ..., T$, and ϵ_t is the error term, while m specifies the number of lags, and T is the sample size (Lee 1991, Engle 1982). Thus the null hypothesis is written as $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_m = 0$ vs $H_A : \alpha_1 = \alpha_2 = ... = \alpha_m = 0$ (one of the α_m is different from zero). Let

$$SSR_0 = \sum_{t=m+1}^{N}(\epsilon_t^2 - \varpi)^2,$$

where $\varpi = \frac{1}{N}\sum_{t=1}^{N}\epsilon_t^2$ is the sample mean of ϵ_t^2, and let

$$SSR_1 = \sum_{t=m+1}^{N} \hat{e}_t^2,$$

where e_t^2 is the least squares residual from the regression in equation 3.80. Therefore under the null hypothesis we can define the F-statistic test as

$$F = \frac{\frac{SSR_0 - SSR_1}{m}}{\frac{SSR_1}{N-2m-1}}$$

which is asymptotically distributed as chi-squared distribution with m degrees of freedom. The null hypothesis is rejected if $F > \chi_m^2(\alpha)$ or if the p-value of F is less than α, where $\chi_m^2(\alpha)$ is the upper $100(1-\alpha)$ percentile of a chi-squared distribution with m degrees of freedom (Andersen et al. 2009, Tsay 2010).

3.1.21 Model checking

Model checking or diagnostic checking is very important after fitting the ARCH and GARCH models to the data. We need to check the adequacy of the fitted GARCH model by examining the series $\hat{\epsilon}_t$, where the different graphical and statistical diagnostics are used. The standardized shocks for ARCH and GARCH model are given in the following equation:

$$\hat{\epsilon}_t = \frac{\epsilon_t}{\sigma_t}, \tag{3.80}$$

and are in the form of independently and identically distributed variables (Gouriéroux 1997, Tsay 2010).

Thus the skewness, Lagrange multiplier test for the ARCH effect, kurtosis, the Jarque-Bera test and quantile-quantile plot (q-q plot) of $\hat{\epsilon}_t$ series are used to check the validity of the distribution assumption. In addition, the Ljung-Box statistics of $\hat{\epsilon}_t$ series can also be used to check the adequacy of the mean equation and the series of $\hat{\epsilon}_t^2$ used to test the validity of the volatility equation (Francq & Zakoian 2010, Tsay 2010).

3.1.22 Chapter Summary

This chapter provided a discussion on the theory of ARCH and GARCH models and their extensions for mining companies listed on the JSE. The introduction of ARCH and GARCH models with their extensions were discussed. The parameter estimates with different distributions were also discussed, which have been applied in Chapter 4. The method of maximum log-likelihood is used to estimate the parameters, and for nonlinear optimization problems the method of Newton-Raphson and its alternative Fisher's Score is used to maximize the conditional log-likelihood. This Chapter also provided and demonstrated different equations on each subsection. The weaknesses in the ARCH model were discussed compared to the GARCH model, and also weaknesses in the GARCH model were discussed compared to its extensions. The model selection criteria were discussed as well as the way of testing for the ARCH effects in the mean. It is known that to use ARCH and GARCH models, the ARCH effect must exist in the mean.

Chapter 4

Application of ARCH, GARCH and extensions of GARCH models

4.1 Introduction

The aim of this chapter is to focus on the application of the ARCH, GARCH and their extension models to the data sets that were introduced in Chapter 3. Hence, we demonstrate how the theory is applied to real data. The ARCH, GARCH and their extension models are fitted using EVIEWS, and R softwares but R software is more focused than EVIEWS because of its appropriateness compared to EVIEWS as discussed in Chapter 3.

4.2 Selection of the Best Model

The selection of the best model, as we mentioned in Chapter 2, can be based on criteria of AIC, BIS and R^2. The larger the R^2, the better the model, and the smaller the AIC or BIS the better the model. In this work, AIC and BIS has been used to select the best model. However, AIC and BIS are not enough to conclude the best fitted model, there are other criteria the model must satisfy such as:

1. The iterative procedures which are used to estimate the parameters in the model have to converge.

2. All the parameters estimated should be significant.

3. The sum of the parameters in the model α_i and β_j, for $i = 1, 2, ..., p$ and $j = 1, 2, .., q$; should be less than or equal to 1.

4.3 Fitting the Model

It is always necessary to check if there is an ARCH effect (heteroscedasticity) in the data set before we proceed to fit the ARCH and GARCH models. The next step is to make sure that the autocorrelation presented in the mean has been removed (Tsay 2005). The daily data from the two data analyzes in Chapter 3, showed the presence of ARCH effect and the presence of autocorrelation in the mean. To check the ARCH effect and autocorrelation, the plot of ACF and PACF, for the returns of the three data sets in Chapter 3, was used and supported by the Box-Jenkins and Lagrange Multiplier test. The autocorrelation has been removed in the mean by modelling the AR (Autoregressive) model, and hence use the residual to run GARCH model. This is also the same as modeling ARMA-GARCH model (Tsay 2005, Wurtz et al. 2009, Ruppert 2011).

4.4 Analysis of Harmony Gold Mining

From Chapter 3, Figure 3.5 and 3.6 respectively showed the presence of autocorrelation in the Harmony Gold Mining Limited data and was supported by the Ljung-Box test in Table 3.2. Table 3.3 also showed that the Null hypothesis, which is: there is no ARCH effect vs alternative, there is ARCH effect was rejected at 5% level of significance, since the p-values were less than 0.05, and the calculated statistics values were too high for all lags, and we concluded there is an ARCH effect in the mean. Thus the use of the ARCH and GARCH models applied since the ARCH effect was found in the data. Hence, we can proceed to remove autocorrelation presented in the mean, where the ACF and PACF of the squared returns suggests an AR (8). Hence we model our ARMA-GARCH model, with various AR up to AR (10) to check the best fitted model. The AR1-GARCH was the best to remove the autocorrelation in the mean since it was the only one which was significant at 5% level of significance. Then the next step is to fit different GARCH models to the data with different distribution to check the best fitted model. In most of the study the order of p, and q is less or equal to 2, and this means that p can take values 0, 1, and 2, with the same being true for q. For more details of others who have done the same, please see Talke (2003) and Mzamane (2013). However, this is not always the case as it can even take value 3 and 4 since this depends on the number of lags which are removing the ARCH effect and autocorrelation in the data after fitting the GARCH model.

Table 4.1: Summary of Fitted Model with Normal Distribution

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+GARCH (1,0)	-4.1434	-4.1381	10091.1200	μ not sig.
AR1+GARCH (2,0)	-4.1615	-4.1548	10136.1600	μ not sig.
AR1+GARCH (1,1)	-4.2316	-4.2250	10306.9500	μ not sig.
AR1+GARCH (1,2)	-4.2356	-4.2276	10317.5900	μ not sig.
AR1+GARCH (2,1)	-4.2313	-4.2233	10307.0900	μ, and α_2 not sig.
AR1+GARCH (2,2)	-4.2352	-4.2259	10317.5900	only $ar1$, and α_1

Table 4.2: Summary of Fitted Model, Condition Normal Distribution

Parameter	Estimate	Std.Error	t-value	p-value
mu	$3.839e^{-04}$	$3.785e^{-04}$	1.0140	$3.1045e^{-01}$
ar1	$8.042e^{-02}$	$1.554e^{-02}$	5.1760	$2.27e^{-07}$
omega	$3.011e^{-05}$	$5.965e^{-06}$	5.0480	$4.47e^{-07}$
alpha1	$1.190e^{-01}$	$1.310e^{-02}$	9.0850	$<2e^{-16}$
beta1	$2.446e^{-01}$	$8.144e^{-02}$	3.0030	$2.67e^{-03}$
beta2	$6.080e^{-01}$	$8.003e^{-02}$	7.5970	$3.02e^{-14}$

From the result in Table 4.1, The AR1+ARCH (2) was the better fitted model rather than the AR1+ARCH (1), when we compare the lowest AIC and BIC, since the parameters' significance is the same, they are both significant except μ which does not affect anything in the result. However, this is not the best model to fit to the data when we compare it with AR1+GARCH (p,q), where p, and q are ≤ 2. Therefore AR1+ARCH (2) is dropped and we can consider the model with AR1+GARCH (p,q) as we said earlier in Chapter 2 that the GARCH model is better than the ARCH model. As a result, it is clearly seen that AR1+GARCH (1,2) is the best fitted model when we compare the smallest AIC and BIC, and is the only one removing the ARCH effect and autocorrelations represented in the residual. The summary of the fitted model is shown in Table 4.2, and all the parameter estimates are significant at 5% level of significance except μ which does not affect anything in the result. The diagnostic test of the fitted model is performed in R using fGARCH package, and we need to check if the standardized residuals are stationary and comes from normal distribution.

Table 4.3: Standardized Residual Tests

			Statistics value	p-value
Jarque-Bera Test	R	Chi^2	1072.15	0.0000
Shampiro-Wilk Test	R	W	0.9807	0.0000
Ljung-Box Test	R	Q (10)	4.0004	0.9473
Ljung-Box Test	R	Q (15)	10.1566	0.8098
Ljung-Box Test	R	Q (20)	12.6140	0.8933
Ljung-Box Test	R^2	Q (10)	14.3809	0.1563
Ljung-Box Test	R^2	Q (15)	18.7557	0.2250
Ljung-Box Test	R^2	Q (20)	25.9313	0.1681
LM Arch Test	R	TR^2	16.3396	0.1762

Table 4.3 reveals that there is no evidence of autocorrelation nor of conditional heteroscedasticity presented in data as was expected. This is confirmed by the Ljung-Box test and Engle's ARCH test, where all statistics values of standardized residuals and their squared values are not

significant. The Shampiro-Wilk and Jarque-Bera tests are considered. Since their statistical values are high, the null hypothesis is rejected. This means that the standardized residuals are not coming from the Gaussian (normal) distribution. Since the normality test is rejected we have to use the model with student t (std), skewed student t (sstd) and generalized error (ged) distributions to see the model best fitted to the data.

Table 4.4: Summary of Fitted Model with STD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+GARCH (1,1)	-4.2911	-4.2831	10452.7600	μ not sig.
AR1+GARCH (1,2)	-4.2938	-4.2844	10460.1400	μ not sig.
AR1+GARCH (2,1)	-4.2907	-4.2814	10452.7700	μ and α_2 not sig.
AR1+GARCH (2,2)	-4.2933	-4.2827	10460.1400	only $ar1$, $\alpha_1$and shape

Table 4.5: Summary of Fitted Model with SSTD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+GARCH (1,1)	-4.2922	-4.2828	10456.2700	μ not sig.
AR1+GARCH (1,2)	-4.2947	-4.2841	10463.5000	μ not sig.
AR1+GARCH (2,1)	-4.2918	-4.2811	10456.2900	μ and α_2 not sig.
AR1+GARCH (2,2)	-4.2943	-4.2823	10463.5000	only $ar1$, α_1, $skew$ and $shape$

Table 4.6: Summary of Fitted Model with GED

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+GARCH (1,1)	-4.2938	-4.2858	10459.3000	μ not sig.
AR1+GARCH (1,2)	-4.2963	-4.2869	10466.2300	μ not sig.
AR1+GARCH (2,1)	-4.2934	-4.2841	10459.3100	μ and α_2 not sig.
AR1+GARCH (2,2)	-4.2958	-4.2852	10466.2300	only $ar1$, α_1 and $shape$

Table 4.4, 4.5 and 4.6 respectively show the std, sstd and ged distributed models, where the AIC and BIC are compared to these of normal distribution. From these tables, it is clearly seen that the GARCH (1,2) model is the best fitted model compared to the GARCH (1,2) model with normal distribution. Hence we checked the best GARCH (1,2) model among the three distributions. Thus the GARCH (1,2) model with ged distribution seems to be the best fitted model to the data since it is the one with the smallest AIC and BIC. However, this is not enough to conclude that it is the best fitted model to the data, because when we check Figure 4.1 reveals that the AR1-GARCH (1,2) model with std and sstd are the best models, since all points seems to fall approximately along one 45-degree reference line compared to others and all parameters are significant. Then between these two models it is clearly seen that the AR1-GARCH (1,2) model with sstd has the smallest AIC and AR1-GARCH (1,2) model has the smallest BIC. In this case the log-likelihood can be used as well, and the best model is the one with the higher log-likelihood. Therefore we can conclude that the AR1-GARCH (1,2) model with sstd seems to be the best fitted model to the data since it is the one with the smallest AIC and BIC. The parameter estimates of the GARCH (1,2) model with sstd are presented in Table 4.7 and the diagnosis of residuals under sstd are presented in Table 4.8. The Ljung-Box test for residuals and their squares for all lags shows that there is no evidence of autocorrelation nor of conditional heteroscedasticity at 5% level of significance which was expected. The equations of the fitted model can be expressed as

$$\hat{y}_t = 3.839e^{-04} + \epsilon_t, \tag{4.1}$$

$$\hat{\sigma}_t^2 = 3.839e^{-04} + 1.242e^{-01}\epsilon_{t-1}^2 + 2.492e^{-01}\sigma_{t-1}^2 + 6.036e^{-01}\sigma_{t-1}^2. \tag{4.2}$$

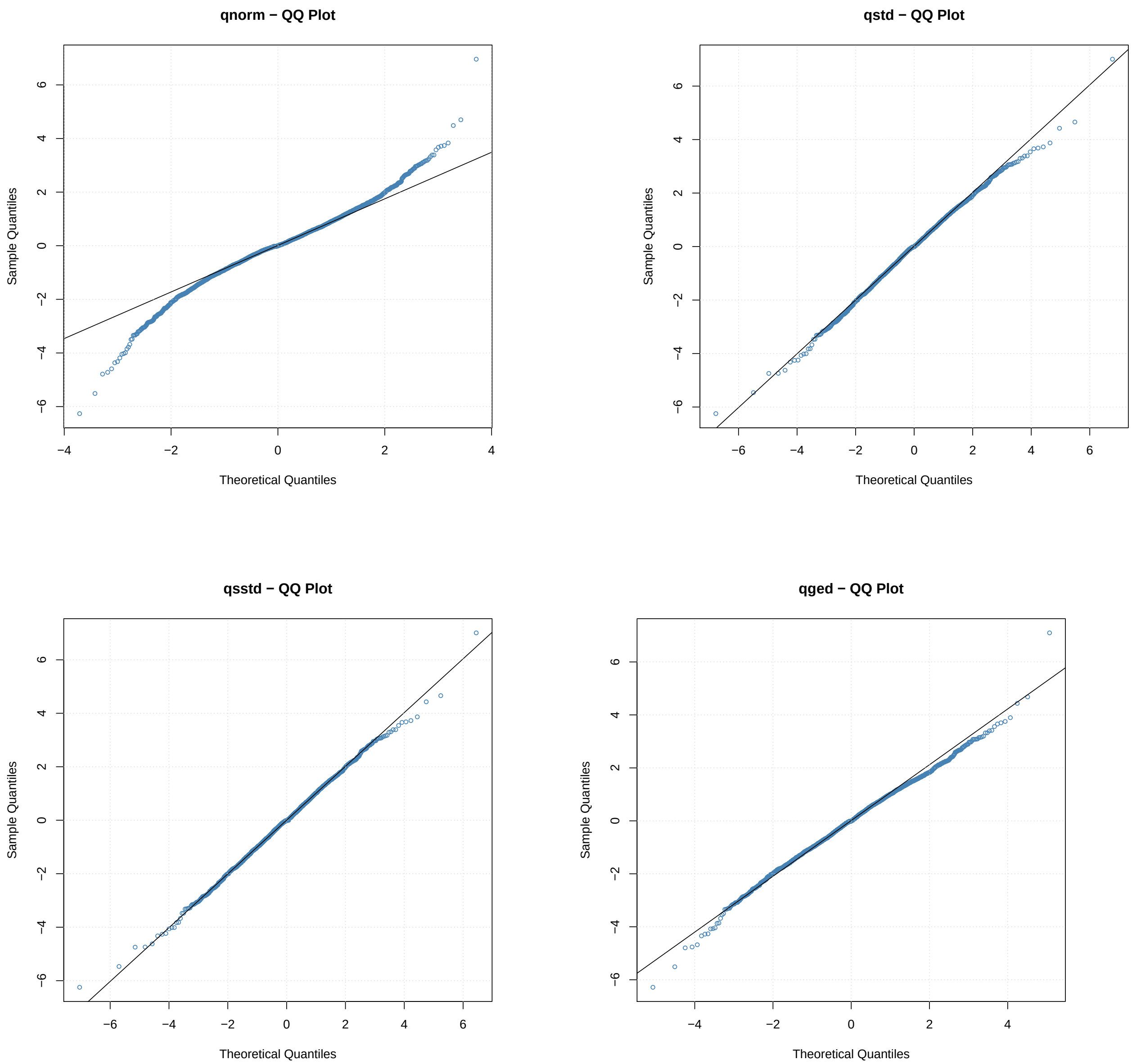


Figure 4.1: Q-Q Plot for All Models

Table 4.7: Summary of Fitted Model, Condition SSTD

Parameter	Estimate	Std.Error	t-value	p-value
mu	-2.303e^{-04}	3.080e^{-04}	-0.7480	4.550e^{-01}
ar1	1.106e^{-01}	1.477e^{-02}	7.4880	6.99e^{-14}
omega	1.851e^{-05}	4.427e^{-06}	4.1820	2.89e^{-05}
alpha1	1.023e^{-01}	1.369e^{-02}	7.4720	7.93e^{-14}
beta1	8.726e^{-01}	1.710e^{-02}	51.0410	$<$2.2e^{-16}
skew	9.746e^{-01}	1.865e^{-02}	52.2560	$<$2.2e^{-16}
shape	7.0960e^{00}	6.988e^{-01}	10.1550	$<$2.2e^{-16}

Table 4.8 shows that the statistical values of standardized residuals and their squared values are greater than zero and were expected to be zero for normality, when the Shampiro-Wilk and Jarque-Bera tests are used. Therefore the null hypothesis is rejected, and we conclude that the residuals do not come from the normal distribution. Then we can proceed to a diagnostics test for standardized residual and squared standardized residual for ACF in R. Figure 4.2 shows that the ACF of standardized residuals are not correlated, which shows the good fit of model. The squared standardized residuals show absence of correlation, and this means that the ARCH effect has been removed. Table 4.9 reveals the forecast of the fitted model up to the lag 10. The forecast of the volatility for the Harmony Gold Mining Limited company on the JSE is for the next 10 days trading. Table 4.9 shows that the AR1+GARCH (1,2) models with sstd produce good results in forecasting since the difference between forecasted volatility and observed volatility is small (Mzamane 2013).

Table 4.8: Standardized Residual Tests

			Statistics value	p-value
Jarque-Bera Test	R	Chi^2	1109.596	0.0000
Shampiro-Wilk Test	R	W	0.9803	0.0000
Ljung-Box Test	R	Q (10)	6.6395	0.7590
Ljung-Box Test	R	Q (15)	12.9200	0.6085
Ljung-Box Test	R	Q (20)	15.4544	0.7498
Ljung-Box Test	R^2	Q (10)	14.3382	0.1581
Ljung-Box Test	R^2	Q (15)	18.9997	0.2137
Ljung-Box Test	R^2	Q (20)	26.7682	0.1419
LM Arch Test	R	TR^2	16.4921	0.1697

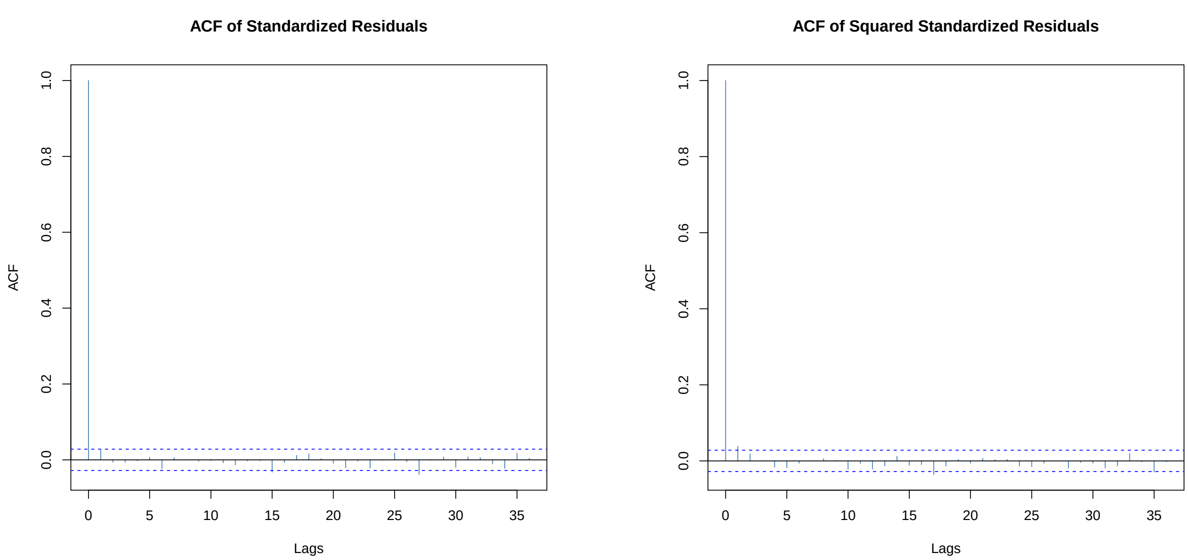


Figure 4.2: Standardized Residual for Fitted Model

Table 4.9: Forecasts of The Fitted Model

Days	mean-forecast	forecasted standard deviation	observed volatility
1	0.0005	0.0285	0.0285
2	0.0004	0.0281	0.0281
3	0.0004	0.0285	0.0285
4	0.0004	0.0284	0.0284
5	0.0004	0.0287	0.0286
6	0.0004	0.0287	0.0286
7	0.0004	0.0288	0.0288
8	0.0004	0.0289	0.0289
9	0.0004	0.0290	0.0290
10	0.0004	0.0291	0.0290

4.4.1 Application for extensions of GARCH models

The EGARCH (1,2), JGR-GARCH (1,2) and APARCH (1,2) models, all with three distributions, which are std, sstd, and ged were compared using AIC and BIC to choose the best fitted model. Table 4.10, 4.11, and 4.12 respectively show that the JGR-GARCH (1,2) model with ged seems to be the best model to fit the data since it is the one with the smallest AIC and BIC. However, this model does not remove the autocorrelation in the squared residuals as the lags increase compared to the other two as shown in Table 4.14. Figure 4.5 reveal the same compare to Figure 4.3, and 4.4 respectively. Hence the JGR-GARCH (1,2) model with sstd is considered since it removes the ARCH effect and autocorrelation in the squared residuals as shown in Table 4.13, and 4.14 respectively. Thus the parameters' estimate of the fitted model is presented in Table 4.28. After that we can proceed with graphs and forecast the fitted model. Moreover, Figure 4.7 exhibits no autocorrelation except on lag 1, 17 and 35.

Figure 4.8 reveals that in Harmony Gold Mining Limited company there is evidence of high volatility clustering and lower volatility clustering. Figure 4.6 shows the leptokurtic distribution (fat tail) in the residuals.

Table 4.10: Summary of Fitted Model with STD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+EGARCH (1,2)	-4.2915	-4.2809	10455.7700	α_1 not sig.
AR1+JGR-GARCH (1,2)	-4.2926	-4.2819	10458.2600	β_1, and γ_1 not sig.
AR1+APARCH (1,2)	-4.2926	-4.2806	10459.4300	ω, β_1 and γ_1 not sig.

Table 4.11: Summary of Fitted Model with SSTD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+EGARCH (1,2)	-4.2918	-4.2798	10457.4700	α_1 not sig.
AR1+JGR-GARCH (1,2)	-4.2930	-4.2810	10460.3100	μ, β_1, γ_1 not sig.
AR1+APARCH (1,2)	-4.2930	-4.2797	10461.3800	μ, ω, β_1 and γ_1 not sig.

Table 4.12: Summary of Fitted Model with GED

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR1+EGARCH (1,2)	-4.2931	-4.2825	10459.6000	α_1 not sig.
AR1+JGR-GARCH (1,2)	-4.2947	-4.2841	10463.5300	γ_1 not sig.
AR1+APARCH (1,2)	-4.2946	-4.2826	10464.2200	ω, β_1 and γ_1 not sig.

Table 4.13: Weighted ARCH LM Tests for Autocorrelation

	JGR-GARCH (1,2) with std	JGR-GARCH (1,2) with sstd	JGR-GARCH (1,2) with ged
Lags	P-Values	P-Values	P-Values
4	0.2465	0.2560	0.2555
6	0.3697	0.3831	0.3881
8	0.5940	0.6096	0.6153

Table 4.14: Weighted Ljung-Box Test on Standardized Squared Residuals

	JGR-GARCH (1,2) with std	JGR-GARCH (1,2) with sstd	JGR-GARCH (1,2) with ged
Lags	P-Values	P-Values	P-Values
1	0.0061	0.0053	0.0031
8	0.0306	0.0261	0.0163
14	0.0752	0.0673	0.0465

Table 4.15: Summary of Fitted Model, Condition STD

Parameter	Estimate	Std.Error	t-value	p-value
mu	0.0011	0.0003	3.1288	0.0018
ar1	0.0583	0.0145	4.0251	0.0001
omega	0.0000	0.0000	2.9215	0.0035
alpha1	0.1396	0.0252	5.5383	0.0000
beta1	0.2597	0.1338	1.9417	0.0522
beta2	0.5946	0.1370	4.3410	0.0000
gamma1	-0.0270	0.0247	-1.0942	0.2739
shape	5.6538	0.4916	11.5008	0.0000

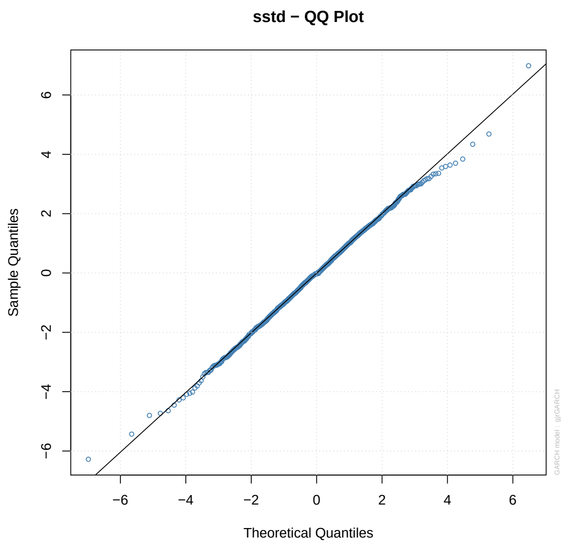


Figure 4.3: Q-Q Plot for JGR-GARCH (1,2) with STD

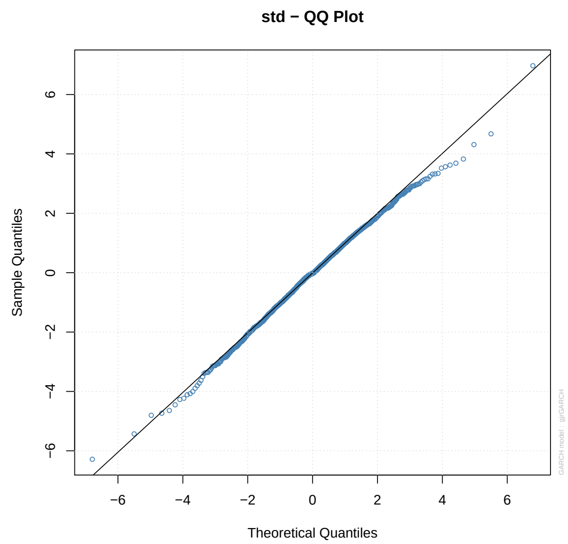


Figure 4.4: Q-Q Plot for JGR-GARCH (1,2) with SSTD

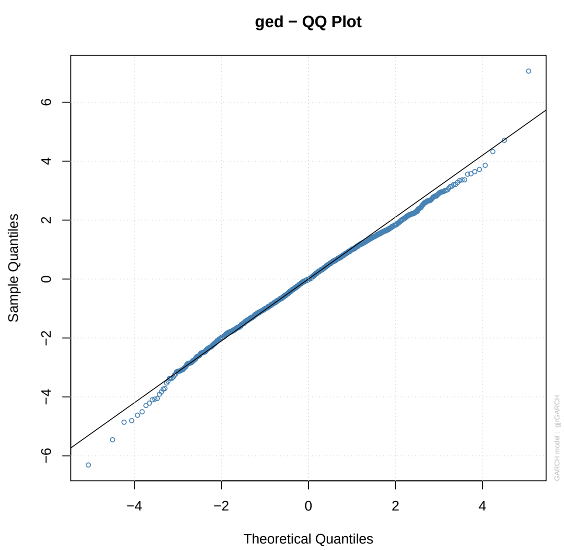


Figure 4.5: Q-Q Plot for JGR-GARCH (1,2) with GED

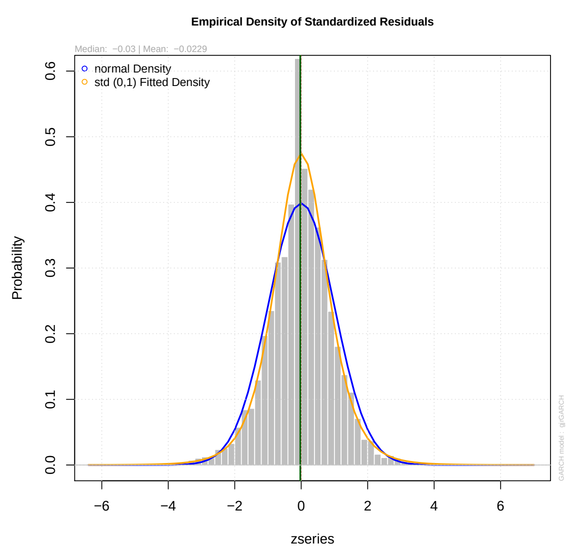


Figure 4.6: Empirical Density of Standardized Residuals

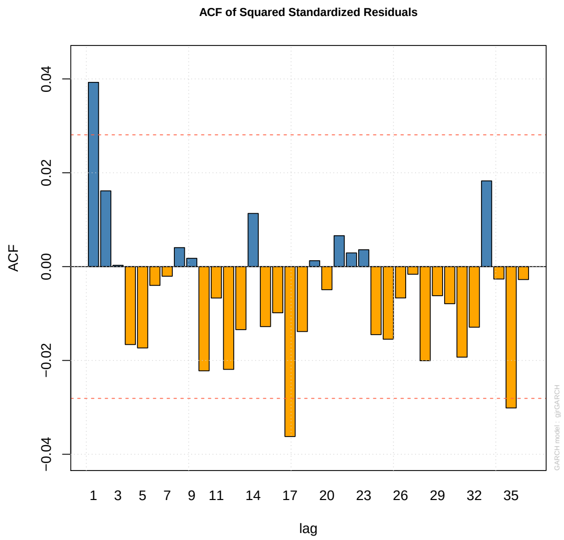


Figure 4.7: ACF of Squared Standardized Residuals

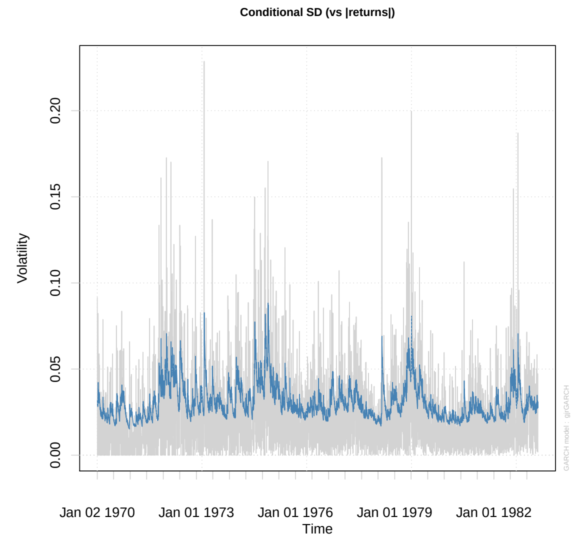


Figure 4.8: Conditional Squared Standardized Residuals

4.5 Analysis of Impala Platinum Holdings Limited

In Chapter 3 Table 3.6 showed the presence of autocorrelation in the Impala Platinum Holdings Limited data. Table 3.7 also showed that the ARCH effect was presented in the Impala Platinum Holdings Limited data. This resulted in the use of the ARCH and GARCH models in Impala Platinum Holdings Limited data. After that we can proceed to remove autocorrelation presented in the mean, where the ACF and PACF of the squared returns suggests an AR (4), hence we model our ARMA-GARCH model, with various AR up to AR (4) to check the best fitted model. The AR2-GARCH was the best to remove the autocorrelation and ARCH effect in the means since it was the only one which was significant at 5% level of significance. The next step is to fit different GARCH models to the data with different error distributions to check the best fitted model, where the order of p and q is less or equal to 2. Table 4.16 shows that AR2+GARCH (1,2) is the best fitted model to the data fit since it is the one with the smallest AIC and BIC compared to others. Also, the sum of α_1, β_1 and β_2 is 0.9795 which is less than 1. The summary of the fitted model is shown in Table 4.17 and the parameter estimates are significant at 5% level of significance. The diagnostic test of the fitted model is performed in R using fGARCH package, and we have to check if the standardized residual are stationary.

Table 4.16: Summary of Fitted Model with Normality Distribution

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR2+GARCH (1,0)	-4.4387	-4.4320	10811.0500	μ not sig.
AR2+GARCH (2,0)	-4.4676	-4.4596	10882.3800	μ not sig.
AR2+GARCH (1,1)	-4.5193	-4.5113	11008.1300	μ not sig.
AR2+GARCH (1,2)	-4.5211	-4.5117	11013.5400	μ not sig.
AR2+GARCH (2,1)	-4.5188	-4.5095	11008.1000	α_2 and μ not sig.
AR2+GARCH (2,2)	-4.5207	-4.5100	11013.5400	only $ar1$, $ar2$ and α_1

Table 4.17: Summary of Fitted Model, Conditional Normal Distribution

Parameter	Estimate	Std.Error	t-value	p-value
mu	-5.691e^{-04}	3.309e^{-04}	-1.7200	8.54e^{-02}
ar1	1.029e^{-01}	1.602e^{-02}	6.4260	1.31e^{-10}
ar2	-5.222e^{-02}	1.511e^{-02}	-3.4550	6e^{-04}
omega	2.741e^{-05}	6.633e^{-06}	4.1320	3.60e^{-05}
alpha1	1.139e^{-01}	1.650e^{-02}	6.9010	5.16e^{-12}
beta1	4.664e^{-01}	9.730e^{-02}	4.7940	1.63e^{-06}
beta2	3.853e^{-01}	9.295e^{-02}	4.1450	3.40e^{-05}

Table 4.18 reveals that there is no evidence of autocorrelation nor of conditional heteroscedasticity presented in data. This is confirmed by the Ljung-Box test and Engle's ARCH test, where all statistics values of standardized residual and their squared values are not significant. The Shampiro-Wilks and Jarque-Bera test are considered, since their statistical values are high the null hypothesis is rejected. This means that the standardized residual do come from the gaussian (normal) distribution. Since the normality test is rejected we have to try the model

with std and sstd. The model with ged is better than normal distribution but not with std and sstd. Thus the std and sstd are used, where the AIC and BIC are compared to these of normal distribution.

Table 4.18: Standardized Residuals Tests

			Statistics value	p-value
Jarque-Bera Test	R	Chi^2	6400.686	0.0000
Shampiro-Wilk Test	R	W	0.9689	0.0000
Ljung-Box Test	R	Q (10)	15.7180	0.1080
Ljung-Box Test	R	Q (15)	22.2533	0.1013
Ljung-Box Test	R	Q (20)	24.3999	0.2254
Ljung-Box Test	R^2	Q (10)	9.5418	0.4816
Ljung-Box Test	R^2	Q (15)	13.2276	0.5847
Ljung-Box Test	R^2	Q (20)	14.7366	0.7913
LM Arch Test	R	TR^2	10.8185	0.5445

Table 4.19: Summary of Fitted Model with STD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR2+GARCH (1,1)	-4.5962	-4.5869	11196.5200	μ not sig.
AR2+GARCH (1,2)	-4.5971	-4.5865	11199.7600	μ not sig.
AR2+GARCH (2,1)	-4.5958	-4.5852	11196.5400	μ and α_2 not sig.
AR2+GARCH (2,2)	-4.5967	-4.5847	11199.7600	only $ar1$, $ar2$, α_1 and $shape$

It is clearly seen from Table 4.19, 4.20 respectively, that the GARCH (1,2) model with std and sstd is the best fitted model compared to the GARCH (1,2) model with normal distribution. Hence we checked the best GARCH (1,2) model among the two distributions. Thus the GARCH (1,2) model with std distribution seems to be the best fitted model to the data since it is the one with the smallest AIC and BIC.

Table 4.20: Summary of Fitted Model with SSTD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR2+GARCH (1,1)	-4.5958	-4.5852	11196.5600	μ not sig.
AR2+GARCH (1,2)	-4.5968	-4.5848	11199.7900	μ not sig.
AR2+GARCH (2,1)	–4.5954	-4.5834	11196.5800	μ and α_2 not sig.
AR2+GARCH (2,2)	-4.5963	-4.5830	11199.7900	only $ar1$, $ar2$, α_1, $skew$ and $shape$

The parameter estimates of the GARCH (1,2) model with std are presented in Table 4.21 and diagnosis of residuals under std is presented in Table 4.22. The Ljung-Box test for residuals and their squares for all lags shows that there is no evidence of autocorrelation nor of conditional heteroscedasticity at 5% level of significance.

Table 4.21: Summary of Fitted Model with STD

Parameter	Estimate	Std.Error	t-value	p-value
mu	-2.627e^{-04}	2.994e^{-04}	-0.8770	3.803e^{-01}
ar1	9.170e^{-02}	1.516e^{-02}	6.0480	1.47e^{-09}
ar2	-5.300e^{-02}	1.439e^{-02}	-3.6820	2e^{-04}
omega	2.071e^{-05}	6.605e^{-02}	3.1350	1.7e^{-03}
alpha1	1.365e^{-01}	2.211e^{-02}	6.1770	6.55e^{-10}
beta1	5.172e^{-01}	1.203e^{-01}	4.2990	1.72e^{-05}
beta2	3.258e^{-01}	1.131e^{-01}	2.8810	4e^{-03}
shape	5.9260e^{00}	4.910e^{-01}	12.0690	<2.2e^{-04}

Table 4.22 shows that the statistical values of standardized residuals and their squared values are greater than zero and were expected to be zero for normality, when the Shampiro-Wilk and Jarque-Bera tests are used. Therefore the null hypothesis is rejected, and we conclude that the residuals do not come from the normal distribution. Afterwards, we can proceed to diagnostics test for standardized residuals and squared standardized residuals for ACF in R. Figure 4.10 shows that the ACF of standardized residuals are not correlated, which indicates the good fit

of model. The squared standardized residuals show absence of correlation, and this means that the ARCH effect has been removed. Table 4.23 shows the forecast of the fitted model up to the lag 10.

Table 4.22: Standardized Residuals Tests

			Statistics value	p-value
Jarque-Bera Test	R	Chiˆ2	7165.759	0.0000
Shampiro-Wilk Test	R	W	0.9667	0
Ljung-Box Test	R	Q (10)	16.2197	0.0935
Ljung-Box Test	R	Q (15)	22.6480	0.0919
Ljung-Box Test	R	Q (20)	24.5548	0.2190
Ljung-Box Test	Rˆ2	Q (10)	10.0171	0.4390
Ljung-Box Test	Rˆ2	Q (15)	14.9959	0.4517
Ljung-Box Test	Rˆ2	Q (20)	16.8457	0.6630
LM Arch Test	R	TRˆ2	11.9065	0.4532

Figure 4.9 reveals that the AR2-GARCH (1,2) model with std and sstd is the best fitted model to the data compared to the model with normal error distribution. The model with std is the best model since all points seems to fall approximately along one 45-degree reference line compared to others and its AIC and BIC are lower than the one for sstd.

The equation of the fitted model is given by

$$\hat{y}_t = -2.627e^{-04} + \epsilon_t, \tag{4.3}$$

$$\hat{\sigma}_t^2 = -2.627e^{-04} + 1.365e^{-01}\epsilon_{t-1}^2 + 5.172e^{-01}\sigma_{t-1}^2 + 3.258e^{-01}\sigma_{t-1}^2 \tag{4.4}$$

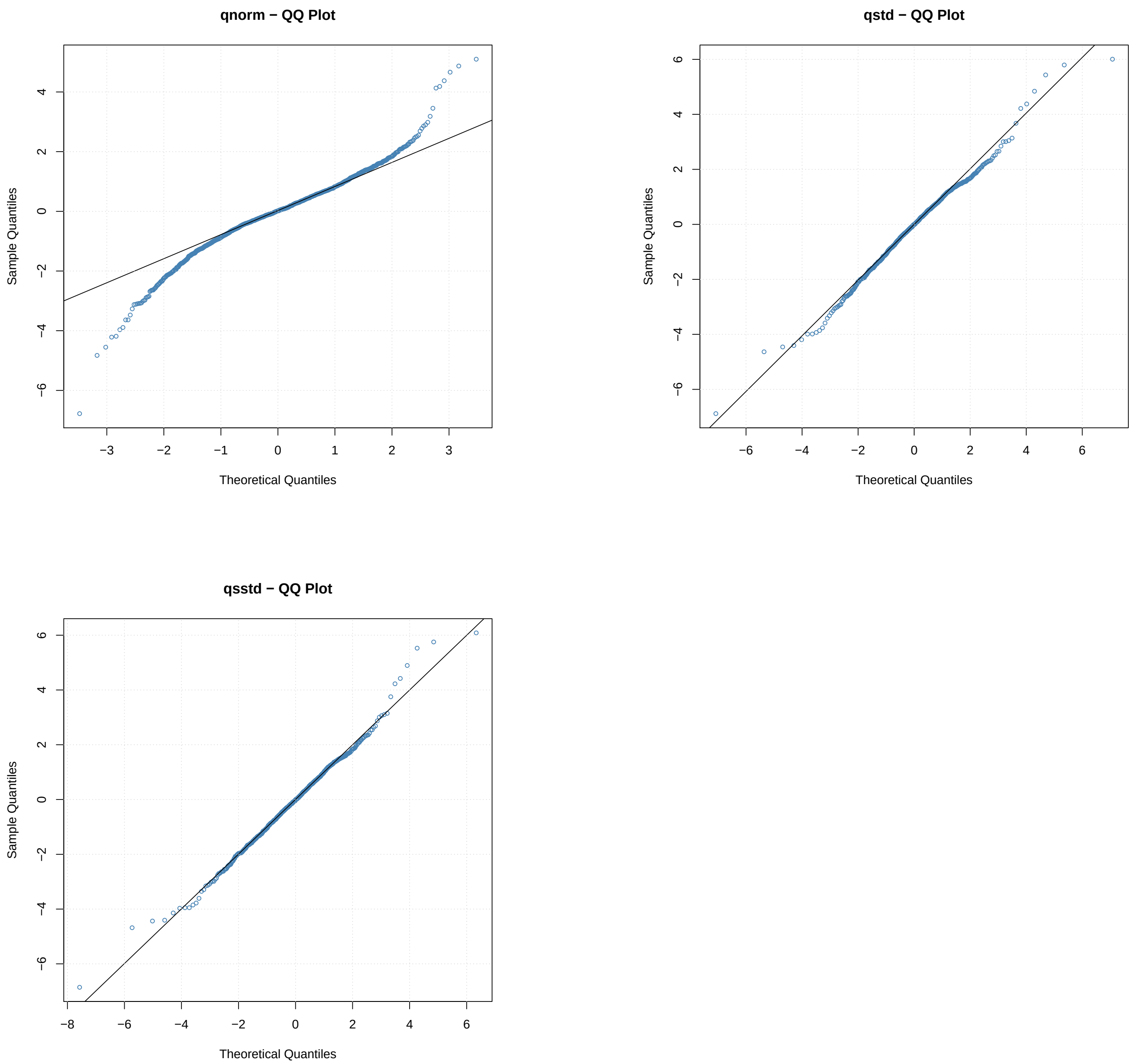


Figure 4.9: Q-Q Plot for Three Distributions

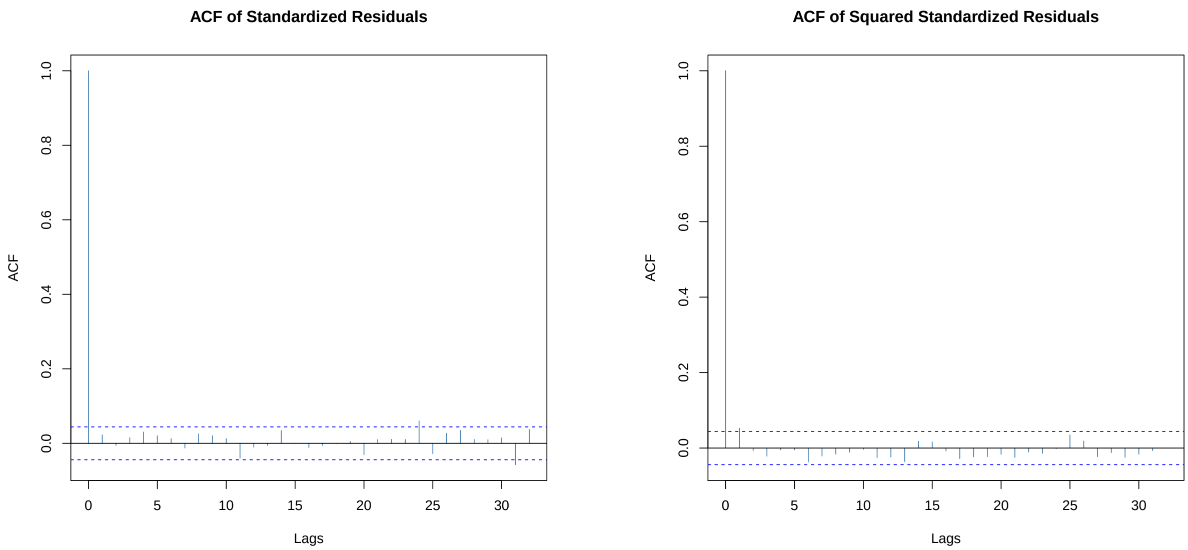


Figure 4.10: Standardized Residuals

Table 4.23: Forecasts of The Fitted Model

Days	mean-forecast	forecasted standard deviation	observed volatility
1	0.0052	0.3680	0.3680
2	-0.0079	0.3729	0.3728
3	-0.0083	0.3776	0.3775
4	-0.0083	0.3823	0.3822
5	-0.0083	0.3870	0.3869
6	-0.0083	0.3917	0.3916
7	-0.0083	0.3964	0.3963
8	-0.0083	0.4010	0.4009
9	-0.0083	0.4057	0.4055
10	-0.0083	0.4103	0.4101

4.5.1 Application for extensions of GARCH models

The EGARCH (1,2), JGR-GARCH (1,2) and APARCH (1,2), all with two distributions, which are std and sstd were compared using AIC and BIC to choose the best fitted model. Table 4.24 and 4.25 respectively show that the EGARCH (1,2) model is the best model to fit the data since it is the one with smallest AIC and BIC. Moreover, we have to choose the best EGARCH model within two distributions. The EGARCH (1,2) model with std seems to be fitted the model to the data well, when we compare the smallest AIC and BIC. Table 4.26 and 4.27 respectively show that the autocorrelation and ARCH effects which were represented in the squared residuals mean have been removed. Thus the parameters estimate of the fitted model are presented in Table 4.28. Figure 4.13 exhibits no autocorrelation except on lag 32. Figure 4.14 reveals that in Impala Platinum Holdings Limited there is evidence of high volatility clustering and lower volatility clustering and Figure 4.12 shows the leptokurtic distribution (fat tail) in the residuals.

Table 4.24: Summary of Fitted Model with STD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR2+EGARCH (1,2)	-4.6046	-4.5926	11218.9700	μ, ω and γ_1 not sig.
AR2+JGR-GARCH (1,2)	-4.5981	-4.5861	11202.9900	μ, and ω not sig.

Table 4.25: Summary of Fitted Model with SSTD

Model	AIC	BIC	Log likelihood	sig.parameters at5%
AR2+EGARCH (1,2)	-4.6042	-4.5909	11218.9700	μ, ω and γ_1 not sig.
AR2+JGR-GARCH (1,2)	-4.5977	-4.5843	11203.0000	μ, and ω not sig.

Table 4.26: Weighted ARCH LM Tests for Autocorrelation

	EGARCH (1,2) with std	EGARCH (1,2) with sstd
Lags	P-Value	P-Value
4	0.7931	0.7931
6	0.4671	0.4670
8	0.3368	0.3368

Table 4.27: Weighted Ljung-Box Test on Standardized Squared Residuals

	EGARCH (1,2) with std	EGARCH(1,2) with sstd
Lags	P-Value	P-Value
1	0.3246	0.3249
8	0.3993	0.3995
14	0.3616	0.3617

Table 4.28: Summary of Fitted Model, Condition STD

Parameter	Estimate	Std.Error	t value	p-value
mu	0.0000	0.0003	0.0297	0.9763
ar1	0.0895	0.0166	5.3802	0.0000
ar2	-0.0549	0.0180	-3.0624	0.0022
omega	-0.2083	0.3951	-0.5271	0.5981
alpha1	0.0293	0.0144	2.0344	0.0419
beta1	0.5327	0.0265	20.0849	0.0000
beta2	0.4389	0.0272	16.1654	0.0000
gamma1	0.2510	0.1464	1.7142	0.0865
shape	5.7504	0.5916	9.7209	0.0000

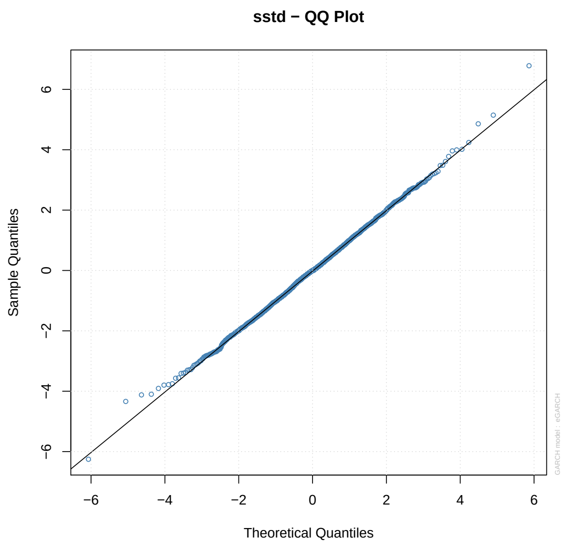


Figure 4.11: Q-Q Plot for EGARCH (1,2) with STD

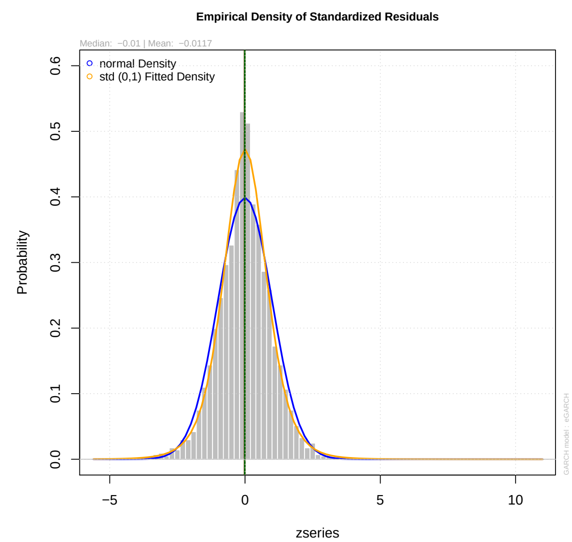


Figure 4.12: Empirical Density of Standardized Residuals

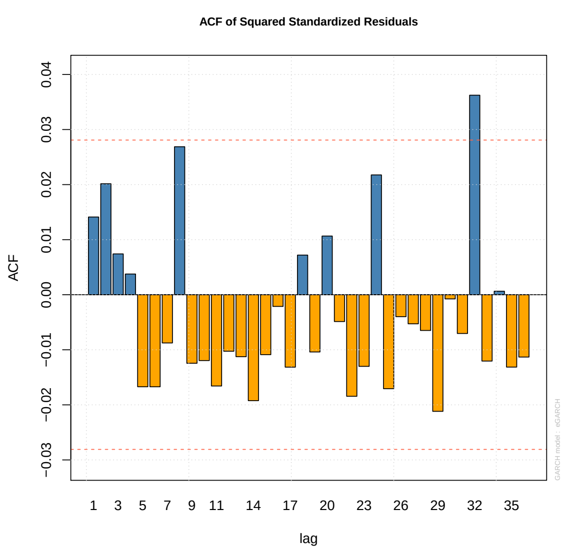


Figure 4.13: ACF of Squared Standardized Residuals

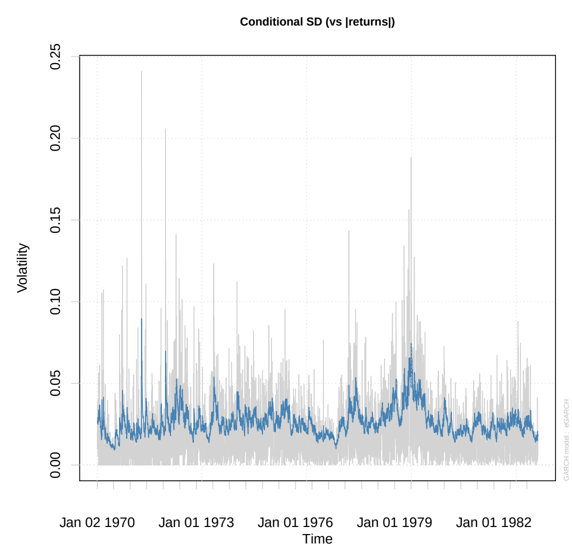


Figure 4.14: Conditional Squared Standardized Residuals

4.5.2 Chapter Summary

In this chapter the application of the GARCH model and its extensions were applied with different distributions for various daily data of mining companies for JSE within the same time period. After trying different GARCH models it was found that in daily returns the autocorrelation and ARCH effects which were presented in mean can be removed by modeling the combination of an Autoregressive (AR) model with GRACH (p,q) models. In the Harmony Gold Mining company the AR1+GARCH (1,2) with normal distribution was found to be the best model to remove the autocorrelations and ARCH effects were presented in the mean. However, the Shampiro-Wilk and Jarque-Bera tests showed that the residuals do not come from the normal distribution, and this results in the use of other distributions. The std, sstd and ged were compared and it was found that the AR1+GARCH (1,2) seemed to be the best fit model to the data. The extensions of the GARCH model compared were, EGARCH (1,2), JGR-GARCH (1,2) and APARCH (1,2) with std, sstd and ged. The AR1+JGR-GARCH (1,2) with sstd was the best model to fit the data. The forecast for the next 10 days of trading for the best fit model produced good results, since the difference between forecasted volatility and observed volatility was small. Impala Platinum Holdings Limited showed that AR2+GARCH (1,2) with normal distribution was the best model to fit the data and remove the autocorrelation and ARCH effects were presented in the mean. The Shapiro-Wilks and Jarque-Bera tests also showed that the residuals do not come from the normal distribution. Furthermore, the two distributions (std and sstd) were used since ged is almost the same as normal distribution, as explained in Chapter 2, and proven with Harmony Gold Mining company that it is not better than std and sstd. The AR2+GARCH (1,2) with std was the best model to fit the data. The extensions of the GARCH model, the AR2+EGARCH (1,2) with std was found to be the best fit model to the data. The forecast was also done the same way as with the Harmony Gold Mining company. The sum of parameter estimates in Harmony Gold Mining Limited company and Impala Platinum Holdings Limited are $\alpha_1 + \beta_1 + \beta_2 = 0.9716$ and $\alpha_1 + \beta_1 + \beta_2 = 0.9656$ respectively. Hence we can conclude that Harmony Gold Mining Limited company is more volatile than Impala Platinum Holdings Limited since it is the one with higher $\alpha_1 + \beta_1 + \beta_2 = 0.9716$.

Chapter 5

Multivariate GARCH model

5.1 Introduction

5.1.1 History of Multivariate GARCH model

Multivariate models were introduced by Bollerslev et al. (1988), to study the conditional covariance and correlation between multiple markets, after realizing that the univariate GARCH models can not handle such situations. The first model they suggested was the Vector (VEC) GARCH model, then later other authors started to develop more multivariate GARCH models such as the BEKK model by Engle & Kroner (1995) which was named after Baba, Engle, Kraft and Kroner (1995); the constant conditional correlation (CCC) by Bollerslev (1990), which was extended later by (Engle 2002, Tse & Tsui 2002) to the dynamic conditional correlation (DCC) model. The reader is referred to (Pesaran & Pesaran 2010, Chang et al. 2011, Wang & Wu 2012) for more multivariate models. Modelling multivariate volatility is an extension of modelling univariate volatility, where the dynamic relation in the volatility of multiple returns series is considered to make decisions regarding asset returns. In the modelling of volatility the crucial part is the covariance matrix, which is the conditional covariance matrix of multiple asset returns. Hence modelling the co-movement of multiple returns series among the different markets is of great significant (Mzamane 2013). Multivariate volatilities play very important roles in financial applications, such as portfolio selection and asset allocation. Furthermore, it can be applied in computation of the value at risk of a financial position with multiple assets (Tsay 2010). However, the multivariate time series faces some challenges in practice due to the complexity of models and the increasing number of parameters in the model.

In this study we focus on the Dynamic Conditional Correlation (DCC-GARCH) model introduced by Engle (2002), however the other multivariate techniques can also be used. We are modelling the conditional variance and correlations rather than modelling the conditional covariance matrix which is used in the Constant Conditional Correlation (CCC-GARCH) model. We consider the multivariate GARCH model with return series $\gamma_t = (\gamma_{1t}, ..., \gamma_{Nt})''$, of dimension conditional mean μ_t and $N \times N$ covariance matrix H_t, such that $\gamma_t = \mu_t + \epsilon_t$, where $\mu_t = E[\gamma_t|\phi_{t-1}]$ is the conditional mean of γ_t given that the past information ϕ_{t-1} is given by

$$\epsilon_t = \sqrt{H_t}\varphi_t,$$

where φ_t is an $N \times 1$ series of independent and identically distributed random vector, and satisfies the following conditions:

- $E[\varphi_t] = 0$
- $E[\varphi_t\varphi_t'] = I_N$,

where I_N is identity matrix of order N (Ijumba 2013). The conditional covariance matrix $\sqrt{H_t}$ is positive definite, with $N \times N$ dimension and is given by equation:

$$H_t = Cov[\gamma_t|\phi_{t-1}] = \begin{bmatrix} \sigma_{11,t} & \cdots & \sigma_{N1,t} \\ \vdots & \ddots & \vdots \\ \sigma_{1N,t} & \cdots & \sigma_{NN,t} \end{bmatrix}, \tag{5.1}$$

where the diagonal element $\sigma_i i, t$ is the variance of the $i^t h$ return, and $(i,j)^t h$ element is the covariance between $\gamma_i t$ and $\gamma_j t$. From the equation above the process is said to be second order stationary if:

1. $E[\gamma_{it}^2] < \infty, \forall t \in \mathbb{Z}, i = 1, ..., N,$

2. $E[\gamma_t] = \mu, \forall t \in \mathbb{Z},$

3. $Cov[\varphi_t \varphi_{t+h}] = E[(\gamma_t - \mu)(\varphi_{t+h} - \mu)'] = \Gamma(h), \forall t, h \in \mathbb{Z},$

where $\Gamma(h)$ is covariance matrix of dimension $N \times N$. The mean is N- dimension vector of conditional expectation of γ_t. This is known that at lag zero, where $\Gamma_\gamma(0) = var(\gamma t)$ is a symmetric matrix.

5.2 Dynamic Conditional Correlation (DCC) model

The DCC model is the extension of the CCC model developed by Bollerslev (1990) to model the conditional covariance matrix. The CCC model does not model the conditional variance and correlations, furthermore the assumption of constant seems unrealistic in real application. Thus Engle (2002) and Tse & Tsui (2002) extend the CCC model to the DCC model to overcome these problems. The idea of developing the DCC model was to model the conditional variance and correlation rather than modelling the conditional covariance matrix.

This model can be expressed as

$$H_t = D_t R_t D_t, \tag{5.2}$$

where $D_t = Diag(\sqrt{h_{ii,t}})$, is an $N \times N$ diagonal matrix, R_t is the unconditional correlation matrix of the standardized residuals and $h_i i$ can be any univariate GARCH model. From the equation 5.2, R_t is time varying and this has been developed by Tse & Tsui (2002) in the following equation:

$$R_t = (1 - \theta_1 - \theta_2)R + \theta_2 R_{t-1} + \theta_1 \varphi_{t-1} \varphi_{t-1}^{'} \tag{5.3}$$

where $\theta_1 + \theta_2 < 1$ and θ_1, θ_2 are both assumed to be non-negative parameters. When $\theta_1 = \theta_2 = 0$ this was the case for the CCC test. In addition $R = \rho_{ij}$ is a time invariant $N \times N$ positive definite correlation matrix with unit diagonal element ($\rho_{ii} = 1$). The matrix form for R_t is expressed by

$$R_t = \begin{bmatrix} 1 & \rho_{12,t} & \rho_{13,t} & \cdots & \rho_{1n,t} \\ \rho_{21,t} & 1 & \rho_{23,t} & \cdots & \rho_{2n,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1,t} & \rho_{n2,t} & \cdots & \sigma_{nn-1,t} & 1 \end{bmatrix} \tag{5.4}$$

and φ_{t-1} is $N \times N$ correlation matrix, which can be expressed as

$$\varphi_{ij,t-1} = \frac{\sum_{l=1}^{t} \lambda_{i,t} - h\lambda_{j,t-h}}{\sqrt{\left(\sum_{l=1}^{t} \lambda_{i,t-h}^2\right)\left(\sum_{l=1}^{t} \lambda_{i,t-l}^2\right)}}, \tag{5.5}$$

where $\lambda_{i,t} = \frac{\epsilon_{i,t}}{\sqrt{\lambda_{ii,t}}}$, for $i = 1, 2...N$ and $1 \leq i < j \leq N$. The necessary condition for φ_{t-1} to be positive definite is $t > N$, hence R_t is also positive definite.

Engle (2002) expressed the DCC model as

$$H_t = D_t R_t D_t, \tag{5.6}$$

where

$$R_t = (P_t)^{-1} P_t (P_t)^{-1} \tag{5.7}$$

and $P_t = diag(P_t)^{\frac{1}{2}}$. P_t is the $N \times N$ symmetric positive definite matrix, with

$$P_t = (1 - \theta_1 - \theta_2)\hat{P} + \theta_2 R_{t-1} + \theta_1 \varphi_{t-1}\varphi_{t-1}^{'} \tag{5.8}$$

where $0 \leq \theta_1 + \theta_2 < 1$

$\hat{P} = cov[\varphi_t \varphi_t^{'}]$ is the $N \times N$ unconditional covariance matrix estimated by

$$\hat{P} = \frac{1}{T}\sum_{t=1}^{T} \varphi_t \varphi_t^{'} \tag{5.9}$$

Therefore the single correlation of R_t is given by

$$\rho_{ij,t} = \frac{p_{ij,t}}{\sqrt{p_{ii,t}p_{jj,t}}} \leq 1, \tag{5.10}$$

where $p_{ij,t}$ are all elements of matrix P_t and this P_t has to be positive definite to ensure that R_t is also positive definite (Andersen et al. 2009).

5.3 Dynamic Conditional Correlation (DCC) model parameter estimation

In this section the parameter estimation of the DCC model is determined using the likelihood, we follow the estimation procedure from Engle (2002), under the multivariate Gaussian distribution for the standardized error ϵ_t. The likelihood function is given by

$$L(\theta) = \prod_{t=1}^{T} \frac{1}{(2\Pi)^{\frac{n}{2}}(|H|)^{\frac{1}{2}}} \exp\left(\frac{-1}{2}\epsilon_t^T H_t^{-1}\epsilon_t\right) \tag{5.11}$$

where T is the time period used to estimate the model, and the parameter of model $\theta = (\eta, \omega) = (\eta_1, ..., \eta_n, \omega)$

and $\eta_i = \mu, \alpha_{11}...., \alpha_{1p}, \beta_{11}, ..., \beta_{nq}$, which are the parameters of univariate GARCH (p,q) model for i^{th} asset series, $i = 1, 2, ..., n$ and $\omega = (\theta_1, \theta_2)$ are the parameters of the correlation in equation 5.8. Thus the log-likelihood from equation 5.11 is given by

$$\begin{aligned} log(L(\theta)) &= \frac{-1}{2}\sum_{t=2}^{T}\left(nlog(2\Pi) + log(|H_t|) + \epsilon_t' H_t^{-1}\epsilon_t\right) \\ &= \frac{-1}{2}\sum_{t=2}^{T}\left(nlog(2\Pi) + log(|D_tR_tD_t|) + \epsilon_t' D_t^{-1}R_t^{-1}D_t^{-1}\epsilon_t\right) \\ &= \frac{-1}{2}\sum_{t=2}^{T}\left(nlog(2\Pi) + 2log(|D_t|) + log|R_t| + \epsilon_t' D_t^{-1}R_t^{-1}D_t^{-1}\epsilon_t\right). \end{aligned} \tag{5.12}$$

This can be broken up into two parts, where the first part is the volatility component, which is the sum of the univariate GARCH likelihood and the correlation component. In the first stage, the parameter η is estimated for each asset series, and the likelihood is estimated by replacing R_t with the identity matrix I_N, while in the second stage the correlation parameter ω is estimated using the correctly specified log-likelihood given the parameter η (Orskaug 2009). Therefore, the volatility component in the first stage is given by

$$\begin{aligned} log(L_v(\eta)) &= \frac{-1}{2}\sum_{t=2}^{T}\Big(nlog(2\Pi) + 2log(|D_t|) + log(|I_N|) + \epsilon_t^{'}D_t^{-1}I_N D_t^{-1}\epsilon_t\Big) \\ &= \frac{-1}{2}\sum_{t=2}^{T}\Big(nlog(2\Pi) + 2log(|D_t|) + \epsilon_t^{'}D_t^{-1}I_N D_t^{-1}\epsilon_t\Big) \\ &= \frac{-1}{2}\sum_{t=2}^{T}\left(nlog(2\Pi) + \sum_{i=1}^{n}[log(h_{it})] + \frac{\epsilon_{it}^2}{h_{it}}\right) \\ &= \sum_{i=1}^{n}\left(\frac{-1}{2}[log(h_{it})] + \frac{\epsilon_{it}^2}{h_{it}} + k\right) \end{aligned} \tag{5.13}$$

where k is a constant. The second stage is estimated following the first stage and the correlation parameters are unknown at this stage. Thus the quasi-likelihood for the correlation component is expressed as

$$log(L_c(\eta/\omega, \epsilon_t)) = \frac{-1}{2}\sum_{t=1}^{T}\Big(nlog(2\Pi) + 2log(|D_t|) + log(|R_t|) + \epsilon_t^{'}D_t^{-1}R_t^{-1}D_t^{-1}\epsilon_t\Big)$$

$$log(L_c(\eta/\omega, \epsilon_t)) = \frac{-1}{2}\sum_{t=1}^{T}\Big(nlog(2\Pi) + 2log(|D_t|) + log(|R_t|) + \epsilon_t^{'}R_t^{-1}\epsilon_t\Big). \tag{5.14}$$

The parameter selection for log-likelihood is influenced by $log(|R_t|) + \epsilon_t^{'}R_t^{-1}\epsilon_t$, since it is based on the conditional of ω where the constant term is removed in the parameter estimation of the DCC model (Engle & Sheppard 2001).

Therefore the maximization of the the equation 5.12 becomes

$$log(L_c(\eta/\omega, \epsilon_t)) = \frac{-1}{2}\sum_{t=1}^{T}\Big(log|R_t| + \epsilon_t^{'}R_t^{-1}\epsilon_t\Big). \tag{5.15}$$

5.4 DCC model diagnostics

It is very important to check the adequacy of the estimated DCC model. The different tests have been proposed for univariate and multivariate GARCH model, such as Portmanteau test of Li & Mak (1994) for testing the adequacy of the univariate GARCH model and Ljung-Box test for McLeod & Li (1983) for both univariate and multivariate models. Hence the Ljung-Box test is the best since we are interested in the multivariate GARCH model. This was proven by (Tse & Tsui 1999), where they showed that Portmanteau test has lower power, since it is based on a single pair of residuals and its asymptotic null distribution is unknown, while its statistics tends to be undersized. However, its autocorrelation of pairs of individuals standardized perform better. In this chapter the Lagrange Multiplier (LM) test for a constant correlation by Engle & Sheppard (2001) has been considered. Thus the test is assumed to be

$$H_0 : R_t = \hat{R} \tag{5.16}$$

$\forall t \in T$ against

$$H_1 : vech^{\alpha}(R_t) = vech(\hat{R}) + \beta_1 vech(R_{t-1}) + \beta_2 vech(R_{t-1}) + ... + \beta_p vech(R_{t-1}), \tag{5.17}$$

where $\hat{R}$ is a constant correlation and *vech* is the operator that stacks a matrix as column. Assume $\hat{\phi}_t = \hat{R}_t^{\frac{-1}{2}} \hat{D}_t^{-1} \hat{\epsilon}_t$ is a $N \times 1$ vector of standardized residual, such that the LM test can follow the regression below.

$$vech^{\alpha}(\hat{\phi}\hat{\phi}' - I_N) = \hat{\beta}_0 + \hat{\beta}_1 vech^{\alpha}(\hat{\phi}_{t-1}\hat{\phi}'_{t-1} - I_N) + ... + \hat{\beta}_p vech^{\alpha}(\hat{\phi}_{t-p}\hat{\phi}'_{t-p} - I_N) + \hat{\eta}_t, \tag{5.18}$$

where $vech^{\alpha}$ is the modified *vech*, that is choosing only the above element of the diagonal. Thus under the null hypothesis, all the estimated parameters from equation 5.14 should be equal to zero, and the test can be then conducted as

$$\frac{\hat{\delta}x'x\hat{\delta}}{\hat{\sigma}^2}$$

which is asymptotically χ^2_{p+1}, while $\hat{\delta}$ are the estimated regression parameters and x is a matrix containing the regressors (Engle & Sheppard 2001).

5.5 Application for DCC-GARCH model and Results

This section provides the output of the DCC-GARCH (1,1) model applied to the two JSE mining companies' daily returns data and check the diagnostic tests result of the fitted model using the two test which are the Ljung-Box and Lagrange multiplier test.

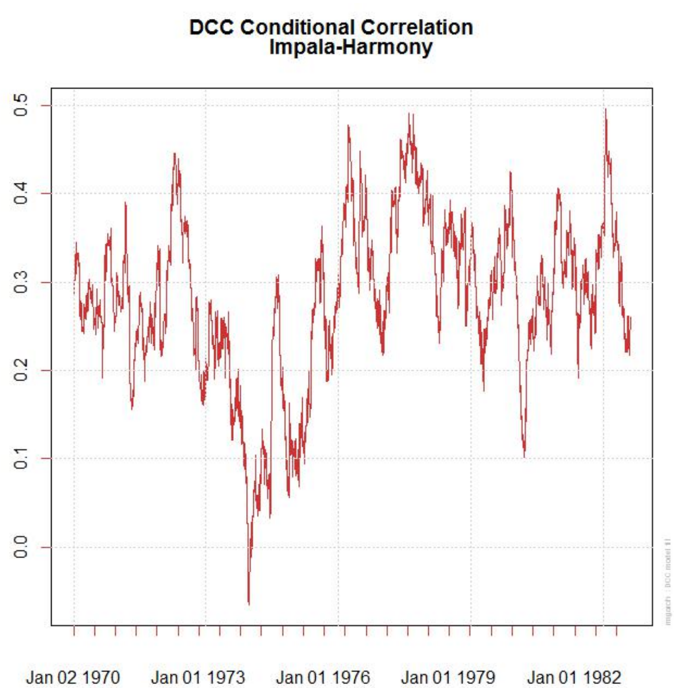
5.5.1 Parameter Estimation for DCC-GARCH (1,1) model

Table 5.1 shows the summary of the parameter estimates for the DCC-GARCH model. The univariate GARCH (1,1) model for each JSE mining company returns describe the diagonal elements of D_t as defined in equation 5.6, while $\hat{\theta}_1$ and $\hat{\theta}_2$ indicate the DCC conditional correlation parameters. The significance of coefficient $\hat{\alpha}_1$ and $\hat{\beta}_1$ for the univariate GARCH model at 5% level, indicates that the conditional volatility in the two JSE mining companies is highly persistent and one company is reacting to different shock received from other company. The sum of $\hat{\alpha}_1$ and $\hat{\beta}_1$ is less than one for both two companies with Harmony Gold Mining Limited having the highest value of 0.9880 compared to the one of Impala 0.9661. The DCC correlation parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ are significant at 5%, which implies that the correlations between the two mining companies are time varying and this shows that there is significant effect between the two companies. Furthermore the values of $\hat{\theta}_1 + \hat{\theta}_2 = 0.9947 < 1$, implies that the correlation matrix P_t is positive definite and the conditional covariance is highly persistent. The same results were found by Ijumba (2013). Figure 5.2 reveals that the estimate correlation has more non-stable pattern, however the moving pattern appears to be centralized around an upward trend correlation. The decline in correlation is between year 1972 to 1973, after which an upward trend is noted.

Table 5.1: Summary of Fitted DCC-GARCH (1,1) for Impala Platinum Holdings Limited and Harmony Gold Limited Mining Companies

Company	Parameter	Estimate	Std Error	P-value
Impala	α_0	-0.0006	0.0003	0.0882
	α_1	0.0966	0.0472	0.0406
	β_1	0.8695	0.0720	0.0000
Harmony	α_0	0.0004	0.0004	0.2423
	α_1	0.0837	0.0197	0.00002
	β_1	0.8953	0.0262	0.0000
Impala & Harmony	θ_1	0.0108	0.0046	0.0188
	θ_2	0.9839	0.0083	0.0000
AIC	BIC		LOG	
-8.8203	-8.8029		21485.9100	

Figure 5.1: Harmony Gold Mining and Impala Platinum Holdings Daily Closing Prices



5.5.2 Diagnostic Checking for the DCC-GARCH (1,1) model

After fitting the model it is always important to explore the adequacy of the standardized residuals. This is done by using the univariate Ljung-Box test on each of the JSE mining company standardized residuals. Moreover, to test the constant correlation the LM test by Engle & Sheppard (2001) is used. Table 5.2 shows that all the p-values are significant since are all greater than 5% level of significance. This means that there is no serial correlation since the null hypothesis of no serial correlation was accepted. Furthermore, the results from Table 5.3 show that the p-values are less than 5% level of significance and this suggests the rejection of null hypothesis of constant correlation. Therefore, the DCC-GARCH model is a good fit and the correlation between the two JSE mining companies develops over the time.

Table 5.2: Ljung-Box Test

Lags	Critical-value	p-value
5	9.5669	0.0885
10	17.8684	0.0572
15	21.6826	0.1164
20	23.0812	0.2848
25	26.9168	0.3601

Table 5.3: Engle's ARCH Test for Heteroscedasticity

Lags	Critical-value	p-value
5	588.3619	$<2.2e^{-16}$
10	657.3726	$< 2.2e^{-16}$
15	685.3642	$< 2.2e^{-16}$
20	703.6231	$< 2.2e^{-16}$

5.6 Chapter Summary

In this Chapter the primary interest was to provide a theory on a multivariate model and its applications to model the volatility in specific JSE-listed mining companies. This procedure of modelling multivariate volatility is an extension of modelling univariate volatility, where the dynamic relation in the volatility of multiple returns series is considered to make decisions related to asset returns. There are different techniques of modelling multivariate volatility, however in this chapter it is limited to the DCC-GARCH model by Engle (2002). The DCC-GARCH model is the extension of the CCC-GARCH model by Bollerslev (1990). In the DCC-GARCH model the conditional variance and correlation are modelled rather than modelling the conditional covariance matrix which is used in the Constant Conditional Correlation (CCC-GARCH) model. The theory and parameter estimations of DCC-GARCH model were discussed together with their diagnostic test. Furthermore the results of the multivariate DCC-GARCH model were presented, where the daily returns series data, which are the same as mining companies from univariate GARCH models, were used and the R software was used to analyze the multivariate DCC-GARCH model. The conditional volatility and correlations of the two mining companies were estimated. Hence the persistence volatility was presented in the two mining companies with Harmony Gold Mining Limited having the highest value of 0.9838. These results of high persistence volatility agreed with the results from the fitted GARCH models described in Chapter 4. The multivariate models have the capacity of capturing simultaneously interaction between one or more time series. The hypothesis of constant correlation between these two mining companies was rejected and the correlation between them seems to vary with time. The two mining companies are positively highly correlated as the values θ_1 and θ_2 of correlation show in Table 5.1. Furthermore the diagnostic test to check the adequacy of the model showed that the DCC model was good enough to estimate the volatility and correlation of the two mining companies.

Chapter 6

Stochastic Volatility (SV) model

6.1 Introduction

The Stochastic Volatility (SV) model is a modern technique that is used to model the time-varying volatility, as in the ARCH and GARCH type models. Although the two models are used to model the time-varying volatility, there is a difference between them. The SV models were introduced by Taylor (1982), and are parameter-driven, while the ARCH family models are observation-driven (Shephard 1996). However, the conditional variance in the SV model is modelled as an unobserved stochastic model, and allows the observations to be the function of some unobserved components (Talke 2003). Stochastic models are difficult to fit to the data, since the log likelihood for these models can not be estimated easily, unlike ARCH family models. Therefore, this renders the SV model not as useful as the ARCH and GARCH models. Thus the use of quasi-likelihood method was applied to estimate the parameters, with the support of the state-space model, together with the Kalman Filter method (Tsay 2005). We are focusing on the quasi-likelihood method for estimating the parameters of a SV model. However, other method can be used such as a Monte Carlo method.

6.2 State-Space Models

Note that the state-space approach to time series analysis was initiated by Kalman (1960) to simplify the maximum likelihood estimation problem and to take care of missing data, due to its flexibility and is based on structural analysis of the problem (Durbin & Koopman 2001, Tsay 2010). The two general equations of the state-space model are given in time t as follows:

$$x_t = \Phi x_{t-1} + \omega_t \tag{6.1}$$

$$y_t = A_t x_{t-1} + \nu_t, \tag{6.2}$$

where the equations 6.1 and 6.2 are state and observation equations, respectively. Thus ω_t and ν_t are assumed to be independent white noise processes, and the observation error and the state error are assumed to be uncorrelated Dralle (2011). From equation 6.1 ω_t is a vector white noise process with covariance matrix Σ_ω, and ν_t from equation 6.2 has variance σ_ν^2. The state equation can be used to generate x_{ti} from the previous states $x_{t-1,j}$, where $j = 1, ..., p$ for $i = 1, ..., p$ and $t = 1, ..., n$. Hence from state equation the observation matrix Φ_t is $p \times p$ transition matrix, whilst ω_t is assumed to be $p \times 1$ independent and identically distributed normal vectors with a mean zero of the zero vector and covariance matrix Q. The observation equation is very important, since we cannot observe the state vector x_t directly. Therefore the observation equation gives a linear transformation of the state vector x_t with additional white noise. Thus the observation matrix A_t is assumed to be $q \times p$ dimension, and the observed vector y_t to be $q \times 1$ dimension, whilst ν_t is assumed to be white noise and Gaussian with $q \times q$ covariance matrix R. The problems associated with the state-space model can be specified in three categories, when we are estimating the state vector x_t data given the $\boldsymbol{Y_s = y_1, ..., y_s}$, where s is a time which can vary compared to t:

- When $s < t$, the problem is forecasting or prediction;
- When $s = t$, the problem is filtering;
- When $s > t$, the problem is smoothing.

This problem can be solved by using the methods of Kalman recursions (Chan 2010). Thus the following definition is considered to derive the Kalman Filter and Kalman Smoothing.

$$x_t^s = E(x - t|Y_s) \tag{6.3}$$

and

$$P^s_{t_1,t_2} = E[(x_{t_1} - x^s_{t_1})(x_{t_2} - x^s_{t_2})'] \tag{6.4}$$

when, $t_1 = t_2$, from equation 6.4, then $P^s_{t_1,t_2}$ becomes P^s_t. The derivative equation of the Kalman Filter and Kalman Smoothing relies on the Gaussian (normality) assumption. This implies that equation 6.4 is the conditional error covariance given by:

$$P^s_{t_1,t_2} = E[(x_{t_1} - x^s_{t_1})(x_{t_2} - x^s_{t_2})'|Y_s] \tag{6.5}$$

This implies that the covariance matrix between $x_t - x_t^s$ and Y_s is zero for any s and t. Hence $x_t - x_t^s$ and Y_s are independent due to the Gaussian assumption, whilst the conditional distribution of $x_t - x_t^s$ given Y_s is the unconditional distribution of $x_t - x_t^s$ (Shumway & Stoffer 2006). Note that when the state-space models are used, the process starts with a vector x_0, which is normal with μ_0 and $p * p$ covariance matrix Σ_0.

6.3 The Kalman Filter

Filtering is a widely used methodology for the incorporation of observed data into time-evolving systems. It provides an on-line approach to state estimation inverse problems when data are acquired sequentially. The Kalman Filter plays a central role in many applications because it is accurate for linear systems subject to Gaussian noise, and because it forms the basis for many approximate filters which are used in high dimensional systems (Lee et al. 2011). The motivation for using the Kalman Filter is to update the state variable recursively as new data become available (Tsay 2010, Einicke 2012). The Kalman Filter is based on the assumption of linearity and Gaussianity and can be used to update the filter from x^{t-1}_{t-1} to x^t_t, when a new y_t

is observed. The Kalman Filter is derived as follows: from the equation 6.3 and using 6.1 we have

$$x_t^{t-1} = E(x_t|Y_{t-1}) = E(\Phi x_{t-1} + w_t|Y_{t-1}) = \Phi x_{t-1}^{t-1} \tag{6.6}$$

and from equation 6.4 we have

$$\begin{aligned} var(x_t|Y_{t-1}) &= P_t^{t-1} \\ &= E[(x_t - x_t^{t-1})(x_t - x_t^{t-1})'] \\ &= E\{[\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t][\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t]'\} \\ &= \Phi P_{t-1}^{t-1}\Phi' + Q. \end{aligned} \tag{6.7}$$

Then we can define the innovations as

$$\epsilon_t = y_t - E(y_t|Y_{t-1}) = y_t - A_t x_t^{t-1}, \tag{6.8}$$

for t=1,2,...,n. Now we can find the expectation and variance of ϵ_t respectively as follows:

$$E(\epsilon_t) = 0 \tag{6.9}$$

and

$$\Sigma_t = var(\epsilon_t) = var[A_t(x_t - x_t^{t-1}) + \nu_t] = A_t P_t^{t-1} A_t' + R. \tag{6.10}$$

We also know that $E(\epsilon_t y_s') = 0$ for $s < t$. This shows that the innovations are independent of past observations since the innovations follow a Gaussian process (Shumway & Stoffer 2006). Hence we can express the covariance between x_t and ϵ_t conditional on Y_{t-1} as follows:

$$\begin{aligned} Cov(x_t, \epsilon_t|Y_{t-1}) &= cov(x_t, y_t - A_t x_t^{t-1}|Y_{t-1}) \\ &= cov(x_t - x_t^{t-1}, y_t - A_t x_t^{t-1}|Y_{t-1}) \\ &= cov[x_t - x_t^{t-1}, A_t(x_t - x_t^{t-1}) + \nu_t] \\ &= P_t^{t-1} A_t'. \end{aligned} \tag{6.11}$$

Thus the joint distribution of x_t and ϵ_t conditional on Y_{t-1} becomes normal and we get the following equation:

$$\begin{pmatrix} x_t \\ \epsilon_t \end{pmatrix} |Y_{t-1} \sim N\left(\begin{bmatrix} x_t^{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} P_t^{t-1} & P_t^{t-1}A_t' \\ A_tP_t^{t-1} & \Sigma_t \end{bmatrix} \right). \tag{6.12}$$

By using Appendix A, we can then write x_t^t as

$$x_t^t = E(x_t|y_1, ..., y_{t-1}, y_t) = Ex_t^t = E(x_t|Y_{t-1}, \epsilon_t) = x_t^{t-1} + K_t\epsilon_t \tag{6.13}$$

where,

$$K_t = P_t^{t-1}A_t'\Sigma_t^{-1} = P_t^{t-1}A_t'(A_tP_t^{t-1}A_t' + R)^{-1}.$$

Afterwards, by using result 1 from Appendix A, P_t^t can be calculated as

$$P_t^t = cov(x_t|Y_{t-1}, \epsilon_t) = P_t^{t-1} - P_t^{t-1}A_t'\Sigma_t^{-1}A_tP_t^{t-1} \tag{6.14}$$

(Shumway & Stoffer 2006).

By using the initial conditions $x_0^0 = \mu$ and $P_0^0 = \Sigma_0$ for $t = 1, ..., n$ for the state-space model given by equation 6.1 and 6.2, the Kalman Filtering algorithm can be written as

$$x_t = \Phi x_{t-1}^{t-1}, \tag{6.15}$$

$$P_t^{t-1} = \Phi x_{t-1}^{t-1}\Phi' + Q, \tag{6.16}$$

with

$$x_t^t = x_t^{t-1} + K_t(y_t - A_tx_t^{t-1}), \tag{6.17}$$

$$P_t^t = (I - K_tA_t)P_t^{t-1}, \tag{6.18}$$

where

$$K_t = P_t^{t-1} A_t' (A_t P_t^{t-1} A_t' + R)^{-1}, \tag{6.19}$$

K_t is known as Kalman. The equations 6.18 and 6.19 are used for prediction when $t > n$ with x_n^n and P_n^n as initial conditions (Shumway & Stoffer 2000).

6.4 The Kalman Smoother

The Kalman Smoother algorithm is another which generates estimates of the unobserved state-space based on the whole sample. The Kalman Smoother is also based on the assumption of linearity and Gaussianity and is performed recursively backward in time depending on the forward-backward smoothing formula. The smoother may be employed for off-line state estimation and this requires estimates of the state-space model parameters (Kim 2005, Einicke 2012). The aim of smoothing is to estimate the state variable x_t supported by all the available information. This means the estimation of the state is based on the sample $y_1, ..., y_n$, where $t < n$ (Tsay 2010).

Thus the Kalman Smoother is derived by using the following definition:

$$Y_{t-1} = (y_1, ..., y_{t-1}) \tag{6.20}$$

and

$$\eta_t = (\nu_t, ..., \nu_n, w_{t+1}, ..., w_n) \tag{6.21}$$

with Y_0 as an empty set, then we can consider

$$q_{t-1} = E(x_{t-1} | Y_{t-1}, x_t - x_t^{t-1}, \eta - t) \tag{6.22}$$

for $1 \leq t \leq n$.

Knowing that $Y_{t-1}, (x_t - x_t^{t-1})$ and $\eta - t$ are mutually independent, and x_{t-1} and η_t are independent, then Result 1 in Appendix A can be used to create the following equation:

$$q_{t-1}x_{t-1}^{t-1} + J_{t-1}(x_t - x_t^{t-1}), \tag{6.23}$$

with

$$J_{t-1} = cov(x_{t-1}, x_t - x_t^{t-1})[P_t^{t-1}]^{-1} = P_{t-1}^{t-1}\Phi'[P_t^{t-1}]^{-1}, \tag{6.24}$$

then we can have

$$x_{t-1}^n = E[x_{t-1}|Y_n] = E[q_{t-1}|Y_n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1})], \tag{6.25}$$

since $Y_{t-1}, x_t - x_t^{t-1}$ and η_t give $Y_n = (y_1, ..., y_n)$.

Therefore, the error covariance, P_{t-1}^n, is obtained from equation 6.25 which gives us the following equations:

$$x_{t-1} - x_{t-1}^n = x_{t-1} - x_{t-1}^{t-1} - J_{t-1}(x_t^n - \Phi x_{t-1}^{t-1}), \tag{6.26}$$

$$x_{t-1} - x_{t-1}^n = x_{t-1} - xt - 1^{t-1} - J_{t-1}x_t^n + J_{t-1}\Phi x_{t-1}^{t-1}, \tag{6.27}$$

$$x_{t-1} - x_{t-1}^n + J_{t-1}x_t^n = x_{t-1} - x_{t-1}^{t-1} + J_{t-1}\Phi x_{t-1}^{t-1}. \tag{6.28}$$

Hence the following equation is obtained by multiplying both sides of equation 6.28 by the transpose of itself and then expect to have the following equation:

$$P_{t-1}^n + J_{t-1}E(x_t^n x_t^n)J_{t-1} = P_{t-1}^{t-1} + Jt - 1\Phi E(x_{t-1}^{t-1}x_{t-1}t - 1)\Phi' J_{t-1}, \tag{6.29}$$

since the cross-product terms are zero, we then have

$$E[x_t^n(x_t^n)^{'}] = E[x_t x_t^{'}] - P_t^n = \Phi E[x_{t-1}x_{t-1}^{'}\Phi^{'} + Q - p_t^n, \tag{6.30}$$

and

$$E[x_{t-1}^{t-1} x_{t-1}^{t-1'}] = E[x_{t-1} x_{t-1}^{'}] - P_{t-1}^{t-1}. \tag{6.31}$$

Thus the Kalman Smoother for the state-space model from the equations 6.1 and 6.2 with the initial conditions x_n^n and P_n^n, also available in the Kalman Filter is

$$x_{t-1}^n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}), \tag{6.32}$$

and

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}^{'}, \tag{6.33}$$

where

$$J_{t-1} = P_{t-1}^{t-1} \Phi^{'} (P_t^{t-1})^{-1} \tag{6.34}$$

(Shumway & Stoffer 2006).

6.5 The Lag One Covariance Smoother

The lag one covariance smoother is very important in a SV model for employing both the Kalman Smoother and Kalman Filtering the output. Furthermore, the lag one covariance smoother can be used to recursively obtain $P_{t,t-1}^n$, which is defined by equation 6.4. Hence the lag one covariance smoother can be derived by defining first:

$$\hat{x}_t^s = x_t - x_t^s, \tag{6.35}$$

Consequently, the equation 6.17 and 6.19 can be used to obtain:

$$P_{t,t-1}^t = E[\hat{x}_t^t \hat{x}_{t-1}^{t'}], \tag{6.36}$$

$$P_{t,t-1}^{t} = E\{[\hat{x}_t^{t-1} - K_t(y_t - A_t x_t^{t-1})][\hat{x}_{t-1}^{t-1} - J_{t-1}K_t(y_t - A_t x_t^{t-1})]'\}, \tag{6.37}$$

$$P_{t,t-1}^{t} = E\{[\hat{x}_t^{t-1} - K_t(A_t \hat{x}_t^{t-1} + \nu_t)][\hat{x}_{t-1}^{t-1} - J_{t-1}K_t(A_t \hat{x}_t^{t-1} + \nu_t)]'\}. \tag{6.38}$$

Thus use of expectation, after expanding equation 6.38 gives us:

$$P_{t,t-1}^{t} = P_{t,t-1}^{t-1} - P_t^{t-1}A_t'K_t'J_{t-1}' - K_tA_tP_{t-1}^{t-1} + K_t(A_tP_t^{t-1}A_t^{'} + R)K_t^{'}J_{t-1}^{'}, \tag{6.39}$$

we also have that

$$K_t(A_tP_t^{t-1}A_t^{'} + R) = P_t^{t-1}A_t^{'}, \tag{6.40}$$

and

$$P_{t,t-1}^{t-1} = \Phi P_{t-1}^{t-1}, \tag{6.41}$$

for any $t = 1, ..., n$. Thus the equation 6.32 can be used to obtain:

$$\hat{x}_{t-1}^{n} + J_{t-1}x_t^n = \hat{x}_{t-1}^{t-1} + J_{t-1}\Phi x_{t-1}^{t-1}, \tag{6.42}$$

$$\hat{x}_{t-2}^{n} + J_{t-2}x_{t-1}^n = \hat{x}_{t-2}^{t-2} + J_{t-2}\Phi x_{t-2}^{t-2}. \tag{6.43}$$

Therefore, we can multiply the left side of the equation 6.42 by the transpose of the left hand side of equation 6.43 and multiply the right hand side of equation 6.42 by the transpose of the right hand side of equation 6.43. Hence we equate the two results and use the expectation. Thus, the left hand side becomes

$$P_{t-1,t-2}^{n} + J_{t-1}E[x_t^n x_{t-1}^{n'}]J_{t-2}^{'}, \tag{6.44}$$

while the right hand side is

$$P_{t-1,t-2}^{t-2} - K_{t-1}A_{t-1}P_{t-1,t-2}^{t-2} + J_{t-1}\Phi K_{t-1}A_{t-1}P_{t-1,t-2}^{t-2} + J_{t-1}\Phi E[x_{t-1}^{t-1}x_{t-2}^{t-2}\Phi' J_{t-1}'] \tag{6.45}$$

(Dralle 2011).

By taking the expectation of $x_t^n x_{t-1}^{n'}$ and $x_{t-1}^{t-1} x_{t-2}^{t-2'}$ respectively, we get the following equations:

$$E[x_t^n x_{t-1}^{n'}] = E[x_t x_{t-1}'] - P_{t,t-1}^n = \phi E[x_{t-1} x_{t-2}'] \phi^{'} + \Phi Q - P_{t,t-1}^n, \tag{6.46}$$

$$E[x_{t-1}^{t-1} x_{t-2}^{t-2'}] = E[x_{t-1}^{t-2} x_{t-2}^{t-2'}] = E[x_{t-1} x_{t-2}'] - P_{t-1,t-2}^{t-2}. \tag{6.47}$$

The state-space model defined by equations 6.1 and 6.2, with K_t and J_t for $t = 1, ..., n$ and P_n^n is accessible from the Kalman Filter and Kalman Smoother. Thus the use of the initial condition

$$P_{n,n-1}^n = (I - K_n A_n) \Phi P_{n-1}^{n-1}. \tag{6.48}$$

For t=n, n-1, n-2,.., 2 from equation 6.44 and 6.45, the lag one covariance smoother is written as

$$P_{t-1,t-2}^n = P_{t-1}^{t-1} J_{t-2}^{'} + J_{t-1}(P_{t,t-1}^n) - \Phi P_{t-1}^{t-1} J_{t-2}^{'} \tag{6.49}$$

(Shumway & Stoffer 2000).

6.6 Maximum Likelihood Estimation

The parameters of the state-space model, specified in equation 6.1 and 6.2, which are the initial mean (μ_0), covariance (Σ_0), the transition matrix (Φ), the state and observation covariance matrix Q and R respectively, must be estimated in favour of the Kalman Filter and Smoothing equations. Thus the maximum likelihood under the assumption of $x_0 \sim N(\mu_0, \Sigma_0)$, and the errors $\omega_1, ..., \omega_n$ and $\nu_1, ..., \nu_n$ respectively; are jointly normal and uncorrelated. The innovation defined in equation 6.8 as $\epsilon_t = y_t - A_t x_t^{t-1}$ can be used to estimate the likelihood. Despite that result of the equation 6.8, the innovation of the likelihood function can be obtained, and note that the innovations are independent normal random vectors with mean zero and covariance matrix defined in equation 6.10 as

$$\Sigma_t = A_t P_t^{t-1} A_t' + R.$$

Hence the log-likelihood can be expressed as

$$\ln L_Y(\theta) = -\frac{1}{2}\sum_{t=1}^{n} \log|\Sigma_t(\theta)| - \frac{1}{2}\sum_{t=1}^{n} \epsilon_t(\theta)' \Sigma_t(\theta)^{-1} \epsilon_t(\theta). \tag{6.50}$$

Note that for simplicity the constant has been ignored, and $\theta = (\mu_0, \Sigma_0, \phi, Q, R)$.

Moreover, the maximization of the log-likelihood is done by fixing the initial x_0, and then improving a set of recursions for the log-likelihood function and its first two derivatives. Thus the use of the Newton-Raphson procedure applied to update the parameter values until the maximization of the log-likelihood has been reached. Hence we can summarize the Newton-Raphson procedure into the following steps:

1. Select initial values for the parameters, means θ_0.

2. Run the Kalman Filter using the initial values, θ_0 to obtain a set of innovations and error covariance.

3. Run iteration of the Newton-Raphson procedure to obtain new estimates of the parameters, means θ_1.

4. By using the parameter estimates obtained from step 3, repeat step 2 to generate a new set of innovations and error covariances. Next run step 3 again; and this process will continue until the difference between successive estimates of parameters or log-likelihood becomes small enough (Shumway & Stoffer 2006).

6.7 The Expectation Maximization Algorithm (EM)

The EM algorithm is well known in statistical analysis and is used for performing maximum likelihood estimates of parameters in statistical models for any incomplete data or complete data that come from the exponential family. The EM algorithms can also be used for solving joint state and parameter estimation problems (Lee et al. 2011). The EM algorithm was explained and given its name by Dempster et al. (1977) in their classic paper, where they first demonstrated the results about the algorithm and its properties. However, they mentioned that the methods had been suggested many times by other authors at earlier points in time, such as (Orchard et al. 1972, Martin-Löf 1973, Sundberg 1974). Furthermore the EM algorithm consists two steps, which are E-step (expectation) for computing the expectation of the log-likelihood evaluated using the current estimate for the parameters and M-step (maximization), to compute parameters and maximizing the expected log-likelihood found on the E-step. Hence, these parameter estimates are then used to determine the distribution of the latent variables in the next E-step (Dempster et al. 1977, Durbin & Koopman 2001).

The idea of using the EM algorithm in a SV model is that by observing the states $X_n = (x_0, x_1, ..., x_n)$ and the observations $Y_t = (y_1, ..., y_n)$, then we can take (X_n, Y_n) to be the completed data with joint density

$$f_\theta(X_n, Y_n) = f_{\mu_0}, \Sigma_0(x_0) \prod_{t=1}^{n} f_{\Phi,Q}(x_t|x_{t-1}) \prod_{t=1}^{n} f_R(y_t|x_t). \tag{6.51}$$

Thus the log-likelihood for the complete data under the assumption of normality is expressed as

$$\begin{aligned} -2\ln L_{X,Y}(\theta) = {} & \ln|\Sigma_0| + (x_0 - \mu_0)^{'} \Sigma_0^{-1} (x_0 - \mu_0) + \ln|Q| \\ & + \sum_{t=1}^{n} (x_t - \Phi x_{t-1})^{'} Q^{-1} (x_t - \Phi x_{t-1}) + \ln|R| \\ & + \sum_{t=1}^{n} (y_t - A_t x_t)^{'} R^{-1} (y_t - A_t x_t) \end{aligned} \tag{6.52}$$

(Shumway & Stoffer 2006).

Thus the maximum likelihood estimates of θ are obtained by using the EM algorithm based on the incomplete data, given by Y_n. However, this is achieved by maximizing the conditional expectation of the complete data likelihood. The iteration j for $j = 1, 2, ...$, where the conditional expectation equation to be maximized is

$$Q(\theta|\theta^{(j-1)}) = E[-2\ln L_{X,Y}(\theta)|Y_n, \theta^{(j-1)})]. \tag{6.53}$$

Hence the Kalman Smoother can be used to obtain the conditional expectations, given the value of the parameters, $\theta^{(j-1)}$ and this induces the following equation

$$\begin{aligned} Q(\theta|\theta^{(j-1)}) = &\ \ln|\Sigma - 0| + tr\Sigma_0^{-1}[P_0^n + (x_0^n - \mu_0)(x_0^n - \mu_0)^{'}] + \ln|Q| \\ &+ trQ^{-1}(S_{11} - S_{10}\Phi^{'} - \Phi S_{10}^{'} + \Phi S_{00}\Phi^{'}) + \ln|R| \\ &+ trR^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)^{'} + A_t P_t^n A_t^{'}], \end{aligned} \tag{6.54}$$

where

$$S_{11} = \sum_{t=1}^{n}(x_t^n x_t^{n'} + P_t^n), \tag{6.55}$$

$$S_{10} = \sum_{t=1}^{n}(x_t^n x_{t-1}^{n'} + P_{t,t-1}^n), \tag{6.56}$$

and

$$S_{00} = \sum_{t=1}^{n}(x_{t-1}^n x_{t-1}^{n'} + P_{t-1}^n). \tag{6.57}$$

From equations 6.54, 6.55, 6.56 and 6.57 respectively, the current values of parameter $\theta^{(j-1)}$ are used for calculations in the smoother equations. In addition, by maximizing equation 6.54 with respect to the parameters at j^{th} iterations, which lead to the update estimates

$$\Phi^{(j)} = S_{10}S_{00}^{-1}, \tag{6.58}$$

$$Q^{(j)} = n^{-1}(S_{11} - S_{10}S_{00}^{-1}S_{10}^{'}, \tag{6.59}$$

and

$$R^{(j)} = n^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)^{'} + A_t P_t^n A_t^{'}. \tag{6.60}$$

Note that the initial means and covariance cannot be estimated simultaneously. The usual convention is to fix both the mean and covariance, or just the covariance matrix, and then use

$$\mu_0^{(j)} = x_0^n, \tag{6.61}$$

and this is the estimator which is obtained from minimizing equation 6.54 under the assumption that the covariance matrix has been fixed (Dralle 2011). Thus the steps involved in the EM algorithm are summarized as follows:

1. Select the starting values for the parameters $\theta^{(0)} = \mu_0, \Phi, Q, R$, and fix Σ_0.
2. Compute the likelihood of the incomplete data as in equation 6.50.
3. Perform the E-step of the algorithm using the Kalman Filter and Kalman Smoothing by using the parameter $\theta^{(j-1)}$ to calculate S_{11}, S_{10}, S_{00} from equations 6.55, 6.56 and 6.57 respectively.
4. Perform the E-step to update the estimates, μ_0, Φ, Q and R, by using equations 6.58 to 6.61 to get $\theta^{(j)}$.
5. Repeat step 2 and 4 until convergence has been achieved

(Shumway & Stoffer 2006).

6.8 The Stochastic Volatility Model (SV)

The Stochastic Volatility models are not different from the ARCH type models, however the SV model adds a stochastic noise term to the equation for σ_t. Recalling from Chapter 2 the equation of GARCH (1,1) model was given by

$$\epsilon_t = \sigma_t y_t, \tag{6.62}$$

$$\Sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{6.63}$$

where y_t is Gaussian white noise. By defining

$$h_t = \ln \sigma_t^2, \tag{6.64}$$

and

$$z_t = \ln \epsilon_t, \tag{6.65}$$

we then deduce from equation 6.63 the following equation:

$$z_t = h_t + \ln y_t^2. \tag{6.66}$$

Thus the equation 6.66 taken as the observation equation, and h_t known as the stochastic variance is assumed to be an unobserved state process. Then from the equation 6.66 the volatility process follows an autoregressive (AR1) process such that h_t can be expressed as

$$h_t = \phi_0 + \phi_1 h_{t-1} + \omega_t, \tag{6.67}$$

where ω_t is the white Gaussian noise with variance σ_w^2. Thus the SV model is made up of equations 6.66 and 6.67 respectively. Hence the ARCH normality assumption on y_t is kept to fit the SV model, since $\ln y_t^2$ is distributed as the log of a chi-squared random variable with one degree of freedom. Thus the probability density function of $\ln y_t^2$ is given by

$$f(\ln y_t^2) = \frac{1}{\sqrt{2\Pi}} \exp\left[\frac{-1}{2}\left(e^{\ln y_t^2} - \ln y_t^2\right)\right], \tag{6.68}$$

for $-\infty < \ln y_t^2 < \infty$, where the mean for the $\ln y_t^2$ is -1.27 and variance $\frac{\Pi^2}{2}$. The stochastic volatility model can be fitted by writing the observation equation 6.65 as

$$z_t = \alpha + h_t + y_t, \tag{6.69}$$

where y_t is the white noise, with mixture distribution of two normals, where one is centered at zero (Shumway & Stoffer 2000). We then write

$$y_t = u_t z_{t0} + (1 - u_t) z_{t1}, \tag{6.70}$$

where u_t is independent and identically distributed Bernoulli process with $Pr(u_t = 0) = \Pi_0$, and $Pr(u_t = 1) = \Pi_1$, with the condition of $\Pi_0 + \Pi_1 = 1$. Then z_{t0} is independent and identically distributed with mean zero and variance σ_0^2, and z_{t1} is an independent and identically distributed with mean μ_1 and variance σ_1^2. The SV model is fitted with the use of the Kalman Filter, which requires a slight modification, as stated by Shumway & Stoffer (2006) in the following equations:

$$h_{t+1}^t = \phi_0 + \phi_1 h_t^{t-1} + \sum_{j=0}^{1} \pi_{tj} K_{tj} \epsilon_{tj}, \tag{6.71}$$

$$P_{t+1}^t = \phi_1^2 P_t^{t-1} + \sigma_w^2 - \sum_{j=0}^{1} \pi_{tj} K_{tj}^2 \Sigma_{tj}, \tag{6.72}$$

$$\epsilon_{t0} = z_t - \alpha - h_t^{t-1}, \tag{6.73}$$

$$\epsilon_{t1} = z_t - \alpha - h_t^{t-1} - \mu_1, \tag{6.74}$$

$$\Sigma_{t0} = P_t^{t-1} + \sigma_0^2, \tag{6.75}$$

$$\Sigma_{t1} = P_t^{t-1} + \sigma_1^2, \tag{6.76}$$

$$K_{t0} = \frac{\phi_1 P_t^{t-1}}{\Sigma_{t0}}, \tag{6.77}$$

$$K_{t1} = \frac{\phi_1 P_t^{t-1}}{\Sigma_{t1}}, \tag{6.78}$$

Note that the filtering equations 6.71 and 6.78 for the model are given by the equations 6.67 and 6.69 respectively. Thus the probabilities $\Pi_{t1} = Pr(u_t = 1|z_1, ..., z_t)$, for $t = 1, ..., n$ need to be determined to complete the filtering equation. Since $\Pi_{t0} + \Pi_{t1} = 1$, then we can find Π_{t0}, which is $\Pi_{t0} = 1 - \Pi_{t1}$. This can be calculated once the Π_{t1} has been achieved, and to get this we let $f_j(t|t-1)$ to be the conditional density of z_t given that $z_1, ..., z_t$, and $u_t = j$ for $j = 0$ or 1. Hence the Bayes rule can be applied and we have the following equation:

$$\Pi_{t1} = \frac{\Pi_1 f_1(t|t-1)}{\Pi_0 f_0(t|t-1) + \Pi_1 f_1(t|t-1)}. \tag{6.79}$$

$\Pi_1 = \frac{1}{2}$,is sufficient, if there is no reason to prefer one state. However, the exact value of $f_j(t|t-1)$ is challenging to obtain (Shumway & Stoffer 2006). Thus the approximate of $f_j(t|t-1)$ is obtained by employing the normal distribution with mean $h_t^{t-1} + \mu_j$ and variance Σ_{tj} for $j = 0, 1$ and $\mu_0 = 0$. Then the estimated parameters for the model are given by maximum likelihood (MLE) with an application of log likelihood as

$$\ln L_Z(\theta) = \sum_{t=0}^{n} \ln \left(\sum_{j=0}^{1} \Pi_j f_j(t|t-1) \right), \tag{6.80}$$

where $f_j(t|t-1)$ is the normal density approximated as $N\left(h_t^{t-1} + \mu_j, \sigma_j^2\right)$. Hence the equation 6.80 can be maximized as a function of the parameters θ by employing the Newton-Raphson method, or an EM algorithm, when the completed data likelihood is considered (Shumway & Stoffer 2006).

6.9 Application of Stochastic Volatility (SV) Models

6.9.1 Introduction

Recall from Chapter 6 that the SV model by Taylor (1986) is defined as

$$z_t = h_t + \ln y_t^2, \tag{6.81}$$

$$h_t = \phi_0 + \phi_1 h_{t-1} + \omega_t, \tag{6.82}$$

where ϕ_0, ϕ_1 are the parameters to be estimated, ω_t is the white Gaussian noise with variance σ_w^2, while the stochastic volatility h_t is the state variable and is modelled as an AR (1) process. The only condition for the SV model is that the parameter estimate $\phi_1 < 1$ should be less than one for the model to be stationary (Taylor 1986).

This chapter provides the results of the SV model, where this model employs the logarithm of the squared residual from an ARMA model rather than residuals themselves. However, this can cause a problem if one of the residuals is zero, and this can be corrected by adding a positive constant to the residuals to confirm that there are no zero values. The modification should lead us in the form of $(\epsilon_t + k)$, where k is a small positive constant. Hence the SV model can be applied to the data using the logarithm of the squared modified residuals which is given by

$$\ln((\epsilon_t + k)^2).$$

The R software has been used for the data set same as the R software used for the GARCH models.

6.9.2 The Stochastic Volatility Model for the Impala Platinum Holdings Limited data

In this section we used R software to analyze the data, where we first used the logarithm of the squared residuals from the ARMA model and then we checked if there were no residuals which zero. This can cause a huge problem since one cannot get the logarithm of zero. However, to deal with this problem the small positive constant k needs to be added, then the transformed logarithm becomes $\ln((\varepsilon_t + k)^2$. In our case we did not have this problem of zero values for the residuals, we then had to fit the data set straight without any transformation. In addition, the estimation of parameters procedure used was a Newton-Raphson method to maximize the likelihood equation 6.80. This procedure invokes the state-space methodology, where the likelihood method is obtained from the combination of the EM algorithm with Kalman Filter and smoother algorithms from the state space approach. Moreover the SV model was applied to the residuals of the AR (2) model for the return series as was done in Chapter 4. Table 7.1 shows the parameter estimates for the SV model, where higher values of $\phi_1 = 0.9697$ indicate the high persistence volatility in the Impala Platinum Holdings Limited company data. This high persistence volatility in the Impala Platinum Holdings Limited company data agreed with the results from the GARCH (1,2) model discussed in Chapter 4. Then the positive value 0.2241 of skewness implies that the distribution has a long right tail and deviation from normality, while the excess value 8.4985 of kurtosis implies that the empirical distribution is not normal but peaked.

Table 6.1: Parameter Estimates for The Impala Platinum Holdings Limited SV Model

Parameters	Estimate	Standard Error
ϕ_0	0.9834	0.0695
ϕ_1	0.9697	0.0687
σ_w	0.1928	0.0728
α	-7.2070	0.0706
σ_0	1.0350	0.0215
μ_1	-2.3608	0.1308
σ_1	4.0316	0.0316
$Skewness$	0.2241	-
$Kurtosis$	8.4985	-

6.9.3 The Stochastic Volatility Model for the Harmony Gold Mining Company data

The software to analyze the data was R, and the estimation of parameters procedure used was a Newton-Raphson method to maximize the likelihood equation 6.80. This procedure invokes the state-space methodology, where the likelihood method is obtained from the combination of the EM algorithm with Kalman Filter and smoother algorithms from the state-space approach. Hence the SV model was applied to the residuals of the AR (1) model for the return series as seen in Chapter 4. Table 7.2 shows the parameter estimates for the SV model, where higher values of $\phi_1 = 0.9769$ indicate the high persistence volatility in the harmony Gold mining data company. This high persistence volatility in Harmony Gold Mining company data agreed with the results from the GARCH (1,2) model discussed in Chapter 4. Then the negative value -0.3060 of skewness implies that the distribution has a long left tail and deviation from normality, while the excess value 6.7758 of kurtosis implies that the empirical distribution is not normal but peaked.

Table 6.2: Parameter Estimates for The Harmony Gold Mining Company SV Model

Parameters	Estimate	Standard Error
ϕ_0	0.9977	0.0731
ϕ_1	0.9769	0.0704
σ_w	0.1871	0.0883
α	-7.6075	0.0883
σ_0	1.1483	0.0398
μ_1	-3.4931	0.1722
σ_1	4.4561	0.1092
$Skewness$	-0.3060	-
$Kurtosis$	6.7758	-

6.10 Chapter Summary

In this chapter the primary interest was to provide a discussion on the theory of SV model and its applications. This model is used to model the time-varying as GARCH-type models, and the only difference is that the SV models are parameter-driven while GARCH-type models are observation-driven. In the SV model, the conditional variance is modelled as an unobserved stochastic model, and this allows the observation to be the function of some unobserved stochastic volatility model. Moreover, the SV models are very hard to fit to the data, and the use of likelihood method is not easy to estimate. Thus, the quasi-likelihood method was applied to estimate the parameters with the support of the State-space model together the Kalman Filter method. The theory and use of state-space were discussed, where the state-space is used to overcome the problem of maximum likelihood estimates and handling missing data. There are problems associated with state-space, which were categorized in three ways, according to the variation between two time intervals(s and t). It was discussed that when $s < t$, $s = t$ and $s > t$ the problem is forecast or prediction, filtering and smoothing respectively. The theory of EM algorithm, which is an alternative method to estimate the parameters for state-space models was discussed.

Moreover, the applications for the time series data, were the daily returns from JSE for the two mining companies, namely Harmony Gold Mining Limited and Impala Platinum Holdings Limited. The model employs the logarithm of the squared residual from an ARMA model and there was no challenge of having a zero for the residuals. Hence the model was fitted straight and it was clearly shown that the empirical distribution was not normal, but peaked for both mining companies as shown in Table 7.1 and 7.2 respectively. Furthermore the high value of ϕ_1 indicates that there is a high persistence volatility in both mining companies' data. There is a higher persistence volatility in Harmony Gold Mining Limited compared to Impala Platinum Holdings Limited since the former is the one with a high value of ϕ_1. These results of high persistence volatility agreed with the results from the fitted GARCH models described in Chapter 4 and the DCC-GARCH model in Chapter 5. The model was fitted using R software and the mean equations for the data sets used were the same as those used in GARCH models.

Chapter 7

Discussion and Conclusion

The primary objective of this study was to ascertain the best GARCH models and to apply them to the volatility of the two JSE mining companies. The volatility is considered as a risk, where it is measured by the standard deviation of the returns, and it can also be used as a measure of uncertainty of the returns. Moreover, the higher the volatility the more variable and uncertain the return can be. This high level of volatility in the mining sector is associated with high levels of risks for investors interested in this sector. Thus investors and portfolio managers have always been ambitious in learning about the market's present performance and possible future performance to make decisions regarding investing in a particular company. Therefore, this study will help the investors to know when and which company they should invest in, since they are interested in maximizing the returns on their investment.

The data sets used for the two mining companies, namely Harmony Gold Mining Limited and Impala Platinum Holdings Limited, were the daily closing prices from 3 January 1995 to 3 July 2014. This yields a total of 4870 time series observations, which yielded 4869 log returns because one observation was lost due to differencing in daily closing price series. The analysis involved the use of univariate GARCH, SV and multivariate GARCH models with the application of R software throughout this thesis. However, the study was limited to the Autoregressive Conditional Heteroscedasticity (ARCH) model by Engle (1982), Generalized ARCH (GARCH) by Bollerslev (1986), Exponential GARCH (EGARCH) model by Nelson (1991), Asymmetric Power ARCH (APARCH) model by Ding et al. (1993), the GJR-GARCH model by Glosten et al. (1993) and the Dynamic Conditional Correlation (DCC) GARCH model by Engle (2002). The univariate GARCH models were applied to different distributions, namely Gaussian, Student-t, Skew Student-t and Generalized Error distributions. In both companies the Gaussian error distributions did not conform to the return data series, and the model with three remaining distribution errors was considered.

The GARCH model was fitted with a combination of Autoregressive (AR) processes also used by Ruppert (2011). The AR1+GARCH (1,2) model with Skewed Student-t distribution (sstd) error was found to be the best fit model to the Harmony Gold Mining Limited data set. This was deemed so, since it was the only one to remove the ARCH effect and the autocorrelation which was presented in the mean. Furthermore, the parameter estimates $\hat{\alpha}_1 > 0$ and $\hat{\beta}_1 > 0$ are statistically significant and their sum is less than one. This implies that the unconditional volatility for Harmony Gold Mining Limited company return series is finite. The asymmetric models, the AR1+EGARCH model with Student-t distribution (std) was found to be the best fit model compared to others, however the ARCH effect was not removed from the model. Therefore the GJR-GARCH model was considered since it was the only one to remove the ARCH effect and autocorrelation was presented in the mean.

The results were consistent with Mzamane (2013) in his study of GARCH modelling of volatility in the JSE index. The leverage parameter $\hat{\gamma} = -0.0247$ is a negative and was expected to be positive, however $\alpha_1 + \gamma \geq 0$, the model is still valid. Furthermore, it is seen that a positive shock (good news) generates less volatility than negative shock (bad news) in the Harmony Gold Mining Limited company returns series (Tsay 2010). In the case of Impala Platinum Holdings Limited the AR2+GARCH (1,2) model with Student-t distribution error was found to be the best fit model to the data, since it uniquely removed the ARCH effect and the autocorrelation was presented in the mean, and all coefficients of the estimated parameters were significant. The parameter estimates $\hat{\alpha}_1 > 0$ and $\hat{\beta}_1 > 0$ are statistically significant in both symmetric and asymmetric GARCH models and their sum is less than one. This implies that the unconditional volatility for Impala Platinum Holdings Limited company return series is finite. The asymmetric models, the AR2+EGARCH model with Student-t distribution was found to be the best fit model to the data compared to others. Similar results were found by Hajizadeh et al. (2012) in his study of a hybrid modelling approach for forecasting the volatility of S&P 500 index return. The leverage parameter $\hat{\gamma} = 0.2510$ is positive which was expected. Furthermore, it is seen that a negative shock (bad) generates less volatility than a positive shock (good news) in the Impala Platinum Holdings Limited returns series (Tsay 2010). The higher volatility persistence was found in the Harmony Holdings Limited mining company compared to the Impala Platinum Holdings Limited mining company with the same distribution (Gaussian distribution). This is proven by the high value of $\alpha_1 + \beta_1 + \beta_2 = 0.9716$ compared to $\alpha_1 + \beta_1 + \beta_2 = 0.9656$ for the Impala Platinum Holdings Limited mining company.

In Chapter 5, the application of the DCC-GARCH model was also based on the daily returns from the univariate models. The persistent volatility was presented in both mining companies where the sum of $\hat{\alpha}_1$ and $\hat{\beta}_1$ is less than one, with the highest value of (0.9880) in the Harmony Gold Mining Limited company compared to that of Impala Platinum Holdings Limited (0.9661). This is in agreement with the results from the univariate GARCH models, where the high volatility was found in the Harmony Gold Mining Limited Company. This is due to the capacity of multivariate models to capture the simultaneous interaction between one or more time series. The correlation parameters of the DCC-GARCH model $\hat{\theta}_1$ and $\hat{\theta}_2$ are significant at 5%, which implies that the correlations between the two mining companies is time varying. Furthermore, the diagnostic test to check the adequacy of the model showed that the DCC model was good enough to estimate the volatility and correlation of the two mining companies.

In Chapter 6, the data sets used were the same as those from the univariate GARCH models where the calculation of returns was done first, then the model was fit for the mean. Hence the last step was to fit the SV model to the residuals from the mean equation. This was done in order to avoid the logarithm of the observation to converge to negative infinity, when there are residuals with zero values. However, in both mining companies' returns data, the study did not encounter this problem, the model was fitted without any transformation to the residuals. The empirical distribution was not normal, but peaked for both the mining companies. Furthermore, the high value of ϕ_1 indicates that there is a high persistence volatility in both mining companies' data. These results were in agreement with the ones from Univariate GARCH models and the DCC GARCH model. The R software did not give us the results of p-values and the diagnostic test. The results were consistent with Dralle (2011) in his study of modelling volatility in financial time series. The SV models are less commonly used to compare to the Univariate GARCH models, due to their complexity requirement in fitting higher order models and there being limited software to fit the parameters. The overall finding indicated the higher volatility persistence in both two mining companies with higher volatility in the Harmony Holdings Limited mining company. Finally the result revealed the correlation between the two mining companies to be varied over time.

Although this study achieved its objectives by analyzing volatility, in selected mining companies, and it has some limitations that may be addressed by future research. The first one was the number of companies used for the analysis and was related to the analysis of the two mining company sectors. Although the conclusion was reached with the analysis of two companies, it may be beneficial for future studies to use more companies in order to generate more generalizable results. Furthermore, future studies can compare the volatility of mining sectors to other sectors of the JSE. The second limitation can be the sample size of each company, where the study used a sample size of 4870 observations for each company. It could be better if the sample size could be increased to 8000 or 10000 observations and comparisons made. Moreover, it is known that the greater the sample size, the more accurate the estimations are. The last limitation was the testing of other methods that may account for volatility and this may relate to the use of measures of volatility other than GARCH models. Although the GARCH models have advantages have advantages in measuring volatility, as indicated in the literature review, it may be beneficial to compare it with another model of volatility. Hence, future research can identify whether there are other models that account for the volatility in the mining sector.

Future research could account for different methods to estimate the parameters of a SV model and modelling the volatility of time series market, such as the Markov chain Monte Carlo (MCMC) methods as well as the Gibb sampling used by Tsay (2010). The study could also use the asymptotic analysis to specify how the presence of the stochastic volatility affects the option prices and investment strategies used by Lorig & Sircar (2014). These models could be compared with the model discussed in Chapter 6 to determine the best fit model. Furthermore,we could use the orthogonal GARCH (OGARCH) model to build a covariance matrix for a large number of assets and modify correlated returns into the uncorrelated portfolio in modelling the interaction and co-movements among the group of time series variables. The non-stationary vector error correlation (VEC) model can also be considered, together with BEKK model. The exponentially weighted moving average (EWMA) model can also be used to count the most recent returns using a decaying factor λ. Then we can compare them with DCC model to see the best model for modelling the volatility and co-movements among the different stock markets.

Appendix A

Theorem 1

Consider x,y and z as three random variables, with multivariate normal joint distribution. Then we can also assume a nonsingular diagonal block covariance Σ_{ww}, such that $w = x, y, zand \sum_{yz} = 0$. Then we can have

$$E(x|y)\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),$$

$$var(x|y) = \Sigma_{xx} - \Sigma_{xx}\Sigma_{yy}^{-1}\Sigma_{xy},$$

$$E(x|y, z) = E(x|y) + \Sigma_{xz}\Sigma_{zz}(z - \mu_z),$$

$$var(x|y, z) = var(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$$

(Tsay, 2010).

A.1 Result1

Suppose x and y are multivariate normal joint, so that

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right), \tag{A.1}$$

hence, the distribution of x and y is multivariate normal with mean

$$\mu_x|y = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \tag{A.2}$$

and the covariance matrix becomes

$$\Sigma_{xx}|y = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}. \tag{A.3}$$

Thus the distribution of y given x is also multivariate normal with mean

$$\mu_y|x = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x), \tag{A.4}$$

the covariance matrix is also given by:

$$\Sigma_{yy}|x = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \tag{A.5}$$

(Durbin & Koopman 2001).

Appendix B

R codes for the Univariate, Multivariate GARCH models and Stochastic Volatility model

B.1 R Code for the symmetric GARCH Models

Normal (norm), student-t (std), skew student-t (sstd) and generalized error (ged) distributions were assumed

```
library(fGarch)
data=read.csv("return.csv",header=T)
data=return[-1]
d1=garchFit( arma(1,0)+garch(1,2),data=return,trace=FALSE,
include.delta=FALSE,cond.dist="norm")
summary(d1)
plot(d1)

d2=garchFit( arma(1,0)+garch(1,2) ,data=return,trace=FALSE,
include.delta=FALSE,cond.dist="std")
```

```
summary(d2)
plot(d2)

d3=garchFit( arma(1,0)+garch(1,2) ,data=return,trace=FALSE,
include.delta=FALSE,cond.dist="sstd")
summary(d3)
plot(d3)

d4=garchFit( arma(1,0)+garch(1,2) ,data=return,trace=FALSE,
include.delta=FALSE,cond.dist="ged")
summary(d4)
plot(d4)
```

B.2 R Code for the Asymmetric GARCH Models

```
library(rugarch)
data=read.csv("return.csv",header=T)
data=return[-1]

spec=ugarchspec(variance.model=list(model="gjrGARCH",garchOrder=c(1,2)),
mean.model=list(armaOrder=c(1,0)),distribution.model="std")
ARMA2=ugarchfit(spec=spec,data=return)
show(ARMA2)
plot(ARMA2)

spec=ugarchspec(variance.model=list(model="apARCH",garchOrder=c(1,2)),
mean.model=list(armaOrder=c(1,0)),distribution.model="std")
```

ARMA3=ugarchfit(spec=spec,data=return)

show(ARMA3)

plot (ARMA3)

spec=ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,2)), mean.model=list(armaOrder=c(1,0)),distribution.model="std")

ARMA4=ugarchfit(spec=spec,data=return)

show(ARMA4)

plot(ARMA4)

B.3 R Code for the DCC-GARCH (1,1) GARCH Models

library(timeSeries)

library(rugarch)

library(rmgarch)

library(fAssets)

library(ccgarch)

$> data = read.csv("returns.csv", header = T)$

$> y1 = data\$y1[-1]$

$> y2 = data\$y2[-1]$

$> y = cbind(y1, y2)$

$> garch11y.spec = ugarchspec(variance.model = list(model = "sGARCH",$
$garchOrder = c(1, 1)), mean.model = list(armaOrder = c(0, 0)), distribution.model = "norm")$

$> dcc.garch11y.spec = dccspec(uspec = multispec(replicate(2, garch11y.spec)),$
$dccOrder = c(1, 1), distribution = "mvnorm")$

$> dcc.garch11y.spec$

$> dcc.fit = dccfit(dcc.garch11y.spec, data = y)$

$> class(dcc.fit)$

$> slotNames(dcc.fit)$

$> names(dcc.fit@mfit)$

$> names(dcc.fit@model)$

$> dcc.fit$

$> plot(dcc.fit)$

B.4 R Code for the Stochastic Volatility Models

```
library(”stochvol”)
data=read.csv(”return.csv”,header=T)
data=return[-1]
arima = arima(return)
n=length(y)
y=log(y^2)
alpha=mean(y)
phi0=0
phi1=0.90
initialQ=0.70
initialSigma0=1
mu=-1
initialSigma1=1
initialparameter=c(phi0,phi1,initialQ,alpha,initialSigma0,mu,initialSigma1)
SV=function(n,y,phi0,phi1,initialQ,alpha,initialSigma0,mu,initialSigma1)
{
y=as.matrix(y)
Q=initialQ^2
Sigma0=initialSigma0^2
```

```
Sigma1=initialSigma1^2
h0=0
P0=initialQ^2/(1-phi1)
P0[P0<0]=0
ht=matrix(0,n,1)
Pt=matrix(0,n,1)
pi0=0.5
pi1=0.5
newpi0=0.5
newpi1=0.5
for(i in 1:n)
{
ht[i]=phi1*h0*phi0
Pt[i]=phi1*P0*phi1+Q
s0=Pt[i]+Sigma0
s1=Pt[i]+Sigma1
kt0=Pt[i]/s0
kt1=Pt[i]/s1
e0=y[i]-ht[i]-alpha
e1=y[i]-ht[i]-mu-alpha
f0=(1/sqrt(s0))*exp(-0.5*e0^2/s0)
f1=(1/sqrt(s1))*exp(-0.5*e1^2/s1)
newpi0=(pi0*f0)/(pi0*f0+pi1*f1)
newpi1=(pi1*f1)/(pi0*f0+pi1*f1)
h0=ht[i]+newpi0*kt0*e0+newpi1*kt1*e1
P0=newpi1*(1-kt1)*Pt[i]+newpi0*(1-kt0)*Pt[i]
like=like-0.5*log(pi0*f0+pi1*f1)
}
list(ht=ht,Pt=Pt,like=like)
```

```
}
Maximize=function(parameter)
{
phi0=parameter[1]
phi1=parameter[2]
initialQ=parameter[3]
alpha=parameter[4]
initialSigma0=parameter[5]
mu=parameter[6]
initialSigma1=parameter[7]
svmodel=SV(n,y,phi0,phi1,initialQ,alpha,initialSigma0,mu,initialSigma1)
return(svmodel$like)
}
estimate = optim (initialparameter, Maximize, NULL, method="BFGS",
hessian=TRUE, control = list (trace = 1, REPORT = 1, maxit = 1000 ))
standarderror=sqrt(diag(solve(estimate$hessian)))
cbind(estimate$par,standarderror)
```

Bibliography

Abdalla, S. Z. S. & Winker, P. (2012), 'Modelling stock market volatility using univariate GARCH models: Evidence from Sudan and Egypt', *International Journal of Economics and Finance* **4**(8), 161–176.

Achia, T., Wangombe, A. & Anyika, E. (2013), Time-series modeling of returns from the NSE 20-share index: An empirical study of the impact of political climate on market volatility. **URL** http://erepository.uonbi.ac.ke/bitstream/handle/11295/38632.

Agresti, A. (2002), *Categorical data analysis,* 2nd ed., Hoboken, New Jersey: Wiley.

Ahmed, A. E. S. & Suliman, S. Z. (2011), 'Modeling stock market volatility using GARCH models evidence from Sudan', *International Journal of Business and Social Science* **2**(23), 114–128.

Alberg, D., Shalit, H. & Yosef, R. (2008), 'Estimating stock market volatility using asymmetric GARCH models', *Applied Financial Economics* **18**(15), 1201–1208.

Andersen, T. G. & Bollerslev, T. (1997), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**(2–3), 115–158.

Andersen, T. G., Davis, R. A., Kreiss, J.-P. & Mikosch, T. V. (2009), *Handbook of financial time series*, Heidelberg: Springer Science & Business Media.

Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**(1), 3–30.

Bekaert, G., Hoerova, M. & Duca, M. L. (2013), 'Risk, uncertainty and monetary policy', *Journal of Monetary Economics* **60**(7), 771–788.

Bentes, S. R., Menezes, R. & Mendes, D. A. (2008), 'Long memory and volatility clustering: Is the empirical evidence consistent across stock markets?', *Physica A: Statistical Mechanics and its Applications* **387**(15), 3826–3830.

Black, F. (1976), 'Studies of stock price volatility changes.', *Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section, American Statistical Association,* 177-181 .

Black, F. & Scholes', M. (1973), 'The pricing of options and corporate liabilities', *The Journal of Political Economy* **81**(3), 637–654.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model', *The Review of Economics and Statistics* **72**(3), 498–505.

Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992), 'ARCH modeling in finance: A review of the theory and empirical evidence', *Journal of Econometrics* **52**(1-2), 5–59.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994), 'ARCH models', In *Handbook of Econometrics* **4**, 2959–3038.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *The Journal of Political Economy* **96**(1), 116–131.

Box, G. E., Jenkins, G. M. & Reinsel, G. C. (2008), *Time series analysis: Forecasting and control,* 4th ed., Hoboken, New Jersey: Wiley.

Brooks, C. (2008), *Introductory econometrics for finance,* 2nd ed., Cambridge: Cambridge University Press.

Chan, N. H. (2010), *Time series: Applications to finance with R and S-Plus,* 2nd ed., Hoboken, New Jersey: Wiley.

Chang, C.-L., McAleer, M. & Tansuchat, R. (2011), 'Crude oil hedging strategies using dynamic multivariate GARCH', *Energy Economics* **33**(5), 912–923.

Cont, R. (2007), *Long Memory in Economics,* Volatility clustering in financial markets: Empirical facts and agent-based model*s*, Berlin: Springer.

Cryer, J. D. & Chan, K.-S. (2008), *Time series analysis: With applications, in R,* 2nd ed., New York: Springer.

Daníelsson, J. (2011), *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*, Chichester: Wiley.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.

Ding, Z., Granger, C. W. & Engle, R. F. (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance* **1**(1), 83–106.

Dralle, B. (2011), Modelling volatility in financial time series, Master's thesis, Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg.

Durbin, J. & Koopman, S. J. (2001), *Time series analysis by state space methods*, Oxford: Oxford University Press.

Einicke, G. A. (2012), *Smoothing, filtering and prediction: Estimating the past, present and future*, Rijeka, Croatia: InTech.

Emenike, K. (2010), Modelling stock returns volatility in Nigeria using GARCH models. **URL** http://mpra.ub.uni-muenchen.de/22723/.

Engle, R. (2001), 'GARCH 101: The use of ARCH/GARCH models in applied econometrics', *The Journal of Economic Perspectives* **15**(4), 157–168.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica: Journal of the Econometric Society* **50**(4), 987–1007.

Engle, R. F. (1983), 'Estimates of the variance of US inflation based upon the ARCH model', *Journal of Money, Credit and Banking* **15**(3), 286–301.

Engle, R. F. (1990), 'Stock volatility and the crash of'87: Discussion', *The Review of Financial Studies* **3**(1), 103–106.

Engle, R. F. & Bollerslev, T. (1986), 'Modelling the persistence of conditional variances', *Econometric Reviews* **5**(1), 1–50.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R. F., Lilien, D. M. & Robins, R. P. (1987), 'Estimating time varying risk premia in the term structure: The ARCH-M model', *Econometrica: Journal of the Econometric Society* **55**(2), 391–407.

Engle, R. F., Patton, A. J. et al. (2001), 'What good is a volatility model ?', *Quantitative Finance* **1**(2), 237–245.

Engle, R. F. & Sheppard, K. (2001), 'Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH', NBER Working Paper 8554, Cambridge Mass: National Bureau of Economic Research .

Fama, E. F. (1965), 'The behavior of stock-market prices', *The Journal of Business* **38**(1), 34–105.

Fernández, C. & Steel, M. F. (1998), 'On Bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association* **93**(441), 359–371.

Francq, C. & Zakoian, J.-M. (2010), *GARCH models: Structure, statistical inference and financial applications*, Chichester: Wiley.

Franke, J., Härdle, W. K. & Hafner, C. M. (2011), *Statistics of financial markets,* 4th ed., Berlin: Springer.

Friedmann, R. & Sanddorf-Köhle, W. G. (2002), 'Volatility clustering and nontrading days in Chinese stock markets', *Journal of Economics and Business* **54**(2), 193–217.

Geweke, J. (1986), 'Modelling the persistence of conditional variances: A comment', *Econometric Reviews* **5**(1), 57–61.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**(5), 1779–1801.

Goudarzi, H. & Ramanarayanan, C. (2011), 'Modeling asymmetric volatility in the Indian stock market', *International Journal of Business and Management* **6**(3), 221.

Gouriéroux, C. (1997), *ARCH models and financial applications*, New York: Springer Science & Business Media.

Gujarati, D. (2004), 'Gujarati: *Basic econometrics*, 4th ed.', New York: McGraw-Hill Companies .

Hajizadeh, E., Seifi, A., Zarandi, M. F. & Turksen, I. (2012), 'A hybrid modeling approach for forecasting the volatility of S&P 500 index return', *Expert Systems with Applications* **39**(1), 431–436.

Higgins, M. L. & Bera, A. K. (1992), 'A class of nonlinear ARCH models', *International Economic Review* **33**(1), 137–158.

Ijumba, C. (2013), Multivariate analysis of the BRICS financial markets, Master's thesis, Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg.

Iori, G. (2002), 'A microsimulation of traders activity in the stock market: The role of heterogeneity, agents interactions and trade frictions', *Journal of Economic Behavior & Organization* **49**(2), 269–285.

Jacobsen, B. & Dannenburg, D. (2003), 'Volatility clustering in monthly stock returns', *Journal of Empirical Finance* **10**(4), 479–503.

Jefferis, K. & Okeahalam, C. (1999), 'International stock market linkages in Southern Africa', *South African Journal of Accounting Research* **13**(2), 27–51.

Jiang, W. (2012), 'Using the GARCH model to analyze and predict the different stock markets. Masters Thesis, Statistics,Uppsala University, Department of Statistics. Uppsala'.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of basic Engineering* **82**(1), 35–45.

Kang, S. H. & Yoon, S.-M. (2009), 'Modeling and forecasting the volatility of Eastern European emerging markets', *Journal of International Economic Studies* **13**(1), 113–134.

Kim, J. (2005), Parameter estimation in stochastic volatility models with missing data using particle methods and the EM algorithm, PhD thesis, Statistics, University of Pittsburgh, Pittsburgh.

Kirchgässner, G., Wolters, J. & Hassler, U. (2012), *Introduction to modern time series analysis*, Berlin: Springer Science & Business Media.

Lambert, P., Laurent, S. et al. (2001), Modelling financial time series using GARCH-type models with a skewed student distribution for the innovations, Stat Discussion Paper–0125. **URL** http://dial.uclouvain.be/pr/boreal/object/boreal:91014.

Lee, J. H. (1991), 'A lagrange multiplier test for GARCH models', *Economics Letters* **37**(3), 265–271.

Lee, W., McDougall, D. & Stuart, A. (2011), 'Kalman filtering and smoothing for linear wave equations with model error', *Inverse Problems* **27**(9), article: 095008.

Li, W. & Mak, T. (1994), 'On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity', *Journal of Time Series Analysis* **15**(6), 627–636.

Liu, H.-C. & Hung, J.-C. (2010), 'Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models', *Expert Systems with Applications* **37**(7), 4928–4934.

Liu, L.-M. (2009), *Time series analysis and forecasting,* 2nd ed., River Forest, Il: Scientific Computing Associates.

Lorig, M. & Sircar, R. (2014), Stochastic volatility: Modeling and asymptotic approaches to option pricing & portfolio selection.
URL: http://princeton.edu/ sircar/Public/ARTICLES/LorigSircar-VolReview˙revised.pdf.

Lux, T. & Marchesi, M. (2000), 'Volatility clustering in financial markets: A microsimulation of interacting agents', *International Journal of Theoretical and Applied Finance* **3**(4), 675–702.

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

Mandimika, N. Z. & Chinzara, Z. (2012), 'Risk–return trade-off and behaviour of volatility on the South African stock market: Evidence from both aggregate and disaggregate data', *South African Journal of Economics* **80**(3), 345–366.

Martin-Löf, P. (1973), 'Statistika modeller (statistical models): Anteckningar fran seminarier läsåret 1969–1970 (notes from seminars in the academic year 1969–1970), Stockholm: Stockholm University'.

McLeod, A. I. & Li, W. K. (1983), 'Diagnostic checking ARMA time series models using squared-residual autocorrelations', *Journal of Time Series Analysis* **4**(4), 269–273.

McMillan, D. G. & Ruiz, I. (2009), 'Volatility persistence, long memory and time-varying unconditional mean: Evidence from 10 equity indices', *The Quarterly Review of Economics and Finance* **49**(2), 578–595.

Muzindutsi, P. F. (2011), Exchange rate shocks and the stock market index: Evidence from the Johannesburg stock exchange, Master's thesis, Commerce, University of KwaZulu-Natal, Pietermaritzburg.

Mzamane, T. P. (2013), GARCH modelling of volatility in the Johannesburg stock exchange index, Master's thesis, Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica: Journal of the Econometric Society* **59**(2), 347–370.

Neokosmidis, I. et al. (2009), Econometric analysis of realized volatility: Evidence of financial crisis, PhD thesis, Economics, Aristotle University of Thessaloniki, Thessaloniki.

Orchard, T., Woodbury, M. A. et al. (1972), 'A missing information principle: Theory and applications', *Proceedings of the 6th Berkeley Symposium on Mathematical, Statistics and Probability,* University of California Press Berkeley, CA. **1**, 697–715.

Orskaug, E. (2009), Multivariate DCC-GARCH model:-with various error distributions, Master's thesis, Physics and Mathematics, Norwegian University of Science and Technology, Trondheim.

Pantula, S. G. (1986), 'Comment', *Econometric Reviews* **5**(1), 71–74.

Pesaran, B. & Pesaran, M. H. (2010), *Time Series econometrics using Microfit 5.0: A user's manual*, Oxford: Oxford University Press.

Peters, J.-P. (2001), *Estimating and forecasting volatility of stock indices using asymmetric GARCH models and skewed student-t densities,* (Preprint, University of Liege, Belgium).

Prno, J. & Slocombe, D. S. (2012), 'Exploring the origins of social license to operate in the mining sector: Perspectives from governance and sustainability theories', *Resources Policy* **37**(3), 346–357.

Ruppert, D. (2011), *Statistics and data analysis for financial engineering*, New York: Springer.

Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *The Journal of Finance* **44**(5), 1115–1153.

Schwert, G. W. (1990), 'Stock volatility and the crash of 87', *Review of Financial Studies* **3**(1), 77–102.

Sentana, E. (1995), 'Quadratic ARCH models', *The Review of Economic Studies* **62**(4), 639–661.

Shephard, N. (1996), 'Statistical aspects of ARCH and stochastic volatility', *Monographs on Statistics and Applied Probability* **65**, 1–68.

Shumway, R. H. & Stoffer, D. S. (2000), *Time series analysis and its applications*, New York: Springer.

Shumway, R. H. & Stoffer, D. S. (2006), *Time series analysis and its applications: with R examples,* 2nd ed., New York: Springer.

Shumway, R. H. & Stoffer, D. S. (2010), *Time series analysis and its applications: with R examples,* 3rd ed., New York: Springer.

Siklos, P. L. & Skoczylas, L. F. (2002), 'Volatility clustering in real interest rates: International evidence', *Journal of Macroeconomics* **24**(2), 193–209.

Sorensen, P. (2011), 'Mining in South Africa: A mature industry?', *International Journal of Environmental Studies* **68**(5), 625–649.

Su, C. (2010), Application of EGARCH model to estimate financial volatility of daily returns: The empirical case of China university of gothenburg, Master's thesis, School of Business, Economics and Law, University of Gothenburg, Gothenburg.

Sundberg, R. (1974), 'Maximum likelihood theory for incomplete data from an exponential family', *Scandinavian Journal of Statistics* **1**(2), 49–58.

Talke, I. S. (2003), Modelling volatility in financial time series data, Master's thesis, Mathematics, Statistics and Information Technology, University of KwaZulu-Natal, Pietermaritzburg.

Taylor, S. (1982), 'Financial returns modelled by the product of two stochastic processes–a study of daily sugar prices', Time Series Analysis: Theory and Practice 1. North-Holland, Amsterdam: Anderson, OD .

Taylor, S. (1986), *Modelling financial time series,* 2nd ed., Chichester: Wiley.

Thupayagale, P. (2010), Essays in long memory: Evidence from African stock markets, PhD thesis, University of St Andrews, Fife.

Tsay, R. S. (2005), *Analysis of financial time series,* 2nd ed., Hoboken, New Jersey: Wiley.

Tsay, R. S. (2010), *Analysis of financial time series,* 3rd ed., Hoboken, New Jersey: Wiley.

Tse, Y. K. & Tsui, A. K. (1999), 'A note on diagnosing multivariate conditional heteroscedasticity models', *Journal of Time Series Analysis* **20**(6), 679–691.

Tse, Y. K. & Tsui, A. K. C. (2002), 'A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business & Economic Statistics* **20**(3), 351–362.

Tseng, J.-J. & Li, S.-P. (2011), 'Asset returns and volatility clustering in financial time series', *Physica A: Statistical Mechanics and its Applications* **390**(7), 1300–1314.

Wang, Y. & Wu, C. (2012), 'Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models?', *Energy Economics* **34**(6), 2167–2181.

Wurtz, D., Chalabi, Y. & Luksan, L. (2006), Parameter estimation of ARMA models with GARCH/APARCH errors an R and SPlus software implementation*, Journal of Statistical Software,* Unpublished manuscript. **URL** http://www.jstatsoft.org.

Wurtz, D., Chalabi, Y. & Luksan, L. (2009), Parameter estimation of ARMA models with GARCH/APARCH errors an R and SPlus software implementation*, Journal of Statistical Software,* Unpublished manuscrip. **URL** http://www.jstatsoft.org.

Yager, T. R. (2004), 'The mineral industry of South Africa', *U.S. Geological Survey Minerals Yearbook* **36**, 1–12.

Zakoian, J.-M. (1994), 'Threshold heteroskedastic models', *Journal of Economic Dynamics and Control* **18**(5), 931–955.

Zivot, E. & Wang, J. (2007), *Modeling financial time series with S-Plus*, New York: Springer.