



**UNIVERSITY OF
KWAZULU-NATAL**

**INYUVESI
YAKWAZULU-NATALI**

UNIVERSITY OF KWAZULU-NATAL
(PMB)

DISSERTATION

STATISTICS

**Modelling Volatility in Stock Exchange Data: A Case
Study Of Three Johannesburg Stock Exchange(JSE)
Companies**

FORBES KASEKE

Thesis submitted
in fulfillment of the requirements for the degree of

Master of Science
in

Statistics

November, 2015

Declaration

The research work is the original work done by the author Forbes Kaseke and it is not a duplicate source of the research work done by other authors. All the references used are duly acknowledged.

Author: Mr. Forbes Kaseke

Date

Supervisor: Dr. S. Ramroop

Date

Co-Supervisor: Prof. H.G. Mwambi

Date

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal

Pietermaritzburg

Acknowledgement

Firstly glory to the almighty God for the gift of life. Secondly I thank my parents Mr and Mrs Kaseke for the continuous support to see that I have the best education. I would also, like to thank my supervisor's Dr. S. Ramroop and Prof. H. Mwambi for their continuous support and encouragement in completing this theses. I would also, like to thank Dr. T. N. O. Achia for helping me when I was stuck with R coding. My colleagues from the department of Statistics your contributions are not forgotten.

Abstract

For investors and policy makers such as governments, the uncertainty of returns on investments is a major problem. The aim of this paper is to study volatility models for financial data for both univariate and multivariate case. The data to be used is monthly and daily asset returns of three different companies. For the univariate case, the main focus is on GARCH models and their subsequent derivatives. ARCH and GARCH models of different orders are fit. For the monthly data, the GARCH(1,1) outperformed the ARCH and higher order GARCH models. For the daily data, the GARCH(1,1) preceded by an appropriate AR model was the best fit. For the Multivariate volatility models, models such as the DCC-GARCH, EMWA and Go-GARCH were used. All three gave similar results. Various distributional assumptions such as the normal and Student t distributions were assumed for the innovations. Student t and Skewed Student t distributions were more effective because of their ability to capture fat tails of the distributions. Fundamental finance terms and concepts are also discussed.

KEYWORDS :volatility, returns, ARCH models, GARCH models

Contents

1	Introduction	1
1.1	Literature Review	3
1.2	Financial Data	6
1.2.1	Returns	6
1.3	Properties of Returns	7
1.3.1	Volatility	9
1.4	Problem Statement	10
1.5	Objectives Of The Study	11
1.6	Summary	11
2	Data Description: Exploratory Data Analysis	12
2.1	Daily Data	12
2.2	Monthly Data	19
2.3	Summary	25
3	ARCH and GARCH Models	26
3.1	Model Structure	26
3.1.1	Model Building	26
3.1.2	Testing for ARCH effects	27
3.2	ARCH Model	28
3.2.1	Properties of ARCH Models	28
3.2.2	Weaknesses of the ARCH Model	30
3.2.3	Parameter Estimation	30
3.2.4	Non Normal Errors	31
3.2.5	Data Checking	32
3.2.6	Model Checking	32
3.2.7	Forecasting	32
3.3	GARCH Model	32
3.3.1	GARCH(1,1)	33
3.4	Forecasting in the GARCH(1,1) Model	35
3.4.1	Persistence and Stationarity	35
3.5	IGARCH	35
3.6	Forecasting IGARCH Models	36
3.7	GARCH-M	36
3.8	Exponential GARCH Model	37
3.9	Model section using Information Criteria	38

3.10	News Impact Curve	38
3.11	Summary	39
4	Model Application	40
4.1	Old Mutual Data	40
4.1.1	Monthly Data	41
4.1.2	GARCH Extensions For Old Mutual Monthly Data	46
4.1.3	Old Mutual Daily Data	48
4.1.4	GARCH Extensions For Old Mutual Daily Data	53
4.2	Standard Bank Data	56
4.2.1	Monthly Data	56
4.2.2	GARCH Extensions For Standard Bank Monthly Data	60
4.2.3	Standard Bank Daily Data	63
4.2.4	GARCH Extensions For Standard Bank Daily Data	69
4.3	Liberty Holdings	73
4.3.1	Daily Data	73
4.3.2	GARCH Extensions For Liberty Holdings Daily Data	78
4.3.3	Monthly Data	82
4.3.4	GARCH Extensions For Liberty Holdings Monthly Data	86
4.4	Summary	89
5	Multivariate Volatility Models	91
5.0.1	Portmanteau Test	92
5.0.2	Rank Based Test	93
5.1	Diagnostic Checks Of Volatility Models	93
5.1.1	Ling and Li Statistics	93
5.2	Volatility Models	94
5.2.1	Exponentially Weighted Moving Average (EWMA)	94
5.2.2	Dynamic Conditional Correlation Models (DCC-GARCH)	95
5.2.3	Go-GARCH Model	98
5.3	Summary	99
6	Multivariate Volatility Models Application	100
6.1	Multivariate Daily Data	100
6.1.1	EMWA Model Fitting	102
6.1.2	DCC Model Fitting	104
6.1.3	Go-GARCH Model Fitting	107
6.2	Multivariate Monthly Data	110
6.2.1	EMWA Model	113
6.2.2	Go-GARCH Model Fitting	115
6.2.3	DCC Model Fitting	116
6.3	Chapter Summary	119
7	Discussion and Conclusion	121

Appendices	124
Appendix	124

Chapter 1

Introduction

As technological advancement continues, the world is also, becoming more and more interconnected. Markets have become so interdependent giving rise to one global market. As much as the global market has come as a blessing to investors by allowing them to invest globally, it has, however, also come with its own shortfalls. Because the global market is made up of many different economies which are affected by uncertain geopolitical and macroeconomic policies and developments, it means at a point in time, the behaviour of one market is likely to differ to the next. These changes can be spontaneous and damaging in the sense that, a failure in one market, especially the bigger economies, can quickly spread across other markets (there is a correlation in market movements) leading to a global crisis. A good example is the 2008 global recession (Imbs 2010) where all economies took a downturn despite being from different areas around the globe. As such, interest in the behavior of financial markets has grown exponentially, with investors and portfolio managers being keener on knowing how the market is and how it will likely behave in the future. Armed with this information they can, thus, make informed decisions on where and when to invest.

Whilst investors are concerned with maximising the returns on their investment, they recognise the trade-off that occurs between the higher returns and risk. Hence, they have a certain level of risk that they are willing to bear. A good forecast of volatility is an ideal tool for assessing investment risk. Our common traditional time series models such as the (Autoregressive Moving Average) ARMA models are based on stationarity assumptions, that is, they are used to model the conditional expectation of a process given the past; in this case the conditional variance given the past is constant. This is not the case with financial data which tend to be more volatile than constant. Such effects cannot be captured by the ARMA model. The ARMA models work on data that has a constant variance of error terms. This data is said to have homoscedasticity. On the other hand, for financial data the variances of the error terms are not equal, that is it has non-constant volatility. Data with the non-constant volatility/variance is said to suffer from heteroskedasticity. In the presence of heteroskedasticity, the regression coefficients for an ordinary least squares regression are still unbiased, but, the standard errors and confidence intervals estimated by conventional procedures will be too narrow, hence, giving a false sense of precision. ARCH and GARCH models treat heteroskedasticity as a variance to be modelled. As a result, not only are the deficiencies of least squares corrected, but a prediction is computed for the variance of each error term. This prediction turns out

often to be of interest, particularly in applications in finance (Engle 1982). Ruppert (2010) gives an example by considering recent daily stock returns which say, have been unusually volatile, one might then expect that tomorrow's return to be also more variable than usual. In such cases, ARMA models cannot capture this type of behavior because for ARMA models conditional variance is constant. (Francq and Zakoian 2011) states that modelling financial time series is a complex problem and this complexity is not only due to the variety of the series in use (stocks, exchange rates, interest rates, etc.), to the importance of the frequency of observation (second, minute, hour, day, etc) or to the availability of very large data sets. It is mainly due to the existence of statistical regularities (stylized facts) which are common to a large number of financial series and are difficult to reproduce artificially using stochastic models. Francq and Zakoian (2011) further points out the many difficulties that arise from financial data modelling and these include:

- Non-stationarity of the price series, that is the series of prices of a certain commodity is not stationary, but it tends to fluctuate over time due to factors such as inflation, supply and demand amongst others.
- Absence of autocorrelation for the price variations. The series of price variations generally displays small autocorrelations, making it close to white noise.
- Autocorrelations of the squared price returns. Squared returns or absolute returns are generally strongly autocorrelated. This property is not incompatible with the white noise assumption for the returns, but shows that the white noise is not strong.
- Volatility clustering whereby large absolute returns tend to appear in clusters. This property is generally visible on the sample paths. High volatility subperiods are followed by quiet (low-volatility) periods. These subperiods are recurrent but do not appear in a periodic way (which might contradict the stationarity assumption). In other words, volatility clustering is not incompatible with a homoscedastic marginal distribution for the returns.
- There exist fat tails in the data (Fat-Tailed Distribution). These fat tails can be understood by looking at the empirical distribution of daily returns, one can generally observe that it does not resemble a typical Gaussian distribution. It rather has more data at the ends, that is it has higher tail probabilities as compared to the normal distribution hence, the term fat tail distribution. Classical tests for normality such as the kurtosis typically lead to rejection of the normality assumption at any reasonable level for such data. In such cases, it is thus, more reasonable to use fat-tailed distributions such as the studentized t distribution for your models.
- Leverage effects were first noted by Black (1976) and they relate to the asymmetry of the impact of past positive and negative values on the current volatility. Negative returns tend to increase volatility by a larger amount than positive returns of the same magnitude.
- Seasonality calendar effects are also worth mentioning. The day of the week, the proximity of holidays, among other seasonalities may have significant effects on

returns. Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break.

Bearing the above-mentioned complexities in mind, a new model for financial data called the Autoregressive Conditional Heteroscedastic (ARCH) model was proposed by Engle (1982) and later generalized by Bollerslev (1986) and Taylor (1986) to give the Generalized Autoregressive Conditional Heteroscedastic (GARCH) models. These models have been proven to sufficiently capture the properties of financial data. This research seeks to apply the ARCH and the various GARCH models to different data sets so as to explore the suitability of the models for the data sets.

1.1 Literature Review

The unique characteristics of financial data were first noted early in the 20th century when various scholars such as Mitchell (1965) and Mills (1927) noticed that financial data did not adhere to the normality assumptions. Fat tails were first discovered by Mitchell (1965), Mitchell (1921), Mills (1927) and other scholars such as Alexander (1961). Kendall and Hill (1953) and Houthakker (1961) were among the first to give empirical evidence that prices were non stationary. Malkiel and Fama (1970) found that 22 out of the 30 stocks in the Dow Jones Industrial Average (DJIA) exhibited positive daily serial correlations. With all these characteristics it was, thus, clear that traditional time series models were inadequate to describe financial data. New type of models had to be developed.

Engle (1982) developed the ARCH model to capture the unusual behaviour of financial data. He derived the ARCH model for capturing what he termed the ARCH process. He describes ARCH processes as those with mean zero and serially uncorrelated processes with nonconstant variances conditional on the past but constant unconditional variances. The main effect of the model is that it allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant. The paper cements the importance of testing for ARCH effects. To establish the reliability of the model by conventional criteria, it must be tested for serial correlation and for coefficient restrictions. After much comparison, it was seen that the ARCH models fared better in modelling data with ARCH effects. It was also, seen that the ARCH model was better with maximum likelihood estimates rather than ordinary least square estimates.

To overcome some of the weakness of the ARCH model Bollerslev (1986) introduced the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model. This model allows a more flexible lag structure whilst, also, allowing in many cases from an ARCH model a significant decrease in number of parameters hence, giving a more parsimonious model. Like the ARCH model, the standard GARCH has limitations, for example, like the ARCH model it captures aspects of financial data such as volatility clustering and leptokurtosis but fails to capture the leverage effect. Also, for instance, the requirements of conditional variance to be positive can be violated upon estimating it, hence, the need for constraints to be put in place. To cater for these limitations variations of the GARCH model were introduced.

Engle and Bollerslev (1986) developed a new version of the GARCH model which they called the I-GARCH model. This model was made to cater for data whose multi-step forecast do not approach unconditional variance. Another extension of GARCH was the GARCH-M. This model is further discussed under the ARCH GARCH section of the thesis. Nelson (1991) identifies 3 problems with ARCH/GARCH in price applications. These are:

- Researchers such as Black (1976) amongst many others found a negative correlation between current returns and future returns volatility. GARCH models rule this out by assumption.
- GARCH models impose parameter restrictions that are often violated by estimated coefficients and that may unduly restrict the dynamics of the conditional variance process.
- Interpreting whether shocks to conditional variance “persist” or not is difficult in GARCH models because the usual norms measuring persistence often do not agree.

Black (1976) then developed the Exponential GARCH model EGARCH which helped to capture the asymmetries of the positive and negative returns. Other models such as the GJR-GARCH model by (Glosten et al. 1993) and the Asymmetric Power ARCH (APARCH) model by (Ding et al. 1993) were developed to cater for different data behaviour.

Ding and Granger (1996) gives new evidence of long-term dependence that exists in speculative returns series. In their paper, five speculative returns series from different places and different markets were examined. They found out that the absolute returns and their power transformations all have long and positive autocorrelations with the property being strongest for absolute returns. An exception, however, is the exchange return which has the strongest property when taken to power $1/4$. They found that theoretical autocorrelation functions for various GARCH(1,1) models are exponentially decreasing, which is rather different from the sample autocorrelation function for the real data. They proposed a general class of long memory models that has no memory in returns themselves, but long memory in absolute returns and their power transformations. With this model, the estimated results showed that this model gives a much better description of the real data.

Bollerslev and Ole Mikkelsen (1996) proposed a new class of more flexible fractionally integrated EGARCH models for characterizing the long-run dependencies in U.S. stock market volatility. Strong evidence was discovered that the conditional variance for the Standard and Poor’s 500 composite index is best modelled as a mean-reverting fractionally integrated process.

Baillie and Bollerslev (2002) found out that the daily spot rate is well represented by a GARCH(1,1) process with near unit roots. Also, they found out that ARCH effects are still strong in weekly data less so on fortnightly data and minimal to monthly data. It was, also, seen that after accounting for ARCH effects, the data fitted the assumption of conditional normality more reasonable on monthly and fortnightly data. With daily

data, however, normality was found to be inappropriate but replacing with conditional t distribution was appropriate.

In a study by Corhay and Rad (1994) where they looked at five European stock markets they found out that conditional heteroskedasticity is a prime feature of daily returns behaviour of the five European equity indices. They exhibited non-linear dependence that could not be captured by the random walk model. However, the ARCH/GARCH models were consistent in modelling the behaviour with the GARCH-T(1,1), i.e. a GARCH model with conditional errors that are t-distributed, fitted their data best. They concluded that GARCH models can indeed provide better forecasts of volatility than the usual historical estimates and lead to improved valuation models.

Poon and Granger (2003) reviewed scientific papers on ARCH/GARCH models, amongst the 93 scientific papers they looked at, they found out that 17 studies compared alternative versions of GARCH. It was clear that GARCH dominates ARCH. In general, models that incorporate volatility asymmetry such as EGARCH and GJR-GARCH perform better than GARCH, a finding that is supported by most research papers. Also, certain specialized specifications, such as fractionally integrated GARCH (FIGARCH) and regime switching GARCH (RSGARCH) do better in some studies. They, also, note that papers compared used different data sets, many kinds of assets, various intervals between readings and for different purposes. Although a suggestion can be made that a particular method of forecasting volatility is the best, no statement is available about the cost-benefit from using it rather than something simpler or how far ahead the benefits will occur. Another debate that remains revolves on how far ahead one could accurately forecast and to what extent could volatility changes be predicted.

Wilhelmsson (2006) explored forecasting the GARCH model under different distribution assumptions. He found that allowing for leptokurtic error distributions improved the model variance estimates. The model estimated with Students t distribution were the best performing model rather than those models with normal assumption. Allowing for non-time-varying kurtosis improved the forecasts. However, allowing non-time-varying skewness or time-varying kurtosis did not improve the forecasts further.

Poon and Granger (2005) interrogates an important issue as to how volatility is measured. They state that it is known that volatility refers to the spread of all likely outcomes of an uncertain variable. Statistically, volatility is often measured as the sample standard deviation. Sometimes, variance, σ^2 is used as a volatility measure. Since variance is simply the square of standard deviation, it makes no difference whichever measure we use when we compare the volatility of two assets. However, variance is much less stable and less desirable than standard deviation as an object for computer estimation and volatility forecast evaluation. Moreover standard deviation has the same unit of measure as the mean, i.e. if the mean is in dollars, then standard deviation is also expressed in dollars whereas variance will be expressed in dollars square.

Poon (2005) studied the practical issues surrounding volatility forecasting. The results

showed that option implied volatility dominates time-series models because the market option price fully incorporates current information and future volatility expectations. There was no clear winner between historical volatility and ARCH models, but both were better than the stochastic volatility model. Despite the added flexibility and complexity of stochastic volatility models, there was no clear evidence that they provide superior volatility forecasts. Also, high-frequency data clearly provided more information and produced better volatility forecasts, particularly over short horizons. He also gave a potentially useful area for future research and this area focuses on whether forecasting power can be enhanced by using exogenous variables.

1.2 Financial Data

Finance basically revolves around buying and selling. Those providing the goods and services need capital so as make their product available. The products have prices that will generally differ at each point in time. A plot of the prices against time gives us a price series. The price series, in general, is not stationary as it tends to rise in the long run. A different measure is, thus, needed in this case, we employ returns which are discussed in detail below.

1.2.1 Returns

Campbell et al. (1997) gives two main reasons for using returns. Firstly, for average investors, the return of an asset is a complete and scale-free summary of the investment opportunity, that is, the investor has an idea of how much he will gain or lose per each dollar. Secondly, for theoretical and empirical reasons that will become apparent as we discuss returns below, returns have more attractive statistical properties than prices, such as stationarity and ergodicity.

The return on an investment is its revenue expressed as a fraction of the initial investment. If one invests at time t in an asset with price P_t and the price later at time $t-1$ is P_{t-1} , then the net return for the holding period from t to $t-1$ is given by

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1.1)$$

By holding the asset for k periods we get k period simple returns these are known as multi-period simple returns.

$$1 + R_t = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \dots \times \frac{P_{t-k+1}}{P_{t-k}} \quad (1.2)$$

$$= (1 + R_t)(1 + R_{t-1})\dots(1 + R_{t-k+1}) \quad (1.3)$$

$$= \prod_{j=0}^{k-1} (1 + R_{t-j}) \quad (1.4)$$

The k -period simple gross return is just the product of the k one-period simple gross returns involved. This is called a compound return. The k -period simple net return is

$$R_t[k] = (P_t - P_{t-k})/P_{t-k}.$$

In general, however, the log of the returns is used more in application than the return. Log returns are also known as continuously compounded returns. Denoting them by r_t then

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = \ln p_t - \ln p_{t-1} \quad (1.5)$$

For multi-periods returns we have

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) \\ &= \ln(1 + R_t)(1 + R_{t-1})\dots(1 + R_{t-k+1}) \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \dots + \ln(1 + R_{t-k+1}) \end{aligned} \quad (1.6)$$

As can be seen above, log returns have an advantage on multiple periods as they are just simple summations and not multiples. It follows, as such, that log returns have statistical properties that are more tractable. This is explained by Quigley and Ramsey (2008) where they state that major benefit is in normalisation. It allows measuring all variables in a comparable metric and thus, enabling evaluation of analytic relationships amongst two or more variables despite originating from price series of unequal values. It is commonly assumed that the simple returns are log-normally distributed. If simple returns are independent and identically distributed (i.i.d.) as log-normal then it follows that the log-returns are i.i.d. normally distributed and this allows great statistical freedom. This is handy given much of classic statistics presumes normality. Another advantage is for additivity as probability theory shows that the product of normally-distributed variables is not normal. Instead, the sum of normally-distributed variables is normal. As seen above, log returns are additive while simple returns are multiplicative.

1.3 Properties of Returns

The most important property of returns is that they are random, the return a week later for an investment is not known. Campbell et al. (1997) elaborate that “it is largely the explicit modelling of the sources and nature of this uncertainty that distinguishes financial economics from other social sciences.” This uncertainty is, hence, the backbone of the studies in finance as such studies would not matter much if everything was certain.

A prominent feature of returns is that extreme values are more common with the extreme negative returns being more common than the extreme positive. This feature on its own gives an insight on the kind of distribution returns data follow. The data clearly does not follow the classical normal distribution model that is commonly used in many models. To describe the relation of financial data with regards to the normal distribution we look at the moments. Consider a continuous random variable X (log returns say) and the

cumulative distribution (CDF)

$$F_X(x) = \int_{-\infty}^x f_x(u)du \quad (1.7)$$

Where f_x is the probability distribution function (pdf). The ℓ th moment (non-central) of a continuous random variable X is defined as

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x)dx \quad (1.8)$$

The first moment is called the mean or expectation of X i.e μ . If we subtract the mean from X we have the central moment i.e

$$m_\ell = E(X - \mu_x)^\ell = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x)dx \quad (1.9)$$

The second central moment is the variance denoted σ_x^2 and measures variability. The third and fourth central moments measure tail behaviour. They are known as kurtosis and skewness.

$$S(x) = E \left[\frac{(X - \mu_x)^3}{\sigma_x^3} \right] \quad K(x) = E \left[\frac{(X - \mu_x)^4}{\sigma_x^4} \right] \quad (1.10)$$

Where $S(x)$ is the skewness and $K(x)$ is kurtosis. Skewness describes the asymmetry of the distribution. For normal distribution, skewness is always zero, that is, it has perfect symmetry. A negative skewness value indicates that the data has a distribution skewed to the left. This means that the left tail is heavier than the right tail in the distribution. Respectively, a positive skewness value indicates a right skewed distribution with a right tail heavier than its left tail.

Kurtosis is a measure of the extent to which observed data fall near the centre of a distribution or the tails thereof. It is a measure of peakedness. If a curve is placed higher relative to the normal distribution it has more kurtosis. This means that the curves tails at both sides are fatter as a result of it being higher (Karlsson 2002). The quantity $K(x) - 3$ is called the excess kurtosis since $K(x) = 3$ for a normal distribution. Thus, the excess kurtosis of a normal random variable is 0. A distribution with positive excess kurtosis is said to have heavy tails, implying that the distribution puts more mass on the tails of its support than a normal distribution does (Tsay 2005). This is usually the case with returns. In finance applications, one is especially concerned when the return distribution has heavy tails because of the possibility of an extremely large negative return, which could, for example, entirely deplete the capital reserves of a firm. Such a distribution is said to be leptokurtic. On the other hand, a distribution with negative excess kurtosis has short tails. Such a distribution is said to be platykurtic. Most financial asset returns will have kurtosis greater than 3 whilst stock market return is in particular featured by negative skewness. Other financial series as interest rates and returns may have positive skewness.

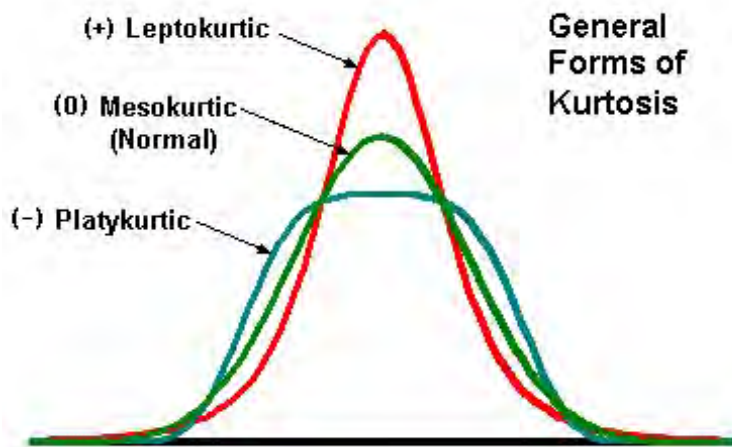


Figure 1.1: General Forms Of Kurtosis



Figure 1.2: Illustrations Of Skewness

1.3.1 Volatility

For most assets, future returns cannot be known exactly and, therefore, are random variables. Risk means uncertainty in future returns from an investment, in particular, that the investment could earn less than the expected return and even result in a loss, that is, a negative return. Risk is often measured by the standard deviation of the return, i.e the volatility (Ruppert 2010).

Sheppard (2013) defines volatility simply as the standard deviation. It is used as a measure of uncertainty of the returns, that is to say, volatility represents that range in which

your returns are likely to fall within. Thus, in essence, volatility quantifies the riskiness of an asset. The higher the volatility the more variable and uncertain the returns are. Characteristics of volatility entail that it is not directly observable. For example, daily volatility can not be seen from return data because there is one observation at the end of a trading day but by subdividing, say, into 10-minute intervals you can estimate daily volatility. It is this ungovernability which makes it difficult to evaluate the forecast of conditional heteroscedastic models. Although volatility is not directly observable, there are characteristics of it that are commonly seen in asset returns. First, the issue of volatility clustering where volatility may be high for certain time periods and low in other periods. Secondly, volatility evolves over time in a continuous manner, meaning that, volatility jumps are rare. Thirdly, volatility does not diverge to infinity but varies within a fixed range. Lastly, volatility reacts differently to a big price increase or drop i.e. leverage effects. With these properties in mind, developed models will seek to capture their effects.

The study of volatility has many advantages comprising of:

- It provides a simple approach to calculating value at risk of a financial position in risk management.
- It plays an important role in asset allocation under the mean-variance framework.
- It improves the efficiency in parameter estimation and the accuracy of interval forecast
- The volatility index of a market has recently become a financial instrument which can be traded.
- Predictability of volatility is important in designing optimal dynamic hedging strategies for options and futures.
- Volatility is often preferred to variance as it is measured in the same units as the original data, unlike variance which is units squared (Engle and Ng 1993).

Another important issue surrounding volatility is serial correlations. Jondeau et al. (2007) states that to test for time dependency in volatility, we need a time varying measure of volatility. Two ways to approach this are to use the mean adjusted square returns and secondly, using the absolute returns. The test for serial dependence is done using the Ljung-Box Q statistic. Non-zero serial correlation in squared or absolute returns is evidence of volatility time dependence.

1.4 Problem Statement

Volatility has become an important part of investment to such an extent that it has become a tradable instrument. As such, we are faced with need to develop or find suitable models that can capture volatility patterns. The aim of this thesis is to run, compare and contrast various ARCH and GARCH models for daily and monthly data in order to capture and predict volatility for both univariate and multivariate case.

1.5 Objectives Of The Study

The objectives of this study are to find the best fitting models and to model volatility of asset returns for both monthly and daily data for both univariate and multivariate case.

1.6 Summary

Studying volatility in the markets is of importance as these market fluctuations, when ungoverned, can affect financial stability of an economy.

This chapter introduced us to the study of volatility. It explains various terms that will be used in this study such as returns and volatility. It, specifically, elaborates on the different types of returns and shows that the best form of return data is the log returns. A detailed study of the properties of returns is done and it is clear that return data has properties that differ from the normal distribution. Specifically, the data possesses fat tails and has excess skewness and hence, does not adhere to the usual normal distribution. It also gave us a brief discussion on past studies on the same topics that have been done hence, giving us an overview of what to look at and expect as this research progresses.

Chapter 2

Data Description: Exploratory Data Analysis

The data to be used is from financial companies that are registered on the Johannesburg Stock Exchange (J.S.E). We consider the following companies: Old Mutual, Liberty Holdings and Standard Bank. We look at daily and monthly returns of each company, this we do, so as to see if different time periods will affect the type of model to be selected. This data was obtained from <http://research.mcgregorbfa.com/Default.aspx> via the University Of Kwazulu Natal website on the price data section where we retrieved the historical prices for the above-listed companies. Returns are then calculated from these prices using Excel before importing the data into R.

2.1 Daily Data

The first step to analyse any time series data is to have a plot of the data as to gain an insight on the data its trends and patterns. The data here is the stock closing prices from January 2003 to December 2013.

Figure 2.1 below shows the plot of the asset closing prices. For all assets, a clear upward trend of closing prices is observed. A more pronounced decline is visible around 2009 which coincided with the global recession after which it resumes to the upward trend. A notable occurrence is that Liberty Holdings prices did not rise as much as other two and also did not get affected by the global recession decline as the other two. This can be seen as an indication that the asset is less risky.

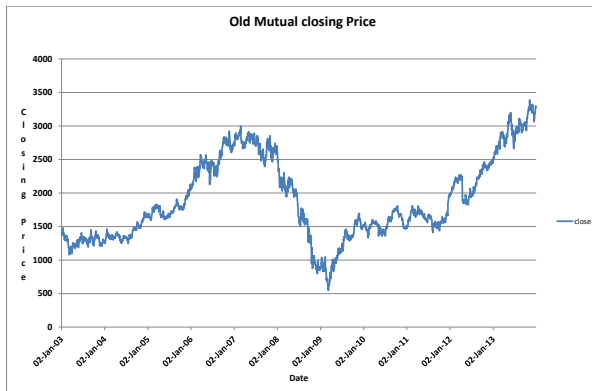
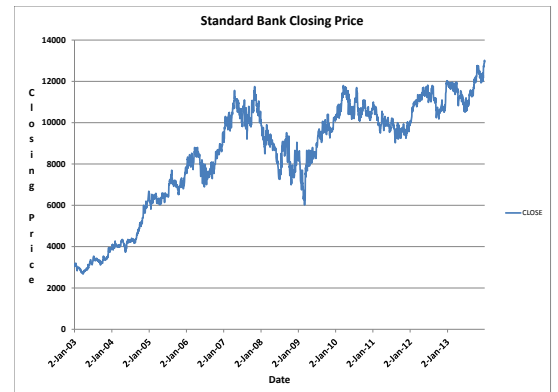


Figure 2.1: Closing Prices For The 3 Companies

Figure 2.2 shows the plots of the return series. These have 1 observation less because one observation is lost during differencing. Unlike the price series which was non-stationary, the return series are stationary, thus, further justifying the use of the return series over the price series. Another distinctive feature in the return series is volatility clustering. We can see that periods of high returns are clustered together and there are followed by small return periods which are also clustered together, this is another feature of financial data. A closer look at the series shows that around 2008 and 2009 there is a cluster of positive and negative returns which is justified because it was a period of uncertainty with most people losing confidence in the markets but, also, others trying to take advantage to purchase stocks, hence, series fluctuations in asset returns.

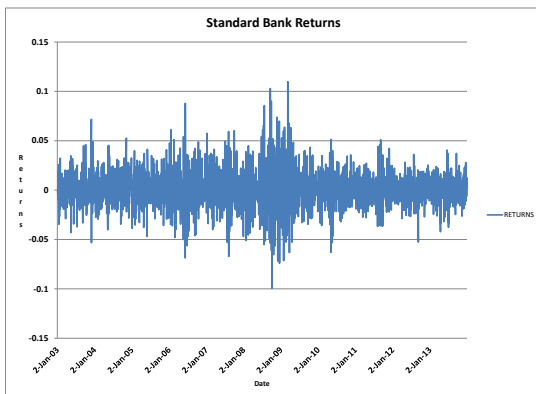
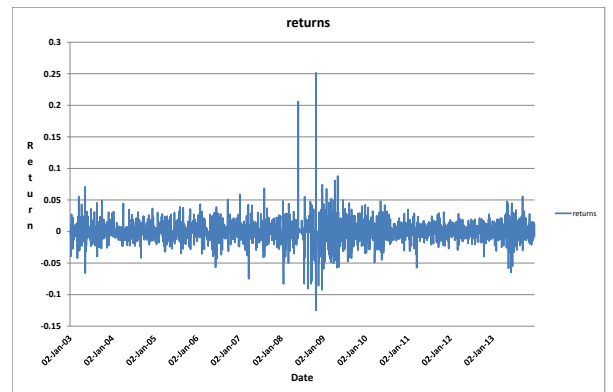
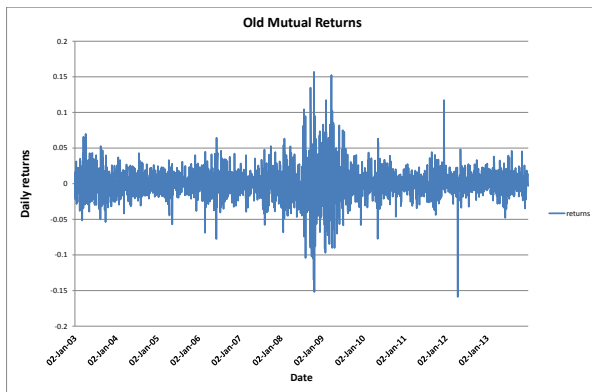


Figure 2.2: Returns Plots For The 3 Companies

Below Figure 2.3 shows the squared return plots for all assets. High squared returns are clustered together and low squared returns are also clustered together. This shows that volatility clustering is present as expected.

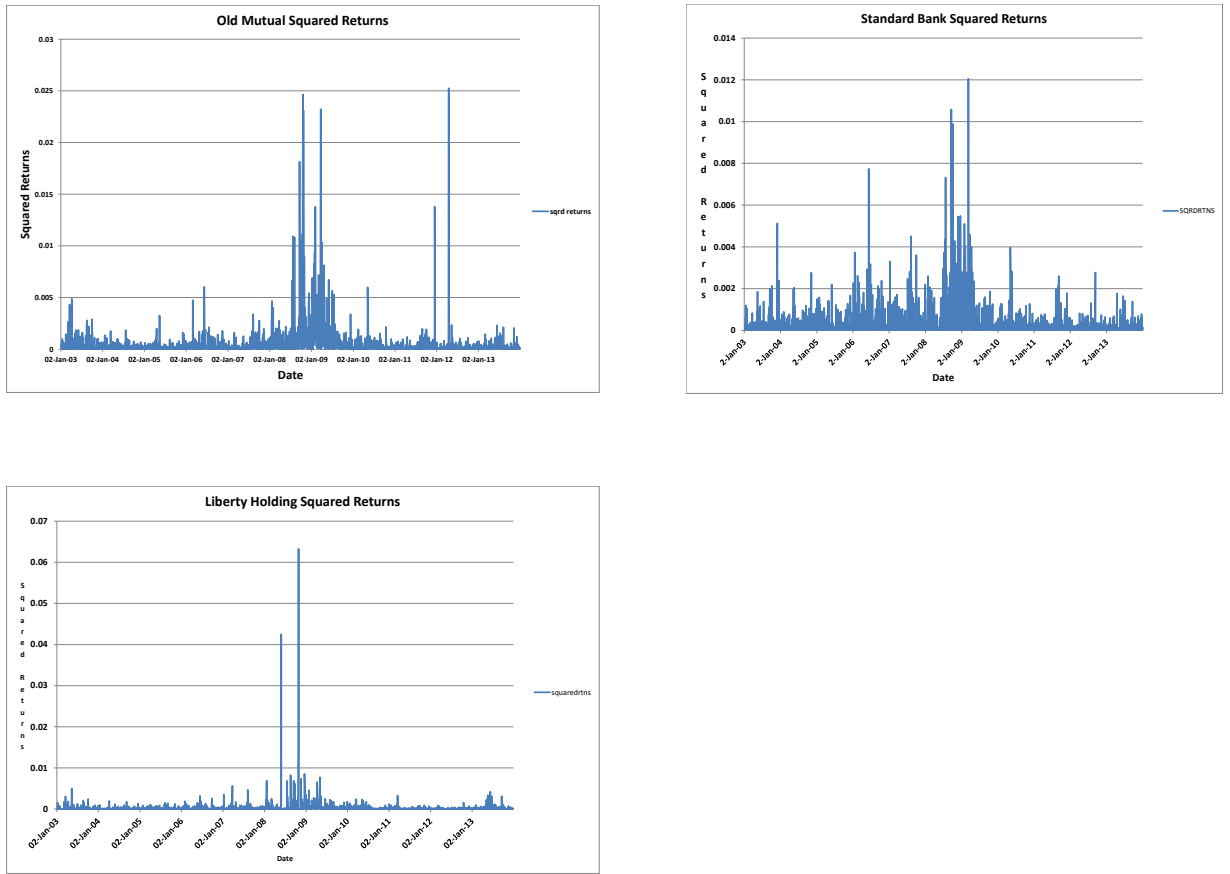


Figure 2.3: Old Mutual Squared Returns

Figure 2.4 shows the ACF and PACF of the Old Mutual return series. The ACF shows small autocorrelations at lags 1, 3, 6 and 10. These low order correlations are expected for daily data. The LjungBox statistics give $Q(1) = 4.7$ $Q(3)=17.1$ $Q(6)=25.5$ and $Q(10) = 40.7$ with all having p-values <0.05 . The p values of these four test statistics are all less than 0.05, suggesting that returns are serially correlated. The squared returns are as expected strongly autocorrelated. The squared return shows slow decay of the lag plots, this indicates there is correlation between the magnitude of change in the return. This means there is serial dependence in the variance of the data. Also, the spikes in PACF are a sign of presence of ARCH effects.

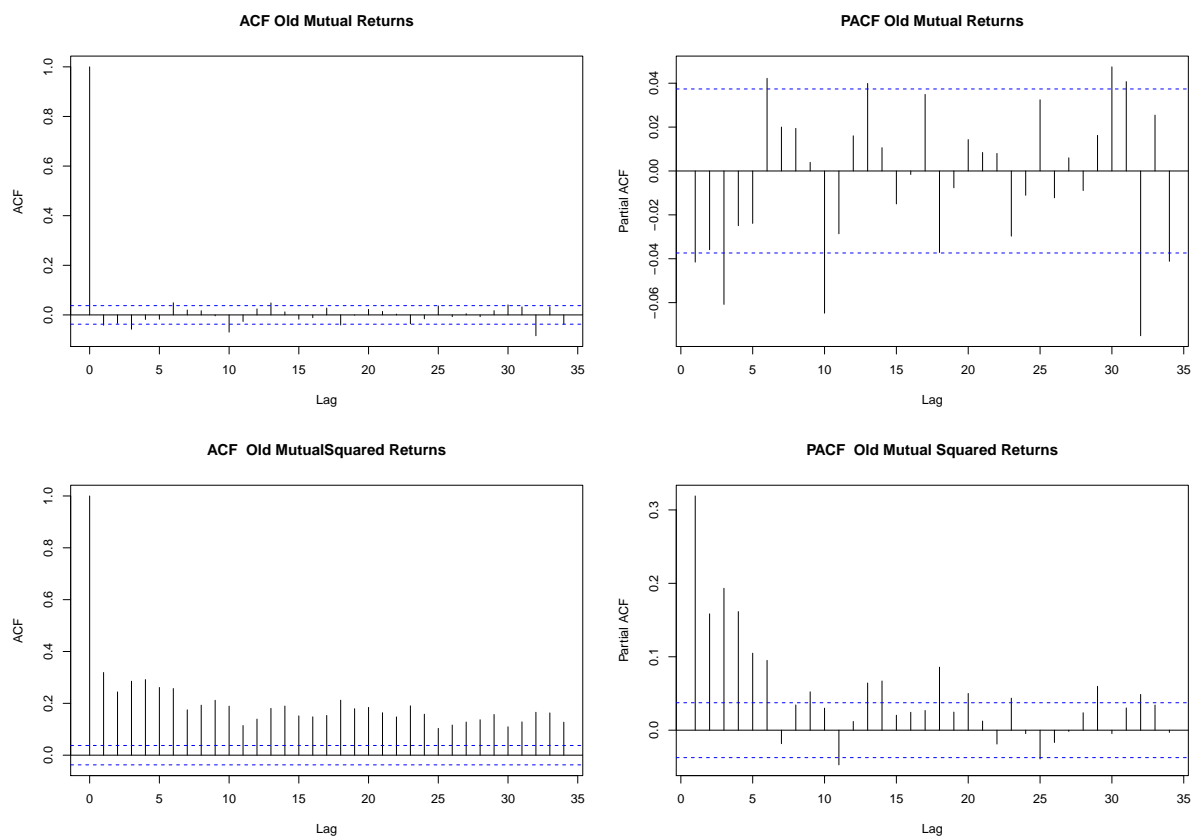


Figure 2.4: Old Mutual ACF and PACF plots

The ACF and PACF of the Standard Bank returns and squared returns are shown in Figure 2.5. For the first few lags, the returns ACF shows the presence of autocorrelations, albeit rather small. LjungBox statistics show that these autocorrelations are not significant. The squared returns are as expected strongly autocorrelated. Together the ACF and PACF suggest the series to be random. The squared return shows slow decay of the lag plots, this indicates there is correlation between the magnitude of change in the return. This means there is serial dependence in the variance of the data. The PACF of the log returns suggests us fitting an AR(3) model.

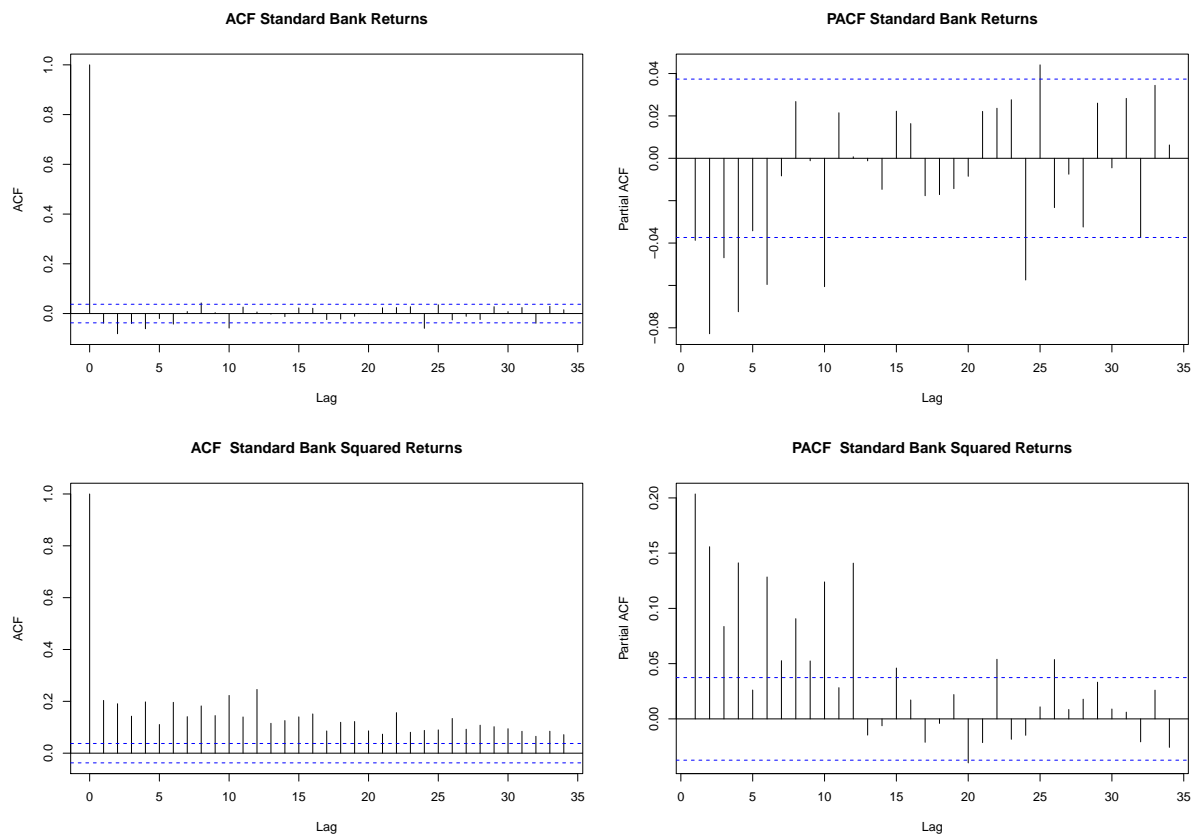


Figure 2.5: Standard Bank ACF and PACF plots

Figure 2.6 below shows the ACF and PACF's of the Liberty Holding returns. The returns ACF has a few significant low order lags meaning there is some correlation in the series. The PACF has significant lags also. The PACF of the squared returns has significant spikes hence, showing the presence of ARCH effects.

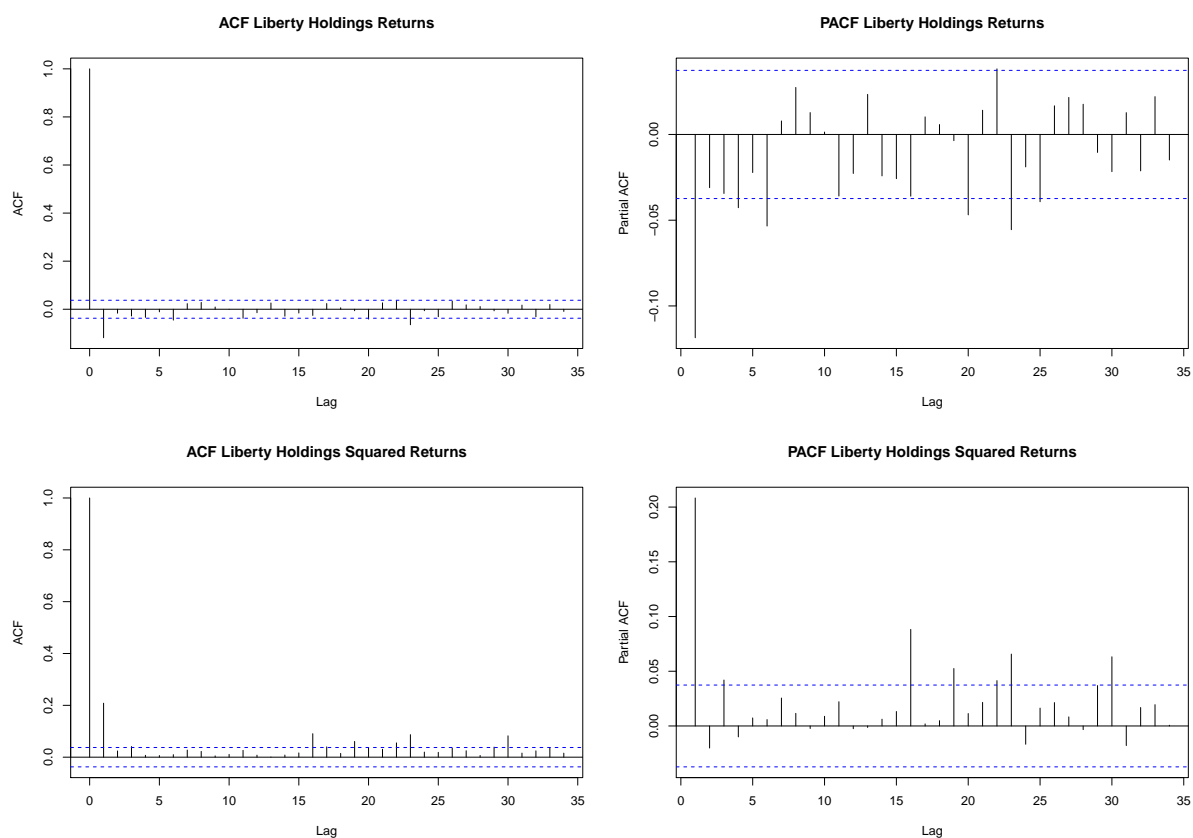


Figure 2.6: Liberty Holdings ACF and PACF plots

Descriptive Statistics of the Daily Data

Table 2.1: Descriptive statistics for daily data of the 3 companies

Stat	Old Mutual	Standard Bank	Liberty Holdings
nobs	2750	2750	2750
Minimum	-0.1731	-0.1047	-0.1336
Maximum	0.1458	0.1042	0.2243
Mean	0.0003	0.0005	0.00032
median	0.0000	0.00015	0.0000
Variance	0.0005	0.00034	0.00029
Stdev	0.0220	0.0185	0.0170
Skewness	-0.2993	0.0967	0.7927
Kurtosis	7.7725	2.5289	20.3092

Table 2.1 above shows basic statistics of the log returns as processed from R. These stats clearly show that the data is not normally distributed as expected. For Standard Bank and Liberty Holdings data is positively skewed but negative for Old Mutual. All, however, have very high excess kurtosis especially the Liberty were it is 20.3 which is almost thrice of Old Mutual and 8 times that of Standard Bank. This shows that the data has heavy

tails. It is also, important to note that R gives the excess kurtosis and not the simple kurtosis, hence, results shown in Table 2.1 are thus, excess kurtosis. Non-zero skewness shows that the distribution is skewed, unlike the normal distribution. To justify further the test for normality Jarque-Bera test is used and it gives p values < 0.05 , hence, the normality for log-return is rejected.

2.2 Monthly Data

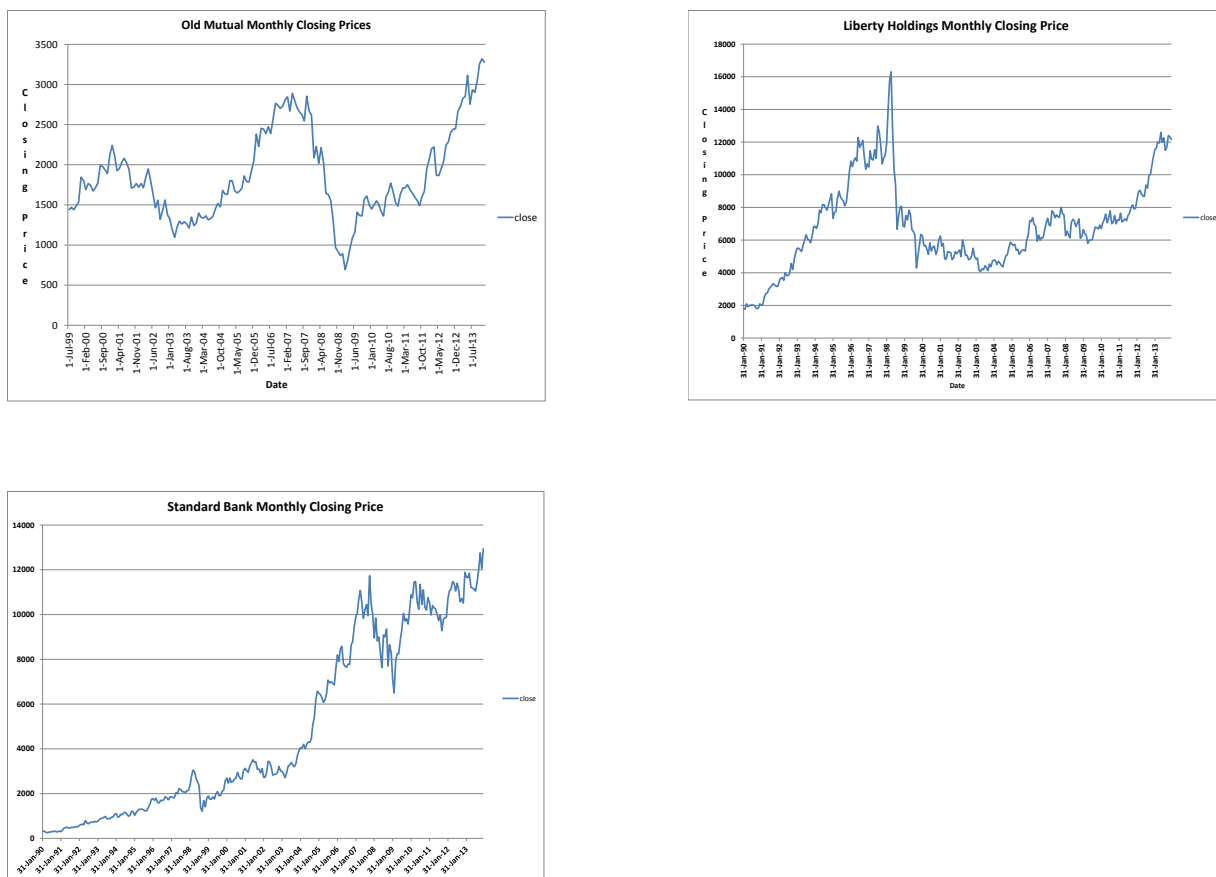


Figure 2.7: Monthly Closing Prices For The 3 Companies

The monthly closing prices are shown above in Figure 2.7. It can be seen that the prices for all assets are generally rising in the long run. A major decline is noted around 2008 which corresponds to the global recession. These sharp distinctive upward and downwards trends show that all series are not stationary. Old Mutual closing prices have two periods of slumping that is the 2008 one which is common across all assets and another one around 2003 a period which decline might have been because of own company policies. Standard Bank monthly closing prices from Jan 1990 to Dec 2013 have a more consistent

upward trend. Liberty Holdings monthly closing prices have a clearly defined increase from 1990 to 1998 then the prices take a very sharp fall. From then they start to rise slowly with the common decline in 2008 then gets back on an upward trend. Altogether for all Companies there is a general upward trend. A slight fall in prices is noted around 2008 which coincides with the global recession.

The return series is given in Figure 2.8 this series is stationary as expected from a return series. Volatility clustering can also be seen in the return series. The squared log return series follows in Figure 2.9 and it also shows volatility clustering.

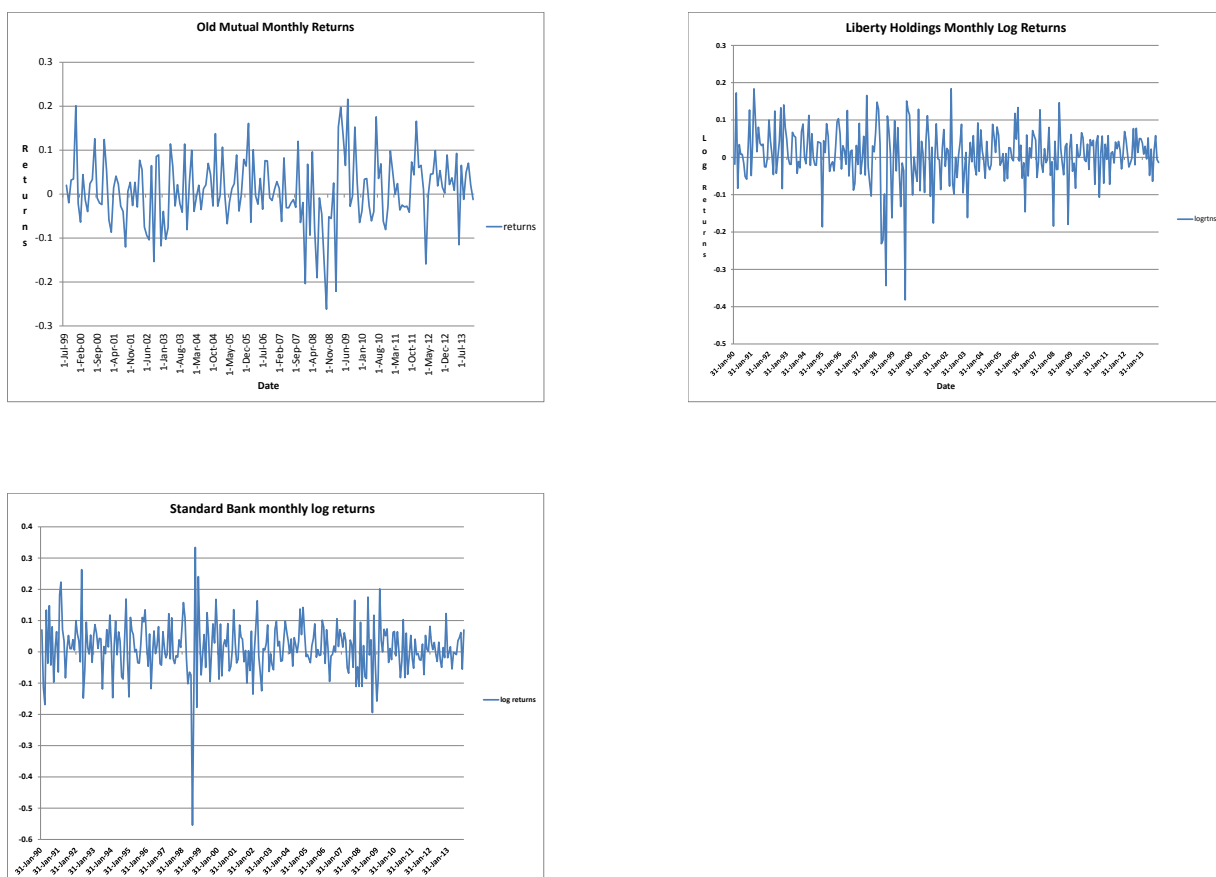


Figure 2.8: Monthly Return Series For The 3 Companies

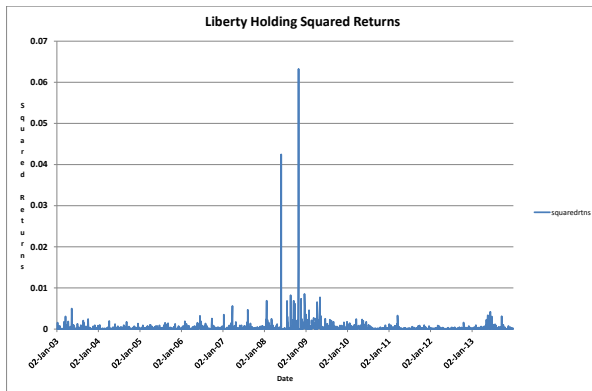
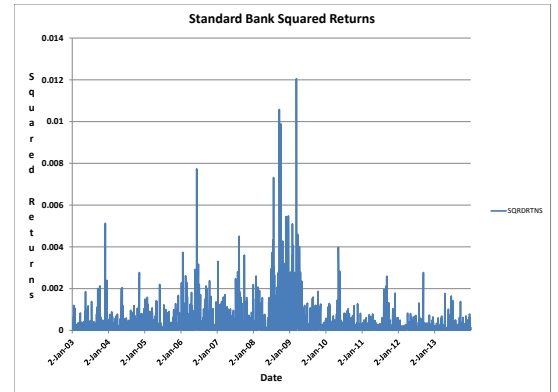
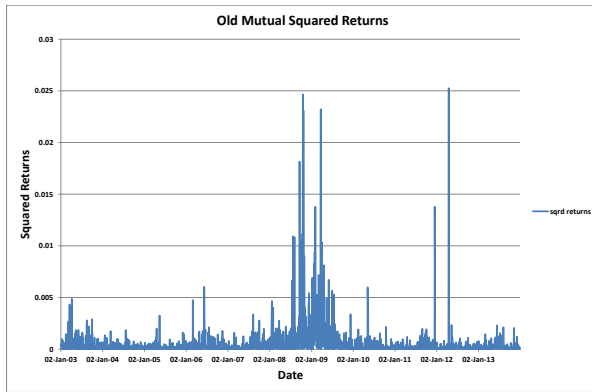


Figure 2.9: Monthly Squared Return Series For The 3 Companies

The ACF and PACF of Old Mutual monthly returns are shown in Figure 2.10 below. The ACF has no significant lags as are all within their two standard error limits, indicating that they are not significantly different from zero at the 5% level. This feature is common in monthly returns. An explanation may be that after one month a large enough period will have passed hence effects of previous return is not as great on the current. ACF of the squared returns has lags that are significant this shows dependence in series. The sample PACF of the squared series shows some big spikes, especially at lags 4 and 5 this suggests presence of ARCH effects.

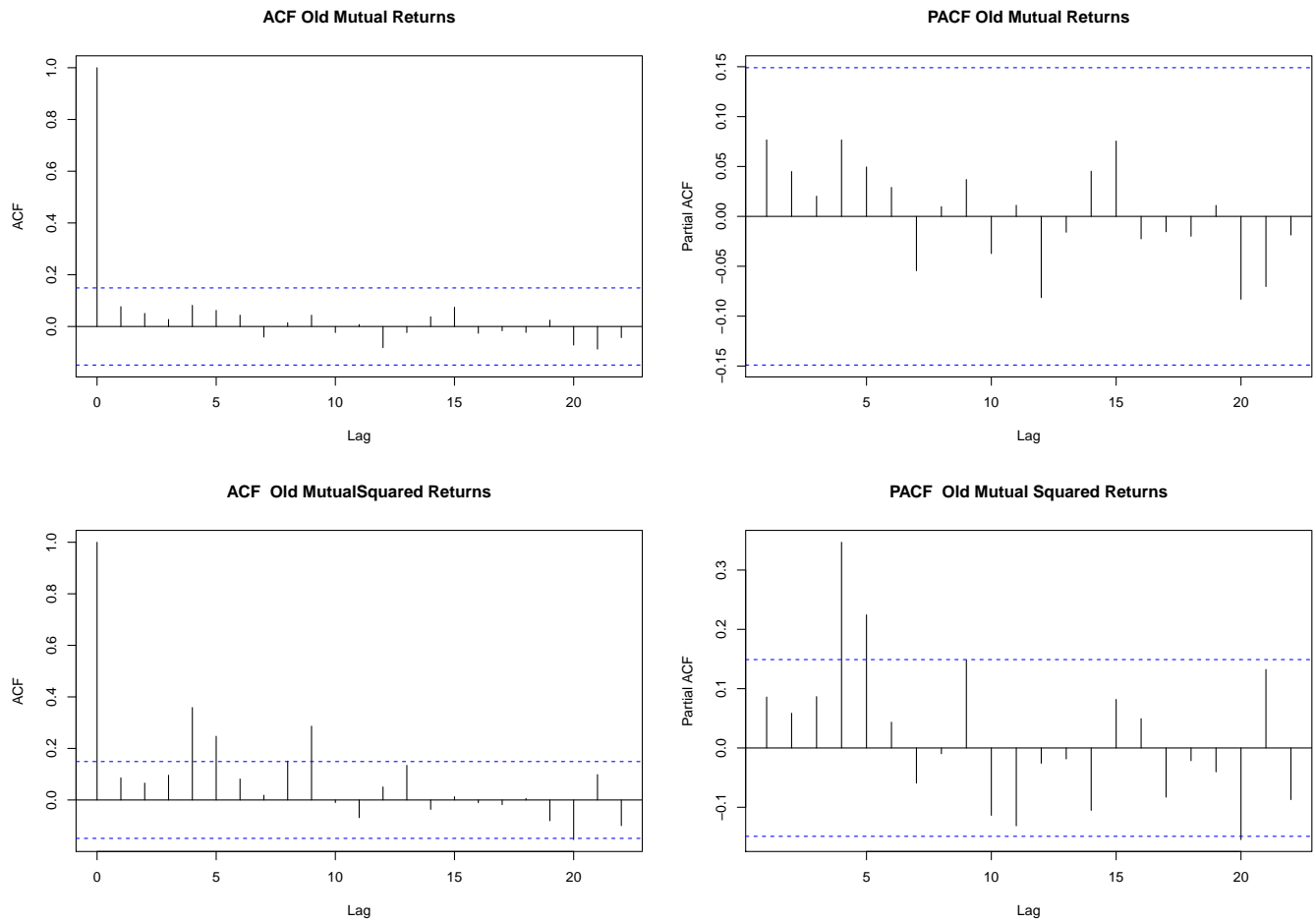


Figure 2.10: ACF and PACF of Old Mutual

The ACF and PACF plots for Standard Bank log returns are shown below in Figure 2.11. The ACF has non-significant lags except only one on lag 5 which is slightly significant otherwise more or less a white noise series. The PACF however, has significant autocorrelations. The PACF of squared returns has lags that are significant this shows dependence in series. Also, the PACF of all the squared returns shows significant lags this suggests presence of ARCH effects. As expected from the ACF and PACF plots we see that the returns are serial uncorrelated but dependent.

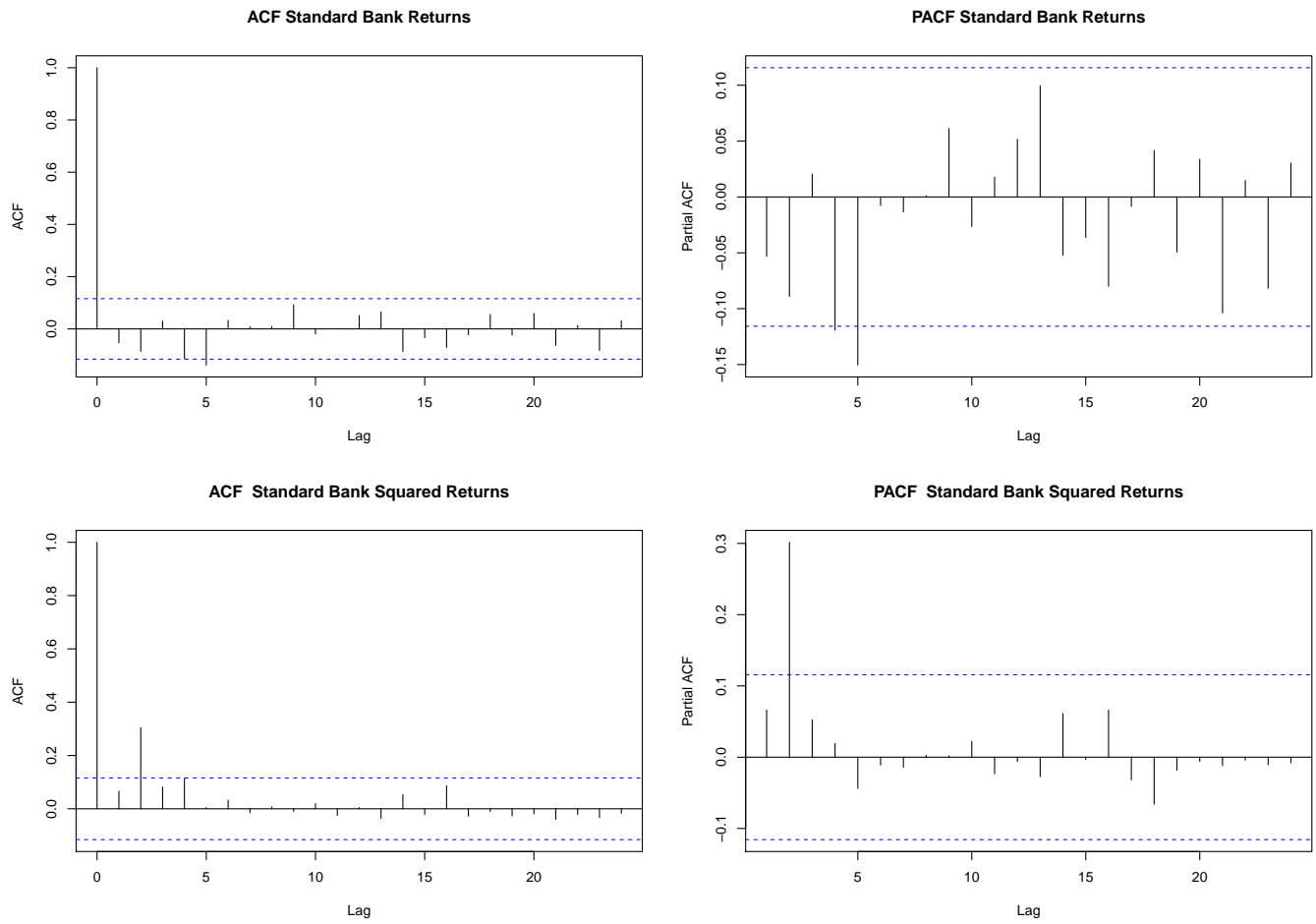


Figure 2.11: ACF and PACF of Standard Bank

Figure 2.12 below shows plots of the ACF and PACF of Liberty Holdings monthly log return and squared log return series. The ACF of the log returns has no correlations as all lags fall within bounds. The PACF however, has significant correlations which justify presence of volatility clustering. The ACF of the squared return series is correlated this shows there is dependence in series. The squared log returns PACF has significant lags which shows that ARCH effects are present.

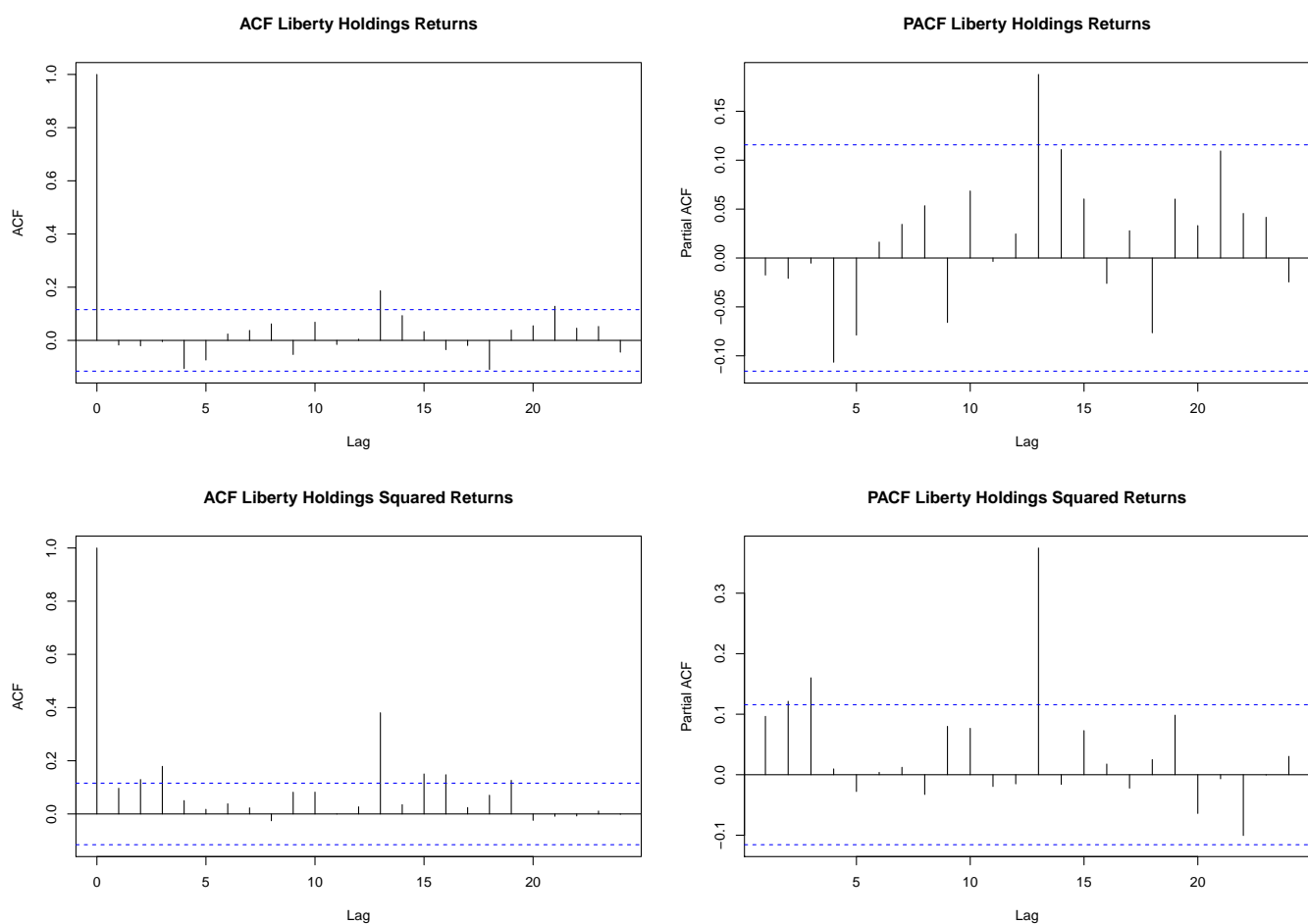


Figure 2.12: ACF and PACF of Liberty Holdings

Descriptive Statistics of the Monthly data

Table 2.2: Descriptive Statistics for Daily data of the 3 Companies

Stat	Old Mutual	Standard Bank	Liberty Holdings
nobs	173	287	287
Minimum	-0.3035	-0.5543	-0.3820
Maximum	0.1955	0.3341	0.1837
Mean	0.00476	0.0131	0.0067
median	0.0085	0.0117	0.0098
Variance	0.00625	0.00686	0.00556
Stdev	0.0790	0.0828	0.0745
Skewness	-0.5807	-0.7918	-1.0414
Kurtosis	1.5669	7.6758	4.0125

Table 2.2 shows the basic statistics of the monthly log returns. Old Mutual has 173 observations unlike others with 287 because for Old Mutual the historical data available starts

only from the year 1999. From the table, it is clear the data is not normal Distributed as seen by the negative skewness and positive excess kurtosis for all assets.

2.3 Summary

Throughout this chapter, we were exploring the properties of our different data sets which were taken from a similar time period. It was clear that all data were non-normal, with all data having positive excess kurtosis thus, showing that the distribution is heavy tailed. This is consistent with the well-known theory of distribution of financial data that we discussed earlier. We made use of the sample ACF to capture the linear dynamic of the data. It was evident that the ACF of monthly data is white noise i.e uncorrelated but that of daily returns were correlated. The PACF of all the squared returns showed significant lags and this suggested presence of ARCH effects.

Chapter 3

ARCH and GARCH Models

In perhaps, one of his most seminar papers Engle (1982) begins by saying “The great workhorse of applied econometrics is the least squares model. This is a natural choice because applied econometricians are typically called upon to determine how much one variable will change in response to a change in some other variable. Increasingly, however, econometricians are being asked to forecast and analyze the size of the errors of the model. In this case, the questions are about volatility and the standard tools have become the ARCH/ GARCH models”. Campbell et al. (1997) cement this by arguing that: “it is both logically inconsistent and statistically inefficient to use volatility measures that are based on the assumption of constant volatility over some period when the resulting series moves through time.” Decades later these statements still hold true as not only econometricians but, also, finance people are more interested in the study of volatility. For this, they still use the traditional ARCH and GARCH models by Engle (1982). This section introduces the ARCH and GARCH model and its variants.

3.1 Model Structure

Following the procedure given by Tsay (2005) for model structure. Let r_t be the log of the return of an asset at time t . The basic idea behind volatility study is that the series r_t is either serially uncorrelated or with minor lower order serial correlations but, it is a dependent series. The volatility models attempt to capture such dependence in the return series. To study these models we consider the conditional mean and variance of r_t given F_{t-1} that is,
 $\mu_t = E(r_t|F_{t-1})$, $\sigma^2 = \text{Var}(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}]$ where F_{t-1} denotes the information set available at time $t - 1$. Typically, F_{t-1} consists of all linear functions of the past returns. The volatility models we are looking at are concerned with the evolution of σ^2 , with the manner under which σ^2 evolves distinguishes one model from another.

3.1.1 Model Building

Tsay (2005) gives the following four steps for ARCH and GARCH model building :

- Specify a mean equation by testing for serial dependence in the data and, if necessary, building an econometric model (e.g., an ARMA model) for the return series

to remove any linear dependence.

- Use the residuals of the mean equation to test for ARCH effects.
- Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
- Check the fitted model carefully and refine it, if necessary, that is change the parameters in the model either by removing the non-significant ones or try another model.

3.1.2 Testing for ARCH effects

For ease in notation, let $a_t = r_t - \mu_t$ be the residuals of the mean equation. The squared series a_t^2 is then used to check for conditional heteroscedasticity, which is also, known as the ARCH effects. Two tests are available that is the LjungBox test and the Lagrange multiplier test of Engle (1982).

LjungBox test

Box and Pierce (1970) developed the Portmanteau statistic given by

$$Q * (m) = T \sum_{\ell=1}^m \hat{\rho}_{\ell}^2 \quad (3.1)$$

as a test statistic for the null hypothesis $H_0 : \rho_1 = \dots = \rho_m = 0$ against the alternative hypothesis $H_a : \rho_i \neq 0$ for some $i \in \{1, \dots, m\}$.

Ljung and Box (1978) modify the $Q^*(m)$ statistic to increase the power of the test in finite samples. This gives what is known as the LjungBox statistic $Q(m)$

$$Q(m) = T(T + 2) \sum_{\ell=1}^m \frac{\hat{\rho}_{\ell}^2}{T - \ell} \quad (3.2)$$

The decision rule is to reject H_0 if $Q(m) > \chi_{\alpha}^2$ where α denotes the $100(1 - \alpha)$ th percentile of a χ^2 distribution with m degrees of freedom.

Lagrange Multiplier Test

Lagrange Multiplier test was developed by Engle (1982) and is equivalent to the usual F statistic for testing $\alpha_i = 0$ ($i = 1, \dots, m$) in the linear regression

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_m a_{t-m}^2 + \epsilon_t, \quad t = m + 1, \dots, T,$$

where ϵ_t denotes the error term, m is a prespecified positive integer and T is the sample size. Specifically, the null hypothesis is $H_0: \alpha_0 = \dots = \alpha_m = 0$.

Let $SSR_0 = \sum_{t=m+1}^T (a_t^2 - \bar{a}^2)$, where \bar{a}^2 is the sample mean of a_t^2 and $SSR_1 = \sum_{t=m+1}^T \hat{\epsilon}_t^2$ where $\hat{\epsilon}_t^2$ is the least squares residual of the prior linear regression. Then we have

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)} \quad (3.3)$$

which is asymptotically distributed as a χ^2 distribution with m degrees of freedom under the null hypothesis. The decision rule is to reject the null hypothesis if $F > \chi_m^2(\alpha)$ where $\chi_m^2(\alpha)$ denotes the upper $100(1 - \alpha)$ th percentile of χ_m^2 , or the p-value of F is less than α (Tsay 2005).

3.2 ARCH Model

A basic feature of asset return data is that large returns tend to be followed by similarly large returns and small ones followed by small ones in both cases, all of either sign. This tells us that volatility is serially correlated. However, for most conventional econometric models the conditional variance does not depend upon the past a_{t-1} . This means there is a general loss of information in the realised forecasts. Engle (1982) proposed the first model with a systematic framework to deal with the serial correlations. He called this model the Auto Regressive Conditional Heteroscedastic Model (ARCH). The variance of this one-period forecast is given by $V(a_t|a_{t-1})$. In this model, such an expression recognizes that the conditional forecast variance depends upon past information and may, therefore, be a random variable. The basic idea of ARCH models is as follows, by letting a_t be the mean correlated return or shock of an asset return which is serially uncorrelated but dependent. The dependence of a_t can then be described by a simple quadratic function of its lagged values. An ARCH(m) model assumes that

$$a_t = \sigma_t \epsilon_t \quad (3.4)$$

$$\sigma^2 = \alpha_0 + \alpha_1 a_t^2 + \dots + \alpha_m a_{t-m}^2 \quad (3.5)$$

where ϵ_t is a sequence of independent and identically distributed (i.i.d) random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i > 0$ for $i > 0$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of a_t is finite. Under the ARCH framework, large shocks tend to be followed by another large shock this can be seen by the model where we see that shock a_t has σ_t which has past squared shocks a_{t-1}^2 from the conditional variance σ_t^2 . Consequently, a_t tends to assume a large value in modulus (Tsay 2005).

3.2.1 Properties of ARCH Models

Tsay (2005) explains that ARCH(m) model properties are best demonstrated through the ARCH(1,1) model as follows:

$$a_t = \sigma_t \epsilon_t \quad (3.6)$$

$$\sigma^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \quad (3.7)$$

where α_0 and α_1 are unknown parameters.

Let A_{t-1} be the information available at time $t-1$ then the unconditional mean is given by $E(a_t) = E(E(a_t|A_{t-1})) = E(\sigma_t E(\epsilon_t)) = 0$ since $E(a_t|A_{t-1}) = 0$ $Cov(a_{t+h}, a_t) =$

$E(a_{t+h}, a_t) = 0$ it is therefore a martingale difference which immediately means a_t has zero mean and is an uncorrelated sequence. The series is however, not independent with dependence relating to the dependence of the conditional variance on past observations Talke (2003).

The unconditional variance of a_t is obtained as follows

$$\begin{aligned} Var(a_t) &= E(a_t^2) = E(E(a_t^2|A_{t-1})) \\ &= E(\alpha_0 + \alpha_1 a_{t-1}^2) = \alpha_0 + \alpha_1 E(a_{t-1}^2) \end{aligned} \quad (3.8)$$

But, a_t is a stationary process with $E(a_t) = 0$, implying $Var(a_t) = Var(a_{t-1}) = E(a_{t-1}^2)$. Therefore, we have $Var(a_t) = \alpha_0 + \alpha_1 Var(a_t)$ and $Var(a_t) = \frac{\alpha_0}{1 - \alpha_1}$. Since variance is positive then we require $0 \leq \alpha_1 < 1$. Higher order moments are also required to exist here α_1 must satisfy other conditions. It follows from $\epsilon_t \sim N(0, 1)$ that all odd moments are zero. Hence, for third moment which defines skewness it follows that

$$\begin{aligned} E[(a_t - E(a_t))^3] &= E[a_t^3] = E[\sigma_t^3 \epsilon_t^3] \\ &= E[\sigma_t^3 E(\epsilon_t^3|A_{t-1})] \\ &= 0 \end{aligned} \quad (3.9)$$

The skewness coefficient for a_t , defined as

$$\frac{E[(a_t - E(a_t))^3]}{Var(a_t)^{\frac{3}{2}}} \quad (3.10)$$

automatically being zero and, therefore, the unconditional distribution of a_t is symmetric (Talke 2003). Also, of interest is the fat tail distribution where we look at the fourth moment of a_t which must be finite. Under normality assumption of ϵ_t we have

$$E(a_t^4|A_{t-1}) = 3[E(a_t|A_{t-1})]^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2 \quad (3.11)$$

Therefore,

$$E(a_t^4) = E[E(a_t^4|A_{t-1})] = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4)$$

Since a_t is fourth order stationary letting $m_4 = E(a_t^4)$ then we have

$$m_4 = 3E[\alpha_0^2 + 2\alpha_0\alpha_1 Var(a_t) + \alpha_1^2 m_4] = 3\alpha_0^2(1 + 2\frac{\alpha_1}{1-\alpha_1}) + 3\alpha_1^2 m_4$$

Hence

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \quad (3.12)$$

Since the fourth order moment should be positive α_1 must satisfy the condition that $1 - 3\alpha_1^2 > 0$ that is $0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$.

The unconditional kurtosis denoted k of a_t is

$$\begin{aligned}
k &= \frac{E(a_t^4)}{[Var(a_t)]^2} \\
&= \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2} \\
&= 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3
\end{aligned} \tag{3.13}$$

The excess kurtosis $k-3$ is thus, positive and the tail distribution of a_t is heavier than that of a normal distribution showing that the model is more likely to produce outliers than Gaussian white noise. This is in agreement with the distribution of asset returns where outliers appear more often than implied by an i.i.d sequence of normal random variates (Tsay 2005).

The properties hold for even higher ARCH models with formulas getting complicated for higher order models. We need a condition that conditional variance σ_t^2 is positive for all t . A natural way to achieve this rewrite model as

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + A'_{m,t-1} \Omega A_{m,t-1} \tag{3.14}$$

Where $A_{m,t-1} = (a_{t-1}, \dots, a_{t-m})$ and Ω is an $m \times m$ non-negative matrix. The ARCH(m) requires Ω to be diagonal.

3.2.2 Weaknesses of the ARCH Model

The ARCH model has the following disadvantages:

- The model assumes that both positive and negative shocks have same effects on volatility as it depends on the square of previous shocks but in reality, it is known that financial asset response is different to both shocks.
- ARCH model is restrictive in the sense that the α_0 must be in the interval $[0, \frac{1}{3}]$ if series has a finite fourth moment. These constraints limit the ability of ARCH model with Gaussian innovations ability to capture excess kurtosis.
- ARCH model provides no insight into the source of variation in the financial time series all it does is provide a way to describe the behaviour of variance but not what causes it.
- ARCH models likely to over-predict volatility because they respond slowly to large isolated shocks to return series.

3.2.3 Parameter Estimation

Under normality assumption, the likelihood function of an ARCH(m) model is

$$\begin{aligned}
f(a_1, \dots, a_t | \alpha) &= f(a_T | A_{T-1}) f(a_{T-1} | A_{T-2}) \dots f(a_{m+1} | A_m) f(a_1, \dots, a_m | \alpha) \\
&= \prod_{t=m+1}^T \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2} \times f(a_1, \dots, a_m | \alpha)\right)
\end{aligned} \tag{3.15}$$

Where $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m)'$ and $f(a_1, \dots, a_m | \alpha)$ is joint probability density function of a_1, \dots, a_m whose exact form is complicated especially when sample size is large. For this it is dropped from prior likelihood function when sample size is sufficiently large. This gives

$$f(a_{m+1}, \dots, a_T | \alpha, a_1, \dots, a_m) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\Pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right) \tag{3.16}$$

The conditional log likelihood ignoring terms with no parameters i.e 2Π is thus

$$\ell(a_{m+1}, \dots, a_T | \alpha, a_1, \dots, a_m) = - \sum_{t=m+1}^T \left[\frac{1}{2} \ln(\sigma_t^2) + \frac{1}{2} \frac{a_t^2}{\sigma_t^2} \right] \tag{3.17}$$

3.2.4 Non Normal Errors

At other times we might assume that error terms do not follow normal distribution but a heavy tailed distribution that captures all outliers standardized students t distribution. Let X be a student distribution with v degrees of freedom. Then $\text{Var}(X) = v/(v-2)$ for $v > 2$ and $\epsilon_t = X/\sqrt{v/(v-2)}$. The probability distribution function of ϵ_t is

$$f(\epsilon_t | v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\Pi}} \left(1 + \frac{\epsilon_t^2}{v-2}\right)^{-(v+1)/2} \tag{3.18}$$

Where $\Gamma(x)$ is gamma distribution. Using $a_t = \sigma_t \epsilon_t$ we get likelihood function of a_t as

$$f(a_{m+1}, \dots, a_T | \alpha, A_m) = \prod_{t=m+1}^T \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\Pi}} \frac{1}{\sigma_t} \left(1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right)^{-(v+1)/2} \tag{3.19}$$

Where $v > 2$ and $A_m = (a_1, \dots, a_m)$ estimates that maximize the prior likelihood function as the conditional MLE's under t-distribution. The degrees of freedom of the t-distribution can be specified a priori or estimated jointly with other parameters. A value between 3 and 6 is often used if it is pre-specified. If the degrees of freedom v of the Student-t distribution is pre-specified, then the conditional log likelihood function is

$$\ell(a_{m+1}, \dots, a_T | \alpha, A_m) = - \sum_{t=m+1}^T \left[\frac{v+1}{2} \ln\left(1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right) + \frac{1}{2} \ln(\sigma_t^2) \right] \tag{3.20}$$

To estimate v together with other parameters then log likelihood changes to

$$\begin{aligned}
\ell(a_{m+1}, \dots, a_T | \alpha, A_m) &= (T-m)[\ln(\Gamma((v+1)/2)) - \ln(\Gamma(v/2)) - 0.5\ln((v-2)\Pi)] + \\
&\quad \ell(a_{m+1}, \dots, a_T | \alpha, A_m)
\end{aligned} \tag{3.21}$$

3.2.5 Data Checking

First step is to find a way to specify the ARCH model. If the ARCH effect is found to be significant, Tsay (2005) states that one can use the PACF of a_t^2 to determine the ARCH order. For a given sample, a_t^2 is an unbiased estimate of σ_t^2 . Therefore, a_t^2 is expected to be linearly related to $a_{t-1}^2, \dots, a_{t-m}^2$ in a manner similar to that of an autoregressive process.

3.2.6 Model Checking

For a properly specified ARCH model, the standardized residuals $\tilde{a}_t = \frac{a_t}{\sigma_t}$ must form a sequence of i.i.d random variables. Hence, one can check the adequacy of fitted ARCH model by examining the series of \tilde{a}_t . In particular the Ljung-Box statistics of \tilde{a}_t can be used to check adequacy of mean equation with that of \tilde{a}_t^2 to test validity of the volatility equation. The skewness, kurtosis and quantile plots used to check distribution assumption validity.

3.2.7 Forecasting

Forecast values are found recursively as those of an AR model. Consider ARCH(m) model a forecast origin h, the one step ahead forecast of σ_{h+1}^2 is

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \dots + \alpha_m a_{h+1-m}^2 \quad (3.22)$$

with the two step

$$\sigma_h^2(2) = \alpha_0 + \alpha_1 a_h^2(1) + \dots + \alpha_m a_{h+2-m}^2 \quad (3.23)$$

In general the ℓ step forecast of $\sigma_{h+\ell}^2$ is

$$\sigma_h^2(\ell) = \alpha_0 + \sum_{i=1}^m \alpha_i \sigma_h^2(\ell - i) \quad (3.24)$$

where $\sigma_h^2(\ell - i) = a_{h+\ell-i}^2$ if $\ell - i \leq 0$

3.3 GARCH Model

The ARCH model often requires many parameters to adequately describe the volatility process of an asset return. As a way to model persistent movements in volatility without estimating a very large number of coefficients in a high-order polynomial $\alpha(L)$, Bollerslev (1986) suggested the Generalized Autoregressive Conditionally Heteroskedastic, or GARCH model

$$\sigma_t^2 = \alpha_0 + \alpha(L) a_{t-1}^2 + \beta(L) \sigma_{t-1}^2 \quad (3.25)$$

Campbell et al. (1997)

where $\beta(L)$ is a polynomial in the lag operator. This is called a GARCH(p,q) model when

the order of the polynomial $\beta(L)$ is p and the order of the polynomial $\alpha(L)$ is q .

Simplified we say a_t follows the GARCH model if

$$a_t = \sigma_t \epsilon_t \quad \text{where} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$$

Where ϵ_t is sequence of Gaussian white noise with mean 0 and variance 1.

$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$. Also $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. The constraint of $\alpha_i + \beta_i$ implies unconditional variance of a_t is finite whilst the conditional variance σ_t^2 evolves over time. α_i for $i = 1, 2, 3, \dots$ determines the reaction to market shocks, the parameter β_i determines the persistence in volatility after a shock and these, together with the parameter α_0 determine the speed of mean reversion and the long run GARCH volatility Alexander (2008). This further simplified by Hillebrand (2003) as follows: by letting $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) = \lambda$, then, the closer λ is to unity the more persistent the volatility is. λ represents the fraction of information that is carried forward per unit time. So based on this $1-\lambda$ is the fraction that is washed out per unit time. Hence, $\frac{1}{1-\lambda}$ is the average time required to revert back to mean when time increments equals to 1. ϵ_t can be Standard Normal or Standardized Student T or General Error Distribution. If $s=0$ the equation reduces to ARCH(m) model. α_i and β_i are ARCH and GARCH parameters respectively.

The intuition of this model is that current volatility is modelled using previous variance that is the $\beta(L)\sigma_{t-1}^2$ part and also the previously observed error $\alpha(L)a_{t-1}^2$. To easily understand the model consider $\eta_t = a_t^2 - \sigma_t^2$ so that $\sigma_t^2 = a_t^2 - \eta_t$. Plugging $\sigma_{t-1}^2 = a_{t-1}^2 - \eta_{t-1}$ ($i = 0, \dots, s$) into equation we have

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

η is a martingale difference series but it is not i.i.d in general. It is an ARMA form but squared thus ARMA can be seen as an ARMA application on squared series $a^2 - t$. Using unconditional mean of ARMA model we have

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}$$

provided the denominator is positive Tsay (2005).

3.3.1 GARCH(1,1)

To demonstrate the properties of the GARCH model, we use the simpler GARCH(1,1) model which is defined as follows

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (3.26)$$

$$0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1 \text{ (Bollerslev 1986).}$$

From the equation we see that a large a_{t-1}^2 or σ_{t-1}^2 gives rise to large σ_t^2 . Thus it means large a_{t-1}^2 tends to be followed by large a_t^2 hence, generating the behaviour of volatility clustering. The conditional mean of a_t is $E(a_t|A_{t-1}) = 0$ and a_t is a martingale difference. This means $E(a_t) = 0$ and a_t is an uncorrelated series. Information set is now $a_t, \sigma_t^2, \dots, a_{t-1}, \sigma_{t-1}^2$.

Properties easily derived by introducing the squared returns a_t^2 which is expressed as

$$a_t^2 = \sigma_t^2 + v_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (a_{t-1}^2 + v_t - \beta_1 v_{t-1}) \quad (3.27)$$

Where $v_t = a_t^2 - \sigma_t^2$. Conditional mean of v_t is 0 and it is a martingale difference therefore $E(v_t) = 0$ and $cov(v_t, v_{t-k}) = 0$ for $k \geq 1$ so that v_t is serially uncorrelated.

$$E(a_t^2) = \alpha_0 + (\alpha_1 + \beta_1)E(a_t^2) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

since ϵ_t is Gaussian white noise and $E(a_t^2) = E(a_{t-1}^2) = E(\sigma_t^2 \epsilon_t^2) = E(\sigma_t^2)$ for variance to be finite $\alpha_1 + \beta_1 < 1$ should hold Talke (2003). For kurtosis assuming 4th order stationarity i.e $E(\epsilon^4) = 3$ then

$$E(a_t^4) = E[E[\sigma_t^4 \epsilon_t^4 | A_{t-1}]] = E[\sigma_t^4 E[\epsilon^4 | A_{t-1}]] = 3[E[\sigma_t^4]]$$

Now $E(\sigma_t^4) = \alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \beta_1^2 E(\sigma_{t-1}^4) + 2\alpha_0 \alpha_1 E(a_{t-1}^2) + 2\alpha_0 \beta_1 E(\sigma_{t-1}^2) + 2\alpha_1 \beta_1 E(\sigma_{t-1}^2 a_{t-1}^2)$ but

$$\begin{aligned} E(\sigma_{t-1}^2 a_{t-1}^2) &= E[E(\sigma_{t-1}^2 a_{t-1}^2 | A_{t-2})] \\ &= E[\sigma_{t-1}^4 E(\epsilon_{t-1}^2) | A_{t-2}] \\ &= E(\sigma_{t-1}^4) \end{aligned} \quad (3.28)$$

Rearranging and noting that $E(a_{t-1}^2) = E(\sigma_{t-1}^2) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$ and

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

$$\text{Thus } E(\sigma^4) = \frac{\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - (\alpha_1 + \beta_1)) [(1 - 2\alpha_1^2) - (\alpha_1 + \beta_1)^2]}$$

$$\text{But, we know } E(a_t^4) = 3E(\sigma_t^4) \text{ hence, we have } E(a_t^4) = \frac{3\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - (\alpha_1 + \beta_1)) [(1 - 2\alpha_1^2) - (\alpha_1 + \beta_1)^2]}$$

The kurtosis is then obtained as

$$k = \frac{E(a_t^4)}{[var(a_t)]^2} = \frac{3(1 - \alpha_1 + \beta_1)^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3 \quad (3.29)$$

given that $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$

Thus, the GARCH(1,1) has a heavy tail compared to the normal distribution. Also, the model provides a simple parametric function that can be used to describe the volatility evolution.

3.4 Forecasting in the GARCH(1,1) Model

Considering the GARCH(1,1) model and that the origin is at h then 1 step ahead forecast is given by

$$\sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \quad (3.30)$$

Where a_h and σ_h^2 are known at time index h hence, we have

$$\sigma_1^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \quad (3.31)$$

Multi-step ahead forecast are given using the from that $a_t^2 = \sigma_t^2 \epsilon_t^2$ and rewriting the volatility equation as

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 + \alpha_1 \sigma_t^2 (\epsilon_t^2 - 1) \quad (3.32)$$

With $t=h+1$ then we have

$$\sigma_{h+2}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2 (\epsilon_{h+1}^2 - 1) \quad (3.33)$$

Since $E(\epsilon_{h+1}^2 - 1 | A_h) = 0$ the 2 step ahead volatility forecast at origin h satisfies

$$\sigma_h^2(2) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2(1) \quad (3.34)$$

In general we have

$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2(\ell - 1) \quad (3.35)$$

$\ell > 1$

3.4.1 Persistence and Stationarity

As seen earlier that when $\alpha + \beta < 1$, the unconditional variance of a_{t+1}^2 , or equivalently the unconditional expectation of σ^2 , is $\frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$ Campbell et al. (1997) shows that by Recursively substituting in (3.31), and using the law of iterated expectations, the conditional expectation of volatility j periods ahead is

$$E(\sigma_{t+j}^2) = (\alpha + \beta < 1)^j \left(\sigma^2 - \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \right) + \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \quad (3.36)$$

The multiperiod volatility forecast reverts to its unconditional mean at rate $\alpha + \beta$.

3.5 IGARCH

When $\alpha + \beta = 1$, the conditional expectation of volatility j periods ahead is instead

$$E(\sigma_{t+j}^2) = \sigma_t^2 + j\alpha_0 \quad (3.37)$$

This implies that today's volatility affects forecasts of volatility into the indefinite future, this feature is termed persistence. Engle and Bollerslev (1986) called this model the Integrated GARCH (IGARCH) model.

Campbell et al. (1997) points out the fact the IGARCH(1,1) process for σ_t^2 looks very much like a linear random walk with drift for $\alpha_0 > 0$ and that without drift for $\alpha_0 = 0$. But, Nelson (1990) is quick to warn that despite such a close appearance the behaviour of the two differs. The structure of the higher moments of σ_t^2 when $\alpha + \beta = 1$ and $\alpha_0 = 0$ implies that the distribution of σ_t^2 becomes more and more concentrated around zero with fatter and fatter tails, which is not the case with a random walk. This special IGARCH(1,1) model is the volatility model used in JP Morgan (1995), which is an approach for calculating the value at risk. In the IGARCH(1,1) model with $\alpha_0 > 0$, σ_t^2 is strictly stationary and ergodic. The innovation a_{t+1} then has a stationary distribution with a zero mean, but with tails that are so thick that no second or higher-order moments exist.

The IGARCH model for the simple IGARCH(1,1) is as shown by Tsay (2005)

$$a_t = \sigma_t \epsilon_t \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) a_{t-1}^2$$

Where a_t is white noise $1 > \beta_1 > 0$. The difference between the GARCH(1,1) and the IGARCH is that the unconditional variance of a_t hence, that of r_t is undefined under the IGARCH(1,1) model. This phenomenon is hard to justify on return series but from theoretical viewpoint it may be caused by occasional level shifts in volatility Tsay (2005).

3.6 Forecasting IGARCH Models

Given that $\alpha_1 + \beta_1 = 1$ then by repeated substitution into general ℓ step ahead forecast of GARCH(1,1) model i.e $\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(\ell - 1)$, $\ell > 1$ then we get

$$\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0 \tag{3.38}$$

$\ell > 1$

Where h is the forecast origin JP Morgan (1995). From above we can see that the effect of $\sigma_h^2(1)$ is persistent on all future volatilities hence, the volatility forecast form a straight line with slope α_0 . Under certain conditions the volatility process is strictly stationary but not weakly stationary as it lacks the first two moments. For $\alpha_0 = 0$ a special IGARCH arises the volatility forecast are just $\sigma_h^2(1)$ for all forecast horizons. This is the volatility model used in RiskMetrics an approach to calculate value at risk.

Francq and Zakoian (2011) for IGARCH(p,q) when $\sum \alpha_i + \sum \beta_i = 1$. Then the ℓ step ahead forecast is

$$\sigma_{h+\ell}^2 = \sigma_h^2 + (\ell - 1)\alpha_0 \tag{3.39}$$

3.7 GARCH-M

Engle et al. (1987) suggest adding a time-varying intercept to the basic univariate model. This resulted in the ARCH-M which was developed into GARCH-M. In finance, the return

of a security can depend on its volatility. (Johnston and Scot, 1999) mention that, one way to account for this is to write the return as a function of the conditional variance. To model this one may consider the GARCH-M model where M is mean. GARCH-M model allows for the conditional variance to have a mean effect. The model is expressed like any GARCH(p,q) but the mean equation changes to

$$r_t = \mu + c\sigma_t^2 + a_t \quad (3.40)$$

Hence, for GARCH(1,1)-M we add $a_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

Where r_t is the logarithm of the ratio of the closing exchange rate prices at time t and t-1 μ and c are constants, parameter c is the risk premium parameter, with a positive c indicating a return that is positively related to its volatility. The formulation of the model implies there exist serial correlations in the return series r_t . These are introduced by those in the volatility process σ_t^2 , hence, existence of risk premiums are another reason that some historical stock prices have serial correlations (Tsay 2005).

3.8 Exponential GARCH Model

The exponential GARCH model was Developed by Nelson (1991) to overcome some of the weaknesses of the GARCH model in handling financial time series. The GARCH model deals with the magnitude of the residuals but does not allow for asymmetric effects between positive and negative asset returns. Statistically, this effect occurs when an unexpected drop in price (bad news) increases predictable volatility more than an unexpected increase in price (good news) of similar magnitude. Nelson (1991) proposed the exponential GARCH model(EGARCH) in which the main focus on volatility depends not only on magnitude but also, the sign of the residuals. The model as given by Nelson (1991) is

$$g(\epsilon_t) = \theta \epsilon_t + \gamma [|\epsilon_t| - E|\epsilon_t|] \quad (3.41)$$

where θ and λ are real constants.

Both ϵ_t and $|\epsilon_t| - E|\epsilon_t|$ are zero-mean i.i.d sequences with continuous distributions.

The asymmetry of $g(\epsilon_t)$ can easily be seen by rewriting it as

$$\begin{aligned} g(\epsilon_t) &= (\theta + \gamma)\epsilon_t & \text{if } \epsilon_t > 0 \\ &= (\theta + \gamma)\epsilon_t & \text{if } \epsilon_t < 0 \end{aligned} \quad (3.42)$$

An EGARCH(m,s) model can be written as

$$a_t = \sigma_t \epsilon_t \quad (3.43)$$

$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \dots + \beta_{s-1} B^{s-1}}{1 - \alpha_1 B - \dots - \alpha_m B^m} g(\epsilon_{t-1}) \quad (3.44)$$

where α_0 is a constant, B is the back-shift (or lag) operator. (Tsay 2005) explains that it uses logged conditional variance to relax the positiveness constraint of model coefficients.

Secondly, the use of $g(\epsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of a_t . Campbell et al. (1997) further points out that this model is appealing because it does not require any parameter restrictions to ensure that the conditional variance of the return is always positive. Also, it becomes both strictly non-stationary and covariance non-stationary when $\alpha_1 + \beta_1 = 1$, so it does not share the unusual statistical properties of the IGARCH (1,1) model. On the other hand, multi-period forecasts of future variances are harder to calculate in the EGARCH model and no closed-form expressions are available.

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^s \alpha_i g(\epsilon_{t-1}) + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2) \quad (3.45)$$

3.9 Model selection using Information Criteria

Several information criterion have been developed to help with model selection. Here we discuss the Akaike information criterion (AIC) and SchwarzBayesian information criterion (BIC). The (AIC) is the most common information criteria it is defined as

$$AIC = \frac{-2}{T} \ln(\text{likelihood}) + \frac{2}{T} (\text{number of parameters}) \quad (3.46)$$

where the likelihood function is evaluated at the maximum-likelihood estimates and T is the sample size. The equation reduces for an a Gaussian $AR(\ell)$ to

$$AIC(\ell) = \ln(\sigma_\ell^2) + \frac{2\ell}{T} \quad (3.47)$$

The first term measures the goodness of fit whilst the second term is the penalty function of the criterion. It is called so because it penalizes a candidate model by the number of parameters used. Therefore the more parameters the higher the penalty.

The BIC is given as

$$BIC(\ell) = \ln(\sigma_\ell^2) + \frac{2\ell \ln(T)}{T} \quad (3.48)$$

From the 2 equation we can see that the difference in the information criteria is their penalty functions.

Given 2 or more models the best model is the one with the lowest AIC or BIC value.

3.10 News Impact Curve

With many different volatility models, another way to differentiate their performance is by looking at how they react to shock i.e news. Engle and Ng (1993) introduce the news impact curve. This measures the effect of a shock in the current period on the conditional variance (volatility) in the subsequent period to facilitate comparison between models. By holding constant the information dated $t - 2$ and earlier, we can examine the implied relation between ϵ_{t-1} and h_t . In the GARCH model, this curve is a quadratic

function centered on $\epsilon_{t-1} = 0$. For the EGARCH, it has its minimum at $\epsilon_{t-1} = 0$ and is exponentially increasing in both directions but with different parameters. This means for the GARCH it treats all news bad or good similarly but EGARCH treats them differently.

Diagnostics Using News Impact Curve

The standard GARCH model has a news impact curve which is symmetric and centered at $\epsilon_{t-1} = 0$. That is, positive and negative return shocks of the same magnitude produce the same amount of volatility. If a negative return shock causes more volatility than a positive return shock of the same size, the GARCH model underpredicts the amount of volatility following bad news and overpredicts the amount of volatility following good news. Furthermore, if large return shocks cause more volatility than a quadratic function allows, then the standard GARCH model underpredicts volatility after a large return shock and overpredicts volatility after a small return shock (Engle and Ng 1993). With these observations Engle and Ng (1993) introduce the Sign Bias Test, the Negative Size Bias Test, and the Positive Size Bias Test. These tests examine whether we can predict the squared normalized residual by some variables observed in the past which are not included in the volatility model being used. If these variables can predict the squared normalized residual, then the variance model is misspecified.

3.11 Summary

This chapter introduces the models to be used for univariate modelling of the data. The model structure and how they capture volatility is discussed. Model selection using AIC and BIC is also discussed. These models will be used in the next section.

Chapter 4

Model Application

The data presented in the previous chapter had the characteristics of financial data as expected. The daily log returns had low order correlations. Monthly data had very low or no correlations hence, their ACF resembled that of a white noise one. The first step before modelling is to perform standard tests to confirm presence of ARCH effects as indicated by the PACF's under data analysis section. Here autocorrelation structure of the residuals and the squared residuals can be inspected. An indication of ARCH presence is that the residuals will be uncorrelated but the squared residuals will show autocorrelation. A variety of software can be used to fit these models for example R, SAS and EVIEWS. For our case we made use of R and EVIEWS with ultimate choice being R as opposed to EVIEWS because

- It comes with many optional packages to use for a variety of models whilst EVIEWS runs only a selected few models, for example, it can not run GJRGARCH, IGARCH to name a few.
- R is free hence, it's a good practice to have it as you can use it anywhere at anytime
- R graphics are simple and precise unlike those from EVIEWS which do not clearly show detail.
- EVIEWS does not have provision for studentized t and skewed t options yet R has many options such as student t and skewed t.

For those models that could be run with EVIEWS it was clear that the estimates and p-values were consistent with the estimates from R with a .00 difference which is not significant. This difference can be brought about by rounding errors and also number of iterations in the calculation procedures used by the particular software in question. Note that omega ω in this section is the constant α_0 of the previous section. R gives the constant using the notation ω .

4.1 Old Mutual Data

The data analysis of the Old Mutual data showed evidence of volatility clustering as witnessed through looking at the plots of the returns. The plots of ACF of squared returns

also suggested presence of ARCH effects. However, to be more certain we perform the ARCH test. The ARCH test here is done using R.

4.1.1 Monthly Data

The ACF of the monthly data showed no autocorrelation, this was backed up by Ljung-Box test on a_t which gave $\chi^2 = 5.9197$, $df = 12$, $p\text{-value} = 0.9201$ hence, therefore we could easily proceed to ARCH effect test. The ACF and PACF of the squared series confirms the presence of conditional heteroscedasticity. The Ljung-Box test on the a_t^2 series gives a chi-square value = 62.7616, $df = 12$, $p\text{-value} = 7.049e-09$. Therefore we conclude that ARCH effects are present. Using the package FGARCH various models are ran and the results are shown in Table 4.1 below.

Table 4.1: Summary of fitted models:cond distribution=norm

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(1,0)	-2.2252	-2.1705	195.4793	$\mu \alpha_1$
GARCH(2,0)	-2.2200	-2.1473	196.034	$\mu \alpha_1 \alpha_2$
GARCH(3,0)	-2.22776	-2.21366	197.2008	$\mu \alpha_1 \alpha_2 \alpha_3$
GARCH(1,1)	-2.2894	-2.2165	202.0358	$\mu \omega$
GARCH(2,1)	-2.290	-2.199	203.14	$\mu \alpha_1 \alpha_2$
GARCH(1,2)	-2.278	-2.187	202.06	GARCH parameters NA
GARCH(2,2)	-2.279	-2.1698	203.14	$\alpha_1 \alpha_2 \beta_1 \beta_2$

¹ NA means value could not be evaluated

Based on the AIC the GARCH(1,1) model is the best model. Although its AIC is not lower than that of the GARCH(2,1) the difference is small and also the GARCH(2,1) parameters are not all significant. We then run the models with t distributed errors and skewed t distributed errors. The results are shown in Table 4.2 and Table 4.3 respectively.

Table 4.2: Summary of fitted models:cond distribution=std

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(1,0)	-2.2641	-2.1912	199.848	$\mu \alpha_1$
GARCH(2,0)	-2.2657	-2.1746	200.9823	$\mu \alpha_1 \alpha_2$
GARCH(3,0)	-2.2690	-2.1600	202.2667	$\mu \alpha_1 \alpha_2 \alpha_3$
GARCH(1,1)	-2.3022	-2.2110	204.138	$\mu \omega \alpha_1 \text{ shape}$
GARCH(2,1)	-2.301	-2.1914	205.0168	$\mu \alpha_1 \alpha_2$
GARCH(1,2)	-2.2902	-2.1809	204.1062	$\mu \beta_2 \text{ shape}$
GARCH(2,2)	-2.2892	-2.1616	205.0168	$\mu \alpha_1 \alpha_2 \beta_2$

Table 4.3: Summary of fitted models:cond distribution=sstd

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(1,0)	-2.2549	-2.1638	200.0482	$\mu \alpha_1$
GARCH(2,0)	-2.2560	-2.1466	201.1401	$\mu \alpha_1 \alpha_2$
GARCH(3,0)	-2.2586	-2.1310	202.3672	$\alpha_1 \alpha_2 \alpha_3$
GARCH(1,1)	-2.2914	-2.1821	204.2097	$\mu \omega \alpha_1$ skew
GARCH(2,1)	-2.2901	-2.1625	205.0937	$\mu \alpha_1 \alpha_2$
GARCH(1,2)	-2.2796	-2.1520	204.1874	$\mu \omega \alpha_1 \beta_2$ shape
GARCH(2,2)	-2.2785	-2.1327	205.0937	$\mu \alpha_1 \alpha_2 \beta_2$

After looking at all models under normal, studentized t and skewed studentized t distribution, it was clear that the GARCH(1,1) was the better model as evident by the lower AIC BIC values. The parameters of the GARCH(1,1) models from the all the error distribution assumption were not all significant at 5%. A detailed output all the GARCH(1,1) estimates as obtained from R are shown in tables 4.4, 4.5 and 4.6 for normal, t and skewed t error distribution respectively.

Table 4.4: Estimates from distribution=norm

	R Estimate	EVIEWS estimate	Pr(> t)in R	Pr(>z)in EVIEWS
mu	0.0097514	0.009761	0.0758	0.1068
omega	0.0006974	0.000697	0.0749	0.2052
alpha1	0.1199413	0.119951	0.0437	0.0891
beta1	0.7632629	0.763274	6.66e-15	0.0000

Table 4.5: Estimates from distribution=std

	R Estimate	EVIEWS estimate	Pr(>t)in R	Pr(>z)in EVIEWS
mu	0.0096322	0.009656	0.0652	0.0851
omega	0.0006760	0.000675	0.1537	0.3548
alpha1	0.1127324	0.112616	0.0856	0.2181
beta1	0.7778690	0.778203	4.18e-12	0.0000
shape	7.2973920	7.295195	0.0794	0.1748

Below we fit model with skewed t which, however, is not provided for in EVIEWS.

Table 4.6: Estimates from distribution=sstd

	Estimate	Std. Error	t value	Pr(>t)	significance code
mu	0.0089068	0.0055846	1.595	0.111	
omega	0.0006597	0.0004710	1.401	0.161	
alpha1	0.1035591	0.0645348	1.605	0.109	
beta1	0.7864759	0.1129801	6.961	3.37e-12	***
skew	0.9536481	0.1183671	8.057	8.88e-16	***
shape	7.8352421	4.9740517	1.575	0.115	

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Based on the above output we can see that despite the GARCH(1,1) models from the t and skewed t distributed errors having better AIC values their parameters are highly insignificant. The normal has estimates which are more significant. All persistence is high in all models. The obtained $(\alpha + \beta)$ values are $(0.120+0.763=0.883)$, $(0.13+0.778=0.908)$ and $(0.104+0.786=0.89)$ for the normal, t and skewed t distributed errors respectively.

Figure 4.1 shows the Q-Q plots of the GARCH(1,1). A comparison of the Q-Q plot shows t and skewed t distribution to be slightly better fit. The tails are much narrower showing that the t and skewed t dealt better with the heavy-tailed data. However, as seen before the parameter estimates of the t and skewed t models were not as significant as those of one fit using normal assumption. Also, by considering that parameter estimates are not very different we conclude that the GARCH(1,1) is acceptable in this case.

NB* Note the use of the phrase “in this case” as in the next section, extensions of the GARCH model will be tried which may fit the data. But for the sake of progress, we select the GARCH(1,1) from this section.

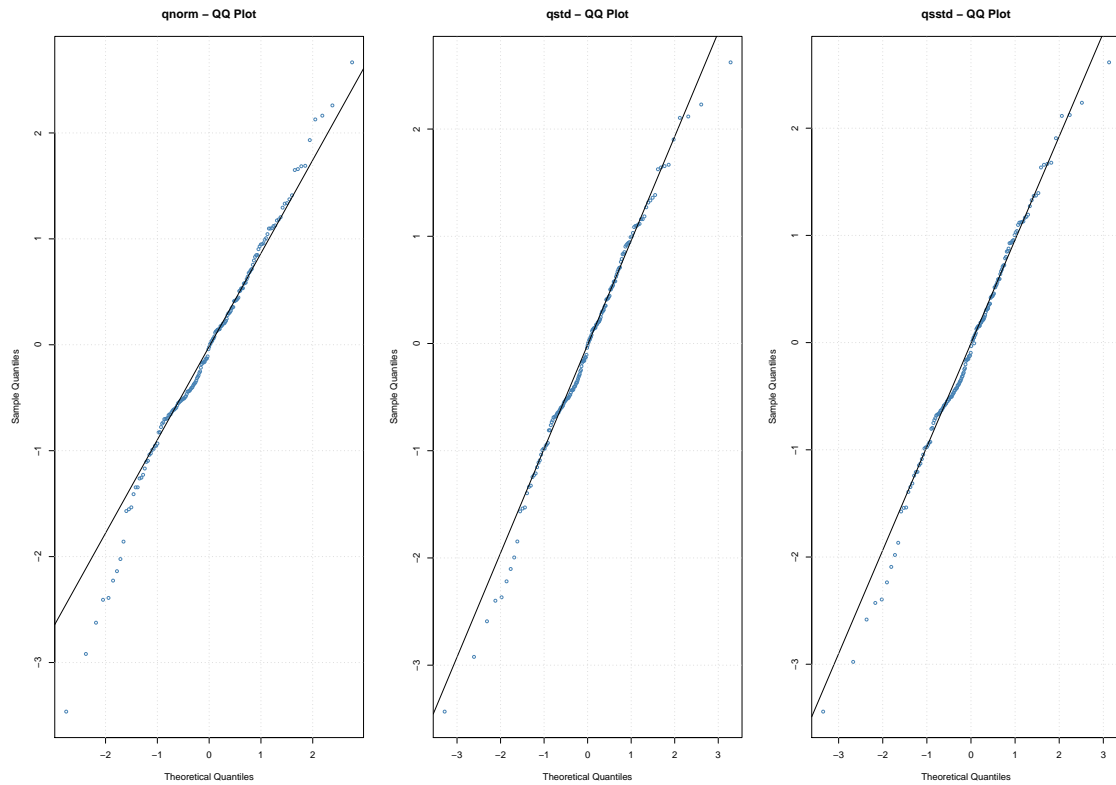


Figure 4.1: Q-Q plots

As a way of model diagnosis R also gives standardized residuals tests which are used to determine if model residuals follow the model assumptions. The output is shown in Table 4.7 below. To assess if the standardized residuals are white noise the Ljung-Box tests up to order 10; 15; 20 are reported both for standardized residuals and their squared values. As can be seen from the p-values < 0.05 nor 0.01 for all these lags. Hence, there is no evidence of autocorrelation nor of conditional heteroscedasticity presence as required. The Jarque-Bera statistic and the Shapiro-Wilk test are reported these test the normality assumption. Both tests do not reject the null hypothesis of normality at 99%. However, the values here are not highly significant hence, we expect the Q-Q plot to have minor deviations from the straight plot.

Table 4.7: Summary of Standard Residuals

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	R Chi^2	6.9312	0.0313
Shapiro-Wilk Test	R W	0.9865	0.0957
Ljung-Box Test	R Q(10)	4.9478	0.8946
Ljung-Box Test	R Q(15)	6.5231	0.9696
Ljung-Box Test	R Q(20)	8.7699	0.9854
Ljung-Box Test	R^2 Q(10)	16.6458	0.0826
Ljung-Box Test	R^2 Q(15)	20.9747	0.1376
Ljung-Box Test	R^2 Q(20)	27.2018	0.1297
LM ARCH Test	R TR^2	17.5389	0.1304

Figure 4.2 below shows the graphical representation of the conditional standard deviations and the ACF of the standardized residual. The standardized residuals have no significant lags hence, showing that no autocorrelations are left. The squared residuals also show that there are no ARCH effects left although a minor significant lag can be seen at 9, which, however, does not affect our conclusion as it is a higher order lag.

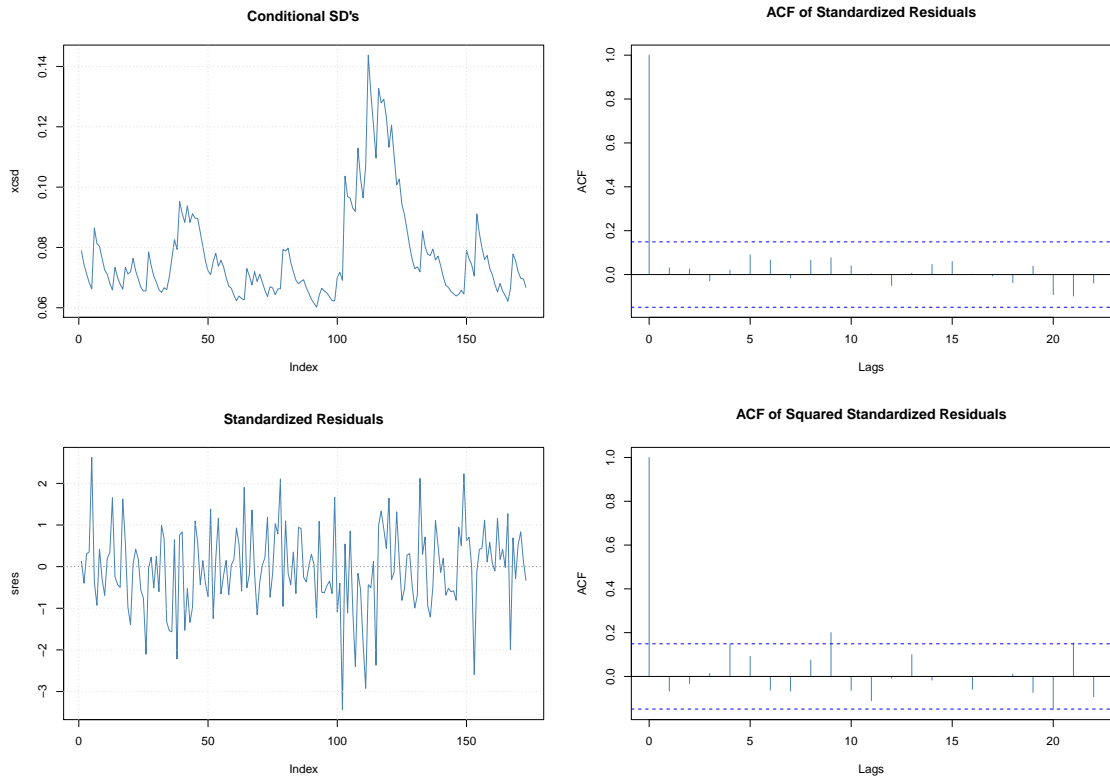


Figure 4.2: Residual Diagnosis plots

4.1.2 GARCH Extensions For Old Mutual Monthly Data

Table 4.8: GARCH Extensions

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
iGARCH(1,1)	-2.2701	-2.2155	199.3675	β_1 is NA, ω
iGARCH(1,2)	-2.2635	-2.1906	199.7903	ω α_1 β_1 β_2
iGARCH(2,1)	-2.2706	-2.1977	200.4042	ω α_1 α_2 β_1
EGARCH(1,1)	-2.3362	-2.2451	207.0853	γ_1 μ
EGARCH(1,2)	-2.3251	-2.2157	207.1222	μ
EGARCH(2,1)	-2.3185	-2.1909	207.5519	μ α_1 α_2 γ_1 γ_2
TGARCH(1,1)	2.251161	-2.160026	199.7254	μ
gjrGARCH(1,1)	-2.314307	-2.223171	205.1875	μ α_1 γ_1

The EGARCH, IGARCH models were fit using the rugarch package in R (see (Ghahlanos 2015)) whilst the TGARCH and gjrGARCH were fit using the Aparch command in FGARCH. None of the fitted models had all of its parameters significant as each model had, at least, one insignificant parameter. However, based on the AIC and BIC the EGARCH model, was the best model. Q-Q plots of the EGARCH(1,1) and the Pearson goodness of fit reject the normal distribution for the error terms. Hence, we fit the model with t distributed errors and obtain an improved Q-Q plot and AIC value. The obtained estimates are shown in Table 4.9 below.

Table 4.9: EGARCH(1,1) estimates from distribution=std

	Estimate	Std. Error	t value	Pr(>t)	QLME p-values
mu	0.007037	0.004467	1.5755	0.115147	0.068923
omega	-0.398103	0.015235	-26.1302	0.000000	0.000000
alpha1	-0.201214	0.050983	-3.9467	0.000079	0.000019
beta1	0.925150	0.000016	57868.7797	0.000000	0.000000
gamma1	-0.065571	0.032941	-1.9905	0.046532	0.019838
shape	8.374201	5.411876	1.5474	0.121773	0.046877

Table 4.9 reveals that μ and the shape parameter are insignificant hence, these can be removed from the model. Noting that the method for fitting is the ML method, if however, we fit using QLME which allows for possible misspecification of the likelihood function (see White (1982)) μ becomes significant at 10% and the shape parameter at 5% whilst estimates remaining more or less the same. These QLME p-values are shown on tab QLME p-values.

Output for Tests for ARCH/GARCH behaviour in standardized residuals is shown below. There is no evidence of serial correlation in squared residuals.

Table 4.10: Weighted Ljung-Box Test on Standardized Squared Residuals: d.o.f=2

	statistic	p-value
Lag[1]	0.4712	0.4924
Lag[2*(p+q)+(p+q)-1][5]	1.4400	0.7547
Lag[4*(p+q)+(p+q)-1][9]	3.2599	0.7154

Below are the Nyblom stability tests. All parameters are not significant at 5% hence no evidence of non-stable parameters.

Table 4.11: Nyblom stability test

Joint Statistic:	1.2792
mu	0.30109
omega	0.15571
alpha1	0.17947
beta1	0.14497
gamma1	0.13770
shape	0.08147

The sign bias test has all entries not significant at 5% hence showing no evidence of leverage effect.

Table 4.12: Sign Bias Test

	t-value	prob sig
Sign Bias	0.82410	0.4111
Negative Sign Bias	0.05441	0.9567
Positive Sign Bias	0.76796	0.4436
Joint Effect	0.88663	0.8287

Below are the fit model diagnostics plots. The ACF plot of the residuals shows no presence of autocorrelation. The plot of squared residuals also shows no significant lags except minor higher order ones at lag 14 and 21 which do not affect the conclusion that the ARCH effects have been removed. The plot of the conditional standard deviations vs returns shows that the returns and the conditional standard deviations take up similar fluctuation as expected.

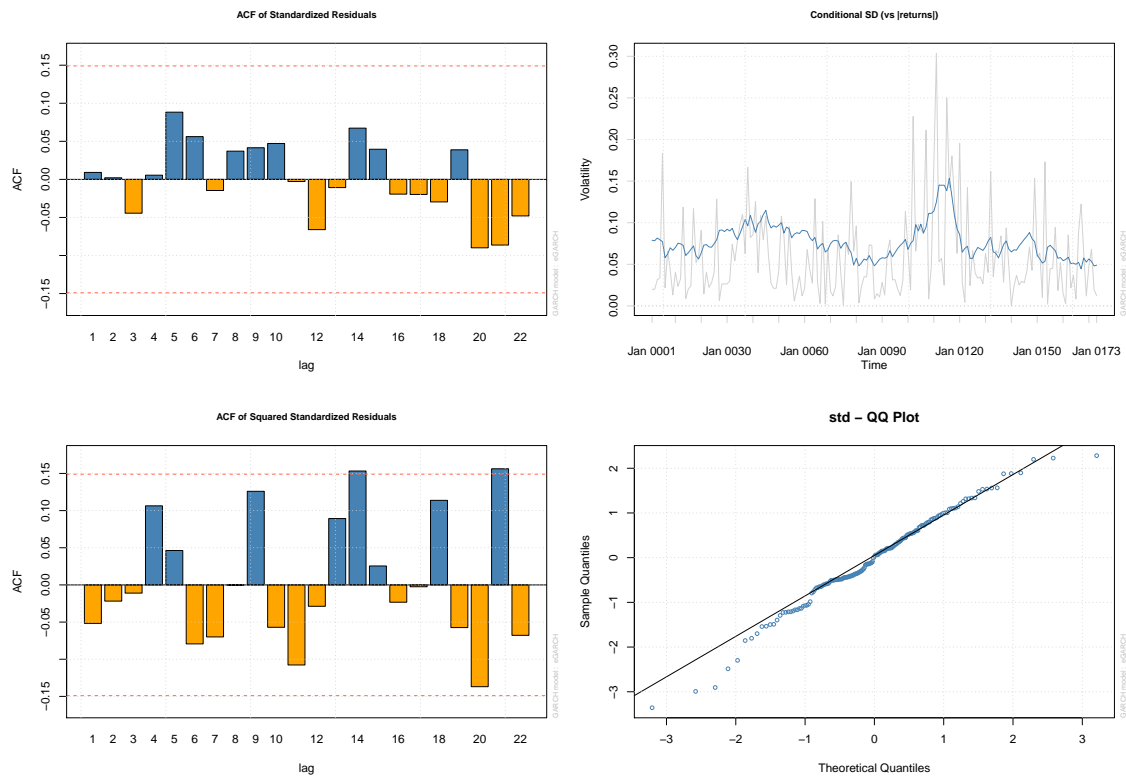


Figure 4.3: Residual Diagnosis plots

4.1.3 Old Mutual Daily Data

From the data analysis section, it was noted that the daily data was autocorrelated as compared to the monthly data. This means that we can not go straight to modelling without removing the autocorrelation. Tsay (2005) suggest the use of AR model to first capture the correlation. This can be done by modelling the AR model than using the residuals to run the GARCH models. This also is the same as jointly modelling an ARMA-GARCH model. This method is also used by Cantaluppi (2013).

The daily data ACF showed the presence of autocorrelations. We proceed to ARCH effect test, The Ljung-Box test on the a_t^2 series gives a $\chi^2 = 1683.85$, $df = 10$, $p\text{-value} < 2.2e-16$. ARCH LM-test from lag 1 to 20 also had p-values of zero showing strong ARCH effects presence.

We proceed to fit models and using package fGARCH in R. The PACF of the squared returns suggests an AR(6) model which seems to be of rather high order we then proceed to look at the PACF of squared residuals which are those after removing the series mean. The result, however, also, suggest an AR(6) hence, we model our ARMA-GARCH model with the ARMA part being an AR(6) initially.

After running various ARMA-GARCH models all with ARMA part as an AR6 it emerged

that in all either none, one or two parameters of the AR component would be significant. Hence, we tried models with AR3, AR2 and AR1 parts. The AR3 and AR2 also showed one parameter being significant which was the AR1 part hence, we settled for the AR1 to take care of the correlations.

Table 4.13: Summary of fitted models:cond distribution=norm

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
AR1+GARCH(1, 0)	-4.9664	-4.9577	6832.739	ω
AR1+GARCH(2,0)	-5.0215	-5.0107	6909.516	All significant
AR1+GARCH(3, 0)	-5.0822	-5.0693	6994.012	AR1
AR1+GARCH(1, 1)	-5.1488	-5.1380	7084.591	AR1
AR1+GARCH(1, 2)	-5.1512	-5.1383	7088.931	AR1 not significant
AR1+GARCH(2, 1)	-5.1480	-5.1351	7084.525	$\mu \alpha_2$
AR1+GARCH(2, 2)	-5.1505	-5.1354	7088.931	AR1 $\beta_1 \alpha_2$

Looking at Table 4.13 AR1+GARCH(1,1), AR1+GARCH(1,1) and AR1+GARCH(2,2) have the 3 best AIC values. We disregard AR1+GARCH(2,2) because it has many insignificant parameters. We also, drop the AR1+GARCH(1,2) despite a lower AIC because the difference in AIC compared to that of AR1+GARCH(1,1) is negligible given that AR1+GARCH(1,1) has 1 less parameter.

Table 4.14 shows that all parameters of the AR1+GARCH(1,1) are significant based on the given codes with the GARCH parameters highly significant.

Table 4.14: GARCH(1,1) model estimates from distribution=norm

Parameter	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	7.649e-04	3.187e-04	2.400	0.0164	*
AR1	-3.499e-02	2.049e-02	-1.708	0.0876	.
omega	7.920e-06	1.961e-06	4.038	5.38e-05	***
alpha1	6.594e-02	1.110e-02	5.942	2.82e-09	***
beta1	9.141e-01	1.434e-02	63.751	< 2e-16	***

¹ Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1.

Diagnosis is performed on residuals in R. We assess if the standardized residuals are white noise. The Ljung-Box tests up to order 10; 15; 20 are reported both for standardized residuals and their squared values. There is no evidence of autocorrelation nor of conditional heteroscedasticity presence as Ljung box statistics are all not significant. The Shapiro and Jarque-Bera test all reject the normality.

Table 4.15: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	R Chi^2	9623.095	0.000
Shapiro-Wilk Test	R W	0.9618	0.000
Ljung-Box Test	R Q(10)	6.8342	0.7410
Ljung-Box Test	R Q(15)	9.9387	0.8236
Ljung-Box Test	R Q(20)	13.8387	0.8386
Ljung-Box Test	R^2 Q(10)	1.8763	0.9972
Ljung-Box Test	R^2 Q(15)	3.0772	0.9995
Ljung-Box Test	R^2 Q(20)	3.6572	0.9999
hline LM ARCH Test	R TR^2	2.9446	0.9959

With normality rejected we try the models with t and skewed t distributed errors. Tables 4.16 and 4.17 below show the t and skewed t distributed error models respectively. The AIC values for both models are lower as compared to those of normally distributed error model indicating an improvement in the model. The t distribution AR1+GARCH(1,1) is the best of all the models as indicated by the lowest AIC and also, all its parameters are significant at 5%.

Table 4.16: Summary of fitted models:cond distribution=std

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
AR1+GARCH(1,0)	-5.1083	-5.0975	7028.917	AR1
AR1+GARCH(2,0)	-5.1413	-5.1283	7075.222	AR1
AR1+GARCH(3,0)	-5.1684	-5.1533	7113.54	AR1
AR1+GARCH(1,1)	-5.2263	-5.2134	7192.213	All significant
AR1+GARCH(2,1)	-5.2256	-5.2106	7192.23	α_2
AR1+GARCH(1,2)	-5.2277	-5.2126	7195.109	AR1 β_1
AR1+GARCH(2,2)	-5.2270	-5.2100	7195.109	AR1 α_2 β_1

Table 4.17: Summary of fitted models:cond distribution=sstd

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
AR1+GARCH(1,0)	-5.1076	-5.0947	7028.940	μ AR1
AR1+GARCH(2,0)	-5.1405	-5.1255	7075.227	AR1
AR1+GARCH(3,0)	-5.1677	-5.1504	7113.543	AR1
AR1+GARCH(1,1)	-5.2256	-5.2105	7192.22	all significant
AR1+GARCH(2,1)	-5.2249	-5.2077	7192.236	α_2
AR1+GARCH(1,2)	-5.2270	-5.2098	7195.117	AR1 β_1
AR1+GARCH(2,2)	-5.2263	-5.2069	7195.117	AR1 α_2 β_1

Estimates of the t error distributed model AR1+GARCH(1,1) are given in Table 4.18 below. The parameters are more significant than those given by normal errors. Hence, the AR1+GARCH(1,1) with t distributed errors is the best model fit. The persistence is

$(0.0543+0.935)=0.98930$ which is high.

Table 4.18: AR1+GARCH(1,1) model estimates from distribution=std

	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	7.642e-04	2.931e-04	2.607	0.00913	**
AR1	-3.837e-02	1.915e-02	-2.003	0.04513	*
omega	3.768e-06	1.284e-06	2.935	0.00333	**
alpha1	5.434e-02	1.005e-02	5.410	6.30e-08	***
beta1	9.353e-01	1.162e-02	80.460	< 2e-16	***
shape	7.198e+00	8.795e-01	8.184	2.22e-16	***

¹ Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1.

Table 4.19 shows the diagnosis of the residuals under t distributed error assumption. It shows that the Ljung-Box tests at lags 10,15, up to 20 of residuals and their squared values show no evidence of autocorrelation nor of conditional heteroscedasticity presence i.e there are all >0.05. The Jarque-Bera and Shapiro-Wilk test are as expected zero showing that the residuals do not follow the normal distribution.

Table 4.19: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	$R\ Chi^2$	12265.64	0
Shapiro-Wilk Test	R W	0.9589	0
Ljung-Box Test	R Q(10)	7.3600	0.6911
Ljung-Box Test	R Q(15)	10.4938	0.7876
Ljung-Box Test	R Q(20)	14.3862	0.81039
Ljung-Box Test	R^2 Q(10)	1.8590	0.9973
Ljung-Box Test	R^2 Q(15)	2.9214	0.9997
Ljung-Box Test	R^2 Q(20)	3.4064	0.9999
hline LM ARCH Test	R TR^2	2.7730	0.9969

Plots of the Q-Q plots for all 3 error distributions are shown in Figure 4.4 below. It can be seen that the t and skewed t distributions assumptions handle the heavy tails better as compared to the normal distribution. This justifies use of the t over the normal especially with estimates varying significantly for the models.

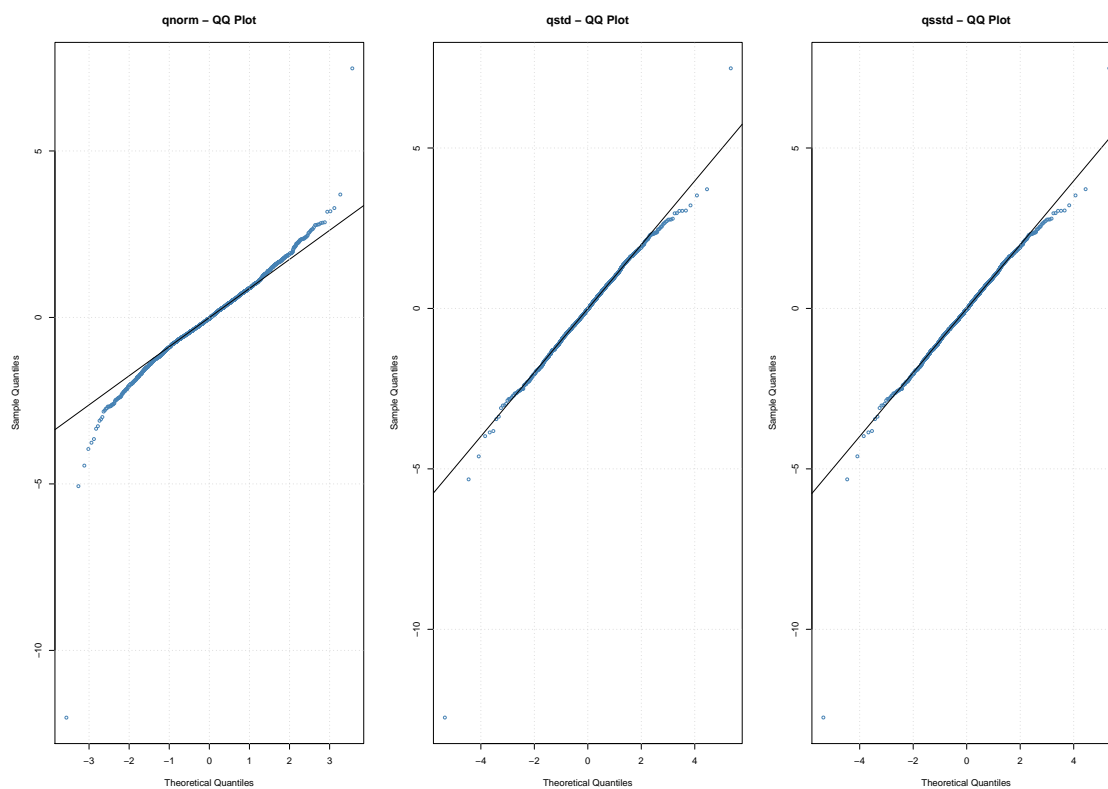


Figure 4.4: Q-Q plots

Below figure 4.5 shows the diagnostic plots obtained from R. The conditional standard deviations plot confirms to pattern of the returns. The ACF of the standardized returns has no autocorrelations left as required and also, the squared returns ACF has no significant lags showing the ARCH effects have been removed.

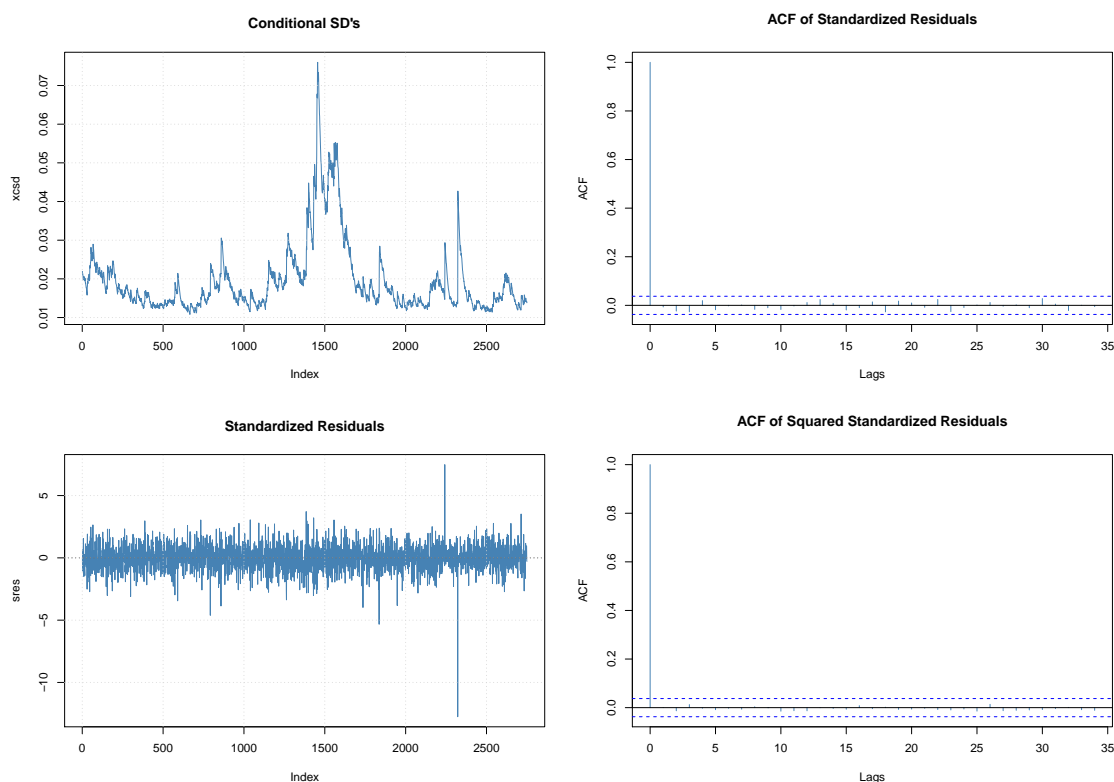


Figure 4.5: Residual Diagnosis plots

4.1.4 GARCH Extensions For Old Mutual Daily Data

Table 4.20: GARCH Extensions

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
iGARCH(1,1)	-5.1424	-5.1360	7073.828	β_1 na ω α_1
iGARCH(2,1)	-5.1417	-5.1331	7074.00	β_1 α_1
iGARCH(1,2)	-5.1466	-5.1379	7081.00	β_1 na ω
EGARCH(1,1)	-5.16090	-5.1502	7101.272	all highly significant
EGARCH(1,2)	-5.1630	-5.1501	7105.084	all significant
EGARCH(2,1)	-5.1621	-5.1470	7104.844	γ_2
AR1+iGARCH(1,1)	-5.1426	-5.1340	7075.105	AR1
AR1+iGARCH(1,2)	-5.1461	-5.1353	7080.894	AR1 β_2 na
AR1+EGARCH(1,2)	-5.1630	-5.1480	7106.186	AR1
AR1+EGARCH(1,1)	-5.1612	-5.1483	7102.612	AR1
TGARCH(1,1)	-4.8882	-4.8775	6726.27	all significant
gjrGARCH(1,1)	-5.1325	-5.1417	7089.687	μ

From the models ran above the EGARCH(1,1) was the best model although the EGARCH(1,2) had better AIC its BIC was not better and also, its parameter were not highly significant as those of EGARCH(1,1). The estimates under the normal and t distribution where all

significant. However, from fit diagnostics, the t distributed errors outperform the normal errors. This is confirmed by the Q-Q plots in Figure 4.6 below.

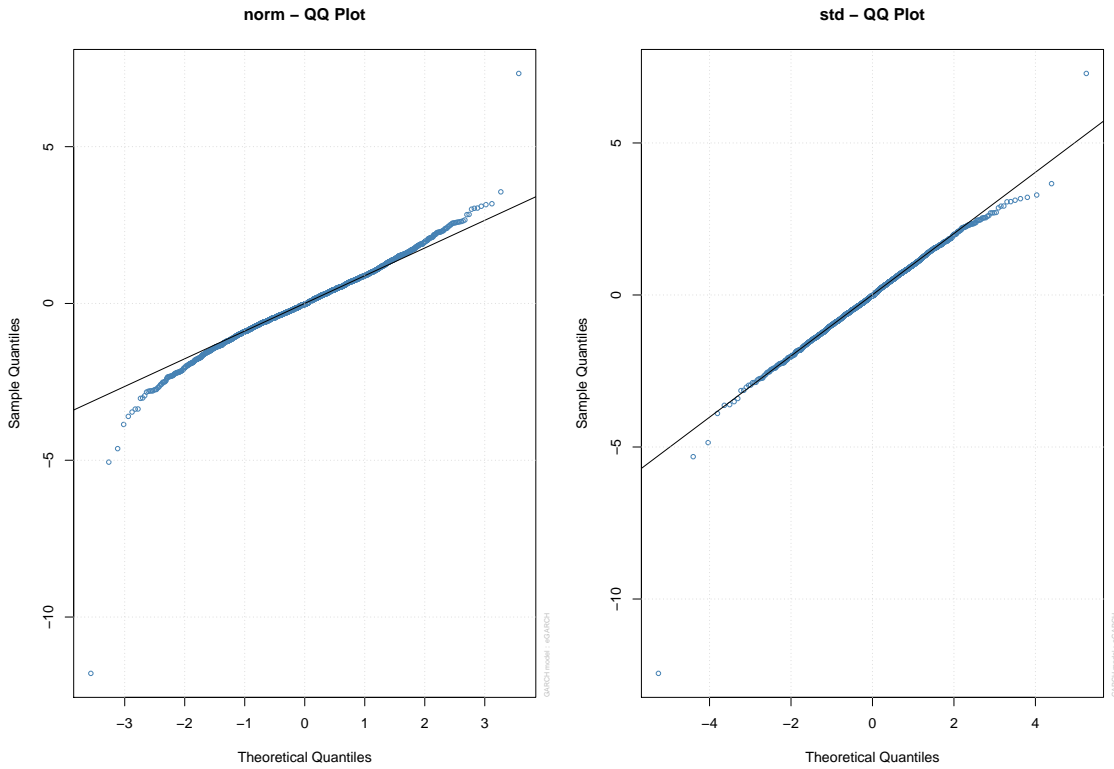


Figure 4.6: Q-Q plots

Table 4.21 shows the model fits from R and also, those for EVIEWS for comparison. Both models had student t distributed error assumption. Clearly both software estimate the model constantly the only difference is the value for omega. The AIC values were -5.2335 for R and -5.233472 for EVIEWS which when round to 4dp is the same as the one from R.

Table 4.21: Summary of fitted model

	Estimate R	Estimate EVIEWS	Pr(>t)in R	Pr(>z)in EVIEWS
mu	0.000604	0.000604	0.040665	0.0368
omega	-0.054424	-0.12347	0.000000	0.000000
alpha1	-0.038703	-0.038710	0.000025	0.0002
beta1	0.993389	0.993386	0.000000	0.0000
gamma1	0.090513	0.090520	0.000000	0.000000
shape	7.485963	7.487485	0.000000	0.000000

We proceed to diagnose our model. The diagnostics are taken from R. The Nymblom Stability test below shows that parameters are stable with μ slightly off.

Table 4.22: Nyblom Stability Test

Joint Statistic	1.1184
mu	0.03304
omega	0.12219
alpha1	0.22642
beta1	0.10565
gamma1	0.19792
shape	0.39563

Output for Tests for ARCH/GARCH behaviour in standardized residuals is shown below. There is no evidence of serial correlation in squared residuals.

Table 4.23: Weighted Ljung-Box Test on Standardized Squared Residuals d.o.f=2

	Statistic	P-Value
Lag[1]	0.1574	0.6916
Lag[2*(p+q)+(p+q)-1][5]	0.8999	0.8819
Lag[4*(p+q)+(p+q)-1][9]	1.1373	0.9798

Figure 4.7 shows the diagnosis plots show that the ACF of standardized residuals are not correlated hence, shows that model is a good fit. The squared standardized residuals show the absence of correlation hence, means that ARCH effects have been removed. The conditional variance and return plot shows the two plots going hand in hand. Periods with extreme returns show higher volatility.

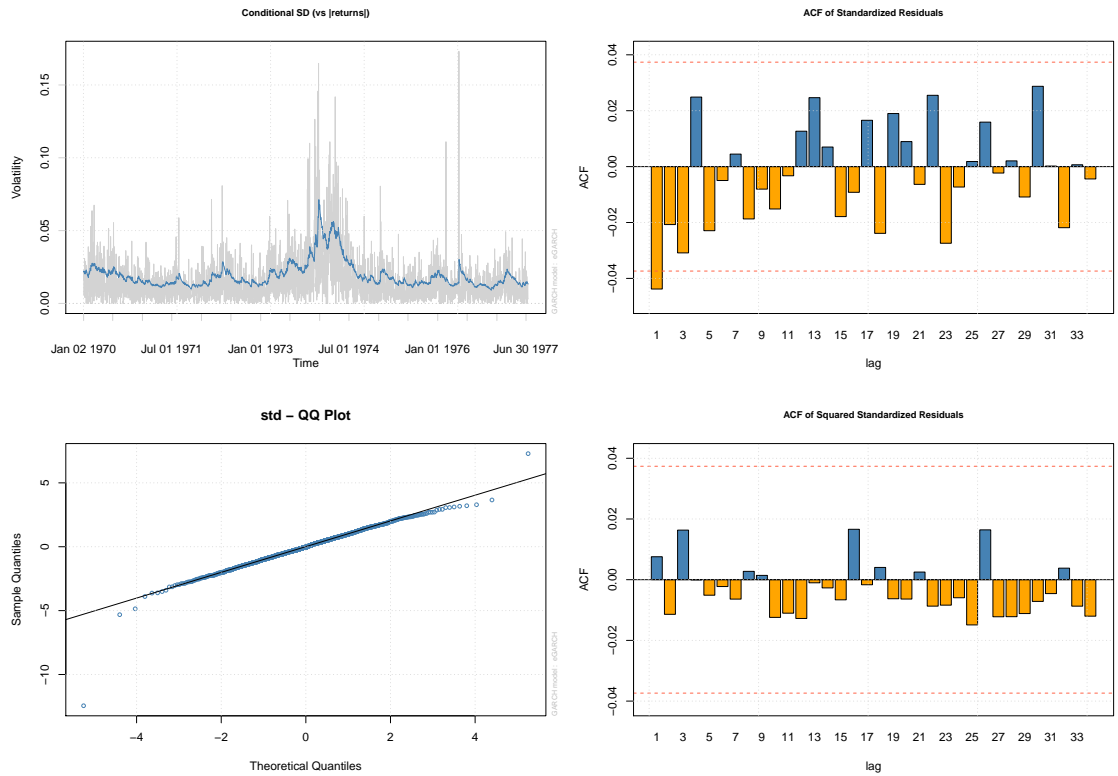


Figure 4.7: Residual Diagnosis plots

4.2 Standard Bank Data

4.2.1 Monthly Data

From the data analysis section, graphical evidence suggested the presence of ARCH effects. Before running models we proceed to test through a statistical test.

The ACF of the log returns had no significant correlations except one at lag 5. Based on statistical tests confirm the series has no significant serial correlations as Ljung box test of the log returns gave $\chi^2 = 16.4438$, $df = 12$, $p\text{-value} = 0.1717$. The Ljung-Box test on the a_t^2 series gives $\chi^2 = 33.1193$, $df = 12$, $p\text{-value} = 0.0009274$. Hence, evidence that there are ARCH effects.

A series of ARCH/GARCH models were run including one with AR2 component as suggested by the PACF of squared returns. All the AR+GARCH models had the AR parts insignificant, therefore, we proceeded to simple ARCH/GARCH.

Table 4.24: Fitted Models

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
ARCH(1)	-2.197	-2.159	318.3205	all significant
ARCH(2)	-2.2398	-2.1888	325.421	α_1
ARCH(3)	-2.2622	-2.1984	329.629	α_1
GARCH(1,1)	-2.2528	-2.2018	327.2788	All significant
GARCH(1,2)	-2.2473	-2.1835	327.4874	$\beta_1 \beta_2$ NA
GARCH(2,1)	-2.2576	-2.1938	328.9626	$\alpha_1 \alpha_2$
GARCH(2,2)	-2.2506	-2.1741	328.9626	$\alpha_1 \alpha_2 \beta_1 \beta_2$
AR1+GARCH(1,0)	-2.2098	-2.1588	321.106	All significant

From Table 4.24 above we see that the ARCH(3) model has the best AIC value but we realise that it has an insignificant parameter hence, we opt for the GARCH(1,1) which has a slightly higher AIC but lowest BIC. It not only has all parameters significant but there are more significant and also, the model is parsimonious compared to the ARCH(3).

With GARCH(1,1) as our model of choice. Its parameters output is shown below

Table 4.25: GARCH(1,1) Model estimates:cond distribution=norm

	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	0.0180618	0.0039673	4.553	5.30e-06	***
omega	0.0012268	0.0005148	2.383	0.01716	*
alpha1	0.2778337	0.0996606	2.788	0.00531	**
beta1	0.5712725	0.1215650	4.699	2.61e-06	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Diagnosis is performed on the residuals in R. We assess if the standardized residuals are white noise. The Ljung-Box tests up to order 10; 15; 20 are reported both for standardized residuals and their squared values. There is no evidence of autocorrelation nor of conditional heteroscedasticity presence as Ljung box statistics are all not significant. The Shapiro and Jarque-Bera test all reject the normality hence, we must change the distribution assumption of the residuals from the normal assumption to a studentized t distribution.

The estimates from the GARCH(1,1) model with t distributed errors are shown below.

Table 4.26: GARCH(1,1) Model estimates:cond distribution=std

	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	0.0153015	0.0040335	3.794	0.000148	***
omega	0.0011855	0.0007009	1.691	0.090787	.
alpha1	0.1732002	0.0839409	2.063	0.039078	*
beta1	0.6352300	0.1574486	4.035	5.47e-05	***
skew	1.0393163	0.0873311	11.901	< 2e-16	***
shape	6.3808254	2.0177405	3.162	0.001565	**

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

The model estimates show that persistence is high as $\alpha + \beta = 0.173 + 0.635 = 0.808$. As a way of diagnosis, the ACF plots of the residuals and squared residual are shown in Figure 4.8 below. The residual plot shows no autocorrelation except a minor one at lag 5 which is not that significant. Hence, we can conclude the white noise assumption is met. The squared residual ACF shows no presence of ARCH effects. Formal tests for the autocorrelation of standardized residuals through Ljung-Box concur with the graphical evidence and do not reject the null of white noise, and tests for autocorrelation of squared residuals do not give evidence of the presence of conditional heteroscedasticity.

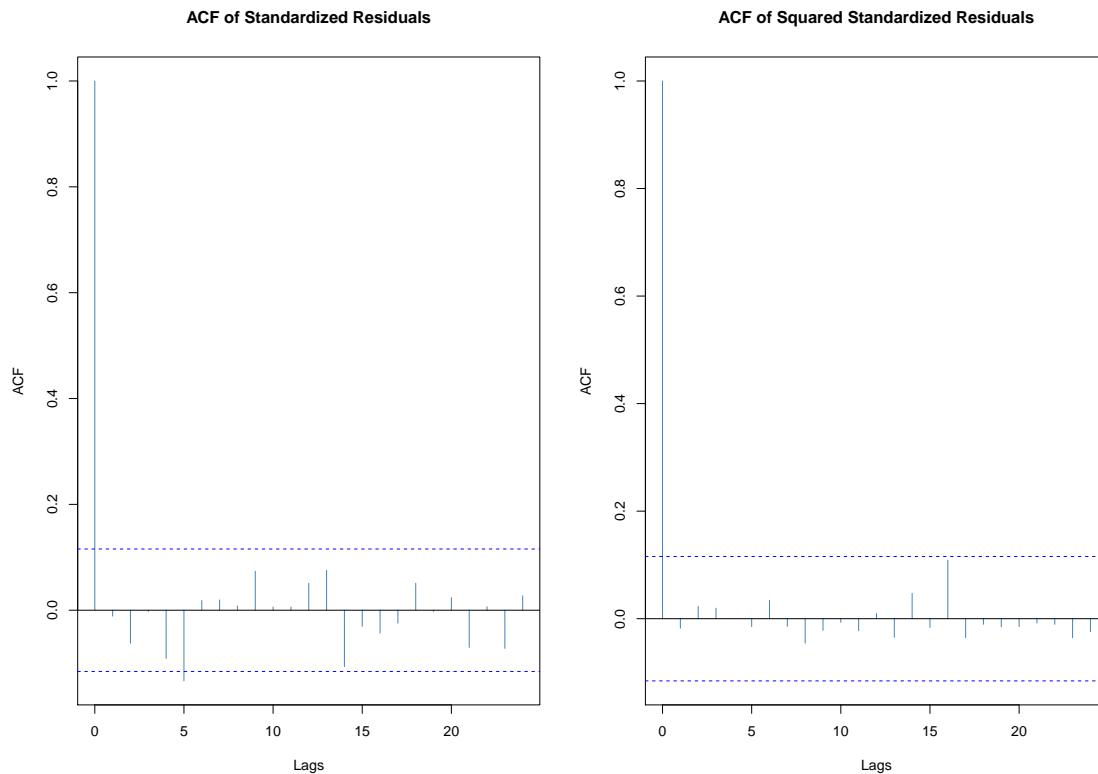


Figure 4.8: ACF plots

Table 4.27: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	$R\ Chi^2$	473.2515	0.00
Shapiro-Wilk Test	R W	0.9483	1.637e-08
Ljung-Box Test	R Q(10)	10.7157	0.3801
Ljung-Box Test	R Q(15)	16.9546	0.3216
Ljung-Box Test	R Q(20)	18.6856	0.5423
Ljung-Box Test	R^2 Q(10)	1.5929	0.9986
Ljung-Box Test	R^2 Q(15)	2.8857	0.9997
Ljung-Box Test	R^2 Q(20)	7.0826	0.9964
LM ARCH Test	R TR^2	1.6171	0.9998

The Jarque-Bera and Shapiro-Wilk statistics reject the null hypothesis of a normal distribution for the standardized residuals.

Q-Q plots are shown in figure 4.9. They show that the t distribution can better capture the behaviour on the tails of standardized residuals. The normal one has its tails drifting further apart from the line whilst the same can be seen with t distributed one, it is, however, to a lesser extent and with fewer observations to have been outliers.

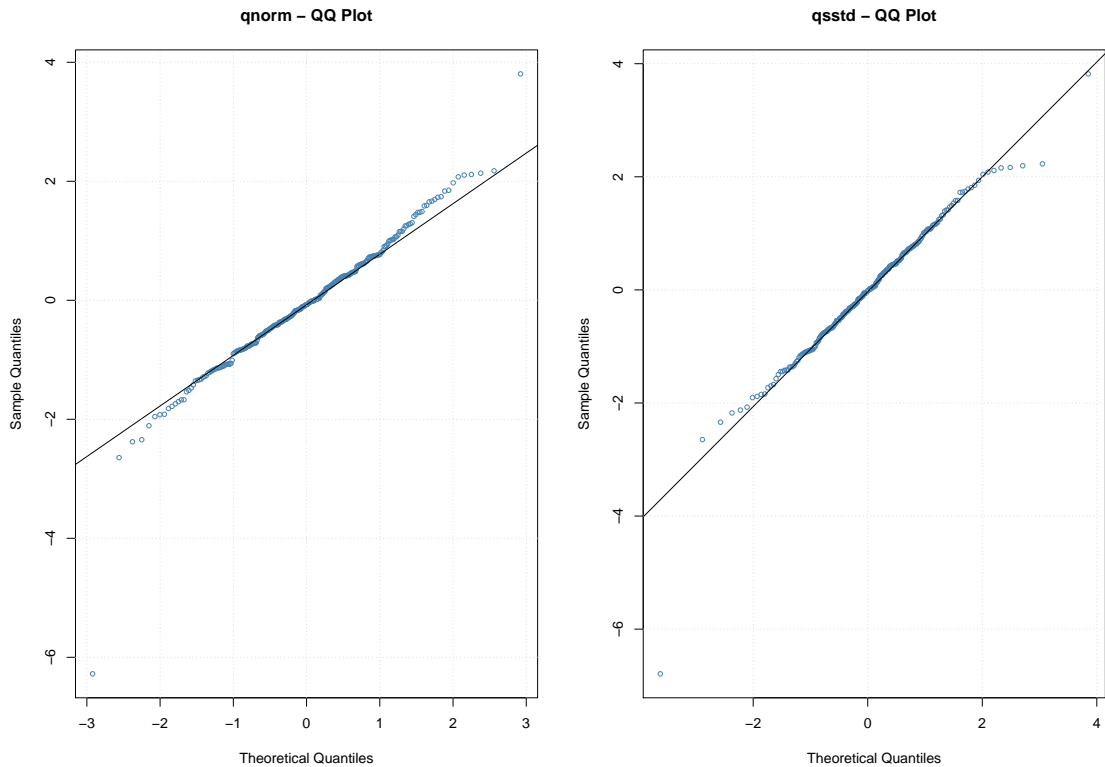


Figure 4.9: Q-Q plots

Figure 4.10 below shows that the ACF standardized residuals have no autocorrelation and

that of squared standardized residuals also has no autocorrelations hence, the model removes the correlations and ARCH effects as expected. The conditional standard deviation plots confirm to the pattern suggested by the return plots.

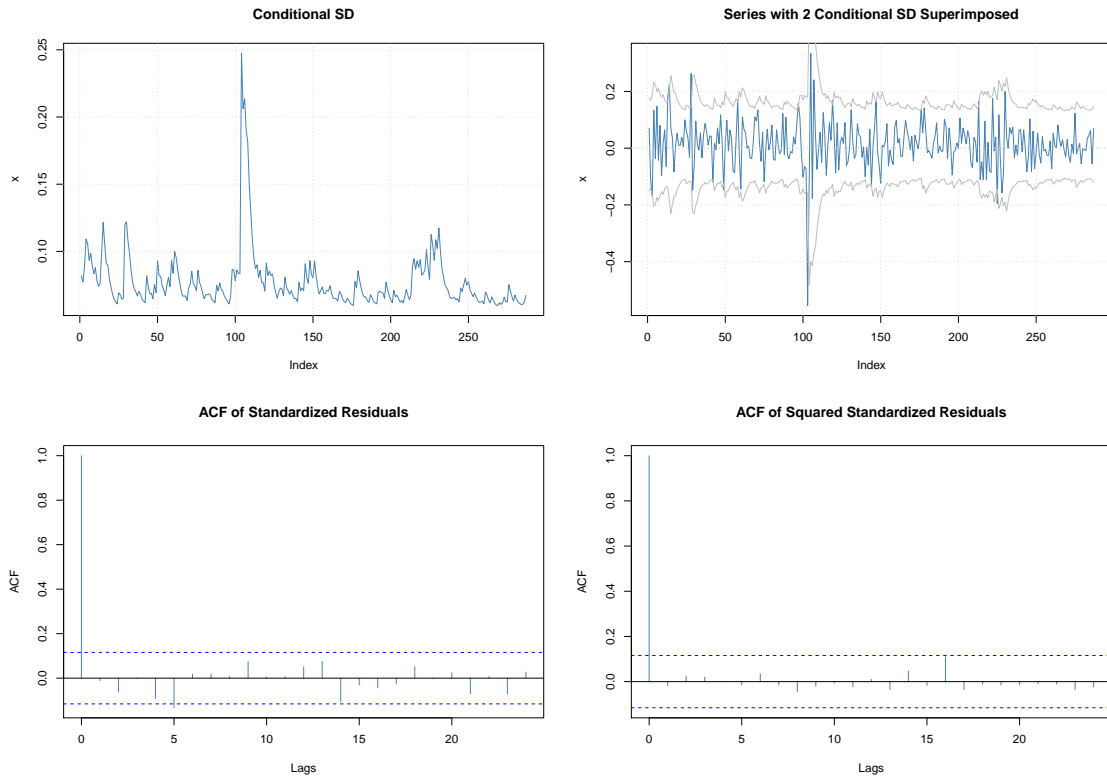


Figure 4.10: Residual Diagnosis plots

4.2.2 GARCH Extensions For Standard Bank Monthly Data

Table 4.28: Fitted Models

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
iGARCH(1,1)	-2.2755	-2.2155	199.3675	beta1 na ω
iGARCH(1,2)	-2.2635	-2.1906	199.7903	ω α_1 β_1 β_2
iGARCH(2,1)	-2.2706	-2.1977	200.4042	ω α_1 α_2 β_1
EGARCH(1,1)	-2.2755	-2.2117	331.5335	all significant
EGARCH(1,2)	-2.2790	-2.2025	333.0432	β_2
EGARCH(2,1)	-2.2776	-2.1883	333.8313	α_1 α_2 γ_1
TGARCH(1,1)	-2.2820	-2.2183	332.4722	all significant
gjrGARCH(1,1)	-2.2708	-2.2071	330.8633	all significant

The fitted models show that the TGARCH(1,1) is the best fit. However, the Q-Q plots below show that the student t distribution is a good fit than the normal distribution.

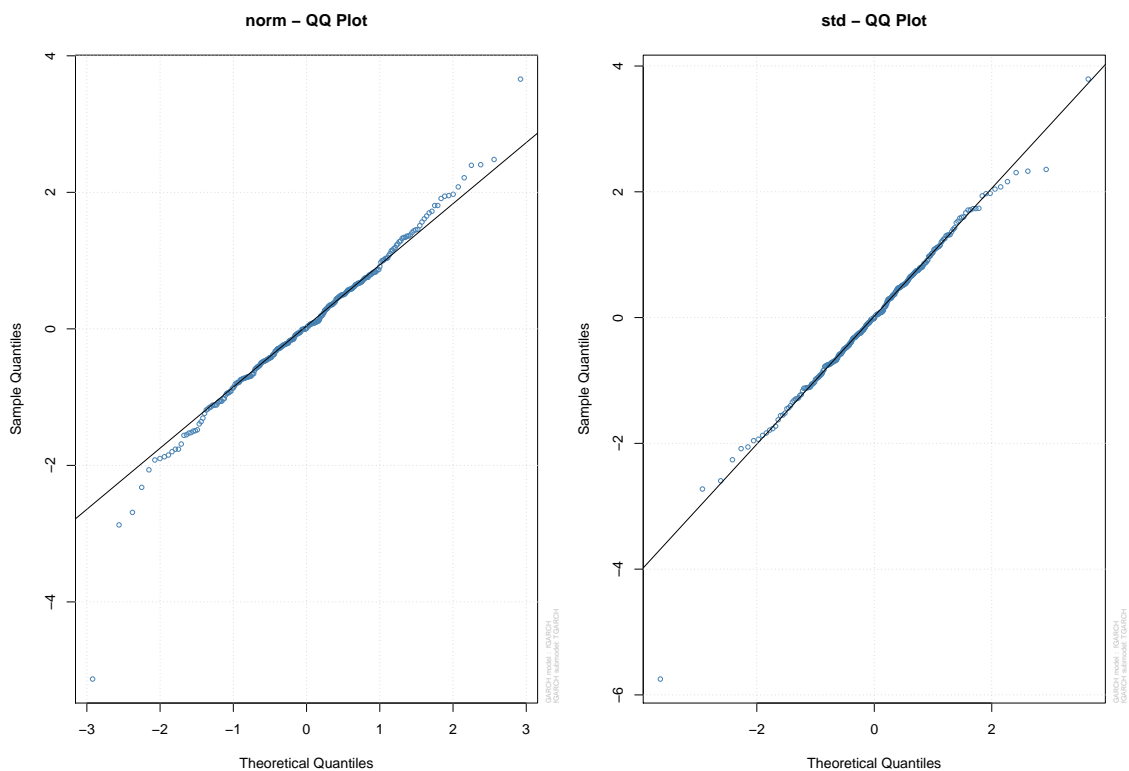


Figure 4.11: Diagnosis plots

The model fitted with student t however, has eta11 parameter to be insignificant, this is the asymmetry parameter in this model. This means on reporting parameters this can be ignored. The estimates are shown in Table 4.29 below.

Table 4.29: TGARCH(1,1) Model estimates:cond distribution=std

	Estimate	Std. Error	t value	Pr(>t)
mu	0.012503	0.004003	3.1229	0.001791
omega	0.017042	0.009111	1.8705	0.061417
alpha1	0.175733	0.068338	2.5715	0.010125
beta1	0.642934	0.147045	4.3723	0.000012
eta11	0.366092	0.295664	1.2382	0.215640
shape	7.061093	2.483225	2.8435	0.004462

The Nyblom test shows that all parameters are stable in the model.

Table 4.30: Nyblom stability test

Joint Statistic:	1.5539
mu	0.29920
omega	0.60762
alpha1	0.34822
beta1	0.46363
eta11	0.08143
shape	0.05193

The sign bias test show no evidence of leverage effect.

Table 4.31: Sign Bias Test

Sign Bias	0.2352	0.8142
Negative Sign Bias	0.3698	0.7118
Positive Sign Bias	0.4080	0.6836
Joint Effect	0.3567	0.9490

The Weighted Ljung-Box Test on Standardized Residuals shows no presence of serial correlation hence, showing our model has no correlations as required.

Table 4.32: Weighted Ljung-Box Test on Standardized Residuals d.o.f=2

	statistic	p-value
Lag[1]	0.05467	0.8151
Lag[2*(p+q)+(p+q)-1][2]	0.71827	0.5994
Lag[4*(p+q)+(p+q)-1][5]	3.59497	0.3091

The Weighted Ljung-Box Test on Standardized Squared Residuals shows no presence of ARCH effects hence, showing our model removed all conditional heteroscedasticity as required.

Table 4.33: Weighted Ljung-Box Test on Standardized Squared Residuals d.o.f=2

	statistic	p-value
Lag[1]	0.05165	0.8202
Lag[2*(p+q)+(p+q)-1][5]	0.57461	0.9456
Lag[4*(p+q)+(p+q)-1][9]	1.14444	0.9795

Figure 4.12 below shows more Diagnosis plots. The Conditional SD (vs —returns—) plot shows that the conditional standard deviation pattern and that of returns are confirming. The ACF of Standardized Residuals has no significant lags except one at lag 5 but as seen from the Ljung Box test above this is insignificant. The Standardized Squared Residuals ACF also has one higher lag which is significant but likewise, the Ljung Box test above showed that this is insignificant.

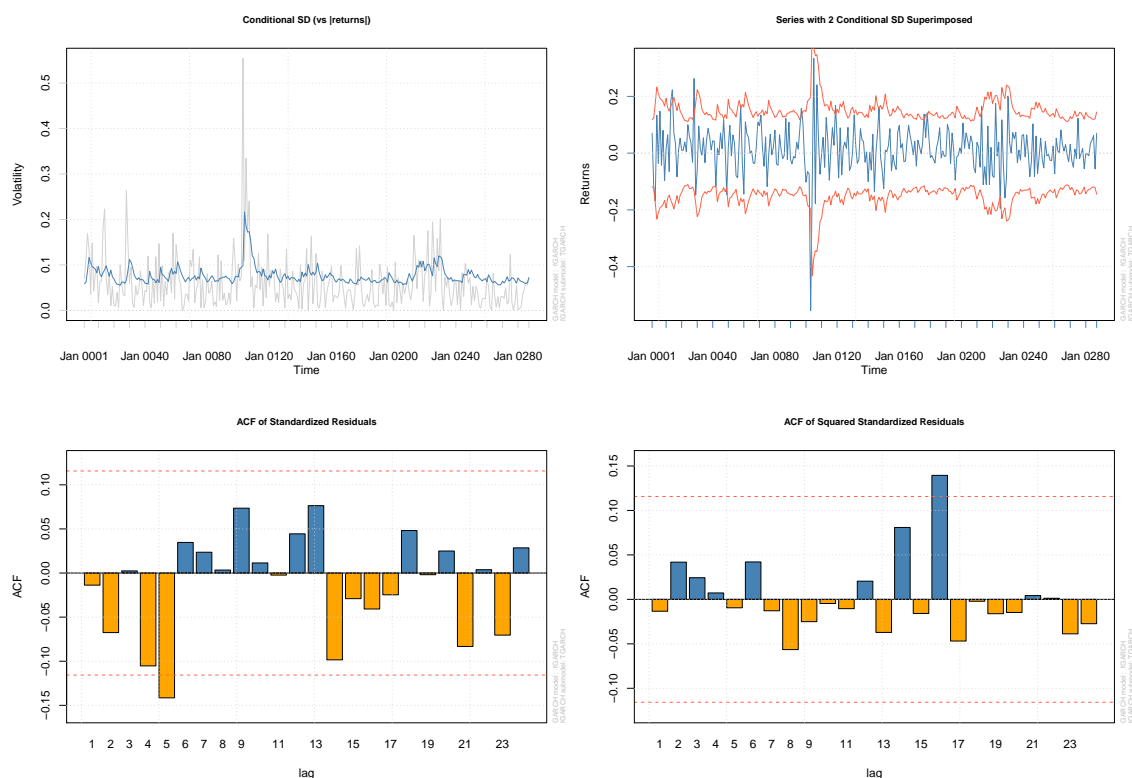


Figure 4.12: Residual Diagnosis plots

4.2.3 Standard Bank Daily Data

Graphical analysis of the ACF in the data analysis section showed that the Standard Bank daily data was strongly correlated. The PACF of the squared returns was also strongly correlated showing the presence of ARCH effects. Below we show results of statistical tests that confirm this.

Ljung box test on the log returns gives $\chi^2 = 57.694$, $df = 10$, $p\text{-value} = 9.871e-09$. This rejects the null of white noise and we conclude that the series is correlated. Now using the mean corrected equation a_t we have $\chi^2 = 57.694$, $df = 10$, $p\text{-value} = 9.871e-09$ which also, rejects the white noise hence, series is correlated. A test for the ARCH effects on a_t^2 gives $\chi^2 = 866.7372$, $df = 10$, $p\text{-value} < 2.2e-16$. Hence, we conclude that ARCH effects are present.

Before fitting models, we rid the data of correlations. This we do by applying an AR model first before the GARCH which is the same as modelling with an AR+GARCH model as seen in the Old Mutual case. The PACF of the squared returns has many significant lags but looking at the lower order significant lags we see that the appropriate model is an AR4. Quite a number of models were deemed adequate for this data. The following table gives the best models and their respective AIC, BIC values.

Table 4.34: Fitted Models

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
AR4+GARCH(1,1)	-5.3358	-5.3185	7344.674	all significant
AR4+GARCH(1,0)	-5.2006	-5.1855	7157.803	all significant
AR3+GARCH(1,0)	-5.1978	-5.1849	7152.982	all significant
AR3+GARCH(1,1)	-5.3331	-5.3181	7340.053	all significant
AR2+GARCH(1,1)	-5.3322	-5.3193	7337.779	all significant
AR2+GARCH(1,0)	-5.196971	-5.186209	7150.836	all significant
AR1+GARCH(1,0)	-5.1916	-5.1830	7142.402	μ
AR1+GARCH(1,0)	-5.3285	-5.3178	7331.707	all significant

The ARMA+GARCH(1,1) models are the best models judging by their lower AIC BIC values. They have very similar AIC and BIC values. Despite the ARMA4+GARCH(1,1) having the lowest AIC among our ARMA+GARCH(1,1). But, we require a parsimonious model. If the decrement in AIC is not significantly different from that of a similar model with fewer parameters then a lower higher AIC can be taken. The first step is to diagnose these models to see if all of them remove ARCH effects and correlations. We will take 3 models ARMA4, ARMA3 and ARMA2 and do diagnosis on residuals.

ACF plots of residuals of ARMA(2,0)+GARCH(1,1), ARMA(3,0)+GARCH(1,1) and ARMA(3,0)+GARCH(1,1) are shown figures 4.13-4.15 below. From the ACF plots of ARMA 2, it can be seen that correlations are still present. The ARMA 3 still has correlations on lag and it, however, the squared residuals show that no ARCH effects remain. The ARMA 4 is the best as it has neither correlations or conditional heteroscedasticity remaining. Further analysis show that the correlations in ARMA 3 are not significant although in this case model of choice is ARMA4.

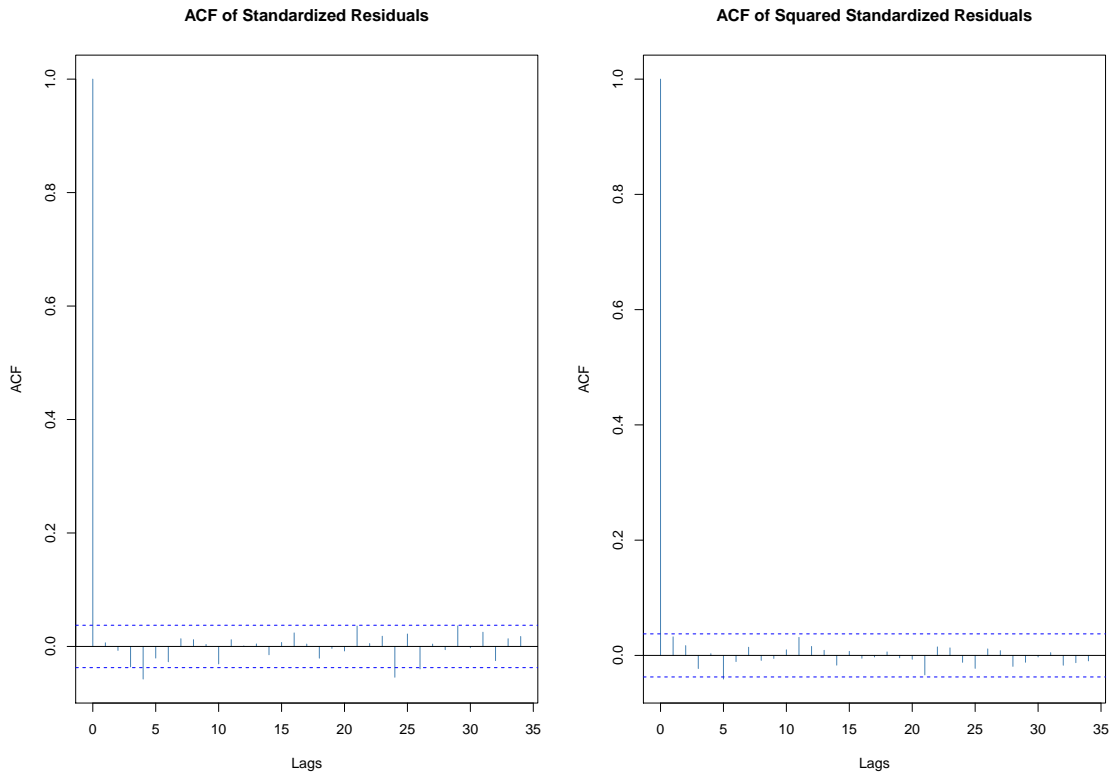


Figure 4.13: ACF of ARMA(2)+GARCH(1,1)

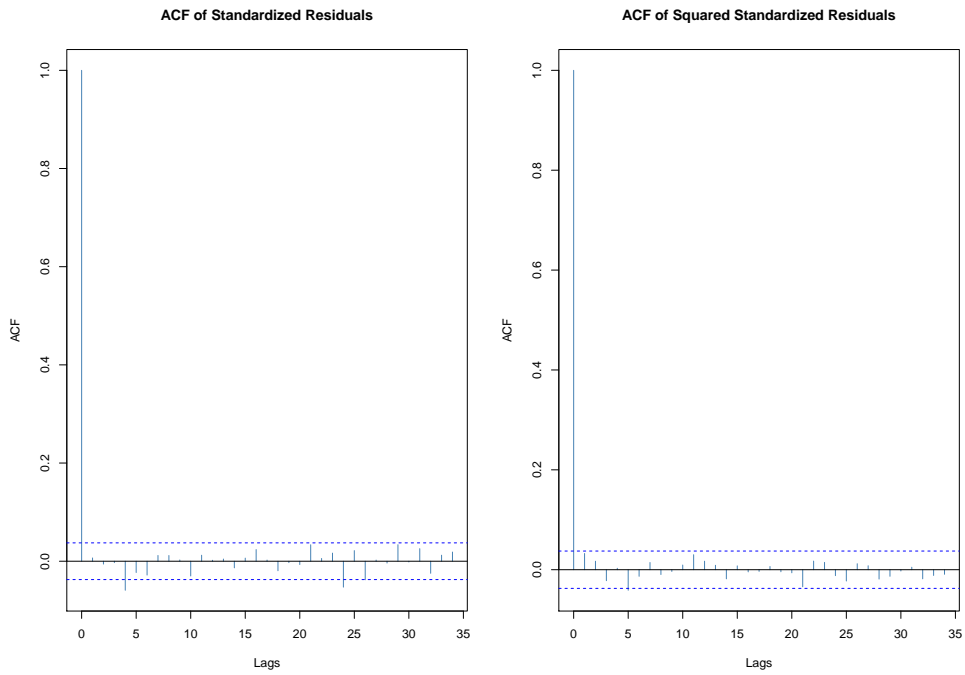


Figure 4.14: ACF of ARMA(3)+GARCH(1,1)

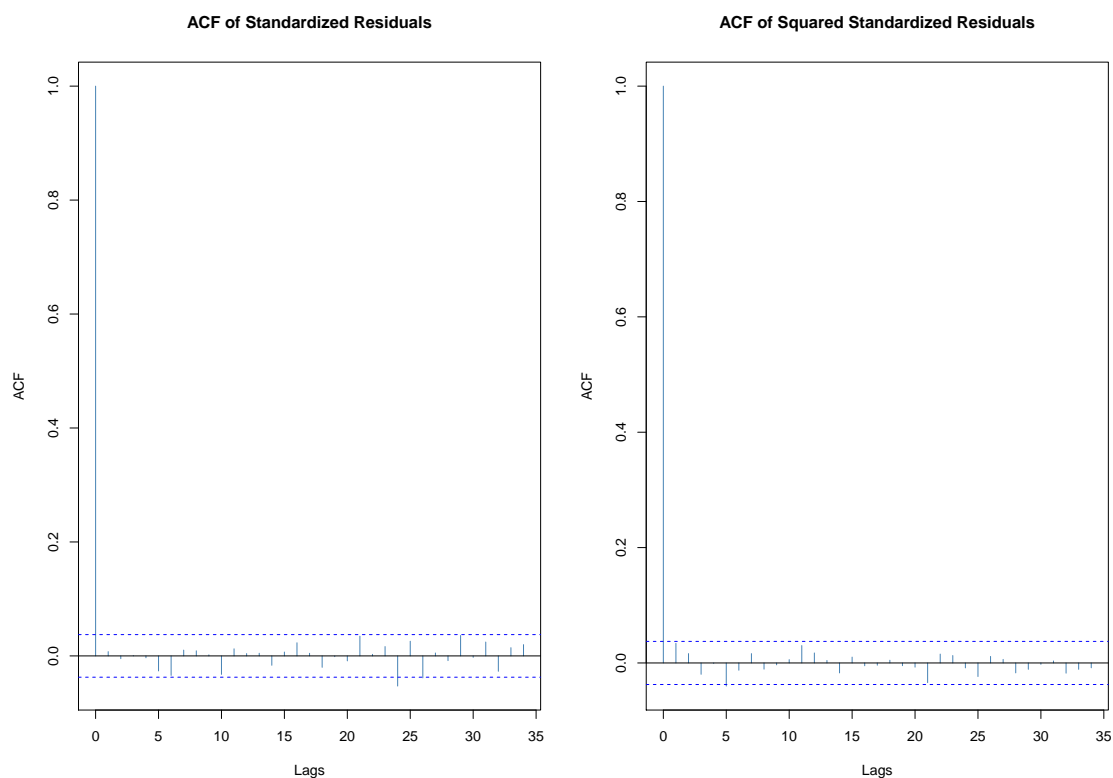


Figure 4.15: ACF of ARMA(4)+GARCH(1,1)

From the diagnosis, we see that we have to keep the ARMA4 model as the best fit. The estimates of the ARMA(4)+GARCH(1,1) as given in R are given below.

Table 4.35: AR(4)+GARCH(1,1) Model estimates:cond distribution=norm

	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	7.454e-04	2.937e-04	2.538	0.011156	*
AR1	-5.719e-02	2.008e-02	-2.848	0.004405	**
AR2	-7.394e-02	1.989e-02	-3.717	0.000202	***
AR3	-4.042e-02	1.989e-02	-2.032	0.042158	*
AR4	-5.964e-02	1.968e-02	-3.030	0.002446	**
omega	7.639e-06	1.918e-06	3.983	6.80e-05	***
alpha1	7.963e-02	1.079e-02	7.380	1.58e-13	***
beta1	8.965e-01	1.423e-02	63.004	< 2e-16	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

All parameter estimates are significant as to given codes.

A look at the residuals shows that the Jarque-Bera and Shapiro-Wilk Test reject normal assumption as expected for this data. The Ljung box, however, shows that correlations are removed as well as ARCH effects.

Table 4.36: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	R Chi^2	61.7870	3.8302e-14
Shapiro-Wilk Test	R W	0.9955	2.234e-07
Ljung-Box Test	R Q(10)	8.828	0.5485
Ljung-Box Test	R Q(15)	10.2422	0.8043
Ljung-Box Test	R Q(20)	13.1061	0.8728
Ljung-Box Test	R^2 Q(10)	11.0148	0.3564
Ljung-Box Test	R^2 Q(15)	15.5217	0.4145
Ljung-Box Test	R^2 Q(20)	15.9026	0.7226
LM ARCH Test	R TR^2	14.0394	0.2982

For certainty, we run the model with t and skewed t distributed errors. The models have better AIC BIC values.

Table 4.37: Estimates for model with t then skewed t

Model	AIC	BIC	Log Likelihood	Non Sig parameters at 5%
AR4+GARCH(1,1)	-5.3498	-5.3304	7364.944	all significant
AR4+GARCH(1,1)	-5.3492	-5.3277	7365.198	all significant

Q-Q plots of the 3 are shown in Figure 4.16 below. The normal distribution clearly did not model the heavy tails well as can be seen by the ends of the plots deviating from the line. On the other hand, the t and skewed t distribution modelled the heavy tails better. The t and skewed t distribution are similar in Q-Q plots and also, AIC BIC values. We choose the t distribution for simplicity.

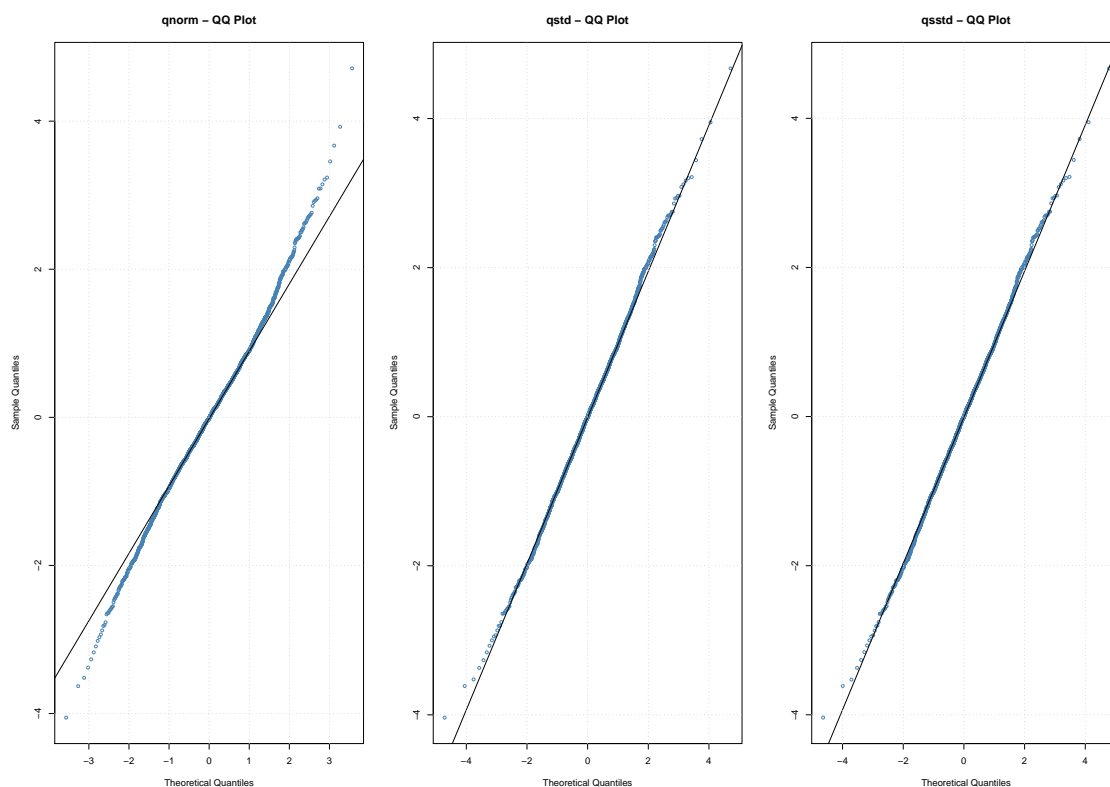


Figure 4.16: Q-Q plots

Estimates from the t are shown in Table 4.38 below. They are all highly significant. The persistence $\alpha + \beta = 0.975$ which is very high.

Table 4.38: AR4+GARCH(1,1) Model estimates:cond distribution=std

	Estimate	Std. Error	t-value	Pr(>t)	Sig code
mu	6.684e-04	2.888e-04	2.314	0.020643	*
AR1	-5.970e-02	1.979e-02	-3.017	0.002555	**
AR2	-7.309e-02	1.965e-02	-3.720	0.000199	***
AR3	-4.194e-02	1.953e-02	-2.147	0.031775	*
AR4	-5.664e-02	1.943e-02	-2.915	0.003560	**
omega	8.216e-06	2.335e-06	3.519	0.000434	***
alpha1	8.298e-02	1.311e-02	6.328	2.49e-10	***
beta1	8.918e-01	1.732e-02	51.477	< 2e-16	***
shape	9.989e+00	1.828e+00	5.465	4.64e-08	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Based on the Jarque-Bera Test and Shapiro-Wilk Test below the normality assumption is rejected. The Ljung-Box tests up to order 10; 15; 20 are reported both for standardized residuals and their squared values. There is no evidence of autocorrelation nor of conditional heteroscedasticity presence as Ljung box statistics are all not significant.

Table 4.39: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	$R\ Chi^2$	62.3117	2.942e-14
Shapiro-Wilk Test	R W	0.9955	1.905e-07
Ljung-Box Test	R Q(10)	9.1069	0.5220
Ljung-Box Test	R Q(15)	10.5256	0.7854
Ljung-Box Test	R Q(20)	13.3909	0.860
Ljung-Box Test	R^2 Q(10)	10.7014	0.3813
Ljung-Box Test	R^2 Q(15)	15.2130	0.4362
Ljung-Box Test	R^2 Q(20)	15.5471	0.7443
LM ARCH Test	R TR^2	13.6531	0.3234

Below in Figure 4.17 are diagnosis plots for residuals. The conditional standard deviation plots confirm to the pattern suggested by the returns plot. The ACF of residuals and squared residuals showing no autocorrelation hence, showing absence of correlation and ARCH effects respectively.

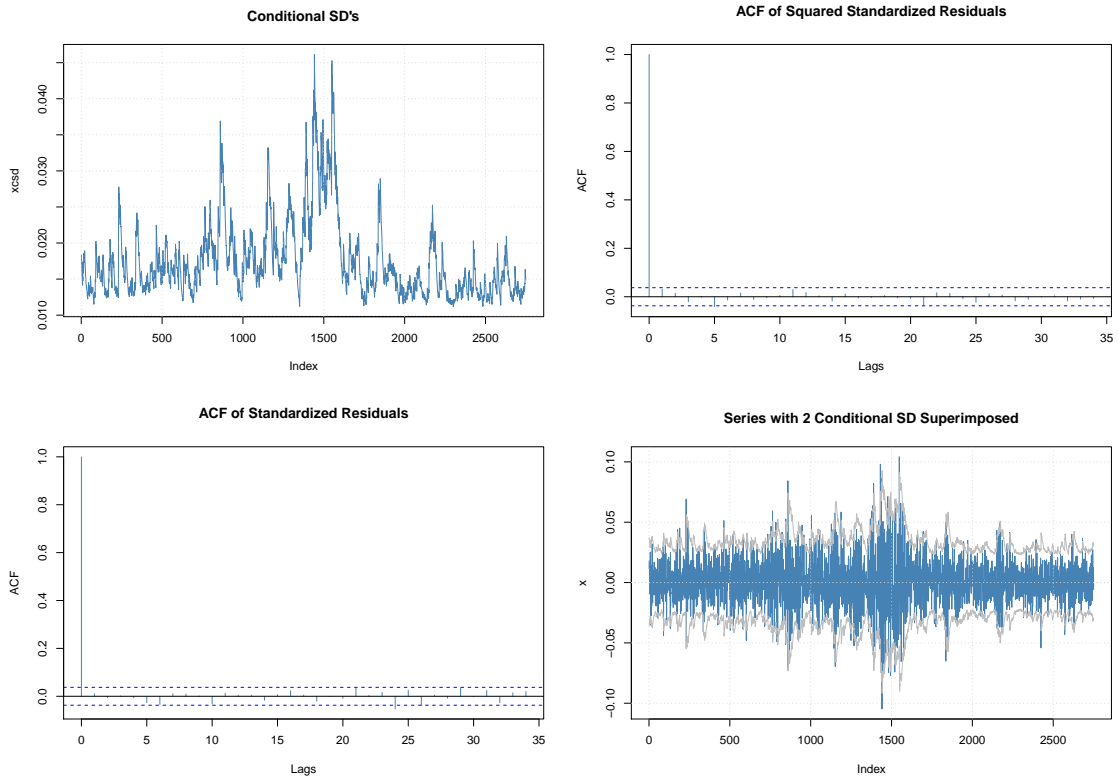


Figure 4.17: Residual Diagnosis plots

4.2.4 GARCH Extensions For Standard Bank Daily Data

Below are the GARCH extension models fitted for the data. All results are for models with student t distributed errors.

Table 4.40: Fitted Models

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
iGARCH(1,1)	-5.3364	-5.3278	7341.499	β_1 is NA
iGARCH(2,1)	-5.3356	-5.3249	7341.499	β_1 is NA, α_2
iGARCH(1,2)	-5.3359	-5.3252	7341.905	β_2
EGARCH(1,1)	-5.3365	-5.3236	7343.691	μ
EGARCH(1,2)	-5.3359	-5.3209	7343.902	$\mu \omega$
EGARCH(2,1)	-5.3358	-5.3185	7344.667	$\mu \alpha_1 \alpha_2 \gamma_2$
AR1+iGARCH(1,1)	-5.3383	-5.3275	7345.172	β_1 is NA
AR1+iGARCH(1,2)	-5.3378	-5.32490	7345.473	β_2 is NA
AR1+EGARCH(1,2)	-5.3376	-5.3204	7347.168	$\mu \omega$
AR1+EGARCH(1,1)	-5.3382	-5.3231	7347.037	μ
AR2+EGARCH(1,1)	-5.3420	-5.3248	7353.298	μ
AR4+EGARCH(1,1)	-5.3448	-5.3232	7359.038	AR3
TGARCH(1,1)	-5.3368	-5.3239	7344.161	μ
gjrGARCH(1,1)	-5.3410	-5.3281	7349.934	μ

The AIC BIC values suggest that the AR4+EGARCH(1,1) is the best model. But, we have to remember that our data had correlations so before we conclude that this model is the best we have to look at the residuals. Also, the best model here has many parameters which is not ideal as we seek a parsimonious model. We thus start by trying out the AR1 and AR2 EGARCH models. After that, we compare both their residuals and see if they fit the data. The plots shown in Figure 4.18 below are for the AR1+EGARCH(1,1) and AR2+EGARCH(1,1), the left-hand plots are for AR1+EGARCH(1,1). Clearly the Q-Q plots show that the t distribution fits the data well. However, the residuals ACF show presence of correlation hence, our model did not completely remove correlation. The squared residual ACF also shows lags that are significant hence, there still remain ARCH effects in the model.

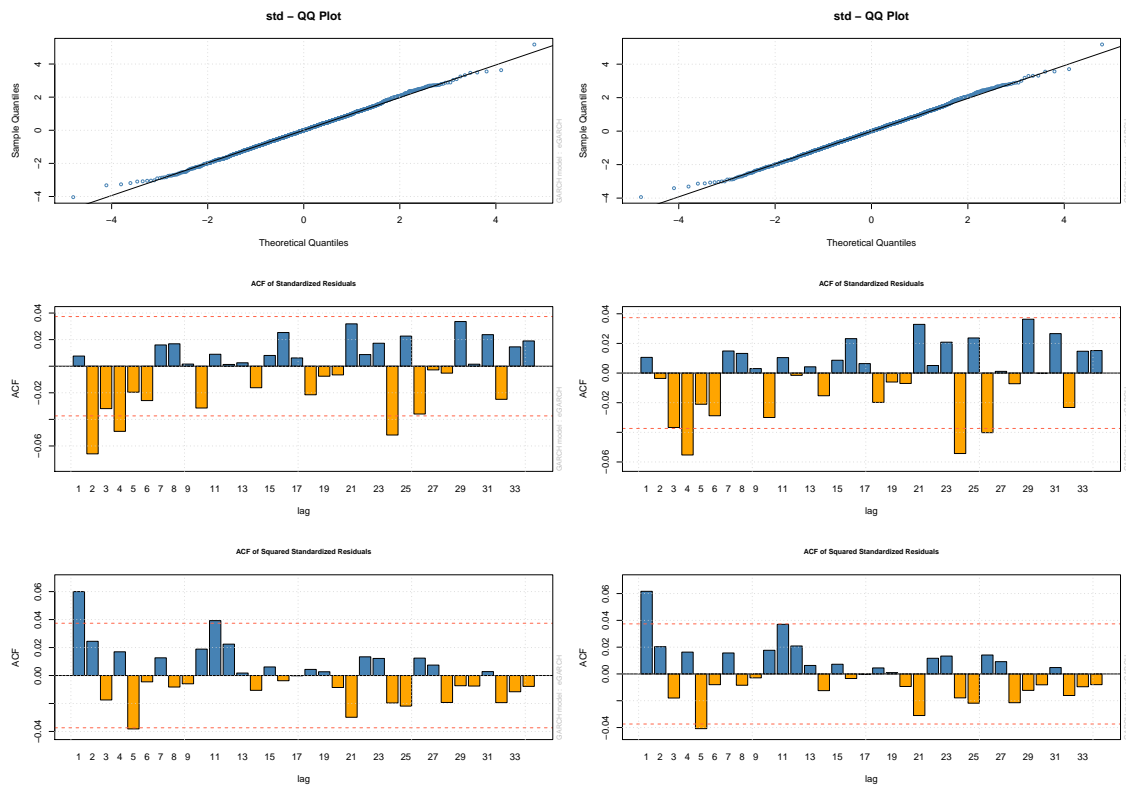


Figure 4.18: Residual Diagnosis plots

We thus, proceed to the suggested AR4+EGARCH(1,1). Estimates of the model are shown below. All parameters are significant at 5% level of significance except the AR3 which is significant at 7%.

Table 4.41: Estimates

	Estimate	Std. Error	t value	Pr(> t)
mu	0.000481	0.000230	2.0920	0.036436
AR1	-0.060957	0.025921	-2.3516	0.018692
AR2	-0.075556	0.020875	-3.6195	0.000295
AR3	-0.037446	0.019407	-1.9296	0.053661
AR4	-0.054638	0.019392	-2.8175	0.004840
omega	-0.170079	0.014270	-11.9190	0.000000
alpha1	-0.027208	0.012416	-2.1914	0.028423
beta1	0.979165	0.001733	564.9714	0.000000
gamma1	0.164110	0.022632	7.2513	0.000000
shape	9.588761	1.807683	5.3044	0.000000

Below are the results of the nyblom stability test. The results show that all parameters are significantly stable.

Table 4.42: Nyblom stability test

Joint Statistic:	3.8467
mu	0.1494
AR1	1.1742
AR2	0.0941
AR3	0.8888
AR4	0.2111
omega	0.5547
alpha1	0.3146
beta1	0.5660
gamma1	0.0688
shape	0.1012

The sign bias test below shows that there are no leverage effects.

Table 4.43: Sign Bias Test

	t-value	prob sig
Sign Bias	0.7599	0.4474
Negative Sign Bias	0.3405	0.7335
Positive Sign Bias	1.5868	0.1127
Joint Effect	2.7585	0.4304

Figure 4.19 below shows the diagnosis plots from the AR4+EGARCH(1,1) model. The conditional standard deviation plot confirms to expected series plot. The standardized residual plots have a few high order significant lags suggesting that not all correlation were removed. The squared standardized residual plot also has a few significant lags showing that our model did not completely remove ARCH effects.

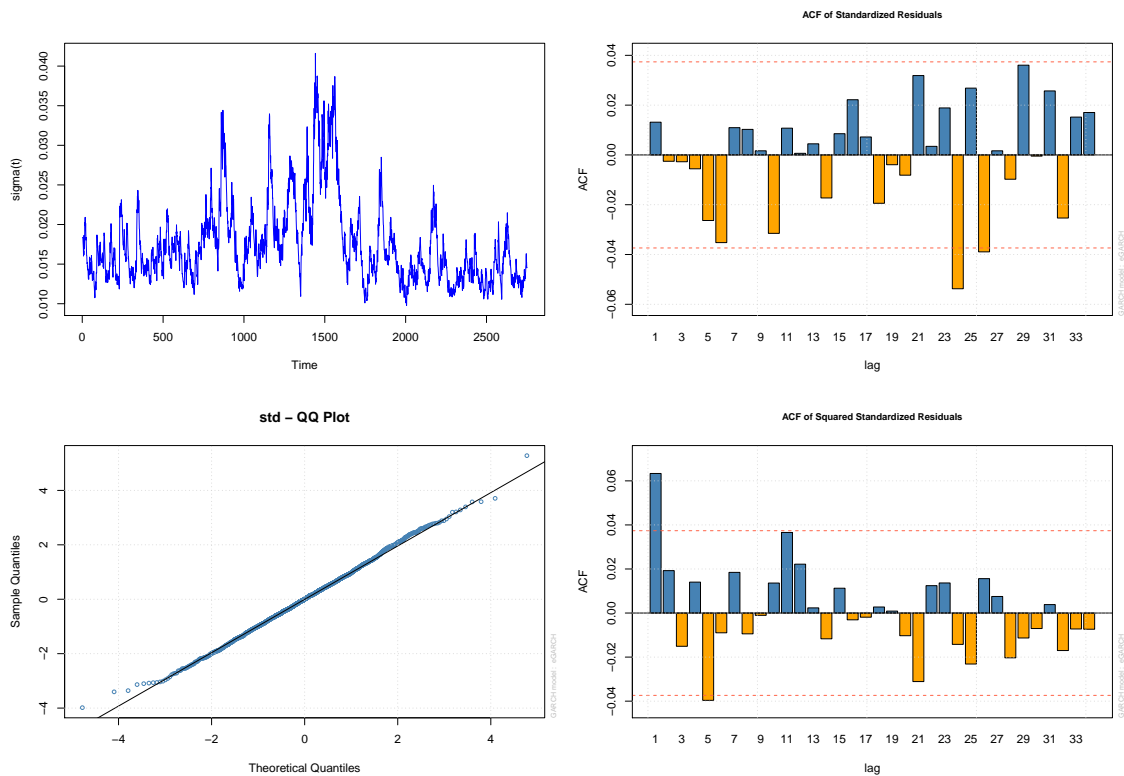


Figure 4.19: Residual Diagnosis plots

In conclusion, we see that our final model of choice gave us mixed outcomes. It gave us a nice Q-Q plot and took care of bias test but we realise it did not complete remove ARCH effects hence, in this data the previous direct AR4+GARCH(1,1) is a better model.

4.3 Liberty Holdings

4.3.1 Daily Data

From the data analysis chapter, the ACF of the returns showed the presence of correlations. The ACF of squared returns also showed the presence of ARCH effects. We look at statistical tests to justify the graphical evidence. Looking at the a_t series, $\chi^2 = 54.7628$, $df = 10$, $p\text{-value} = 3.497e-08$ thus, rejecting the null hypothesis of no correlations and we conclude that the data is indeed correlated. For ARCH effects a_t^2 we have $\chi^2 = 131.3941$, $df = 10$, $p\text{-value} < 2.2e-16$. Hence, there are ARCH effects present. The PACF of the squared returns suggest an AR3 plot to remove the correlations hence, we will try models with such amongst simpler others.

Table 4.44: Summary of fitted models:cond distribution=norm

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(3,0)	-5.4689	-5.4581	7524.689	All significant
GARCH(4,0)	-5.4683	-5.4554	7524.88	α_4 is NA
GARCH(1,1)	-5.5162	-5.5076	7588.754	All significant
GARCH(1,2)	-5.5245	-5.5138	7601.251	All significant
GARCH(2,1)	-5.5155	-5.5045	7588.753	α_2
AR3+GARCH(3,0)	-5.4730	-5.4558	7533.349	AR2 AR3
AR3+GARCH(1,1)	-5.5259	-5.5108	7605.093	AR2 AR3
AR2+GARCH(1,1)	-5.5243	-5.5113	7601.857	AR2
AR1+GARCH(1,1)	-5.5237	5.5130	7600.15	all significant
AR2+GARCH(1,2)	-5.5312	-5.5161	7612.423	AR2
AR1+GARCH(1,2)	-5.5305	-5.5176	7610.46	All significant
AR2+GARCH(2,1)	-5.5235	-5.5085	7601.859	α_2
AR1+GARCH(2,1)	-5.5230	-5.5101	7600.152	α_2

Based on the AIC BIC and parameter we select the AR1+GARCH(1,2) and AR1+GARCH(1,1) the values for these models are close with that of AR1+GARCH(1,2). The AIC values of the AR1+GARCH(1,2) is lower than the AR1+GARCH(1,1) but bearing in mind that it has more parameters such a small difference in AIC values can not offer much gain in model precision. An ideal choice would be the AR1+GARCH(1,1) which is more parsimonious but we compare the two models further to justify our choice.

Table 4.45: GARCH(1,1) model estimates from distribution=norm

Parameter	Estimate	Std. Error	t value	Pr(>t)	sig code
mu	6.867e-04	2.663e-04	2.579	0.0099	**
AR1	-9.512e-02	1.988e-02	-4.786	1.70e-06	***
omega	7.570e-07	3.023e-07	2.504	0.0123	*
alpha1	2.595e-02	3.352e-03	7.740	9.99e-15	***
beta1	9.720e-01	3.240e-03	300.026	< 2e-16	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Estimates shown in Table 4.45 above are for the AR1+GARCH(1,1) with normally distributed errors. All parameters are highly significant. The same applies to AR1+GARCH(1,2) (estimates not shown). Q-Q plots obtained from the two models are shown in Figure 4.20 below. The data points are way off from the axis giving a clear indication that the normal distribution assumption is inappropriate for the data at hand. Hence, no need to further write down AR1+GARCH(1,2) estimates. We hence, rerun the models with t distributed errors.

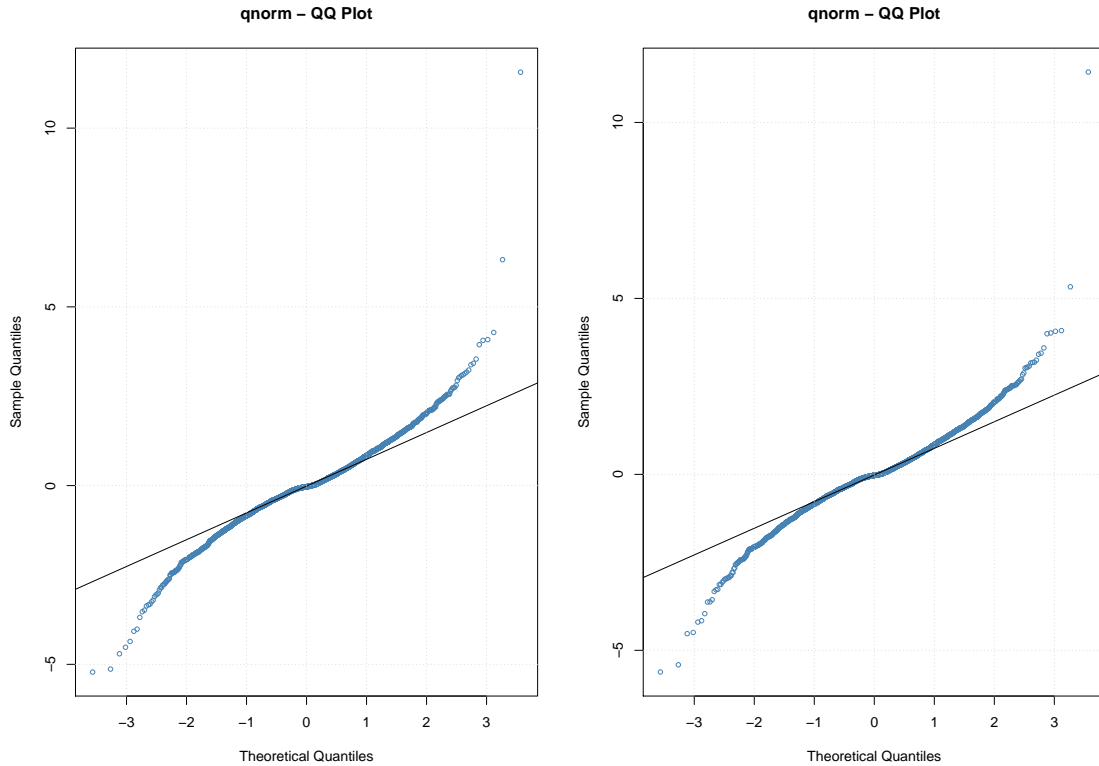


Figure 4.20: Q-Q plots

Table 4.46: Summary of fitted models:cond distribution=std

	AIC	BIC	Log likelihood	Non Sig. parameters at 5%
AR1+GARCH(1,1)	-5.6760	-5.6631	7810.504	μ AR1 sig. at 10%
AR1+GARCH(1,2)	-5.6756	-5.6605	7810.891	β_2 whilst μ AR1 β_1 sig. at 10%

After using student t distribution the AIC BIC values improved for both models. However, some of the parameters significance were sacrificed especially for the AR1+GARCH(1,2) were some parameters became insignificant even at 10%. For the AR1+GARCH(1,1) model, however, all parameters remained highly significant with only μ and the AR1 parameter dropping to significance at 10%. We, therefore, opt for the AR1+GARCH(1,1) as the model of choice. The estimates from the model are shown in Table 4.47 below.

Table 4.47: AR1+GARCH(1,1) Model estimates:cond distribution=std

	Estimate	Std. Error	t value	Pr(>—t—)	sig code
mu	3.852e-04	2.231e-04	1.727	0.08420	.
AR1	-3.379e-02	1.986e-02	-1.701	0.08887	.
omega	4.090e-05	1.355e-05	3.018	0.00254	**
alpha1	2.437e-01	6.251e-02	3.899	9.67e-05	***
beta1	6.567e-01	8.071e-02	8.136	4.44e-16	***
shape	3.529e+00	2.909e-01	12.133	< 2e-16	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

The $\alpha + \beta=0.9741$ is high and it goes hand in hand with that exhibited by the Old Mutual and Standard Bank returns. Table 4.48 below shows the results of the residual analysis.

Table 4.48: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	R Chi^2	64460.31	0
Shapiro-Wilk Test	R W	0.9000	0
Ljung-Box Test	R Q(10)	11.3105	0.3334
Ljung-Box Test	R Q(15)	21.9989	0.1078
Ljung-Box Test	R Q(20)	23.8511	0.2490
Ljung-Box Test	R^2 Q(10)	1.7057	0.9981
Ljung-Box Test	R^2 Q(15)	2.2628	0.9999
Ljung-Box Test	R^2 Q(20)	6.4907	0.9980
LM ARCH Test	R TR^2	1.9260	0.9995

The Jarque-Bera Test and the Shapiro-Wilk Test reject the normality assumption on residuals hence, justifies our leaving of the normality assumption. We will check the Q-Q plot of the student t distribution to see if the t distribution can model the residual better. The Ljung-Box Test on residuals up to lag 20 show that there is no correlation left in the series and the Ljung-Box Test on squared residuals also shows that the ARCH effects have been removed. This shows that our model is a good fit.

The Q-Q plots in Figure 4.21 below are for the model fitted with t distributed errors. Here the data is more confirmed to the axis except a few points at the extremes. Overall being an improvement to the normally distributed errors.

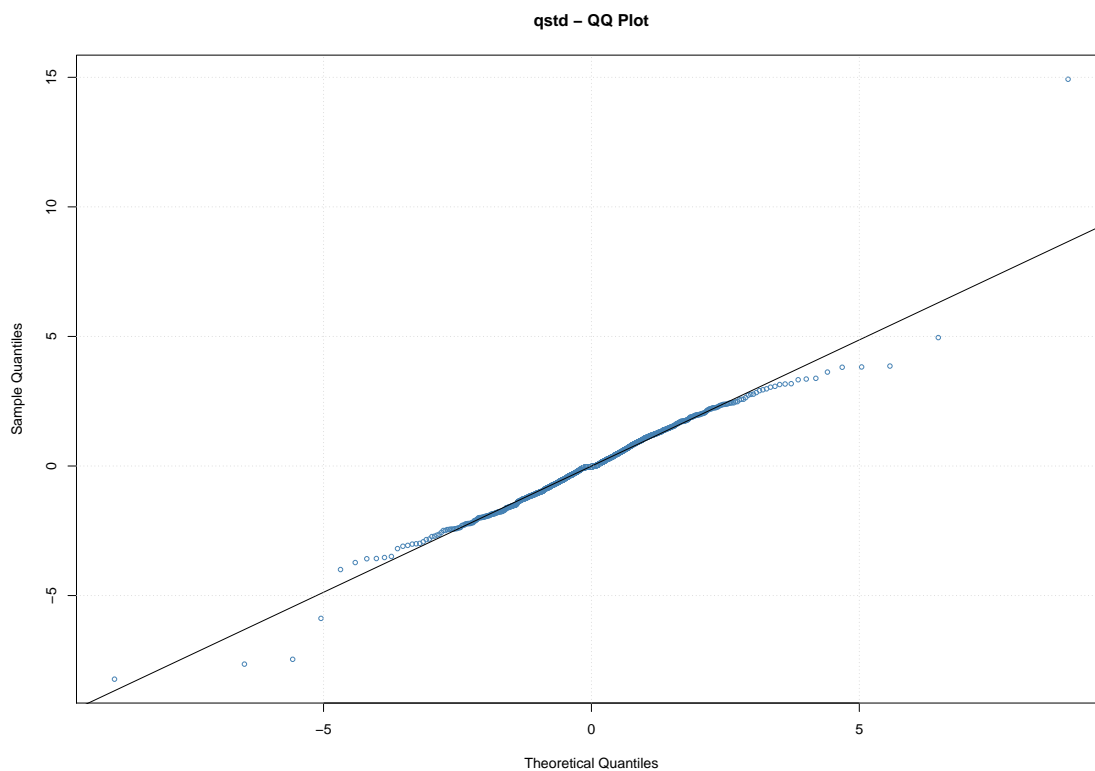


Figure 4.21: Q-Q plots for t distributed errors

Figure 4.22 shows the residual analysis plots. The ACF of residuals is not correlated as expected. The squared residual ACF also has no significant lags which show that the ARCH effects have been removed. The conditional standard deviation plots also, confirm to pattern suggested by the returns. Hence, all these justify our fit model.

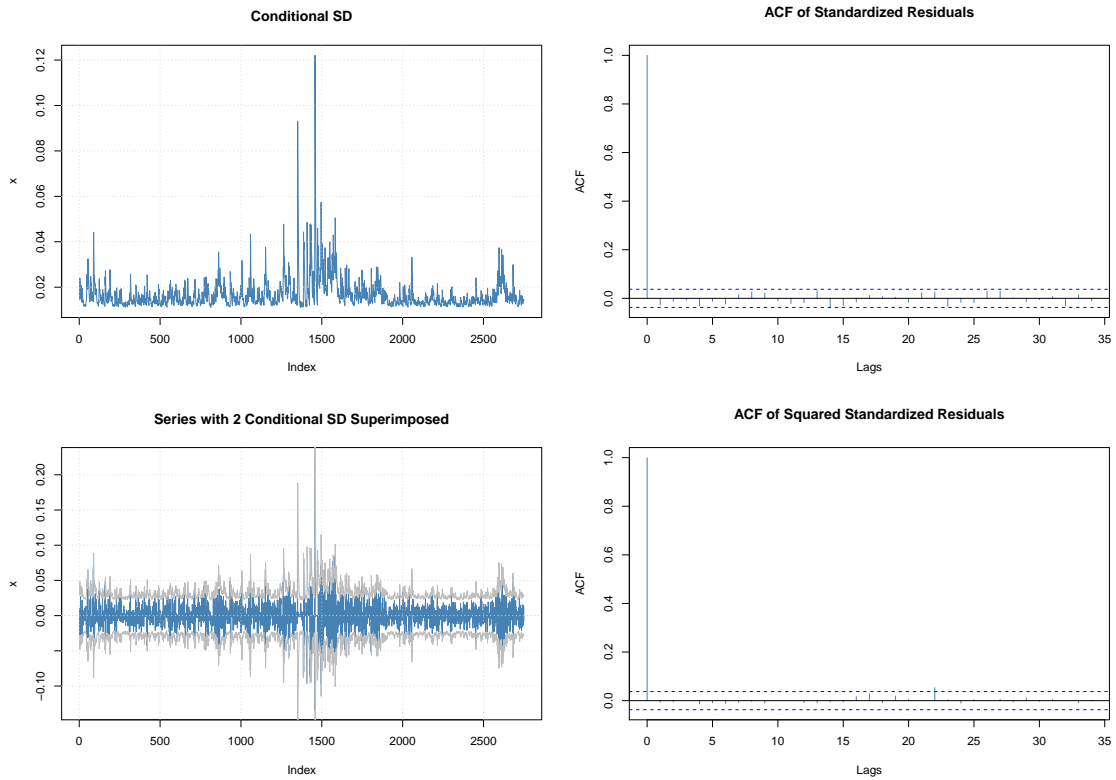


Figure 4.22: Residual Diagnosis plots

4.3.2 GARCH Extensions For Liberty Holdings Daily Data

Table 4.49: Fitted Models

Model	AIC	BIC	Log Likelihood	Non Significant parameters at 5%
iGARCH(1,1)	-5.5164	-5.5100	7588.11	ω β_1 NA
iGARCH(1,2)	-5.5246	-5.5160	7600.291	β_2 is NA
iGARCH(2,1)	-5.5157	-5.5071	7588.104	β_1 is NA α_2
iGARCH(1,1)	-5.4850	-5.4742	7546.82	all significant
EGARCH(1,2)	-5.4990	-5.4861	7567.085	all significant
EGARCH(2,1)	-5.5074	-5.4923	7579.62	α_1 α_2
AR1+iGARCH(1,1)	-5.5239	-5.5153	7599.429	β_1 is NA
AR1+EGARCH(1,1)	-5.4976	-5.4847	7565.153	all significant
AR2+EGARCH(1,1)	-5.4990	-5.4839	7568.087	all significant
gjrGARCH(1,1))	-5.5248	-5.5140	7601.546	μ
TGARCH(1,1)	-5.4742	-5.4635	7532.062	all significant

The best model from he models above is the AR2+EGARCH(1,1). IGARCH(1,2) had better AIC values but its parameters were not significant. We proceed to diagnose our model. The Q-Q plot in Figure 4.23 below shows that the tails are not well modelled by the normal assumption.

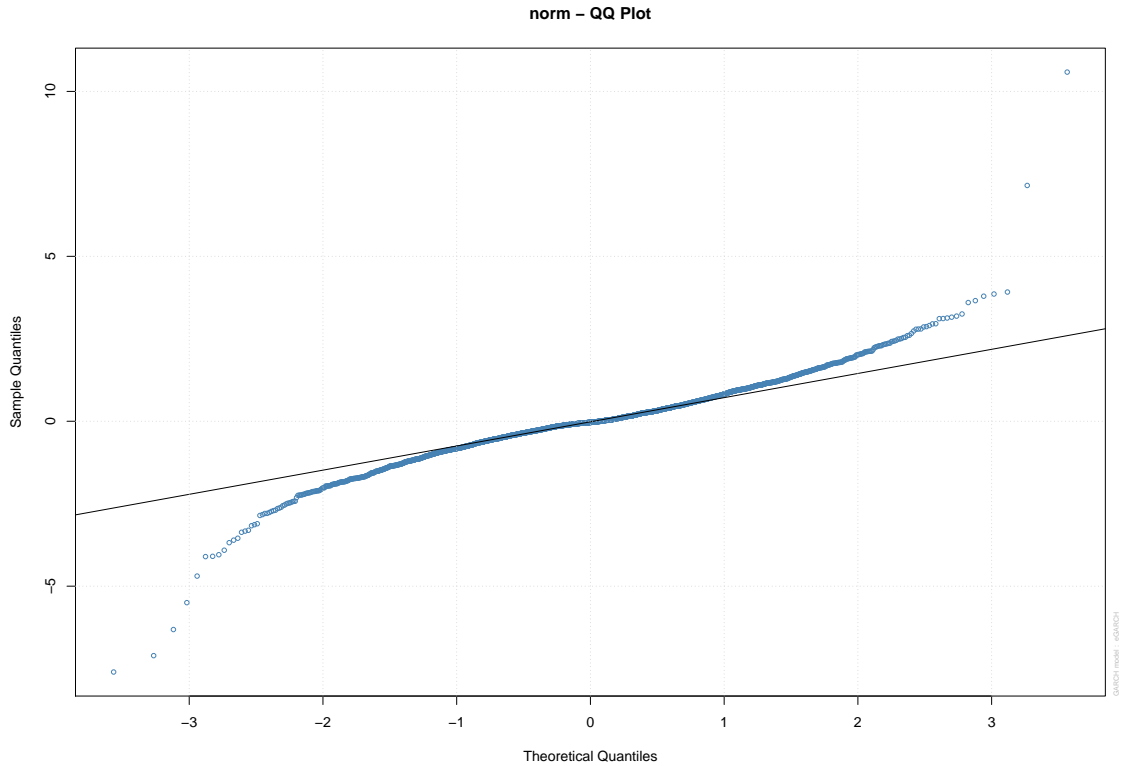


Figure 4.23: QQ plots

With the tails not well modelled, we then try the t distribution. Results gave the AR2 component to be insignificant so we dropped the AR2 component and modelled an AR1+EGARCH(1,1) model. The following diagnostics are for the AR1 model.

Table 4.50: AR1+EGARCH(1,1): model estimates from distribution=std

	Estimate	Std. Error	t value	Pr(> t)
mu	0.000379	0.000216	1.7569	0.078929
AR1	-0.034349	0.017477	-1.9653	0.049374
omega	-0.169312	0.007242	-23.3798	0.000000
alpha1	-0.038807	0.014347	-2.7049	0.006832
beta1	0.979338	0.000845	1159.6148	0.000000
gamma1	0.127384	0.019276	6.6084	0.000000
shape	3.480252	0.274296	12.6879	0.000000

The nyblom stability test shown below gives no evidence of unstable parameters hence, justifying all parameters in the model.

Table 4.51: Nyblom stability test

Joint Statistic	2.6154
mu	0.34844
AR1	0.33677
omega	0.09120
alpha1	0.45518
beta1	0.08489
gamma1	0.19228
shape	0.46269

Sign Bias Test below shows little evidence of leverage effect. Only the negative bias test is significant at 10%.

Table 4.52: Sign Bias Test

	t-value	prob sig
Sign Bias	0.4144	0.67860
Negative Sign Bias	1.7841	0.07452
Positive Sign Bias	1.2583	0.20839
Joint Effect	5.3317	0.14906

Below are the adjusted Pearson Goodness of fit test results. The normal assumption as expected rejected. This justifies the t distribution as shown also by the Q-Q plots.

Table 4.53: Adjusted Pearson Goodness-of-Fit Test:

group	statistic	p-value(g-1)
20	371.5	3.679e-67
30	494.2	4.541e-86
40	506.7	1.073e-82
50	535.0	6.370e-83

Figure 4.24 shows the diagnosis plots. The QQ plots show improvements when compared to the QQ plots of normal error assumption. The ACF of the residuals shows that correlations have been removed for all lags except a minor one at lag 1. ACF of squared residuals shows no significant lags hence, showing that ARCH effects have been removed. Only one higher order lag 22 is significant but this does not matter as we focus only on lower lags. The news impact curve shows that the model treats negative and positive shocks differently as expected. Negative shocks have a greater impact than positive shocks. The effect also increases with the magnitude of the shock which is good as in reality investor reaction changes with the magnitude of the bad or good news.

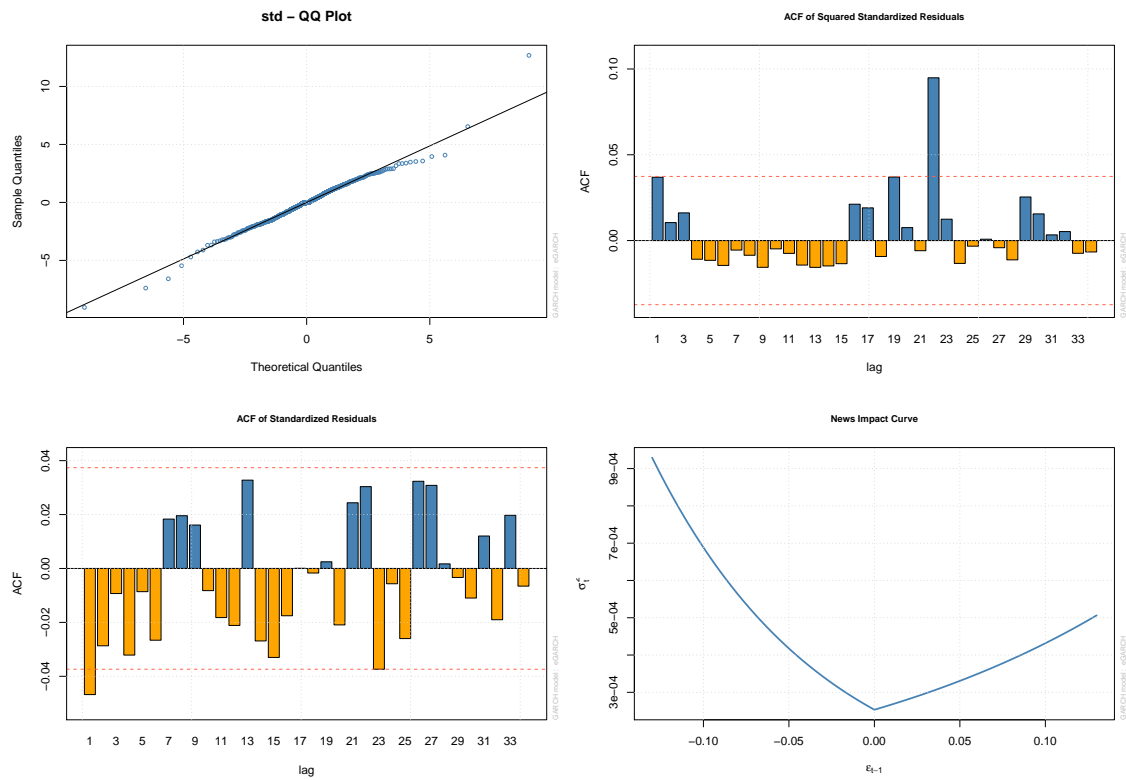


Figure 4.24: Residual Diagnosis plots

The conditional volatility plot is shown below. Here σ_t is typical around 0.01-0.025. Large volatility swings corresponding with the global crises can be seen with σ_t getting as high as 0.045.

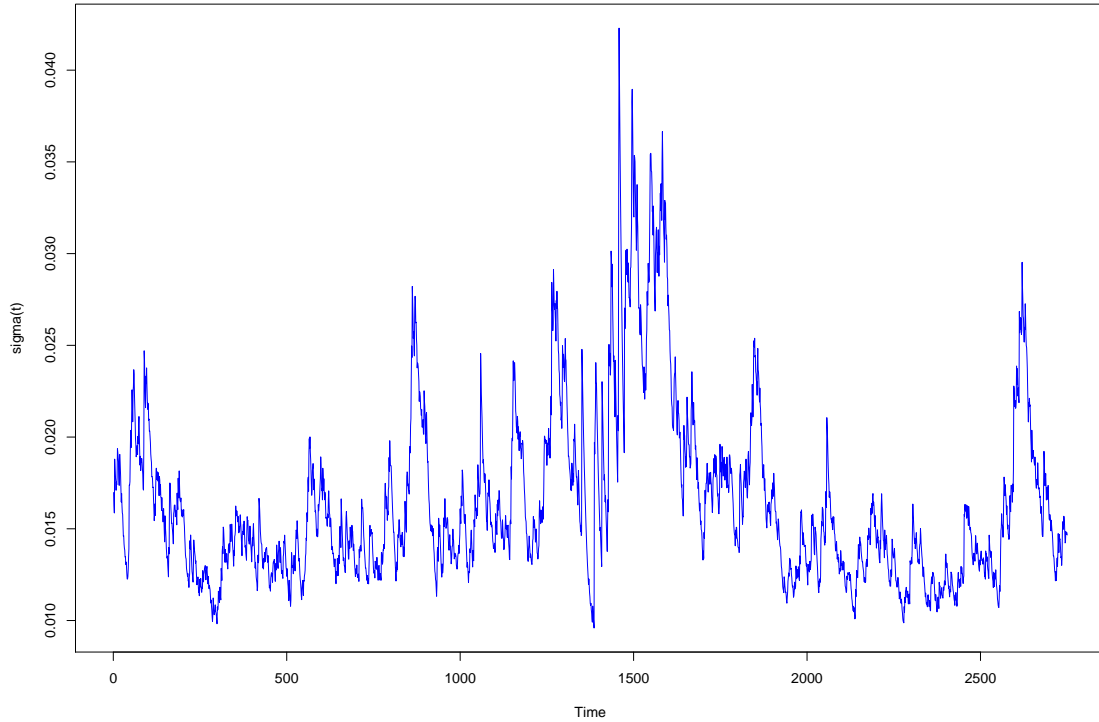


Figure 4.25: Conditional volatility plot

4.3.3 Monthly Data

The ACF for the log return series showed no correlations. Statistical test on return series confirm absence of serial correlations were a_t is $\chi^2 = 9.1622$, $df = 12$, $p\text{-value} = 0.689$. For ARCH effects a_t^2 we have $\chi^2 = 23.6574$, $df = 12$, $p\text{-value} = 0.02264$. Hence, we conclude that ARCH effects are present. We can proceed to fit our ARCH/GARCH models.

The PACF of the squared log returns suggests ARCH(3) model hence, we start with ARCH(3) model then keep modifying till we find the best fit. The table below shows the various models ran.

Table 4.54: Summary of fitted models:cond distribution=norm

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(1,0)	-2.3506	-2.3124	340.3176	$\alpha_1 \mu$
GARCH(2,0)	-2.3724	-2.3214	344.4464	α_1
GARCH(3,0)	-2.4357	-2.3720	354.5239	$\alpha_1 \alpha_2$
GARCH(1,1)	-2.4546	-2.4036	356.2373	ω
GARCH(2,1)	-2.4533	-2.3895	-2.3895	$\omega \alpha_1 \alpha_2$
GARCH(1,2)	-2.4625	-2.3988	358.3754	$\omega \beta_1$
GARCH(2,2)	-2.4463	-2.3698	357.0424	$\alpha_1 \alpha_2 \beta_1 \beta_2$

The GARCH(1,2) has the lowest AIC followed by the GARCH(1,1). However, the

GARCH(1,1) has a better BIC and also, in addition, has only one insignificant parameter. The table below shows the obtained estimates under normal error assumption.

Table 4.55: GARCH(1,1) model estimates from distribution=norm

mu	0.0097910	0.0038228	2.561	0.01043	*
omega	0.0001062	0.0000873	1.217	0.22353	
alpha1	0.0677737	0.0209588	3.234	0.00122	**
beta1	0.9113585	0.0256979	35.464	1.2e-16	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4.55 shows that all parameters are significant except for the parameter omega which is not significant at 5%. Figure 4.26 below shows the QQ plots under the normal error assumption. Clearly the tails are not well captured by this assumption. Hence, the model is re-fit with t and skewed t distributed errors.

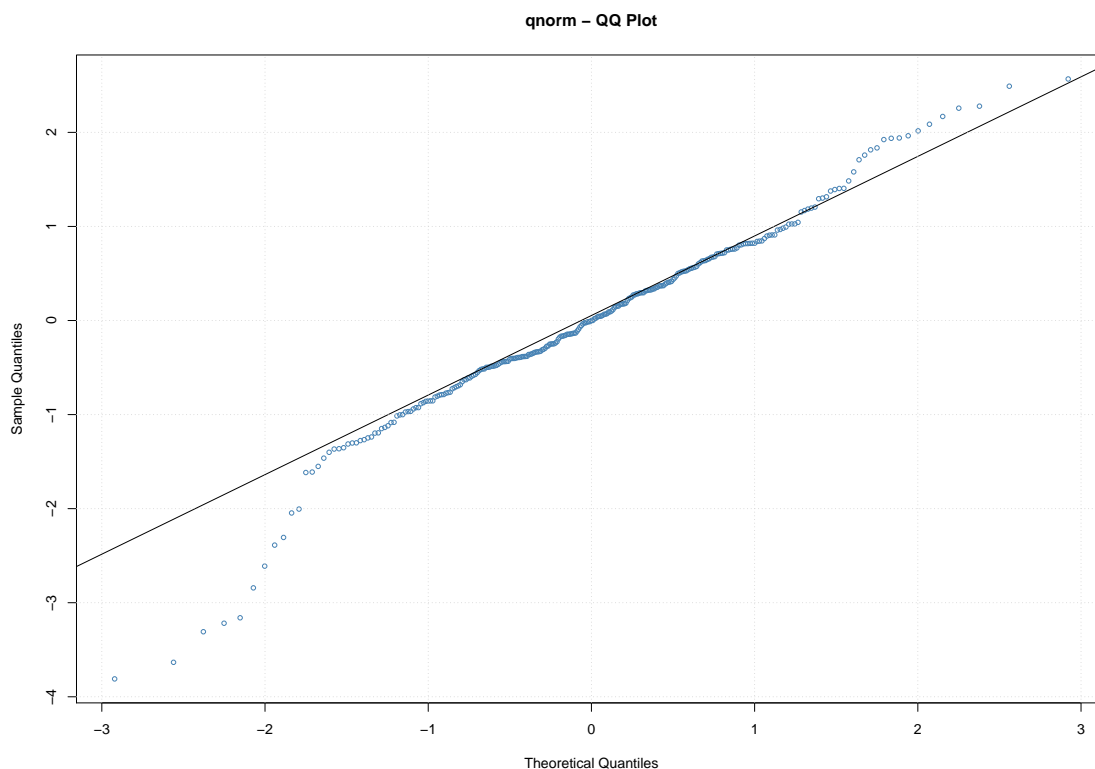


Figure 4.26: QQ plots

Table 4.56 below shows the GARCH(1,1) for the t and skewed t respectively.

Table 4.56: GARCH(1,1) with t and with skewed t

Model	AIC	BIC	Log Likelihood	Non Sig. parameters at 5%
GARCH(1,1)	-2.5139	-2.4502	365.7482	ω
GARCH(1,1)	-2.5085	-2.4320	365.9642	ω

The t and skewed t GARCH(1,1) also, have a non-significant omega but show an improvement in the AIC and BIC values. The t distributed model is the best model.

Figure 4.27 shows the QQ plots of the t and skewed t distributed error assumption models. The plots show an improvement in model fit. The heavy tails are better capture than there were under the normal assumption.

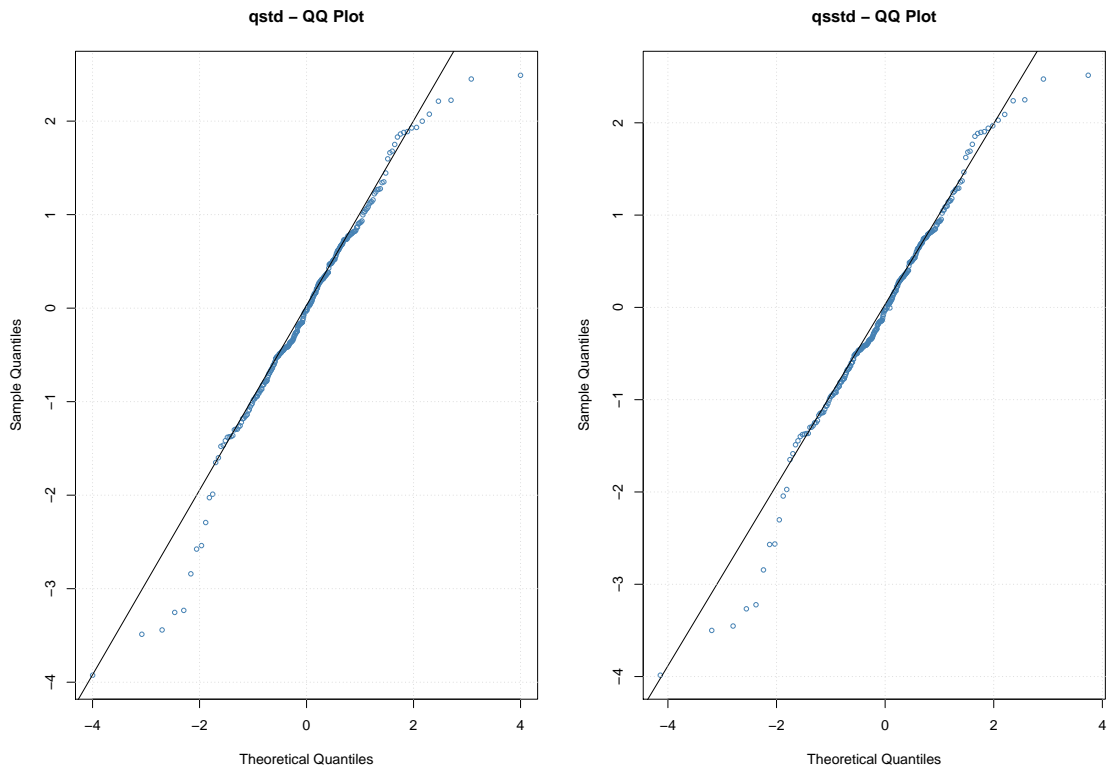


Figure 4.27: QQ plots

Table 4.57 below shows the estimates obtained from the GARCH(1,1) with t distributed errors. Omega is clearly insignificant. We can thus exclude it when reporting model. The persistence is $\alpha + \beta=0.6814$ which is lower than the one exhibited by Old Mutual and Standard Bank. This means that the Liberty Holdings volatility will take less time to revert back to its mean whenever it diverts away from it.

Table 4.57: GARCH(1,1) model estimates from distribution=std

	Estimate	Std. Error	t value	Pr(> t)	sig. code
mu	0.0111197	0.0035579	3.125	0.001776	**
omega	0.0001273	0.0001198	1.062	0.288144	
alpha1	0.0531257	0.0252567	2.103	0.035428	*
beta1	0.9213554	0.0350057	26.320	< 2e-16	***
shape	5.0903915	1.4591863	3.489	0.000486	***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

The residual tests below show that the Jarque-Bera Test and Shapiro-Wilk Test all reject the normality assumption hence, justifies the use of a t distribution. Ljung-Box tests for residual and residuals up to lags 10, 15 and 20 confirm that there is no correlation left and also, ARCH effects have been removed.

Table 4.58: Standardized Residuals Tests:

Test	Test Statistic	Test Statistic Value	P-value
Jarque-Bera Test	R Chi^2	52.075	4.92e-12
Shapiro-Wilk Test	R W	0.9667	3.75e-06
Ljung-Box Test	R Q(10)	10.0720	0.4342
Ljung-Box Test	R Q(15)	17.3010	0.3007
Ljung-Box Test	R Q(20)	19.1571	0.5116
Ljung-Box Test	R^2 Q(10)	9.9773	0.4425
Ljung-Box Test	R^2 Q(15)	22.1848	0.1030
Ljung-Box Test	R^2 Q(20)	30.3922	0.0640
LM ARCH Test	R TR^2	10.377	0.5829

The residual plots in Figure 4.28 show the ACF has no significant correlations as expected. The squared residual ACF also shows the removal of ARCH effects. The conditional standard deviation plots also conform to the pattern suggested by the returns. Hence, all these justify our fit model.

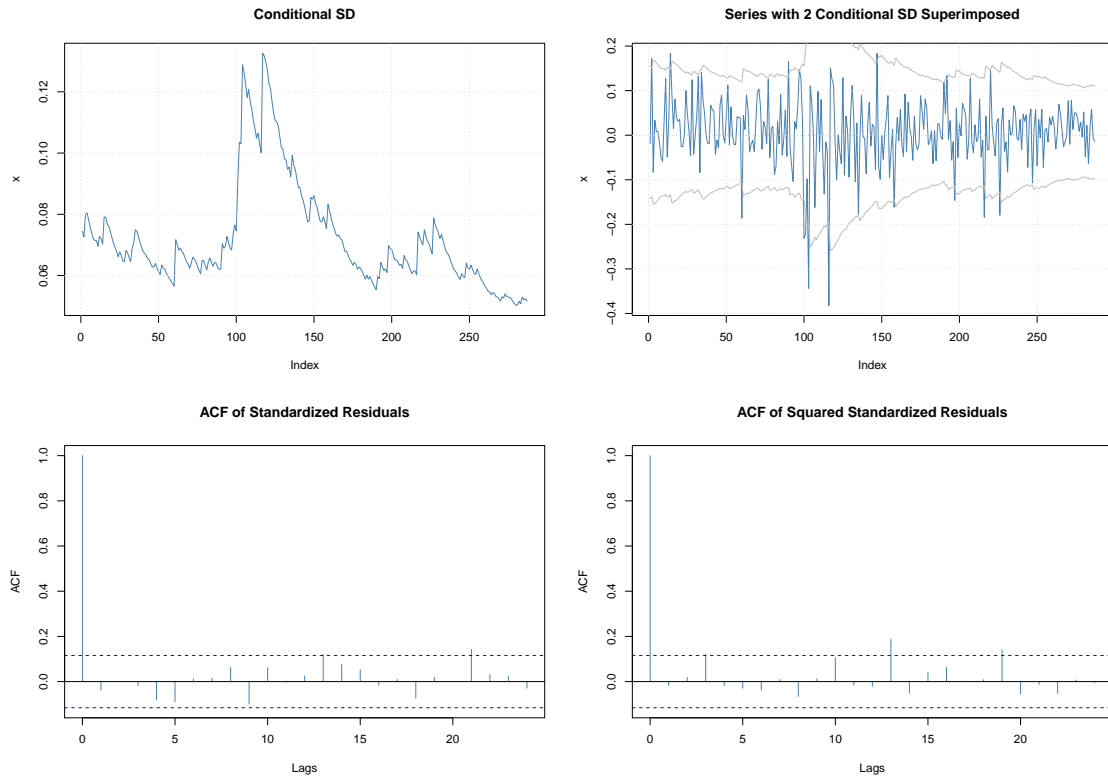


Figure 4.28: Residual Diagnosis plots

4.3.4 GARCH Extensions For Liberty Holdings Monthly Data

Model	AIC	BIC	Log Likelihood	Non Significant parameters at 5%
iGARCH(1,1)	-2.4568	-2.4185	355.5481	$\omega \beta_1$ NA
iGARCH(1,2)	-2.4638	-2.4128	357.558	$\omega \beta_1 \beta_2$ NA
iGARCH(2,1)	-2.4571	-2.4061	356.5931	$\alpha_1 \alpha_2 \omega \beta_1$ NA
EGARCH(1,1)	-2.4528	-2.3890	356.9729	$\omega \alpha_1$
EGARCH(1,2)	-2.5362	-2.4597	369.9446	$\mu \alpha_1 \alpha_2 \beta_1$
EGARCH(2,1)	-2.4572	-2.3680	359.6137	$\mu \alpha_1 \alpha_2 \gamma_1$
TGARCH(1,1)	-2.4432	-2.3795	355.606	$\omega \eta_1$
gjrGARCH(1,1)	-2.4538	-2.3901	357.1241	$\omega \eta_1$

Like other company data, the best model when considering AIC values and significance of parameters was the EGARCH(1,1). The EGARCH(1,2) had the best AIC values but it had too many insignificant parameters as compared to the EGARCH(1,1) which also, is better because of parsimony. The Q-Q plot shown in Figure 4.29 below and the Pearson goodness of fit rejected the normal distribution for the error terms. Hence, we fit the model with t distributed errors.

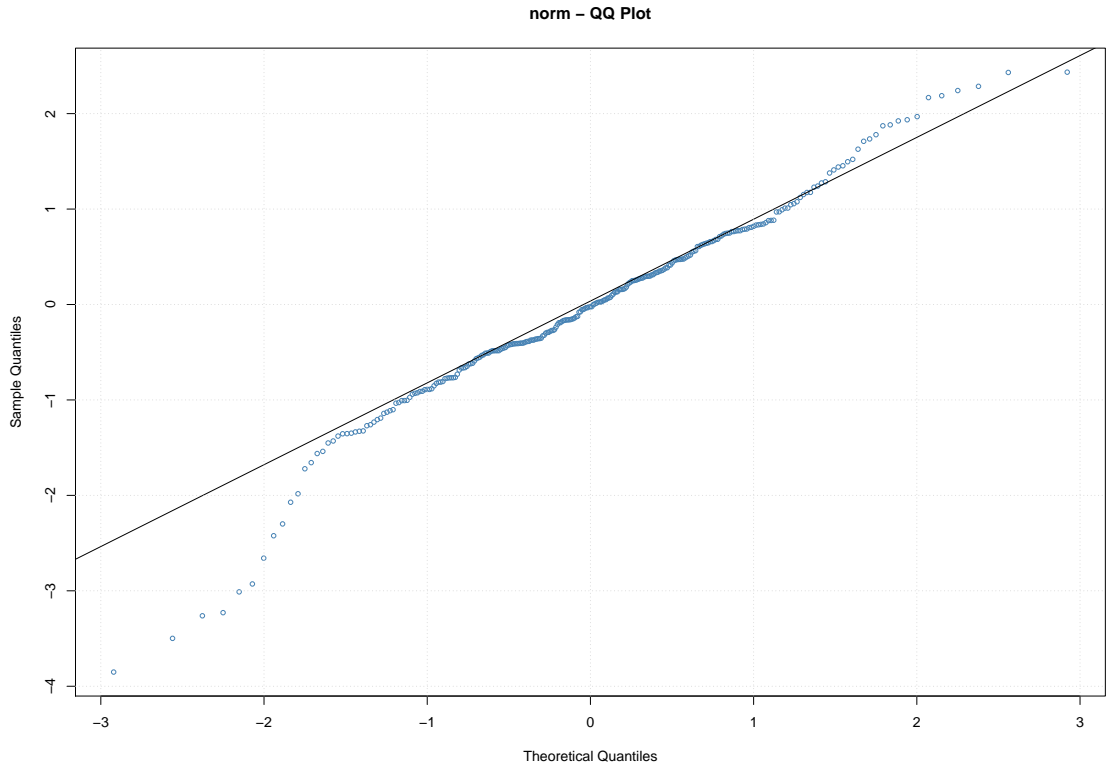


Figure 4.29: Q-Q plots

Table 4.59 shows the resulting model with t distributed errors, clearly the parameters remained as non-significant.

Table 4.59: Estimates from distribution=std

	Estimate	Std. Error	t value	Pr(>t)
mu	0.011707	0.003552	3.29582	0.000981
omega	-0.062621	0.054076	-1.15801	0.246860
alpha1	0.019416	0.050989	0.38078	0.703366
beta1	0.988824	0.010061	98.28175	0.000000
gamma1	0.131017	0.103505	1.26580	0.205584
shape	5.231721	1.554805	3.36487	0.000766

There is no evidence of serial correlation in squared residuals as seen by the non-significant of Weighted Ljung-Box Test on Standardized Squared Residuals below.

Table 4.60: Weighted Ljung-Box Test on Standardized Squared Residuals: d.o.f=2

	statistic	p-value
Lag[1]	0.07733	0.7809
Lag[2*(p+q)+(p+q)-1][5]	2.54860	0.4956
Lag[4*(p+q)+(p+q)-1][9]	3.75588	0.6304

Below are the Nyblom stability tests show that all parameters are stable except for the shape parameter.

Table 4.61: Nyblom stability test

Joint Statistic:	1.2792
mu	0.14853
omega	0.20388
alpha1	0.29453
beta1	0.21972
gamma1	0.12637
shape	0.02897

The sign bias test show no evidence of leverage effect.

Table 4.62: Sign Bias Test

	t-value	prob sig
Sign Bias	0.6111	0.5417
Negative Sign Bias	0.4122	0.6805
Positive Sign Bias	1.2227	0.2225
Joint Effect	1.8605	0.6019

Figure 4.30 shows the model diagnostics plots. The ACF plot of the residuals shows no presence of autocorrelation on most lags except lag 13 and 21 which is not worrying as they are fairly upper lags. The plot of squared residuals also shows no significant lags except, minor higher order ones at lag 13 and 21 which do not affect the conclusion that the ARCH effects have been removed. The significant lags are almost identical with ones for Old Mutual monthly lags of EGARCH(1,1) model. The plot of the conditional standard deviations vs returns shows that the returns and the conditional standard deviations take up similar fluctuation as expected. The Q-Q plot shows failing tail distribution which shows that even the t distribution failed to capture the heavy tails. However, it outperformed the normal distribution assumption.

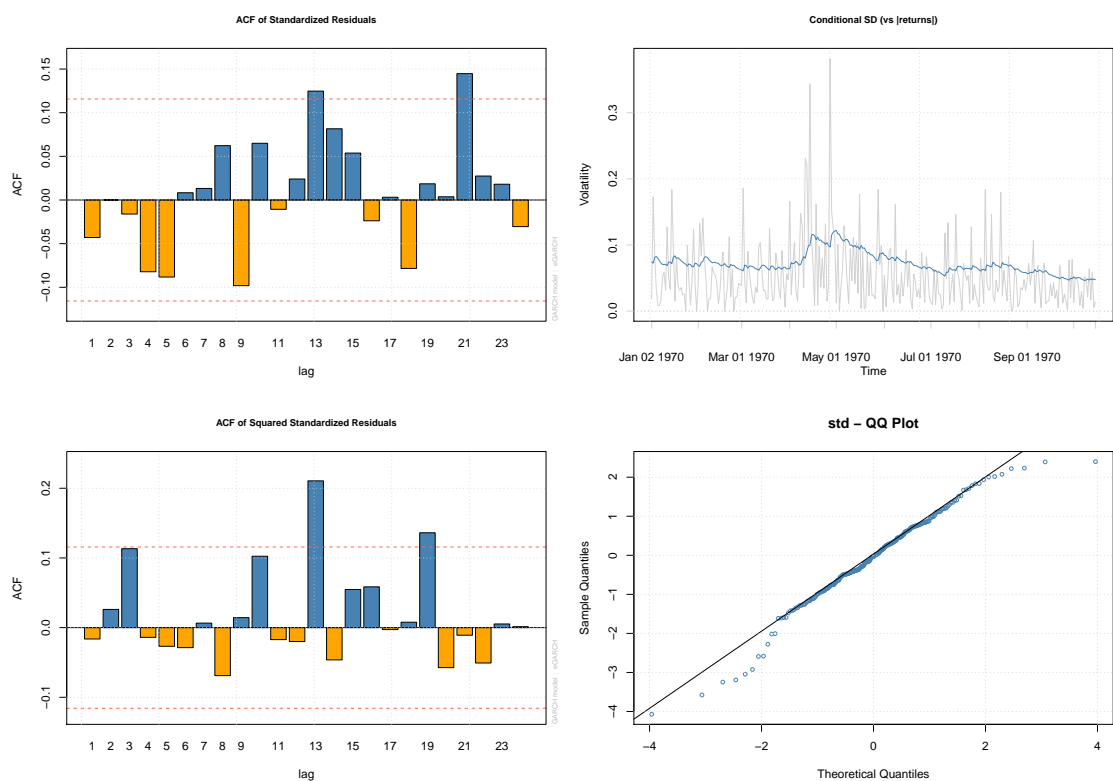


Figure 4.30: Diagnostics plots

4.4 Summary

This section dealt with the actual modelling of the data. For each data series, the first thing was to check for the presence of conditional heteroscedasticity. All of the series above had conditional heteroscedasticity hence, we proceeded to fit the ARCH and GARCH models. The daily data had some significant correlations as seen from the ACF plots hence, the need for autoregressive model (AR) to remove the autocorrelation before applying the GARCH model.

Initial models were ran assuming that ϵ_t are iid standard normal, however, model diagnosis rejected the normal error assumption which came as no surprise as from Chapter 3 we saw that the data is not normal distributed. Past studies have also shown that the financial data has heavy tails which require heavy-tailed distribution such as the student-t and the skew student-t distributions for error modelling. For all data, the GARCH model outperformed the ARCH models. For the GARCH models, the higher order GARCH models did not significantly improve the modelling ability. The additional parameters for the higher order GARCH models were in most cases insignificant. This observation of the GARCH(1,1) outperforming higher order GARCH is consistent with results of previous studies such as that of Ijumba (2013) and Talke (2003) amongst others. The GARCH model showed that the daily data volatility was more persistent than the monthly data volatility for Old Mutual and Standard Bank. For all data sets the α parameter which

determines reaction to shocks ranged between 0.02-0.08 for daily data and 0.05-0.1 for monthly data. For the GARCH extensions, the exponential GARCH was the best fit with cases of the Tgarch going hand to hand with the EGARCH.

The table below summarises the best ARMA-GARCH fit models for each specific data.

Table 4.63: Fitted GARCH models

Company	Model for monthly data	Model for daily data
Old Mutual	GARCH(1,1)	AR1+GARCH(1,1)
Standard Bank	GARCH(1,1)	AR4+GARCH(1,1)
Liberty Holdings	GARCH(1,1)	AR1+GARCH(1,1)

The table below shows the extension of the GARCH models that were fit for each data set. It can be seen that for the daily data we required an AR model to remove correlations before fitting the models as was done for the ARMA-GARCH models. These results show that the GARCH(1,1) and EGARCH are the best models.

Table 4.64: Fitted GARCH extension models

Company	Model for monthly data	Model for daily data
Old Mutual	EGARCH(1,1)	AR1+GARCH(1,1)
Standard Bank	TGARCH(1,1)	AR4+EGARCH(1,1)
Liberty Holdings	EGARCH(1,1)	AR1+EGARCH(1,1)

Chapter 5

Multivariate Volatility Models

Multivariate time series analysis considers the analysis of two or more time series simultaneously that is, two variables collected in time simultaneously. It is used when one wants to model and explain the interactions and co-movements among a group of time series variables. This is very important in the financial world where the world markets are becoming more and more integrated i.e formation of one global market. It is such that a sudden change in one market (especially the big markets) easily spreads and affects other markets. With this in mind, the study of multivariate time series to model the global markets has grown exponentially. (Caporin and McAleer 2010) attributes this increased attention to multivariate modelling due to increased availability of financial data, the increased computational powers of computers and the fact that the financial industry has begun to realize the possible advantages of these models.

The main focus of this chapter and those following is the modelling of the volatility of the multivariate time series. Here we look at time series that has conditional heteroscedasticity a phenomenon that is concurrent with financial data. We seek to be able to use these models to model the volatility across markets and hence, gain an understanding on how market factors push together market volatility.

Let \mathbf{z}_t be the $N \times 1$ multivariate time series defined as

$$\mathbf{z}_t = \mu_t + a_t \quad (5.1)$$

where $\mu_t = E(\mathbf{z}_t|F_{t-1})$ is the conditional expectation of \mathbf{z}_t given F_{t-1} . Then μ_t can follow one of the multivariate models such as the VAR model with nonlinear models also an option where necessary. The innovation component a_t (the shock) is the unpredictable component and it is given as

$$a_t = \Sigma_t^{\frac{1}{2}} \epsilon_t \quad (5.2)$$

where ϵ_t is a sequence of independent and identically distributed random vectors such that $E(\epsilon_t) = 0$ and $Cov(\epsilon_t) = \mathbf{I}_k$ and $\Sigma_t^{\frac{1}{2}}$ denotes the positive-definite square-root matrix of Σ_t Tsay (2014).

For series with conditional heteroscedasticity the $\Sigma_t = cov(a_t|F_{t-1})$ which is the conditional covariance matrix is time dependent. This time dependence is the subject matter of the multivariate volatility models. For a three series case the Σ_t is:

$$\Sigma_t = \begin{vmatrix} \sigma_{t,11} & \sigma_{t,12} & \sigma_{t,13} \\ \sigma_{t,12} & \sigma_{t,22} & \sigma_{t,23} \\ \sigma_{t,13} & \sigma_{t,23} & \sigma_{t,33} \end{vmatrix}.$$

Clearly, all variances are collected on the main diagonal, while all covariances are collected off the main diagonal. Moreover, because $\text{Cov}(x,y) = \text{Cov}(y,x)$ then from simple properties of expectations, Σ_t is by construction a symmetric matrix. This Σ_t is very important in finance, for example, any portfolio choice methodology is clearly based on knowledge or estimation of Σ_t as it is well known that optimal portfolio shares will also depend on the covariance of asset returns considered in pairs.

Tsay (2014) explains the difficulty of multivariate modelling as being in two ways. First he mentions the curse of dimensionality whereby he explains that as the dimension of the multivariate series increases then the more variance and covariance elements are required i.e the number of elements increases quadratically. The second being the difficulty of maintaining the positive definite assumption on the volatility matrix Σ_t as k increases. This assumption is important as it ensures that the portfolio variance is always positive regardless of the underlying portfolio. Therefore, there is a need for special attention as k increases.

In short volatility modelling typically consists of two sets of equations, with, the first set of equations governing the time evolution of the conditional mean μ_t , whereas the second set describes the dynamic dependence of the volatility matrix Σ_t . These two sets of equations are referred to as the mean and volatility equations respectively. In practice, the volatility matrix plays a crucial role in asset allocation and risk management.

5.0.1 Portmanteau Test

Just like the univariate case the portmanteau test exists for the multivariate time series case. If a_t has no conditional heteroscedasticity i.e linear dynamic dependence in the data, then its conditional covariance matrix Σ_t is time-invariant. This implies that Σ_t , hence, a_t^2 , does not depend on the a_{t-i}^2 for $i > 0$. Hence, we test the hypothesis

$$H_0 : \rho_1 = \dots = \rho_n = 0$$

against the alternative hypothesis

$$H_a : \rho_i \neq 0$$

for some i ($1 \leq i \leq m$), where ρ_i is the lag- i cross-correlation matrix of a_t^2 .

$$Q^*(m) = T^2 \sum_{i=1}^m \frac{1}{T-i} b_i'(\hat{\rho}_0^{-1} \otimes \hat{\rho}_0^{-1}) b_i \quad (5.3)$$

where T denotes the sample size, k is the dimension of a_t , and $b_i = \text{vec}(\hat{\rho}_i)$, \otimes is the kronecker product. $Q^*(m)$ is asymptotically distributed as χ^2 df.

5.0.2 Rank Based Test

The heavy tails shown by asset returns may in some cases affect the power of the Portmanteau test hence, the need for another test to compare with. Dufour and Roy (1985) developed a rank based test to overcome this. It considers the rank series of the standardized series ϵ_t . Let R_t be the rank of ϵ_t . The lag ℓ rank autocorrelation of ϵ_t can be defined as

$$\tilde{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (R_t - \bar{R})(R_{t-\ell} - \bar{R})}{\sum_{t=1}^T (R_t - \bar{R})^2}$$

where $\ell=1,2,\dots$ $\bar{R}_t = \sum_{t=1}^T R_t/T = (T+1)/2$
The test static being

$$Q_R(m) = \sum_{i=1}^m \frac{[\tilde{\rho}_i - E(\tilde{\rho}_i)]^2}{Var(\tilde{\rho}_i)} \quad (5.4)$$

This is distributed χ_m^2 asymptotically if ϵ_t has no serially dependance.

5.1 Diagnostic Checks Of Volatility Models

Diagnostics are done on the residuals $\hat{a}_t = z_t - \hat{\mu}_t$ so as to check the adequacy of the model. Where $\hat{\mu}_t$ is the fitted conditional mean of z_t .

5.1.1 Ling and Li Statistics

Ling and Li (1997) developed a method of diagnosis of multivariate time series which unlike other approaches such as (Box and Pierce) and (Mcleod and Li) they did not use residuals autocorrelations but rather used the sum of the squared residual autocorrelations to develop several new portmanteau statistics.

Assume that the innovation ϵ_t also satisfies

- (a) $E(\epsilon_{it}^3) = 0$ and $E(\epsilon_{it}^4) = c1 < \infty$ for $i=1, \dots, k$, and
- (b) ϵ_{it} and ϵ_{jt} are mutually uncorrelated up to the fourth order for $i = j$. Ling and Li (1997) employed \hat{a}_t to propose a model checking statistic for volatility models. Let

$$\hat{\epsilon}_t = \hat{a}_t' \hat{\Sigma}_t^{-1} \hat{a}_t \quad (5.5)$$

be a transformed quadratic residual series. If the fitted model is correctly specified, then, by the ergodic theorem,

$$\frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t = \frac{1}{T} \sum_{t=1}^T \hat{a}_t' \hat{\Sigma}_t^{-1} \hat{a}_t \rightarrow a.s E(\hat{a}_t' \hat{\Sigma}_t^{-1} \hat{a}_t) = E(\epsilon_t' \epsilon_t) = k \quad (5.6)$$

where $\rightarrow a.s$ denotes almost sure convergence or convergence with probability 1. The lag-1 sample autocorrelation of ϵ_t , therefore, can be defined as

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (\hat{\epsilon}_t - k)(\hat{\epsilon}_{t-\ell} - k)}{\sum_{t=1}^T (\hat{\epsilon}_t - k)^2} \quad (5.7)$$

If the model is correctly specified

$$\frac{1}{T} \sum_{t=1}^T (\hat{\epsilon}_t - k)^2 \rightarrow \text{a.s. } E(a_t' \Sigma_t^{-1} a_t - k)^2 \text{ as } n \rightarrow \infty$$

and

$$E(a_t' \Sigma_t^{-1} a_t - k)^2 = E(\epsilon_t' \epsilon_t)^2 - k^2 = E(\epsilon_{it}^4 - 1)k = ck$$

where $c = E(\epsilon_{it}^4) - 1$. Since the denominator of $\hat{\rho}_\ell$ in Equation (5.6) converges to a constant, it suffices to consider the numerator in studying the limiting properties of $\hat{\rho}_\ell$. By letting

$$\hat{C}_\ell = \frac{1}{T} \sum_{t=\ell+1}^T (\hat{\epsilon}_t - k)(\epsilon_{t-1} - k) \quad (5.8)$$

be the lag ℓ sample autocovariance of the transformed residual $\hat{\epsilon}_t$ and C_ℓ be its theoretical counterpart with $\hat{\epsilon}_t$ replaced by $\epsilon_t = a_t' \Sigma_t^{-1} a_t$. To investigate the properties of $\hat{\rho}_\ell$ as a function of the estimate $\hat{\theta}$, Taylor series of expansion is used on $\hat{C}_\ell \approx C_\ell + \frac{\partial C_\ell}{\partial \theta}$ as seen in Ling and Li (1997) and Tsay (2014).

Finally the test static is given as

$$Q_u(m) = T \rho_m' \hat{\Omega}^{-1} \hat{\rho}_m \quad (5.9)$$

5.2 Volatility Models

Various models for modelling multiple time series exist we discuss the most common models which are :

- Exponentially Weighted Moving Average (EWMA).
- Dynamic Conditional Correlation Models (DCC-GARCH).
- Go-GARCH Model.

5.2.1 Exponentially Weighted Moving Average (EWMA)

Another way to capture the dynamic features of volatility is to use an exponential moving average of historical observations where the latest observations carry the highest weight in the volatility estimate. This approach has two important advantages over the equally weighted model. Firstly, volatility reacts faster to shocks in the market as recent data carry more weight than data in the distant past. Secondly, following a shock (a large abnormal return that stands out from the rest), the volatility declines exponentially as the weight of the shock observation falls (JP Morgan 1995).

The EWMA covariance is defined recursively by Sheppard (2013) as

$$\Sigma_t = (1 - \lambda) a_{t-1} \hat{a}_{t-1}' + \lambda \Sigma_{t-1} \quad (5.10)$$

Correlations of the EWMA are given as

$$\rho_{t,ij} = \frac{\lambda\sigma_{t-1,ij} + (1-\lambda)r_{t-1,i}r_{t-1,j}}{\sqrt{[\lambda\sigma_{t-1,i}^2 + (1-\lambda)r_{t-1,i}^2][\lambda\sigma_{t-1,j}^2 + (1-\lambda)r_{t-1,j}^2]}} \quad (5.11)$$

where $0 < \lambda < 1$ denotes the decaying rate or the persistence parameter, where $\sigma_{t,i}^2$ and $\sigma_{t,j}^2$ are the variances for the individual series $r_{t,i}$ and $r_{t,j}$ respectively.

If one starts the recursion with a positive-definite matrix Σ_0 , then the volatility matrix Σ_t is positive-definite for all t (Tsay 2014). A suitable choice is to set to the average covariance over the first m days for some $m > k$ or could be set to the full sample covariance (Sheppard 2013). The single parameter, λ is usually set to .94 for daily data and .97 for monthly data based on recommendations from Risk Metrics (JP Morgan 1995). (Tsay 2014) states that this model is parsimonious and the resulting volatility matrices are easy to update, but, it tends to be rejected by diagnostic checks in an application. This he says is not surprising since it relies on a single decaying parameter to adequately govern the time decay of all conditional variances and covariances. Ijumba (2013) further states that the limitation of the EWMA model is its restrictiveness due to the simplicity of its structure and the assumption of non-estimated λ . Moreover, the fact that λ is identical to all assets and time periods is not realistic.

5.2.2 Dynamic Conditional Correlation Models (DCC-GARCH)

Considering the fact that constant conditional correlations over time is not realistic, researchers seek to generalize Bollerslevs Constant Conditional Correlation (CCC) model. Engle (2002) proposed a new class of estimator that both preserves the ease of estimation of Bollerslev's constant correlation model yet allows for correlations to change over time. Let $\Sigma_t = [\sigma_{ij,t}]$ be the volatility matrix of a_t given $F_{t,1}$, which denotes the information available at time $t - 1$. Engle's dynamic conditional correlation structure is defined as follows:

$$r_t = \mu_t + a_t \quad (5.12)$$

$$a_t = \Sigma_t^{\frac{1}{2}} \epsilon_t \quad (5.13)$$

$$\Sigma_t = D_t R_t D_t \quad (5.14)$$

where :

r_t : $n \times 1$ vector of log returns of n assets at time t .

a_t : $n \times 1$ vector of mean-corrected returns of n assets at time t , i.e. $E[at]=0$. $Cov[at] = Ht$.

μ_t : $n \times 1$ vector of the expected value of the conditional r_t .

Σ_t : $n \times n$ matrix of conditional variances of a_t at time t .

D_t : $n \times n$, diagonal matrix of time varying standard deviations from univariate GARCH models with $\sqrt{\sigma_{i,t}}$ on the on the i th diagonal

R_t : $n \times n$ conditional correlation matrix of a_t at time t .

ϵ_t : $n \times 1$ vector of iid errors such that $E[\epsilon_t]=0$ and $E[\epsilon_t\epsilon_t^T] = I$.

$$D_t = \begin{vmatrix} \sigma_{1t} & 0 & 0 \\ 0 & \sigma_{2t} & 0 \\ 0 & 0 & \sigma_{3t} \end{vmatrix}$$

where

$$\sigma_{it}^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-1}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (5.15)$$

Note that the univariate GARCH models can have different orders. Often the simplest model, GARCH(1,1), is adequate. The specification of the univariate GARCH models is not limited to the standard univariate GARCH(p,q) in Chapter 3, but can include any GARCH process with Gaussian-distributed errors that satisfies appropriate stationarity conditions that ensure the unconditional variance to exist.

Let $\gamma_t = (\gamma_1, \dots, \gamma_k)'$ be the marginally standardized innovation vector, where $\gamma_{it} = a_{it} \sqrt{\sigma_{ii,t}}$. Then, R_t is the conditional correlation matrix of the standardized disturbances γ_t , i.e: $\gamma_t = D_t^{-1} a_t \sim N(0, R_t)$.

Since R_t is a correlation matrix it is symmetric.

$$R_t = \begin{vmatrix} 1 & \rho_{12,t} & \rho_{13,t} \\ \rho_{21,t} & 1 & \rho_{23,t} \\ \rho_{13,t} & \rho_{23,t} & 1 \end{vmatrix} = \sigma_{it} I$$

Hence, elements of $\Sigma_t = D_t R_t D_t$ are as follows : $\Sigma_t = \sqrt{\sigma_{it} \sigma_{jt} \rho_{ij}}$ where $\rho_{ii} = 1$.

Orskaug (2009) explains that, since R_t exists in different forms hence, when specifying a form of R_t two requirements have to be considered:

1. Σ_t has to be positive definite because it is a covariance matrix. To ensure Σ_t to be positive definite, R_t has to be positive definite (D_t is positive definite since all the diagonal elements are positive).
2. All the elements in the correlation matrix R_t have to be equal to or less than one by definition.

To ensure both of these requirements in the DCC-GARCH model, R_t is decomposed into:

$$R_t = Q_t^* Q_t Q_t^* \quad (5.16)$$

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 Q_{t-1} + \theta_2 \gamma_{t-1} - \gamma_{t-1}^T \quad (5.17)$$

where $Q = \text{Cov}[\gamma_t \gamma_{t-1}^T] = E[\gamma_{t-1} \gamma_{t-1}^T]$ is the unconditional covariance matrix of the standardized errors γ_{t-1} . Q can be estimated as :

$$Q = \frac{1}{T} \sum_{t=1}^T \gamma_{t-1} \gamma_{t-1}^T \quad (5.18)$$

The parameters a and b are scalars, and Q_t^* is a diagonal matrix with the square root of the diagonal elements of Q_t at the diagonal. Q_t^* rescales the elements in Q_t to ensure the second requirement $|\rho_{ij}| = \left| \frac{q_{ijt}}{\sqrt{q_{iit}q_{jtt}}} \right| \leq 1$

In addition to the conditions for the univariate GARCH model to ensure positive unconditional variances, given earlier, the scalars a and b must satisfy: $\theta_1 \geq 0$, $\theta_2 \geq 0$ and $\theta_1 + \theta_2 < 1$.

Tse and Tsui (2002) proposed a second type of DCC models which are given as:

$$R_t = (1 - \theta_1 - \theta_2) \bar{R}_t + \theta_1 R_{t-1} + \theta_2 \psi t - 1 \quad (5.19)$$

where $\bar{\rho}_t$ is the unconditional correlation matrix of η_t . θ_i are non-negative real numbers satisfying additional constraint $0 < \theta_1 + \theta_2 < 1$, and $\psi t - 1$ is a $n \times n$ matrix whose elements are functions of the lagged observations of y_t .

Tsay (2014) explains that both models start with the unconditional covariance matrix of γ_t . However, they differ in the way local information at time $t-1$ is used. The DCC model of Engle (2002) uses γ_t only so that Q_t must be re-normalized at each time index t . On the other hand, the DCC model of Tse and Tsui (2002) uses local correlations to update the conditional correlation matrices. Tsay (2014) further points out that DCC models are extremely parsimonious as they only use two parameters θ_1 and θ_2 to govern the time evolution of all conditional correlations regardless of the number of assets k . This simplicity is both an advantage and a weakness of the DCC models. It is an advantage because the resulting models are relatively easy to estimate. It is a weakness of the model because it is hard to justify that all correlations evolve in the same manner regardless of the assets involved. Tsay (2014) goes on to say “experience albeit limited, indicates that a fitted DCC model is often rejected by diagnostic checking”.

Estimation of DCC-GARCH

Orskaug (2009) gives the estimation procedure for DCC models. When the standardized errors, z_t , are multivariate Gaussian distributed, the joint distribution of z_1, \dots, z_T is

$$f(z_t) = \prod_{t=1}^T \frac{1}{(2\Pi)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} z_t z_t^T\right\} \quad (5.20)$$

Here $t = 1, \dots, T$ is the time period used to estimate the model.

Using the rule for linear transformation of variables, the likelihood function for $a_t = \Sigma_t^{\frac{1}{2}} z_t$ is

$$L(\theta) = \prod_{t=1}^T \frac{1}{(2\Pi)^{\frac{1}{2}}} |\Sigma_t^{\frac{1}{2}}| \exp\left\{-\frac{1}{2} a_t^T \Sigma_t^{-1} a_t\right\} \quad (5.21)$$

where θ denotes the parameters of the model. Let the parameters, θ , be divided in two groups; $(\phi, \psi) = (\phi_1, \dots, \phi_n, \psi)$, where $\phi_i = (\alpha_{0i}, \alpha_{1i}, \dots, \alpha_{qi}, \beta_{1i}, \dots, \beta_{pi})$ are the parameters of the univariate GARCH model for the i th asset series, $i = 1, \dots, n$. $\psi = (a, b)$ are the parameters of the correlation structure in 5.17.

By taking the logarithm of (5.21) and substituting $\Sigma_t = D_t R_t D_t$ we get the log-likelihood:

$$\begin{aligned}
\ln(L(\theta)) &= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\Pi) + \ln(|\Sigma_t| + a_t^T \Sigma_t^{-1} a_t)) \\
&= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\Pi) + \ln(|D_t R_t D_t|) + a_t^T D_t^{-1} R_t^{-1} D_t^{-1} a_t) \\
&= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\Pi) + 2 \ln(|D_t|) + \ln(R_t) + a_t^T D_t^{-1} R_t^{-1} D_t^{-1} a_t)
\end{aligned} \tag{5.22}$$

This method of exact estimation is difficult, and hence, the DCC model was designed to allow for two-stage estimation. In the first stage the parameter ϕ of the univariate GARCH models are estimated for each asset series. The likelihood used in the first stage results in replacing R_t with the identity matrix I_n . In the second stage, the parameter ψ are estimated using the correctly specified log-likelihood in (5.22), given the parameter ϕ . Further details of the estimation are presented in Orskaug (2009). Estimation with Multivariate Students t-distributed errors and Multivariate skew Students t-distributed errors are also presented.

5.2.3 Go-GARCH Model

This model employs the concept of orthogonal transformation. This concept seeks to reduce the curse of dimensionality since it is usually very hard to estimate multivariate GARCH models. In practice, alternative methodologies for obtaining the covariance matrix are needed. The orthogonal approach addresses this problem by linearly transforming the observed returns matrix into a set of portfolios with the key property that they are uncorrelated, implying that, we can forecast their volatilities separately. The most commonly used orthogonal transformation in statistics is the principal component analysis (PCA) and for non-Gaussian data, the independent component analysis (ICA) is available to perform the transformation.

Van der Weide (2002) adopts the concept of ICA to propose a class of generalized orthogonal GARCH (Go-GARCH) models for volatility modelling. An important aspect of the model is that the transformation matrix M is time invariant. Hence, the volatility of innovations becomes

$$\Sigma_t = M V_t M' \tag{5.23}$$

where V_t is the volatility matrix of b_t , that is, $V_t = Cov(b_t | F_{t-1})$ with F_{t-1} denoting the information available at $t-1$. Second assumption of the Go-GARCH model is that V_t is a diagonal matrix for all t .

The unconditional covariance of b_t is $\text{Cov}(b_t) = I_k$ hence, by transformation equation $\text{cov}(a_t)$ becomes $\text{Cov}(a_t) = MM'$. Van der Weide (2002) uses these results and 2 other lemmas (see Van der Weide (2002), Tsay (2014)) to uniquely determine M . Tsay (2014) highlights that in theory the orthogonal transformation above assumes that $\{a_t\}$ forms a random sample but in practice, asset returns, the innovation $\{a_t\}$ is serially uncorrelated, but dependent. This gap between theory and practice raises the issue of applicability of Go-GARCH models in analyzing asset returns. However, Go-GARCH models are relatively simple and conceptually appealing nonetheless.

5.3 Summary

This chapter introduced the multivariate time series. It focused on the theory of three specific models that are going to be employed in this thesis. The models are the EMWA, DCC, and the Go-GARCH models. For the DCC model, there are two types and both were discussed in detail. These models will be applied to our data in the next section using R software.

Chapter 6

Multivariate Volatility Models Application

In this section, we combine all our data to have a multivariate return series. We then analyse the data and then run different models described in the previous section. The data analysis is carried out in R as follows

- First use the concatenate command to join the vectors (individual time series) to form the matrix(multiple time series).
- The next step is to then calculate the returns from the closing prices and then have a multiple time series of returns which we will then plot.
- Using the Multi Time Series package (MTS) we then test for the presence of heteroscedasticity.
- Having checked for conditional heteroscedasticity and if found, we then proceed to fit the models described in the previous chapter as shall be discussed in the following sections.

6.1 Multivariate Daily Data

The first step is to do a preliminary analysis of the data so as to understand its properties. Individual plots of the data sets are the same as those seen in chapter 3. These are used to give the multivariate series. A plot of the individual assets now together is given in Figure 6.1 below.

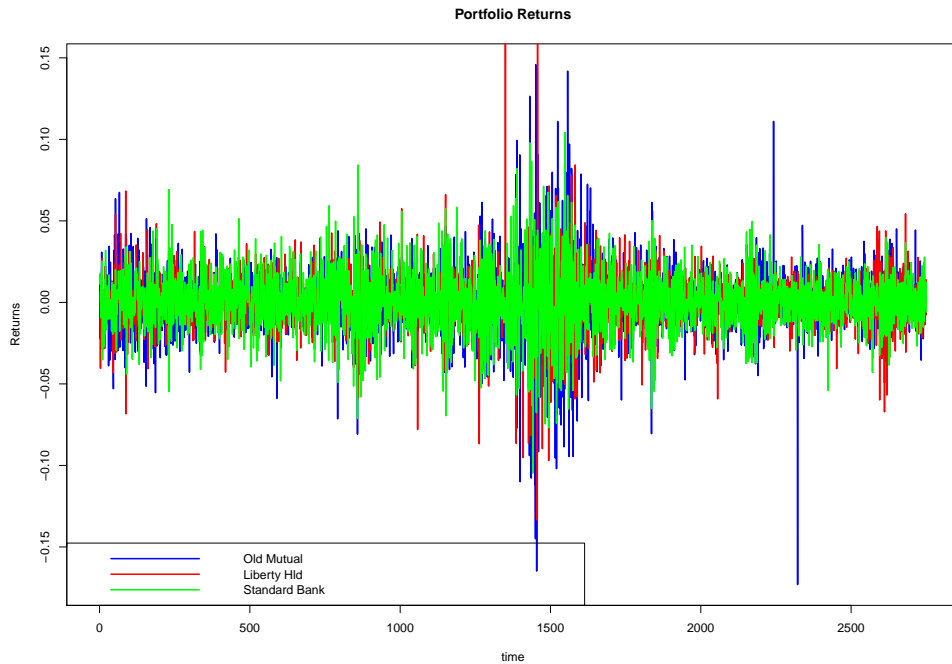


Figure 6.1: Time Series Plot Of The Portfolio Returns

From the returns plots, we can see that in general periods of high positive returns and periods of high negative returns are similar across assets. This makes sense as these assets are taken from the same market. Old Mutual returns, however, have periods where their loss or gains are way above others.

Next we test for volatility presence. This is done through the use of the MarchTest command of the MTS package. The test is carried out on the mean corrected series.

The results obtained were as follows:

Table 6.1: Diagnosis Of Mean Corrected Series

Test	Test statistic	P-value
Q(m) of squared series(LM test)	1817.814	0
Rank-based Test	1411.621	0
$Q_k(m)$ of squared series	3718.072	0
Robust Test(5%)	1102.466	0

As seen from the results all form of tests of presence of conditional heteroscedasticity were able to reject the null hypothesis and accept the presence of conditional heteroscedasticity in the mean corrected series. The same conclusion can be drawn by looking at the cross correlation matrices (CCM) plots shown below in Figure 6.2.

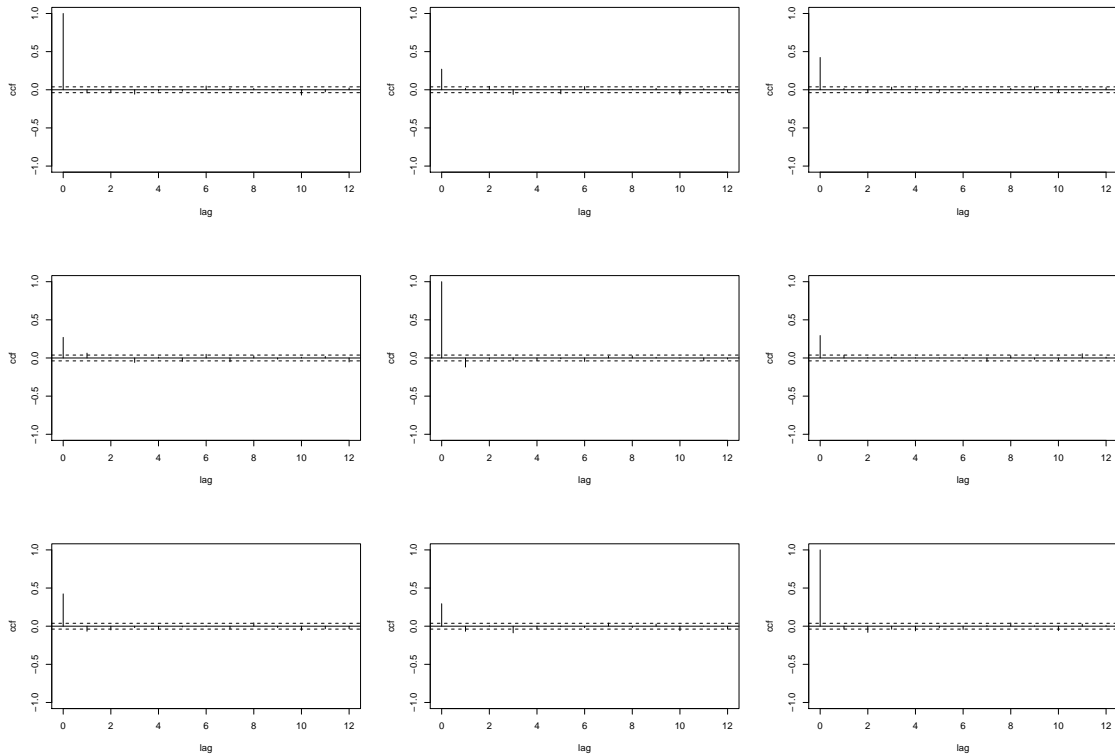


Figure 6.2: Cross Correlation Matrices Plots

The above CCM plots suggest that the returns have 1 lag dependence, hence, when fitting models this dependence must be taken care of through the use of Vector Auto-Regressive (VAR) models.

6.1.1 EMWA Model Fitting

Before fitting the EMWA model the VAR(1) model is used to take care of the lag one correlations as was suggested by the CCM plots. The residuals of the VAR(1) model are then used on the EMWA model.

The results obtained after fitting the EMWA model are shown in Table 6.2 below

Table 6.2: Fitted EMWA model

Coefficient(s)	Estimate	Std. Error	t value	Pr(> t)
lambda	0.9774	0.0013	776.7	<2e-16 ***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

The results gave an estimate of λ , $\hat{\lambda} = 0.977$. This value lies in the expected range as suggested by results of previous studies. Clearly our model uses most recent data more than data that has been there for long which is good (the closer λ is to 1 the more weights are put to the most recent data). It is known that one can choose a value for λ before

running the model since λ is a constant, but here we ran the model with a negative λ so that the λ can be estimated from the data. Model diagnostics on the residual are as follows.

Table 6.3: Diagnosis Of Fitted EMWA model

Test	Test Statistic	P-Value
Q(m) of et	11.43888	0.324362
Rank-based test	135.3322	0
Qk(m) of et	87.95407	0.5413576
Robust Qk(m)	204.1437	7.544609e-11

From the results we see that the Qk(m) and Q(m) accept the null hypothesis but the stronger tests which are the rank based and robust test fail to accept suggesting remaining conditional heteroscedasticity. The robust test rejection is, however, expected from the EMWA model (see (Tsay 2014)).

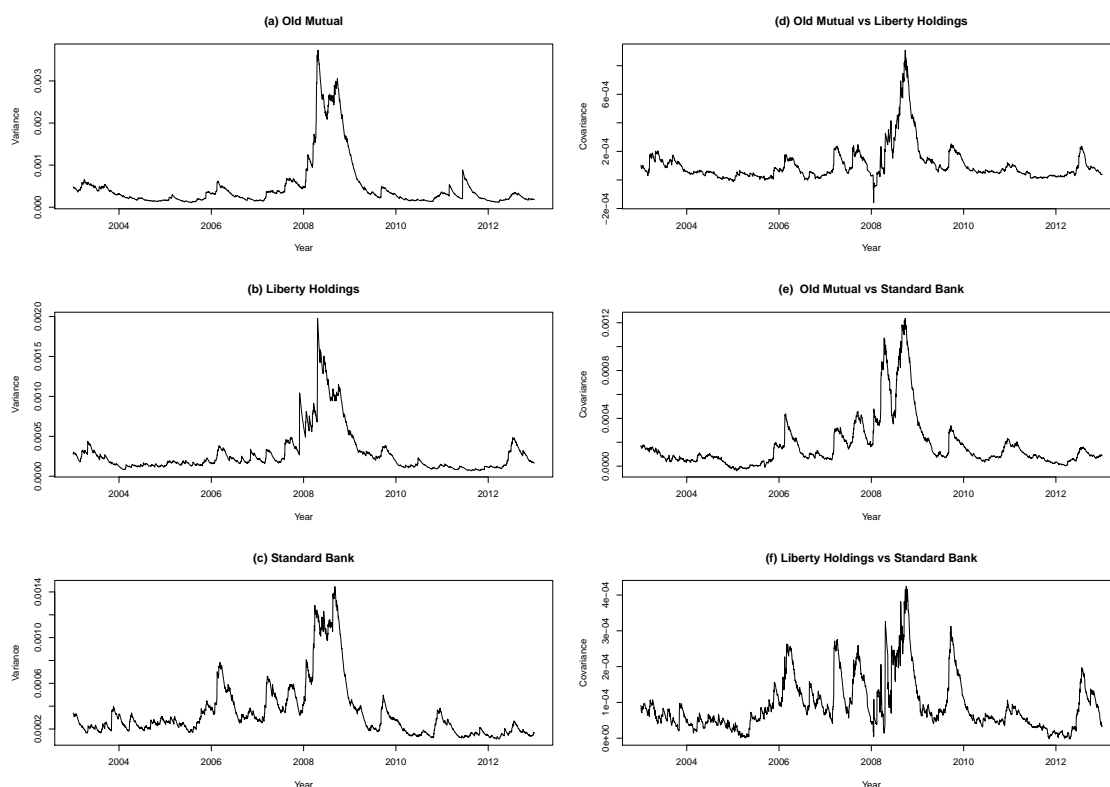


Figure 6.3: Plots From The Volatility Matrix

The volatility plots are shown in Figure 6.3 all have a similar trend but vary in intensity. All volatility plots exhibit persistence which in itself is a stylized fact of financial data. It can be seen that from around the year 2008 to end of 2009 the volatility is very high for all assets, this is explained by the global crisis which affected all assets performance. The Old mutual series is the most volatile of the 3 followed by Liberty Holdings although in a

few select points in time order may be different. The covariance of the assets is greatest between Old Mutual and Liberty Holdings.

6.1.2 DCC Model Fitting

Like the previous sections we use the VAR(1) residuals to model our data. The procedure here is to first fit univariate GARCH(1,1) models to the component series and obtain the marginally standardized series for DCC estimation. This is done using the command dccPre of the MTS package. We then use the dccFit (also from MTS package) command which estimates the specified DCC model using the marginally standardized series. Since our data has heavy tails we use the default students t innovations in our model. Output from the dccPre is skipped and only the important information is given. The resulting univariate GARCH(1,1) are given as follows:

$$\begin{aligned}\sigma_{11,t} &= 8e - 06 + 0.065a_{1,t-1}^2 + 0.915\sigma_{11,t-1} \\ \sigma_{22,t} &= 1e - 06 + 0.025a_{2,t-1}^2 + 0.973\sigma_{22,t-1} \\ \sigma_{33,t} &= 8e - 06 + 0.081a_{3,t-1}^2 + 0.895\sigma_{33,t-1}\end{aligned}$$

We proceed to model the 2 DCC models using the resulting univariate GARCH models.

Table 6.4: Fitted Tse and Tsui model with Student-t innovations

Parameter	Estimate	st.errors	t-values
θ_1	0.92	NaN	NaN
θ_2	0.0248933	0.001145087	21.73922
Degrees Of Freedom	7.221554	0.4355408	16.58066

Table 6.4 above shows the fitted the Tsei and Tsui model. The coefficient estimate for θ_1 standard error could not be solved as the resulting hessian gave NaN which means not a number, this occurs when the hessian fails to converge or when taking square root and the number is negative. θ_2 estimate had a t-ratio of 21.74 hence, it was highly significant. The estimated degrees of freedom for the multivariate Student-t innovations is 7.22 and it is significant.

The resulting DCC model is as follows:

$$\rho_t = (1 - 0.92 - 0.0249)\bar{\rho} + 0.92\rho_{t-1} + 0.0249\psi_{t-1} \quad (6.1)$$

Table 6.5: Diagnosis Of Fitted Tse and Tsui DCC model

Test	test statistic	P-Value
Q(m) of ϵ_t	5.319914	0.8688089
Rank-based test	28.78021	0.001352299
Qk(m) of ϵ_t	60.24456	0.9932952
Robust Qk(m)	130.0516	0.003694302

Table 6.5 gives the results of the diagnosis of the model. We can see that the stronger rank based test and the robust test all confirm the presence of remaining conditional heteroscedasticity which as discussed earlier is not surprising for a DCC model.

Table 6.6: Fitted Engle DCC model with Student-t innovations

Parameter	Estimate	St.errors	t-values
θ_1	0.92	NaN	NaN
θ_2	0.03237957	NaN	NaN
Degrees Of Freedom	7.402539	0.4544672	16.28839

The Engle model is shown above in Table 6.6, here again, NaNs were produced upon evaluating the hessian. This may be because of negative numbers resulting and hence, square root can't be taken on calculation. However, since we have the coefficients we can write down the model and proceed to diagnosis.

The resulting equation for Engle model is as follows:

$$Q_t = (1 - 0.92 - 0.0324)\bar{\rho} + 0.92Q_{t-1} + 0.0324\eta_{t-1}\eta'_{t-1} \quad (6.2)$$

Table 6.7: Diagnosis Of Fitted DCC model

Test	Test statistic	P-value
Q(m) of ϵ_t	31.27671	0.0005280545
Rank-based test	31.27671	0.0005280545
Qk(m) of ϵ_t	55.03853	0.9986402
Robust Qk(m)	130.6981	0.003298202

The model diagnosis where as expected similar to those of the Tse and Tsui model. The robust and rank based tests confirm that there is still some conditional heteroscedastic in the model. The two models have same coefficients for θ_1 whilst others were not that different. The coefficients of the equations suggest that the DCC of Engle appears to have stronger persistence in the time varying correlations.

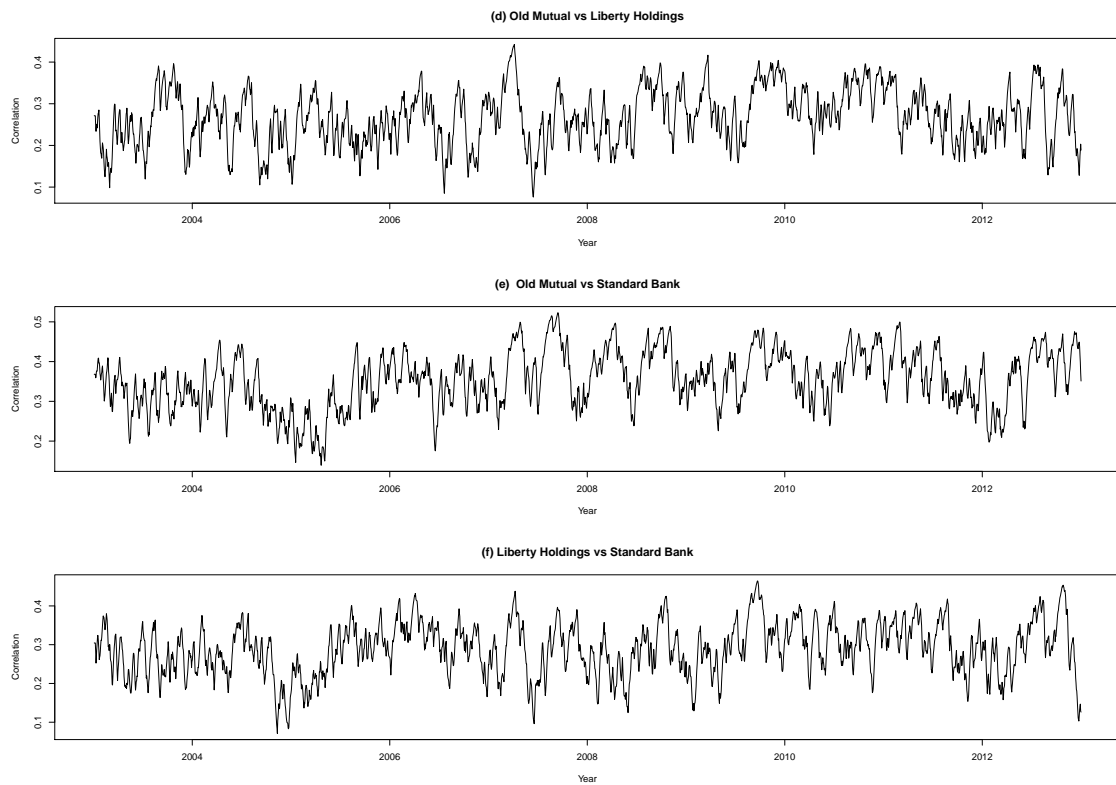


Figure 6.4: Plots of Correlations

Above is the plots of the correlations of the three portfolio assets from the Tse and Tsui model. The correlations are highest between Old Mutual and Liberty Holdings whilst being lowest between Liberty Holdings and Standard Bank. The package does not give the volatility matrix with the volatility component hence, we just have the time-varying correlation plot for the Tse and Tsui model.

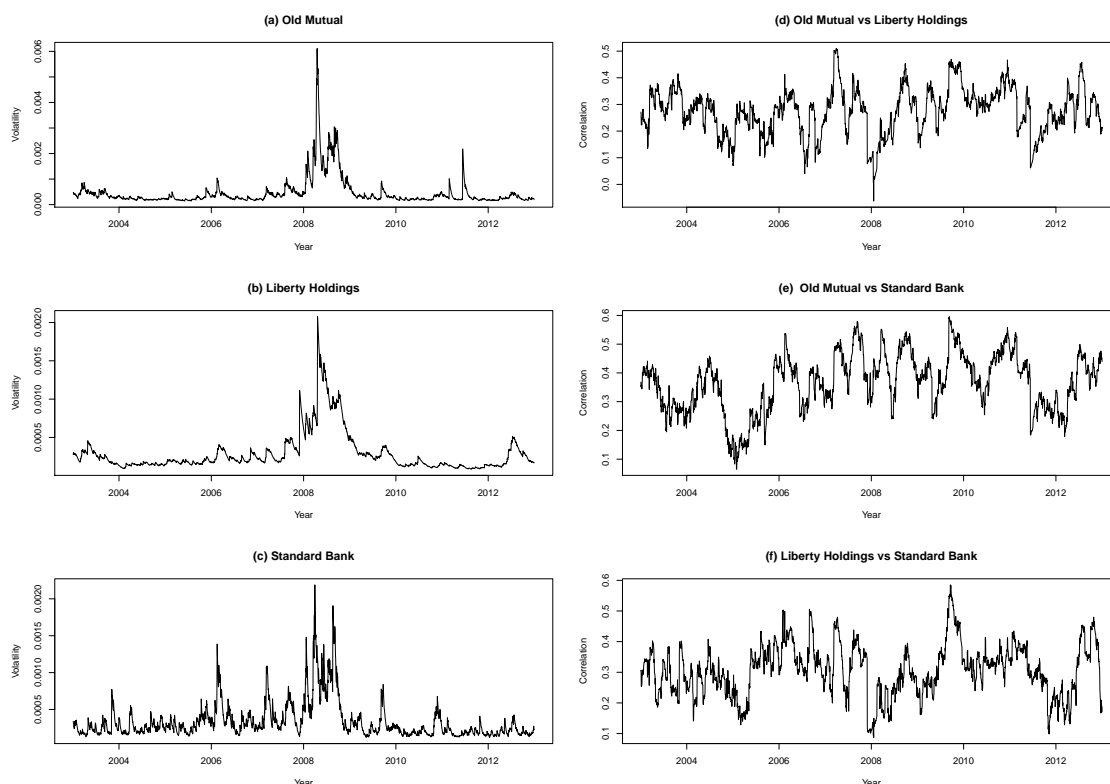


Figure 6.5: Plots of Correlations

Above we have the Engle volatility and time-varying correlation plots. All assets have a similar plot and also, they exhibit persistence and volatility clustering. The time varying correlation plots are similar to those of the Tse and Tsui model hence, these two models are qualitatively similar according to these results.

6.1.3 Go-GARCH Model Fitting

Like all models in this section we fit model on the VAR(1) residuals of the data. We apply the goGARCH(1,1) using the fast ICA estimation method by the gogarch command of the gogarch package. Most of the less relevant estimation section of the results is skipped.

The estimated transformation matrix is as follows:

$$\hat{M} = \begin{pmatrix} 0.015767707 & -0.007959807 & -0.01301738 \\ -0.007752288 & -0.005916669 & -0.01370323 \\ 0.004744601 & 0.010886237 & -0.01408516 \end{pmatrix}.$$

The fitted GARCH(1,1) models for the latent variables b_{it} from the resulting GARCH coefficient estimates section are as follows:

$$\sigma_{1,t}^2 = 0.00526 + 0.0287b_{1,t-1}^2 + 0.962\sigma_{1,t-1}^2$$

$$\sigma_{2,t}^2 = 0.0415 + 0.0747b_{2,t-1}^2 + 0.906\sigma_{2,t-1}^2$$

$$\sigma_{3,t}^2 = 0.00213 + 0.0210b_{3,t-1}^2 + 0.977\sigma_{3,t-1}^2$$

Table 6.8: Diagnosis Of Fitted Go-GARCH model

Test	Test statistic	P-value
Q(m) of ϵ_t	8.758543	0.5551619
Rank-based test	75.39031	3.994027e-12
Qk(m) of ϵ_t	173.802	2.811786e-07
Robust Test(5%)	164.906	2.507871e-06

Model diagnosis is done through the test for conditional heteroscedasticity on the standardized innovations of the fitted Go-GARCH model and based on the above results they still have strong conditional heteroscedasticity. This is not surprising as explained by Tsay (2014) because the squares of the transformed latent variables b_{it} still have significant cross-dependence. This can be seen on the correlation matrix below where the off diagonal elements are non zero. These non zero correlations between the squares of fitted latent variables remain an issue in all other Go-GARCH estimation methods.

$$cor(bt^2) = \begin{vmatrix} 1.0000000 & 0.3957759 & 0.5582137 \\ 0.3957759 & 1.0000000 & 0.2980521 \\ 0.5582137 & 0.2980521 & 1.0000000 \end{vmatrix}.$$

The resulting volatility plots are as shown in Figure 6.6. Here series 1, series 2 and series 3 are Old Mutual, Liberty Holdings and Standard Bank return series respectively.

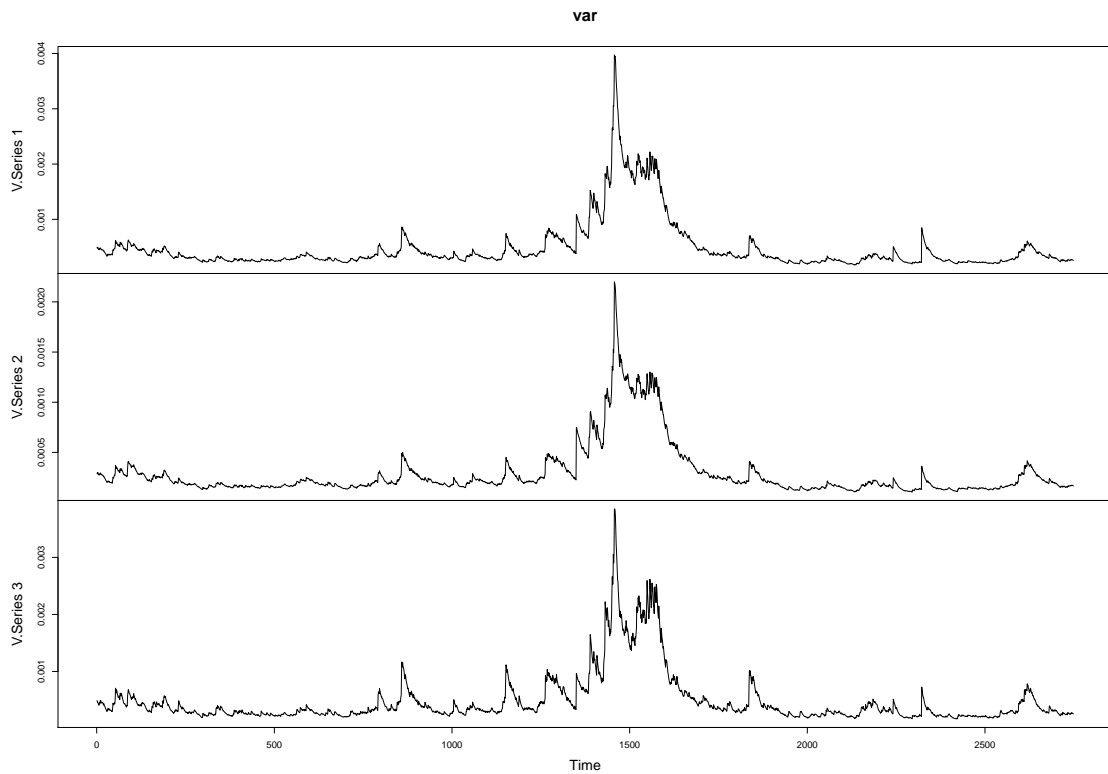


Figure 6.6: Volatility Plots

From Figure 6.6 above we can see that the general volatility pattern is the same for all assets. The similar pattern is due to market factors as these assets all come from the same market. The intensity of the volatilities is different as can be seen by the y-axis values of the plots volatility from series is two is higher than the rest, hence, is a very volatile asset in the portfolio. All 3 Plots have a very high volatility period this corresponds to the global recession of 2008 as markets became very volatile in that period. A drop follows that period. This is a period were economies recovered and the market became less volatile. We deduce that for a risk averse investor Old Mutual will not be a good choice but those who are willing to take more risk can invest in the volatile Old Mutual.

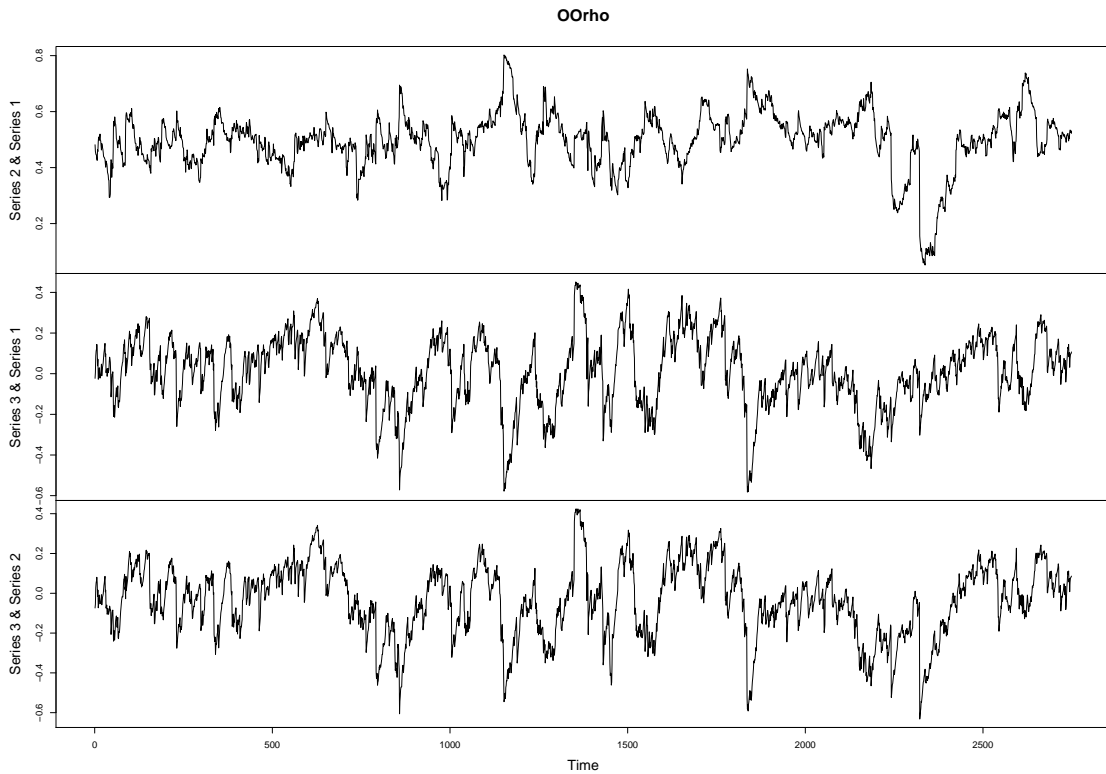


Figure 6.7: Cross Correlation Matrices Plots

The correlation series is shown above. Clearly we can tell that the correlation between Old Mutual and Liberty Holdings is very high as compared to the other two which are quite similar. This result is similar to the one obtained from the EMWA model.

6.2 Multivariate Monthly Data

The monthly data is derived from the individual monthly data of our 3 companies. The data has been trimmed to 173 observations each so that we can have an equal length data frame.

Exploration of the resultant multivariate series is carried out in R. The results are given below.

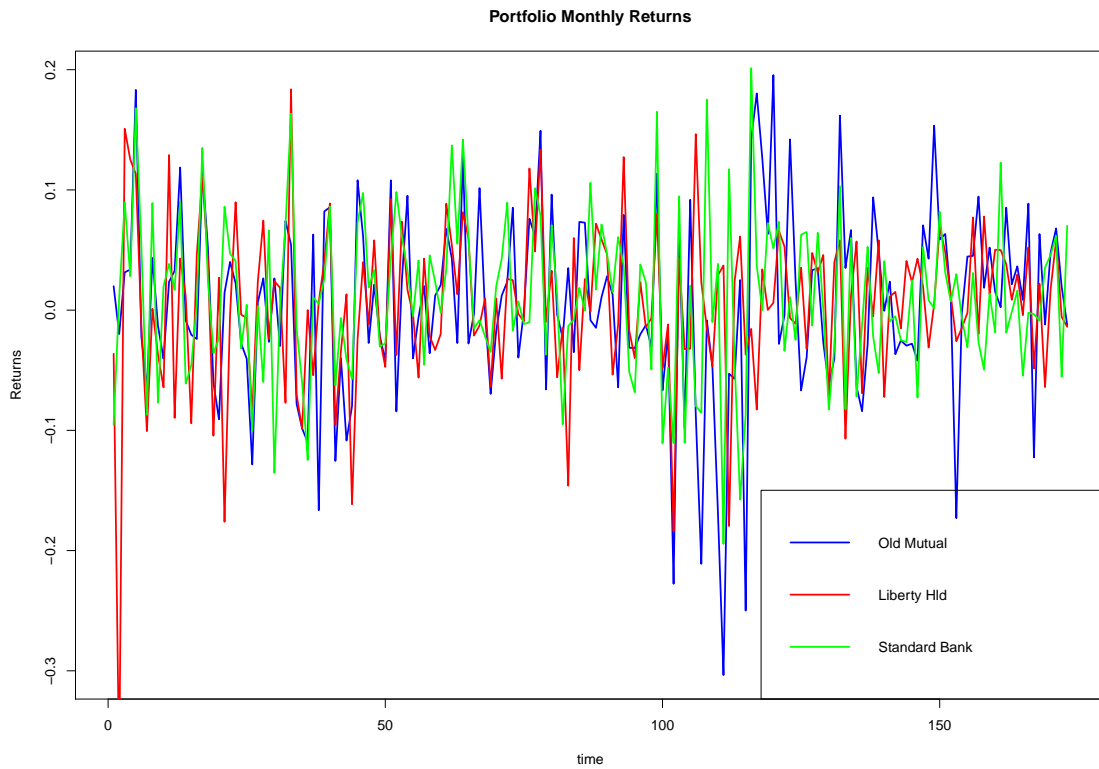


Figure 6.8: Portfolio Returns

The plots of the individual returns are shown in Figure 6.8 above. It can be seen that Old Mutual and Liberty Holdings had a few periods of huge negative returns whilst Standard Bank loss boundary did not fluctuate much. Standard Bank however, had also, huge positive returns in 2008 during which other assets returns fell because of the recession. This shows that Standard Bank performed well amid the crisis.

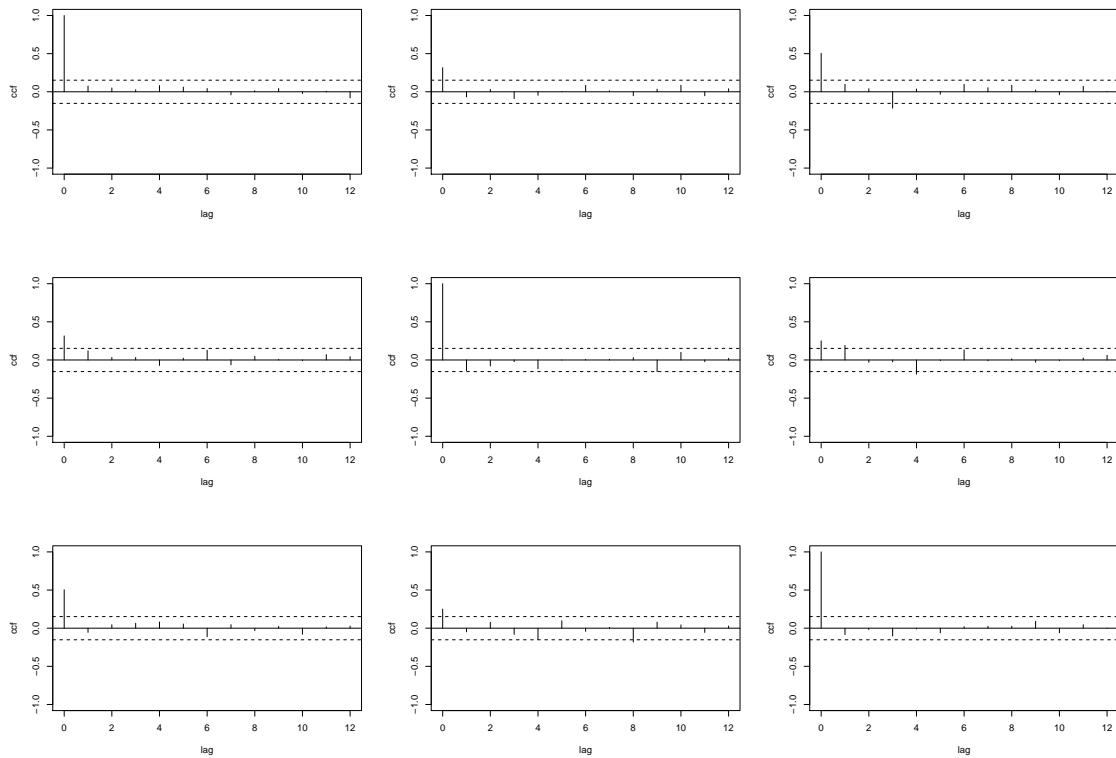


Figure 6.9: Cross Correlation Matrices Plots

The CCM plots in Figure 6.9 above show that the data is not correlated. This is as expected in theory for monthly data. This means that we can easily proceed to fit our multivariate model volatility models with $\mu_t = \mu$ rather than estimate a VAR(p) model to remove correlations. Figure 6.10 shows Significance of CCM Plots, here p-values are greater than 0.05 confirming that the series has no zero CCMs.

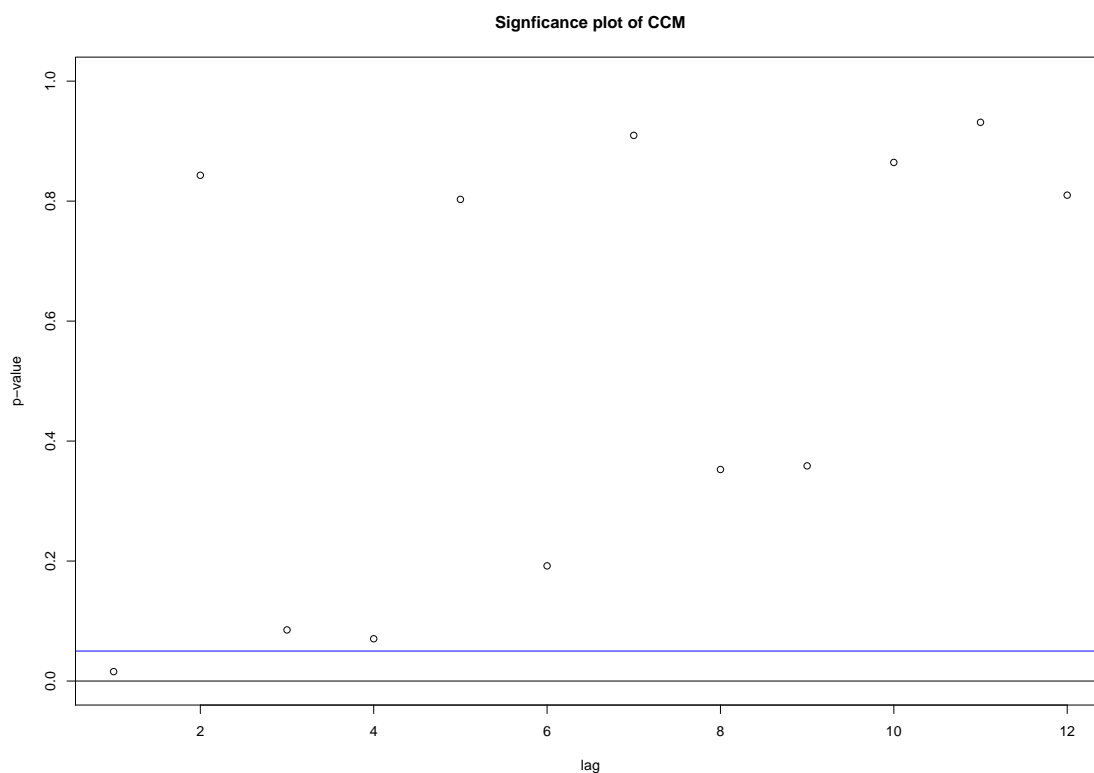


Figure 6.10: Significance of CCM Plots

We apply the four tests for conditional heteroscedasticity in the mean corrected series and the obtained results are as follows:

Table 6.9: Diagnosis tests for data

test	test statistic	p-value
Q(m) of squared series(LM test)	50.84501	1.865268e-07
Rank-based Test	24.5122	0.006350934
Qk(m) of squared series	184.1865	1.906998e-08
Robust Test(5%)	146.1551	0.0001681621

All tests have very small p-values < 0.05 hence, confirm the presence of conditional heteroscedasticity in the monthly log return series. We then proceed to fit our models to the data.

6.2.1 EMWA Model

Table 6.10: Fitted EMWA model

Coefficient(s)	Estimate	Std. Error	t value	Pr(> t)
lambda	0.95695	0.01193	80.22	<2e-16 ***

¹ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

The above Table gives results of the EMWA model. The model was ran with a negative initial value of lambda. This allows the model to estimate its own λ as λ must be positive. The obtained estimate of λ , $\hat{\lambda} = 0.957$ lies in the expected range. The closer λ is to 1 the more weight is emphasized on the most recent returns. A λ closer to 0 will have emphasis on older information hence, poor predictions results Ijumba (2013). Model diagnostics on the residuals are as follows:

Table 6.11: Diagnosis Of Fitted EMWA model

Test	Test Statistic	P-Value
Q(m) of et	28.88411	0.001301018
Rank-based test	10.12224	0.4298353
Qk(m) of epsilont	148.1921	0.0001096714
Robust Qk(m)	79.70564	0.7729107

The diagnosis shows that the trimmed 5% test says that the conditional heteroscedastic still remains whilst the robust and rank based tests reject the presence of conditional heteroscedasticity. This is not shocking as in practice the diagnosis tests on EMWA model residuals usually show remaining heteroscedasticity.

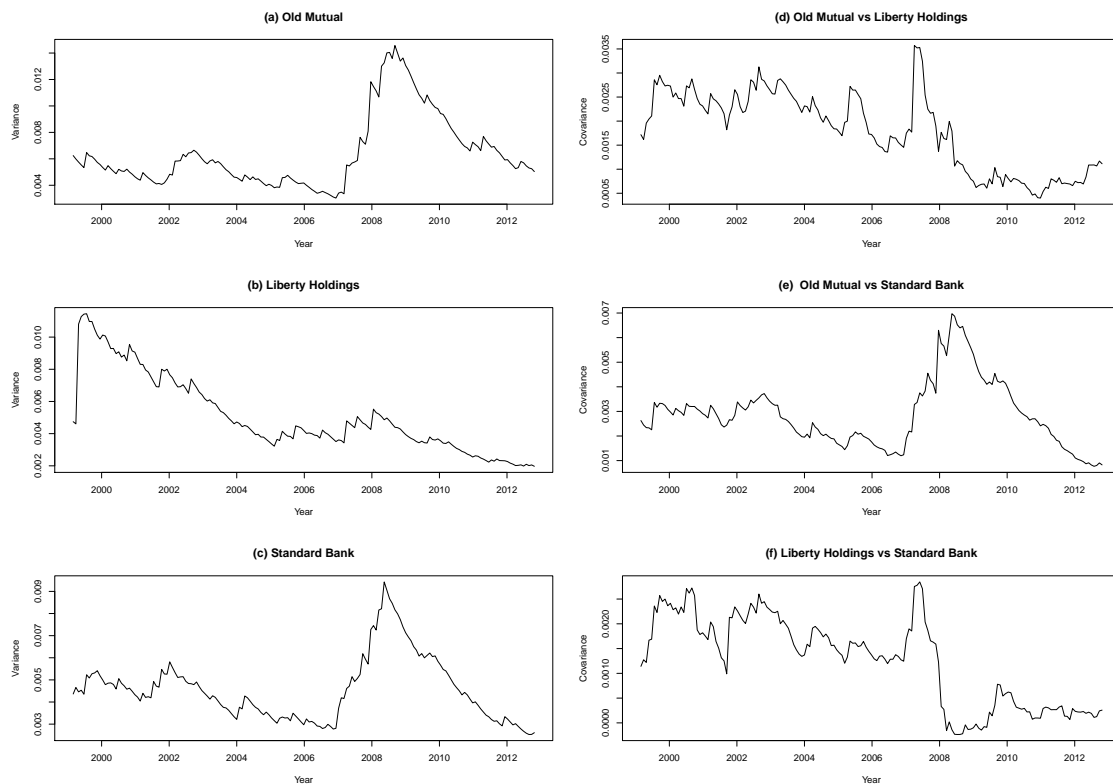


Figure 6.11: Plots From The Volatility Matrix

Figure 6.11 shows that the volatility pattern is the same for Old Mutual and Standard Bank. It is seen that volatility is high during 2008 period up to just after 2009 this is so because of the global recession which occurred then. However, the Liberty Holding seems to be way different from others this can be because of model failure or because the data represented such as monthly data fails to capture all intermediate dependencies during the month. The covariance plots show that the covariance is large for Old Mutual and Standard Bank. For all plots the covariance is stronger during the recession period as the market factor pulled all individual assets together.

6.2.2 Go-GARCH Model Fitting

We apply the Go-GARCH(1,1) using the fast ICA estimation method by the `gogarch` command of the `gogarch` package. Most of the less relevant estimation output is skipped. The estimated transformation matrix is given as follows:

$$\hat{M} = \begin{vmatrix} -0.04871390 & -0.06188057 & -0.002655813 \\ -0.06275217 & 0.02101321 & 0.018332736 \\ -0.01345811 & -0.03397808 & 0.054792084 \end{vmatrix}.$$

The fitted GARCH(1,1) models for the latent variables b_{it} from the resulting GARCH coefficient estimates section are as follows:

$$\begin{aligned} \sigma_{1,t}^2 &= 0.0638 + 0.0665b_{1,t-1}^2 + 0.891\sigma_{1,t-1}^2 \\ \sigma_{2,t}^2 &= 0.119 + 0.305b_{2,t-1}^2 + 0.565\sigma_{2,t-1}^2 \\ \sigma_{3,t}^2 &= 0.114 + 0.134b_{3,t-1}^2 + 0.785\sigma_{3,t-1}^2 \end{aligned}$$

Table 6.12: Diagnosis Of Fitted Go-GARCH model

Test	Test statistic	P-value
Q(m) of ϵ_t	2.875849	0.9841965
Rank-based test	3.078824	0.9795077
Qk(m) of ϵ_t	67.71094	0.9617756
Robust Test(5%)	79.64782	0.774317

Table 6.12 shows the model diagnosis. The standardized innovations of the fitted Go-GARCH model no longer have conditional heteroscedasticity. However, the squares of the transformed latent variables b_{it} still have significant cross-dependence hence, the innovations may still have conditional heteroscedasticity which the test failed to pick. This can be seen in the correlation matrix below as the off-diagonal elements are non-zero. These non-zero correlations between the squares of fitted latent variables remain an issue in all other Go-GARCH estimation methods.

$$\text{cor}(bt^2) = \begin{vmatrix} 1.0000000 & 0.33192773 & 0.06567796 \\ 0.33192773 & 1.0000000 & 0.02611284 \\ 0.06567796 & 0.02611284 & 1.0000000 \end{vmatrix}.$$

The resulting volatility plots are as follows:

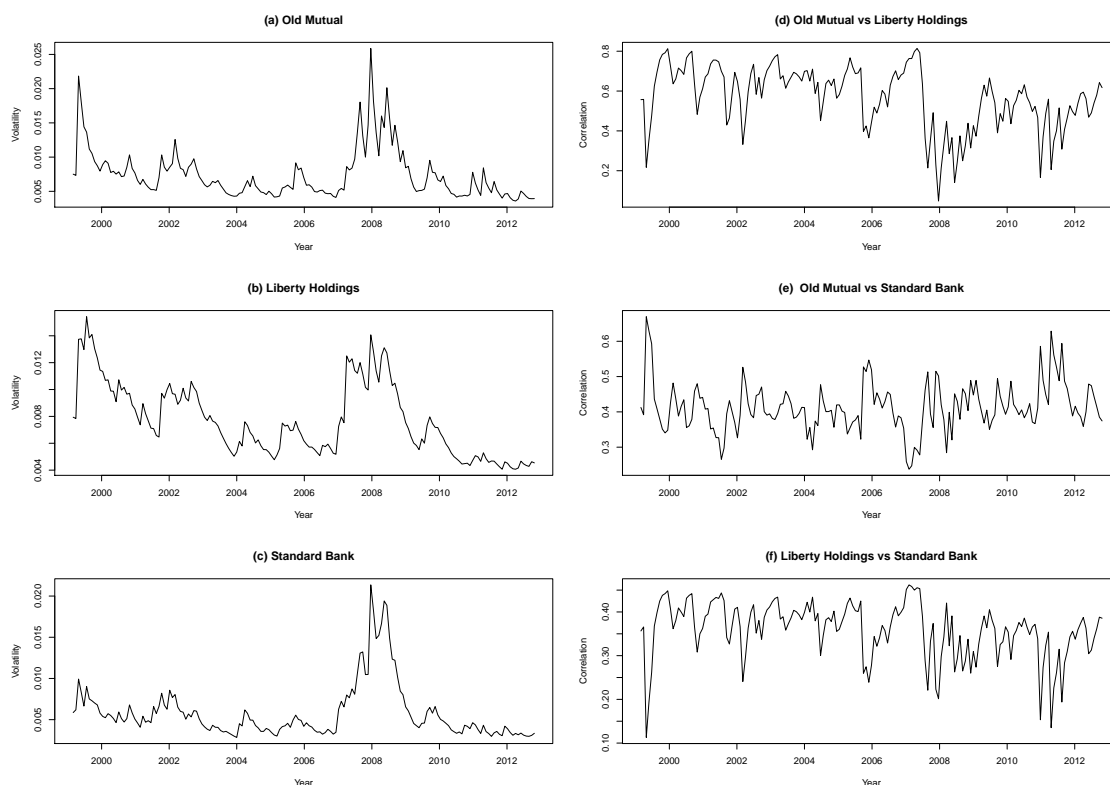


Figure 6.12: Plots from Volatility Matrix

Figure 6.12 shows the volatility series plot and the time-varying correlations plots. We can see that the general volatility pattern is the same for all assets, that is, they rose and dropped at similar periods but with varying intensity. The similar pattern is because of market factors as these assets all come from the same market. Persistence of volatility is clearly visible on all plots hence, this stylized fact is incorporated into the model. Old Mutual and Liberty Holdings have a more similar plot, this is different from EMWA where the Old Mutual was more similar to Standard Bank. Like all other models the Old Mutual asset is more volatile. All 3 Plots have a very high volatility period this corresponds to the global recession of 2008 as markets became very volatile in that period. A drop follows that period as the economies recovered and the market became less volatile. The time varying correlation series is also shown. Clearly we can tell that the correlation between Old Mutual and Liberty Holdings is very high as compared to the other two which are quite similar. This result is similar to the one obtained from the EMWA model.

6.2.3 DCC Model Fitting

Different packages exist to model the DCC model and for our case `dccpre` and `dccFit` command of the MTS package are employed. The `dccPre` resulting univariate GARCH(1,1) are given as follows:

$$\begin{aligned}\sigma_{11,t} &= 7e - 04 + 0.111a_{1,t-1}^2 + 0.771\sigma_{11,t-1} \\ \sigma_{22,t} &= 0 + 0.009a_{2,t-1}^2 + 0.984\sigma_{22,t-1} \\ \sigma_{33,t} &= 0.000374 + 0.102a_{3,t-1}^2 + 0.8061\sigma_{33,t-1}\end{aligned}$$

The next step is to take the volatility series of the 3 estimated models and carry on the DCC model estimation. Recalling that we can either fit Tse and Tsui (2002) kind of model or the Engle model we will, in this case, fit both models and compare. We assume student t distribution for our data so as to capture the heavy tails.

Table 6.13: Fitted Tse and Tsui model with Student-t innovations

Parameter	Estimate	St.errors	t-values
θ_1	0.4	0.5632	0.7102
θ_2	0.0705	0.0949	0.7435
Degrees Of Freedom	11.6698	4.7645	2.4493

The above is the resulting output for the DCC model of Tse and Tsui. The t-ratios however are small hence, the θ 's are not that significant. The resulting model is given as

$$\rho_t = (1 - 0.4 - 0.0705)\bar{\rho} + 0.4\rho_{t-1} + 0.0705\psi_{t-1} \quad (6.3)$$

For diagnosis of the model the following results were found:

Table 6.14: Diagnosis Of Fitted DCC model

Test	Test Statistic	P-Value
Q(m) of et	7.007968	0.7246925
Rank-based test	3.084196	0.9793726
Qk(m) of epsilon	87.54274	0.5537008
Robust Qk(m)	80.65202	0.7493024

The diagnosis rejects the presence of conditional heteroscedasticity hence, our model was a good fit. In general, though it has been seen that diagnosis tests usually reject DCC models. In our case, however, it happens that the model fit the data well.

For the DCC of Engle the results are as follows:

Table 6.15: Fitted Engle DCC model with Student-t innovations

Parameter	Estimate	St.errors	t-values
θ_1	0.4	0.4158	0.9620
θ_2	0.0799	0.0595	1.3441
Degrees Of Freedom	9.5361	2.7282	3.4953

The t-ratios are better than that of Tse and Tsui model but are still not significant. The resulting model from the above results is thus:

$$Q_t = (1 - 0.4 - 0.08)\bar{\rho} + 0.4Q_{t-1} + 0.08\eta_{t-1}\eta'_{t-1} \quad (6.4)$$

Table 6.16: Diagnosis Of Fitted DCC model

Test	Test Statistic	P-Value
Q(m) of et	6.954551	0.7297286
Rank-based test	3.39329	0.970598
Qk(m) of epsilon	88.00226	0.5399113
Robust Qk(m)	99.00904	0.2420799

Like the Tse and Tsui DCC model, the diagnosis tests reject conditional heteroscedasticity and hence, our model fits the data well. A comparison of the models shows that the estimates of the two models are almost similar. The DCC of Engle, however, has stronger persistence than the Tse and Tsui model.

The MTS package only gives the time-varying correlation matrices. Each row contains elements of a cross-correlation matrix which has no variance component, so to obtain the variance plots for the Engle the dcc.estimation command from the ccgarch package and fitted the same DCC model.

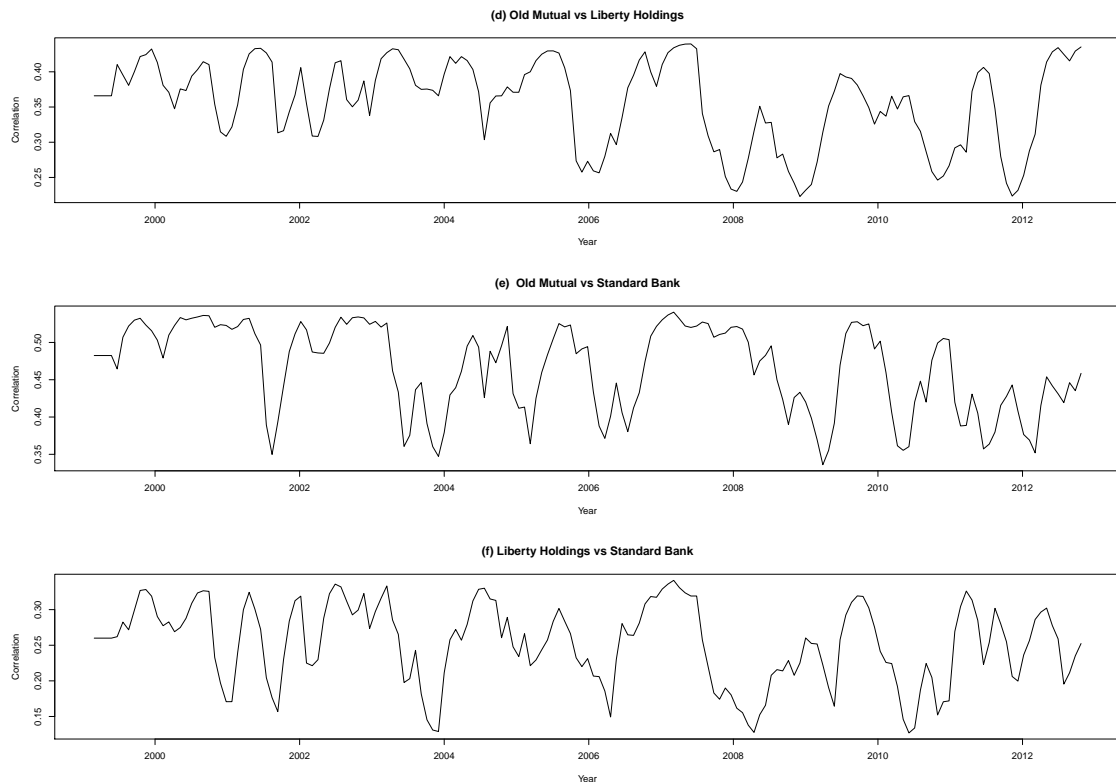


Figure 6.13: Plots of Correlations

Figure 6.13 above shows the plots of the correlations of the three portfolio assets from the Tse and Tsui model. The correlations are highest between Old Mutual and Liberty Holdings whilst being lowest between Liberty Holdings and Standard Bank.

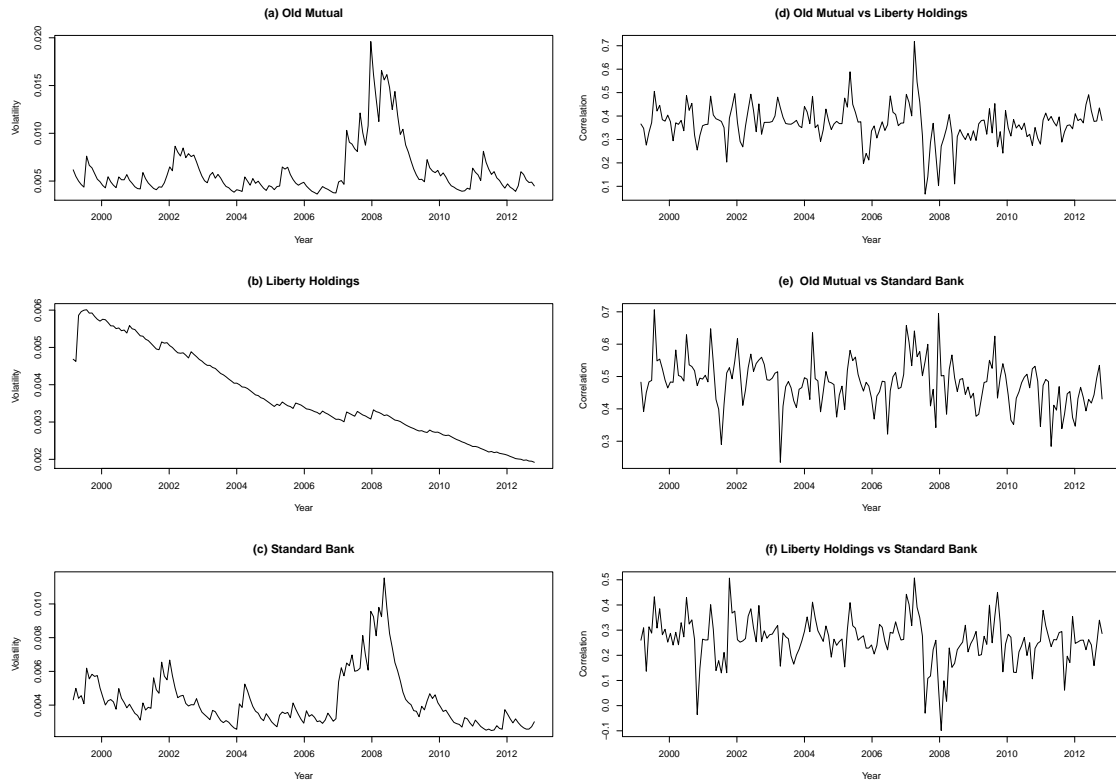


Figure 6.14: Plots from Volatility Matrix

The plots above are for the DCC of Engle as given by R. The volatility series of Old Mutual and that of Standard Bank are very similar but the Liberty Holdings plots seem to be off this may be a sign of a fault in the model or data. The volatility exhibits persistence and volatility clustering as expected. The correlation plots, however, are similar to those obtained from the Tse and Tsui model. Hence, showing as seen in the daily data section that the two models are qualitatively similar.

6.3 Chapter Summary

This chapter we ran 3 types of multivariate models on both monthly and daily data. Like the univariate case the daily data first needed to have correlations removed. The results showed that volatility was similarly high during the global crisis across assets. The Old Mutual was the most volatile asset but in general, all assets volatility followed a similar trend. The similar patterns across assets are explained by market factors as the companies are from the same market. The DCC Liberty plot for monthly data was not convincing as it strayed from the pattern exhibited by other assets and also, by it on other models. This is unusual, the reason may be a failure of the model to incorporate

or capture its variances. However, it may be so as each company can have a different reaction to similar market occurrences. For all other models, however, there was a clear co-movement pattern amongst assets showing that these assets do move together in the same pattern as expected under similar conditions. The differences are the magnitudes of the effects.

Chapter 7

Discussion and Conclusion

The focus of the study was to model asset volatility from returns on investment using volatility models for both the univariate and multivariate case. The first section of the study dealt with the application of ARCH and GARCH models for the univariate case. It also, dealt with the subsequent derivatives of the GARCH models. The second section explored the use of multivariate volatility models to compare the volatility trends of all assets. This was helpful in determining if there was a similar trend in the assets volatility which would have been caused by exposure to similar market factors. In data analysis, the daily data showed that it had serial correlations which had to be taken care of using the Autoregressive models before fitting any models. The monthly data was, however, uncorrelated and hence, we could easily proceed to modelling.

On fitting our univariate models the ARCH models required many parameters to give an acceptable fit for the data unlike for the GARCH model were the simple GARCH(1,1) model with t distribute errors gave the best fit. Not only was the GARCH(1,1) model parsimonious but it also had lower AIC values which further justified it as a model of choice. The t distributed errors were used instead of the Gaussian errors because our data had heavy tails, therefore the normal assumption would easily distort our results. This is coherent with other studies, for example, those by (Talke 2003) and Ijumba (2013) were they also discovered that t or skewed t distributed errors were a better fit. For daily data, the GARCH model was preceded by an AR part so as to remove the correlations. Under the GARCH extensions, it was seen that the EGARCH with t distributed errors was the best fit. All volatility plots for our 3 assets revealed that they were extremely volatile during the 2008-2009 Global recession. The fact that all companies were affected shows that the cause was a market factor.

Another important aspect in volatility is its mean reverting tendency Poterba and Summers (1988), that is, volatility after swinging wildly high or low it will always come back to a certain mean or average value. High volatility periods offer a wide possibility of mispriced stocks because during this period there is investor fear and overreaction in the markets. This results in others following the crowd, that is, they will be making decisions based on what others are doing. Armed with this, wise investors can decide to buy stock when prices start to fall in anticipation of a reversion upwards back to the mean or beyond. Once the prices start to fall again they sell the stocks at a higher price than

they would have purchased with. As an example, an investor who bought the stocks just during the financial crisis and sold them after recovery would have got a high return on his investment. This mean reverting tendency is thus, the backbone of volatility trading for volatility traders i.e they follow the general concept "buy low sell high" but they are armed with forecasts. However, this aspect must be dealt with whilst bearing the knowledge of the stock you are investing in. For example, a change in returns could be a sign that the company no longer has the same prospects it once did, in which case, it is less likely that mean reversion will occur. Also, care must be taken on considering the general persistence of the market or company. That is, some persist on high or low volatility longer than others before reverting back to the average volatility. The upside of this is that persistence and reversion are two sides of the same coin. This is so because high persistence simply means the volatility will take more time to revert back to its mean. In our case for the monthly data, Liberty Holdings was the most persistent with α and β adding up to 0.9778 followed by Old Mutual and Standard Bank with 0.908 and 0.809 respectively. For the daily data however, Liberty Holdings had the least persistence of 0.6814 as compared to Old Mutual and Standard Bank with 0.9799 and 0.976 respectively. This low persistence of Liberty Holdings is very rare. The $\alpha + \beta$ for the GARCH(1,1) models are supposed to be typically close to 1 for all GARCH(1,1) models. Other than that one result, all other values were close to 1 a phenomenon which is supported by other studies such as one by Hillebrand (2003).

For an investor, it is often wise to consider many stocks at once as this helps in pooling risk. This entails spreading the risk of loss from stocks by investing in a portfolio of stock such that in the event of a less fortunate return the fortunate returns can cover up for them. Hence, with a thorough understanding of multivariate volatility the best combination for one's portfolio can be made. The multivariate section introduced models for multivariate volatility. In this section, the major concern was in analysing the co-movements of the volatility of the assets. We wanted to check if the assets followed a similar trend along the given time frame. The first thing to do was to check for dependence in the multivariate series. The daily data had 1 lag correlations hence, VAR models were run first before modelling using the residuals. The results of all multivariate plots showed that all assets were extremely volatile during the 2008-2009 Global crisis, a result that was also confirmed in the univariate case. A previous study by Ijumba (2013) where she used countries in the BRICS also showed that the volatility was high during the Global crisis. Old mutual was the most volatile of the three companies. This suggests that for risk-averse investors for that period had to shy away from investing in Old Mutual. Standard Bank was the least volatile for the period in question. All three models confirmed that correlations were highest between Old Mutual and Liberty Holdings and lowest between Liberty and Standard Bank. For monthly data, just like the univariate case, the multivariate series did not show evidence of correlations. The plots, however, did not clearly exhibit the persistence behaviour of volatility and this was likely because we had only 173 observations.

Lastly, the models covered in this thesis do not cover all models that are available for modelling volatility. For the univariate case, we have models such as the asymmetric

power GARCH (APARCH), conditional heteroscedastic ARMA (CHARMA), whilst for the multivariate case we have models such as the VEC model and the BEKK amongst others. These models have their own strengths and weaknesses. A good study area would be to try and create a single model that encompasses most if not all the strengths of the various models. Also, new ways for diagnosis of multivariate models would be ideal, especially the EMWA and DCC which are usually rejected by diagnosis tests. This a fact supported by Tsay (2014). Other than that, new models are the best way forward for multivariate volatility modelling.

Bibliography

- Alexander, P. C. (2008), *Market Risk Analysis: Pricing, Hedging and Trading Financial Instruments*, John Wiley and Sons.
- Alexander, S. S. (1961), ‘Price movements in speculative markets: Trends or random walks’, *Industrial Management Review (pre-1986)* **2**(2), 7.
- Baillie, R. T. and Bollerslev, T. (2002), “The message in daily exchange rates: a conditional-variance tale”, *Journal of Business and Economic Statistics* **20**(1), 60–68.
- Black, F. (1976), Studies of stock price volatility changes, proceedings of the 1976 meetings of the business and economic statistics section. 177-191, in ‘American Statistical association’.
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics* **31**(3), 307–327.
- Bollerslev, T. and Ole Mikkelsen, H. (1996), “Modeling and pricing long memory in stock market volatility”, *Journal of Econometrics* **73**(1), 151–184.
- Campbell, J. Y., Lo, A. W.-C., MacKinlay, A. C. et al. (1997), *The econometrics of financial markets*, Vol. 2, Princeton University Press, Princeton, NJ.
- Cantaluppi, G. (2013), *Computational Laboratory for Economics*, EDUCatt-Ente per il diritto allo studio universitario dell’Università Cattolica.
- Caporin, M. and McAleer, M. (2010), Ranking multivariate GARCH models by problem dimension, Technical report, Econometric Institute Research Papers.
- Corhay, A. and Rad, A. T. (1994), “Statistical properties of daily returns: Evidence from European stock markets”, *Journal of Business Finance and Accounting* **21**(2), 271–282.
- Ding, Z. and Granger, C. W. (1996), “Modeling volatility persistence of speculative returns: a new approach”, *Journal of Econometrics* **73**(1), 185–215.
- Dufour, J.-M. and Roy, R. (1985), “Some robust exact results on sample autocorrelations and tests of randomness”, *Journal of Econometrics* **29**(3), 257–273.
- Engle, R. (2002), “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models”, *Journal of Business and 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**, 987–1007.
- Engle, R. F. and Bollerslev, T. (1986), “Modelling the persistence of conditional variances”, *Econometric Reviews* **5**(1), 1–50.
- Engle, R. F., Lilien, D. M. and Robins, R. P. (1987), “Estimation of Time Varying Risk Premia in the Term Structure: the ARCH-M Model”, *Econometrica* **55**(2), 391–407.
- Engle, R. F. and Ng, V. K. (1993), “Measuring and testing the impact of news on volatility”, *The Journal of finance* **48**(5), 1749–1778.
- Francq, C. and Zakoian, J.-M. (2011), *GARCH models: structure, statistical inference and financial applications*, John Wiley and Sons, England.
- Ghalanos, A. (2015), *Rugarch: Univariate GARCH models*. R package version 1.3-6.
- Hillebrand, E. (2003), Mean reversion models for financial markets, PhD thesis, Universitätsbibliothek Bremen.
- Houthakker, H. S. (1961), “Systematic and random elements in short-term price movements”, *The American Economic Review* **51**, 164–172.
- Ijumba, C. (2013), Multivariate time series analysis of the brics returns, Master’s thesis, Ukzn Msc, <http://researchspace.ukzn.ac.za>.
- Imbs, J. (2010), “The first Global Recession in decades”, *IMF Economic Review* **58**(2), 327–354.
- Jondeau, E., Poon, S.-H. and Rockinger, M. (2007), *Financial modeling under non-Gaussian distributions*, Springer Science and Business Media, New York.
- JP Morgan, B. (1995), “Risk-Metrics technical manual”, *JP Morgan Bank. New York* .
- Karlsson, L. (2002), “GARCH modelling: Theoretical survey, model implementation and robustness analysis”, *Unpublished Masters Thesis, Kungl Tekniska Högskolan* .
- Kendall, M. G. and Hill, A. B. (1953), “The analysis of economic time-series-part i: Prices”, *Journal of the Royal Statistical Society. Series A (General)* **116**(1), 11–34.
- Ling, S. and Li, W.-K. (1997), “Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors”, *Journal of Time Series Analysis* **18**(5), 447–464.
- Malkiel, B. G. and Fama, E. F. (1970), “Efficient capital markets: A review of theory and empirical work”, *The Journal of Finance* **25**(2), 383–417.
- Mills, F. C. (1927), *The behavior of prices*, Vol. 3, National Bureau of Economic Research, Inc.
- Mitchell, W. C. (1965), *The making and using of index numbers, 1938*, AM Kelley, New York.

- Nelson, D. B. (1990), "Stationarity and persistence in the GARCH(1, 1) model", *Econometric Theory* **6**(03), 318–334.
- Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: A new approach", *Econometrica: Journal of the Econometric Society* **59**, 347–370.
- Orskaug, E. (2009), Multivariate dcc-garch model:-with various error distributions, Master's thesis, Norwegian University of Science and Technology Department of Mathematical Sciences.
- Poon, S.-H. (2005), *A practical guide to forecasting financial market volatility*, John Wiley and Sons, England.
- Poon, S.-H. and Granger, C. (2005), "Practical issues in forecasting volatility", *Financial Analysts Journal* **61**(1), 45–56.
- Poon, S.-H. and Granger, C. W. (2003), "Forecasting volatility in financial markets: A review", *Journal of Economic Literature* **41**(2), 478–539.
- Poterba, J. M. and Summers, L. H. (1988), "Mean reversion in stock prices: Evidence and implications", *Journal of Financial Economics* **22**(1), 27–59.
- Quigley, L. and Ramsey, D. (2008), "Statistical analysis of the log returns of financial assets", *Financial mathematic, University of Limerick* **32**.
- Ruppert, D. (2010), *Statistics and data analysis for financial engineering*, Springer, New York.
- Sheppard, K. (2013), *Financial Econometrics Notes*, University of Oxford.
- Talke, I. S. (2003), Modelling volatility in time series data, Master's thesis, Ukzn Mscs, <http://researchspace.ukzn.ac.za>.
- Taylor, S. (1986), 'Modelling financial time series', *New York: Vlfiley* .
- Tsay, R. S. (2005), *Analysis of financial time series*, Vol. 543, John Wiley and Sons, New Jersey.
- Tsay, R. S. (2014), *Multivariate Time Series Analysis: With R and Financial Applications*, John Wiley and Sons, New Jersey.
- Tse, Y. K. and Tsui, A. K. C. (2002), "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations", *Journal of Business and Economic Statistics* **20**(3), 351–362.
- Van der Weide, R. (2002), "GO-GARCH: A multivariate generalized orthogonal GARCH model", *Journal of Applied Econometrics* **17**(5), 549–564.
- White, H. (1982), "Maximum likelihood estimation of misspecified models", *Econometrica* **50**(1), 1–25.
- Wilhelmsson, A. (2006), "GARCH forecasting performance under different distribution assumptions", *Journal of Forecasting* **25**(8), 561–578.