University of KwaZulu-Natal

School of Mathematics, Statistics and Computer Science
South Africa

GARCH MODELLING OF VOLATILITY IN THE JOHANNESBURG STOCK EXCHANGE INDEX

Tsepang Patrick Mzamane

Thesis Submitted Fulfillment of an Academic Requirement for the Degree of
Master of Science in Statistics

May 2013

Abstract

Modelling and forecasting stock market volatility is a critical issue in various fields of finance and economics. Forecasting volatility in stock markets find extensive use in portfolio management, risk management and option pricing. The primary objective of this study was to describe the volatility in the Johannesburg Stock Exchange (JSE) index using univariate and multivariate GARCH models.

We used daily log-returns of the JSE index over the period 6 June 1995 to 30 June 2012. In the univariate GARCH modelling, both asymmetric and symmetric GARCH models were employed. We investigated volatility in the market using the simple GARCH, GJR-GARCH, EGARCH and APARCH models assuming different distributional assumptions in the error terms. The study indicated that the volatility in the residuals and the leverage effect was present in the JSE index returns.

Secondly, we explored the dynamics of the correlation between the JSE index, FTSE-100 and NASDAQ-100 index on the basis of weekly returns over the period 6 June 1995 to 30 June 2012. The DCC-GARCH (1,1) model was employed to study the correlation dynamics. These results suggested that the correlation between the JSE index and the other two indices varied over time.

Declaration

The work described in this dissertation was carried out under the supervision and direction of Dr. T. Achia and Prof. H.G. Mwambi, School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Pietermaritzburg, from May 2012 to May 2013.
The dissertation represents original work of the author and has not been otherwise been submitted in any form for any degree or diploma to any University. Where use has been made of the work of others it is duly acknowledged in the text.

....
Signature (Student) Date

....
Signature (Supervisor) Date

....
Signature (Supervisor) Date

Dedication

This work is dedicated to my late parents Nomagugu Mavis and Elia Lekhua Mzamane. I know you would have been very proud of me.

Acknowledgments

I would like to express my deep gratitude to Dr. Thomas Achia and Prof. Henry Mwambi, my research supervisors, for their patient guidance, enthusiastic encouragement and useful critiques for this research.

My grateful thanks are also extended to the staff members in the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg for the conducive environment for studying they have build and to my fellow postgraduate students for their moral support which kept me focussed.

Finally, I would like to thank my family (Mzamane and Matola) for their support and encouragement throughout my study.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Modelling and forecasting stock market volatility is a very critical issue in various fields of finance and economics. There have been numerous studies on the volatility of financial markets using time-series, econometric and other relevant techniques. Volatility is defined as a measure of dispersion of returns for a given security or market index [Tsay, 2010]. In simple terms, volatility can be defined as a relative rate at which the price of a market oscillate around its expected value. Volatility can also be measured by computing the variance or standard deviation of returns from the same stock market index. Generally, the higher the volatility, the riskier the market index.

Stock returns can be characterized by the following stylized facts:

- The closing prices are generally unstable, and returns are usually stationary;
- The series of returns exhibit no or little autocorrelation;
- Serial independence between the square values of the series is often rejected, pointing to the existence of non-linear relationships between the subsequent observations;

- Volatility of returns appears to be clustered;
- Normality is rejected in favour of some leptokurtic distribution;
- Some series portray so called leverage effect.

Another key aspect of stock market volatility is the so-called *leverage effect* noted by Black [1976]. This involves an asymmetry of the impact of past positive and negative values on the current volatility. Negative returns (corresponding to price decrease) tends to increase the volatility by larger amount than positive returns (corresponding to price increase) of the same magnitude.

The problem of modeling time series with time varying variance and heteroscedasticity was always obscure. A first attempt to surmount these complications was through the Autoregressive Conditional Heteroscedastic (ARCH) models introduced by Engle [1982]. The Generalized Autoregressive Conditional Heteroscedastic (GARCH) models introduced by Bollerslev [1986] went a step further with aim of capturing leptokurtic returns and volatility clustering. However, despite the success of GARCH models, they have been criticized for failing to capture the leverage effect present in squared [Liu and Hung, 2010].

It is clear that conditional variance is a function only of the magnitudes of the lagged mean corrected returns. This limitation is overcome by introducing more flexible volatility modelling by accommodating the asymmetric responses of volatility to positive and negative mean corrected returns. This more recent class of asymmetric GARCH models includes the Exponential GARCH (EGARCH) model of Nelson [1991] and GJR-GARCH model of Glosten et al. [1993].

This study interrogates patterns in the volatility of returns from the Johannesburg stock exchange (JSE) all share index. The JSE all share index series represents the performance of South African companies, providing investors with a compendious and complementary set of indices, which measures the performance of the major capital and industry sectors of the South African market. Further, the JSE all share index constitutes 99% of the full market capital value, prior to the consideration of any investability weighings, of all ordinary securities listed on the main board of the JSE based on free-float and liquidity benchmark.

The JSE combines the buyers and sellers of four different financial markets, specifically equities, equity derivatives, commodity derivatives and interest rate instruments. The market was established in 1887 after the discovery of gold fields in Johannesburg. In terms of market capitalization, JSE is among the top 20 largest equities exchanges in the world. Despite the fact that JSE changed the methodology of index computation on the 24^{th} of June 2002 from the JSE Actuaries Index Series to the FTSE/JSE Index Series. It recalculated the new index dating back to July 1995 [Ferreira and Krige, 2011]. At the end year 2007, JSE had 411 listed companies with an accumulated market capital of US$828 billion. The significant increase in global flows along with the increasing globalized economic activity has resulted in increased interdependence of major of financial markets all over the world.

The interdependence between global financial markets compels investors and portfolio managers to focus on the movement of not only the domestic markets but also the international markets in order to cautiously project their global investment strategy. Globalization has become a major concern in economic circles since the mid-1990s as it became increasingly clear that the trend toward more integrated world markets has opened a wide potential for greater growth, and presents an unparalleled favourably circumstance for developing countries to enhance their standards of living. The term globalization has many definitions but it differs with the context. According to Mittelman [2000], globalization is a network of processes and activities, to some extent than a single unified phenomenon. The processes and activities, in general, refer to the reduction of barriers between countries.

There are numerous studies that focus on the stock market linkage across countries. Sariannidis et al. [2010] analyzed the volatility linkages among three Asian stock markets, namely India, Singapore and Hong Kong. The results indicated that the markets portray a strong GARCH effect and are highly integrated reacting to information which induce not only the mean returns but their volatility as well. Horng et al. [2009], found that the South Korean and Japanese stock price market volatilities had an asymmetrical relationship in the same period. In this study we shall investigate the co-movement between South Africa and key players in the global markets. We considered the FSTE1-00 and NASDAQ-100 as proxies of trends and performance of global markets. The FTSE-100 is a market-capitalization weighted index representing the performance of the 100 largest UK listed blue chip companies, which pass screening for size and liquidity. The index epitomizes almost 84.35% of the UK's market capitalization and is appropriate as the basis for investment products, such as, derivatives and exchange-traded funds. The FSTE-100 also account for 8.02% of the worlds market capitalization. FTSE-100 constituents are all traded on the London Stock Exchange's SETS trading system. The NASDAQ-100 index comprises 100 of the largest United States and international securities listed on the NASDAQ Stock Market based on market capitalization. The NASDAQ-100 index reflects companies across major industry groups covering hardware and software, telecommunications, retail/wholesale trade and biotechnology. However, there has been few studies that modeled the volatility of stock returns in the emerging stock markets, especially in the South African stock market [Onwukwe et al., 2011, Makhwiting et al., 2012, Chinzara and Aziakpono, 2009, Cifter, 2012].

1.2 Literature Review

Mandelbrot [1963] and Fama [1965] played a significant role in detecting that the uncertainty of stock prices as measured by variances vary with time. Fama [1965] observed further that volatility clustering and leptokurtosis are commonly observed in financial time series. In addition to these features, Black [1976] noted that another phenomenon often observed in a return series is the so called *leverage effect*, which occurs when stock prices are negatively correlated with changes in volatility. The so called *leverage effect* was further investigated by Christie [1982]. According to them the leverage effect suggested that a reduction in the equity value leads to a rise in the debt-to-equity ratio hence raising the riskiness of the firm as manifested by an increase in the future volatility [Bollerslev et al., 1992]. Consequently, observing volatility clustering, the postulate of homoscedasticity becomes irrelevant prompting researchers to investigate how to model volatility clustering or time-varying variance.

Granger and Andersen [1978] developed the bilinear model which enables the conditional variance to rely upon the past realization of the series. Nevertheless, the unconditional variance is either zero or infinity which makes it an unattractive specification [Engle, 1982].

In order to capture the characteristics of financial time series, Engle [1982] proposed the autoregressive conditional heteroscedastic (ARCH) model using lagged disturbances. They considered the residuals of a fundamental ARCH model, employing the Lagrange multiplier (LM) test to explore for the autoregressive heteroscedastic errors and to detect ARCH errors. The importance of adjusting for the ARCH effects in the residuals is discussed extensively in the literature [Bera et al., 1988, Connolly, 1989, Schwert and Seguin, 1990]. It is argued that inferences can be adversely influenced by ignoring the ARCH error structure [Bollerslev et al., 1992]. Empirical evidence based on the study by Bollerslev et al. [1992] revealed that high order ARCH model is required to capture the dynamic behaviour of conditional variance.

To circumvent the deficiencies of ARCH model, Bollerslev [1986] proposed a generalized autoregressive heteroscedastic (GARCH) model. Both the ARCH and GARCH models accommodate volatility clustering and leptokurtosis. However, they fall astray to capture the leverage effect. This impediment is dealt with by considering more tractable volatility models. This is achieved by fitting models for asymmetric responses of positive and negative residuals. Other extended class of asymmetric GARCH models include the Exponential GARCH (EGARCH) model by Nelson [1991], the GJR-GARCH model by Glosten et al. [1993] and the Asymmetric Power ARCH (APARCH) model by Ding et al. [1993].

Baillie and Bollerslev [1989] applied the Student-t distribution whereas Nelson [1991] suggested the Generalized Error Distribution (GED). Most of the studies for modelling volatility have been applied on data from the developed countries while there is rare literature on work that have been conducted in emerging markets. Olweny and Omondi [2011] considered the effect of macro-economic factors on the stock return volatility on the Nairobi Securities Exchange (NSE), Kenya. The attention of the study was on the effect of foreign exchange rate, interest rate and inflation fluctuation on stock return volatility at the Nairobi Securities Exchange. The study used monthly time series data for a 10-year period between January 2001 and December 2010. EGARCH and Threshold Generalized Autoregressive Conditional Heteroscedastic (TGARCH) model was used in the study. In the study the returns were found to be leptokurtic and followed a non-normal distribution. The results exhibited substantiation that foreign exchange, interest rate and inflation have an impact on the Nairobi stock return volatility.

Onwukwe et al. [2011] considered a time-series behaviour of daily stock returns of four firms listed in the Nigerian Stock Market from 2 January 2002 and 31 December 2006, employing three heteroscedastic models, particularly GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models respectively. The four firms whose share prices were explored in the study were UBA, Unilever, Guiness and Mobil. The return series of the four firms exhibited a leverage effect, leptokurtic, volatility clustering and negative skewness which are frequent characteristics of financial time-series. The results showed that the GJR-GARCH(1,1) produces better fit to the data.

Olowe [2009] investigated the volatility of Naira/Dollar exchange rates in Nigeria using GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1), APARCH(1,1), IGARCH(1,1) and TS-GARCH(1,1) models. The monthly time series data over the period January 1970 to December 2007. The TS-GARCH and APARCH were found to be the best fitting models.

Makhwiting et al. [2012] developed ARMA-GARCH type models for modelling volatility and financial risk of shares on the Johannesburg Stock Exchange under the assumption of skewed Student-t distribution. The daily data was used for the period January 2002 to December 2010. The GARCH type models that were employed included TGARCH, GARCH-in mean and EGARCH. The results showed that the ARMA(0,1)-GARCH(1,1) model produces the most accurate forecasts.

Cifter [2012] investigated the relative performance of the asymmetric normal mixture generalized conditional heteroscedastic (NM-GARCH) benchmarked GARCH models with the daily stock market returns of the Johannesburg Stock Exchange, South Africa. The predictive performance of the NM-GARCH was compared against a set of the GARCH models with the assumption of normal,student-t and skewed student-t distributions. The results showed that mixture errors enhances the predictive performance of volatility models.

Mangani [2008] explored the structure of the JSE by employing ARCH-type models. In the analysis volatility was found to be prevalent in this market. The results showed that the standard GARCH(1,1) model provides the best description of the return dynamics relative to its complex augmentations.

Alagidede and Panagiotidis [2009] investigated the behaviour of stock returns in Africa's largest stock market particularly, Egypt, Kenya, Morocco, Nigeria, South Africa, Tunisia and Zimbabwe. The results showed that the empirical stylized facts of volatility clustering, leptokurtosis and leverage effect are present in the African data. Leading world stock markets have become more closely linked in recent years, and this has brought deep interest in the impact of those linkages. A major concern is that stock price movements and other shocks are likely to be transmitted promptly between markets, which implies that interdependence between markets may result to the transmission of national financial disturbances, with wide-ranging implications for other markets [Jefferis et al., 1999]. There have been several studies recently conducted for the transmission of volatility between the markets. Nevertheless, substantially, insufficient work has been done in volatility transmission and return co-movement between matured stock markets and emerging African markets.

Sariannidis et al. [2010] investigated linkages among three Asia stock exchange markets namely, India, Hong Kong and Singapore, during the period July 1997 to October 2005. In the study the multivariate GARCH model was employed. The results indicated that there was ARCH effects among the markets and are highly integrated reacting to information which induce not only the mean returns but their volatility as well.

Gupta and Mollik [2008] studied the varying correlations between equity returns of Australia and the emerging equity markets. The Dynamic Conditional Correlation (DCC) model, which enables correlations to vary with time, have been employed to test if the volatilities of individual markets have any influence on the change in correlations. The results suggested that the correlations between Australia's equity and emerging markets' equity returns change in time and the variation in correlations was influenced by the volatility of the emerging markets.

Anaraki [2011] examined how the European stock market responds to the US fundamentals including the Federal Fund Rate (FFR), the Euro-dollar exchange rate, and the US stock market indices. The Johansen [1988] cointegration technique was employed, and the result suggested that a long-term relationship exists between the European stock market, and the US fundamentals.

Chinzara and Aziakpono [2009] explored returns and volatility linkages between the South African (SA) equity market and the world major equity markets using daily data for the period January 1999 to December 2007. The univariate and multivariate Vector Autoregressive (VAR) models were employed. The results showed that both returns and volatility linkages exist between the South African and the major world stock markets, with Australia, China and the United States portraying most influence on SA returns and volatility.

1.3 Comment on the review

With regard to modelling and forecasting volatility, most of the studies have been conducted on the developed markets. To improve the literature in the emerging markets, GARCH models and their extensions need to be employed within the emerging markets to provide a better understanding of dynamics therein. Another important issue is that of the correlation between the markets. Due to the shortage of software packages for the multivariate GARCH for modelling correlation, there is need for further research on methods that can facilitate development of statistical software and tools to analyse relevant data sets. Furthermore, to accommodate some of the regularities of the returns in the multivariate GARCH, different statistical distributions can be considered for the errors to better describe their statistical properties.

1.4 Problem Statement

In financial markets, large price changes are likely to be followed by large price changes and small changes by small price changes. This characteristic of financial time-series data is known as volatility clustering. Volatility clustering is incompatible with homoscedastic (that is, with a constant variance) marginal distribution for the returns. Moreover, one of the idiosyncrasies of this volatility is its uncertainty. As a result, there has been a lot of empirical studies on modelling and forecasting volatility. Modelling volatility is essential for portfolio management, risk management and option pricing.

The interrelation between international financial markets is a very important issue that is linked with the study of correlation dynamics between markets. The leading world stock markets have become more closely linked in recent years, and this has brought very strong interest in studying and understanding the impact of those linkages. A major concern is that the stock price movements are likely to be transmitted promptly between markets, which implies that the markets may lead to rapid transmission of national disturbances, with wide-ranging implications for other markets. Understanding the linkage between markets is becoming increasingly important because of the emergence of regional and world economic blocks such as the BRICS (Brazil, Russia, India, China and South Africa). The chance that a financial market in one country will influence the other is more probable that it was before the creation such economic blocks.

1.5 Objectives of the Study

1.5.1 Broad Objectives

The primary objective of this study is to describe the volatility in the Johannesburg Stock Exchange (JSE) index using univariate and multivariate GARCH models.

1.5.2 Specific Objectives

The specific objectives of this study are:

1. To review statistical properties of the univariate GARCH models and their extensions;
2. To review statistical properties of the Multivariate GARCH models;
3. To investigate volatility in the Johannesburg Stock Exchange (JSE) using the univariate and multivariate GARCH.

Chapter 2

ARCH AND GARCH Models

2.1 Introduction

Modelling and forecasting stock market volatility has been the subject of attention in recent years. Volatility can be employed as a barometer of risk in financial markets. Most of the econometric models assume the that the variance or volatility is time invariant. Nevertheless, many of the empirical studies that have been carried out concerning volatility refutes this assumption. In financial markets, large price changes tend to be followed by large price changes and small prices changes by small price changes. Thus, the assumption of constant variance (homoscedasticity) is inappropriate. In a seminal paper, Engle [1982], introduced a time-varying conditional variance model called the Auto-Regressive Conditional Heteroscedastic (ARCH) model. The ARCH model employs past errors to model the variance of the series and enables the variance to oscillate over time. In this Chapter, we describe the important statistical issues concerning the ARCH and GARCH Models. We then employ the methods discussed to analyse volatity in the JSE index returns.

2.2 The ARCH (p) process

2.2.1 The ARCH (1) process

The elementary and very useful model for financial time series with time varying volatility is the Autoregressive Conditional Heteroscedastic model of order one, which is abbreviated as ARCH(1). This model was introduced by Engle [1982].

Now let us assume that the continuously compounded return of an asset is given by

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \\ &= \mu_t + \sigma_t z_t, \end{aligned}$$

where z_t is a sequence of independent and identical distributed random variables with mean zero and variance of one. Let $\mathbf{\Phi}_{t-1}$ represent the information set at $t-1$, $\mu_t = E[r_t \mid \mathbf{\Phi}_{t-1}]$ the conditional mean function and $\sigma_t^2 = Var[r_t|\mathbf{\Phi}_{t-1}]$ the conditional variance function. Then the residual return or shock at time t can is defined as

$$\begin{aligned} \varepsilon_t &= r_t - \mu_t, \\ &= \sigma_t z_t. \end{aligned}$$

Model description

Definition 1. *A process $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t\}$ is called an autoregressive conditional heteroscedastic process of order one $ARCH(1)$ if it can be written as*

$$\varepsilon_t = \sigma_t z_t, \tag{2.1}$$

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

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \tag{2.2}$$

where α_0 *and* $\alpha_1 \geq 0$*.*

To gain the insight into the ARCH models, we first explore the statistical properties of the fundamental ARCH(1) model and then proceed to cross examine properties of extensions of the model that exist in literature. Under the normality assumption of ε_t, the process can expressed conditional on $\mathbf{\Phi}_{t-1}$ as

$$\varepsilon_t | \mathbf{\Phi}_{t-1} = \varepsilon_t | \varepsilon_{t-1} \sim N(0, \sigma_t^2).$$

From the structure of the ARCH (1) model in Definition 1, it is clear that a large past squared mean-corrected return or shock implies a large conditional variance σ_t^2 for the mean-corrected return ε_t. Hence, ε_t tends to possess a large value in absolute value [Tsay, 2010]. This means that, in Definition 1, large shocks tend to be followed by another large shock. This characteristic is identical to the volatility clusterings observed in asset returns.

Theorem 1. *Let* $\{\varepsilon_t\}$ *be an ARCH (1) process with* $Var[\varepsilon_t] = \sigma^2 < \infty$*, then it follows that* $\{\varepsilon_t\}$ *is a white noise process.*

Proof. From the conditional expectation

$$E[\varepsilon_t | \mathbf{\Phi}_{t-1}] = E[\sigma_t z_t | \mathbf{\Phi}_{t-1}] = \sigma_t E[z_t | \mathbf{\Phi}_{t-1}]] = \sigma_t(0) = 0$$

it follows that $E[\varepsilon_t] = 0$ and

$$\begin{aligned} Cov[\varepsilon_t, \varepsilon_{t-k}] &= E[\varepsilon_t \varepsilon_{t-k}] - E[\varepsilon_t].E[\varepsilon_{t-k}], \\ &= E[\varepsilon_t \varepsilon_{t-k}], \\ &= E[E[\varepsilon_t \varepsilon_{t-k} | \mathbf{\Phi}_{t-1}]], \\ &= E[\varepsilon_{t-k} E[\varepsilon_t | \mathbf{\Phi}_{t-1}]], \\ &= 0. \end{aligned}$$

□

Since $\{\varepsilon_t\}$ is a martingale difference sequence, then it is an uncorrelated sequence process.

Theorem 2. *Suppose that the process $\{\varepsilon_t\}$ is a second-order stationary ARCH (1) process with $Var[\varepsilon_t] = \sigma^2 < \infty$. Then it follows that,*

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

Proof. From the definition of variance of ε_t we have

$$Var[\varepsilon_t] = E[\varepsilon_t^2] - (E[\varepsilon_t])^2 = E[\varepsilon_t^2].$$

It then follows that

$$\begin{aligned} Var[\varepsilon_t] &= E[E[\varepsilon^2|\mathbf{\Phi}_{t-1}]], \\ &= E[\sigma_t^2], \\ &= E[\alpha_0 + \alpha_1\varepsilon_{t-1}^2], \\ &= \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2], \\ &= \alpha_0 + \alpha_1 E[\varepsilon_t^2]. \end{aligned}$$

Further, since ε_t portrays a second-order stationarity, that is $E[\varepsilon_t^2] = E[\varepsilon_{t-1}^2]$, we have

$$Var[\varepsilon_t] = \alpha_0 + \alpha_1 Var[\varepsilon_t],$$

which implies that

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

when $\alpha_1 < 1$. □

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

Theorem 3. *Suppose that the process $\{\varepsilon_t\}$ is an ARCH (1) process, $z_t \sim N(0,1)$ and $E[\varepsilon_t^4] = c < \infty$. Then*

1. *The fourth moment of ε_t about zero is*

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

with $3\alpha_1^2 < 1$.

2. *The unconditional distribution of ε_t is leptokurtic.*

Proof.

1. If we assume that the series ε_t is fourth-order stationary, then

$$E[\varepsilon_t^4] = E[\varepsilon_{t-1}^4].$$

The fourth moment of ε_t about zero then becomes

$$\begin{aligned} E[\varepsilon_t^4] &= E[E[\varepsilon_t^4|\mathbf{\Phi}_{t-1}]], \\ &= E[E[\sigma_t^4 z_t^4|\mathbf{\Phi}_{t-1}]], \\ &= E[\sigma_t^4 E[z_t^4|\mathbf{\Phi}_{t-1}]], \\ &= E[3\sigma_t^4], \\ &= 3E[\sigma_t^4], \end{aligned}$$

since $z_t \sim N(0,1)$ and $E[z_t^4] = 3$. Further,

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

Making $E[\varepsilon_t^4]$ the subject we find

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

and so

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

Replacing $E[\varepsilon_{t-1}^2] = Var[\varepsilon_t] = \frac{\alpha_0}{1-\alpha_1}$ in this expression and simplifying we get

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

Hence the desired result.

2. The kurtosis of ε_t is given by

$$\begin{aligned}
Kurt[\varepsilon_t] &= \frac{E[\varepsilon_t^4]}{\left(E[\varepsilon_t^2]\right)^2}, \\
&= \frac{\frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)}}{(\frac{\alpha_0}{1-\alpha_1})^2}, \\
&= \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)}, \\
&= 3 + \frac{6\alpha_1^2}{(1-3\alpha_1^2)} > 3.
\end{aligned}$$

□

Thus for the ARCH (1) process it is required that $0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$ for the fourth-order moment and the conditional kurtosis to exist. Furthermore, the excess kurtosis of ε_t is heavier than that of normal distribution [Tsay, 2010]. The variance σ_t^2 is thus a serially correlated random variable with expected value

$$\begin{aligned}
E[\sigma_t^2] &= \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2], \\
&= \alpha_0 + \alpha_1 E[\varepsilon_t^2].
\end{aligned}$$

The series of squared mean-corrected ε_t^2 exhibits important properties, and one of them is that it has a stationary autoregressive representation of order one. We know that

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \tag{2.3}$$

Adding ε_t^2 both sides on equation (2.3) we get

$$\sigma_t^2 + \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 \tag{2.4}$$

which implies that

$$\begin{aligned} \varepsilon_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2, \\ &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \nu_t, \end{aligned}$$

where

$$\begin{aligned} \nu_t &= \varepsilon_t^2 - \sigma_t^2, \\ &= \sigma_t^2 z_t^2 - \sigma_t^2, \\ &= \sigma_t^2 (z_t^2 - 1) \end{aligned}$$

is a conditional heteroscedasticity martingale difference sequence.
Furthermore the series of squared mean-corrected returns ε_t^2 exhibits volatility mean reversion.

Estimation of parameters of the ARCH (1) model

The parameters of the ARCH (1) model can be estimated by implementing different estimation procedures. In general, however, the estimation of the ARCH (1) model is normally carried out using the maximum likelihood [Berndt et al., 1974].

Suppose that a time series $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ is assumed to be a realization of an ARCH (1) process. Under the normality assumption of ε_t, the likelihood function on ARCH (1) model can be expressed as

$$\begin{aligned} f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T) &= f(\varepsilon_T | \mathbf{\Phi}_{T-1}) \times f(\varepsilon_{T-1} | \mathbf{\Phi}_{T-2}) \times \ldots \times f(\varepsilon_2 | \mathbf{\Phi}_1) \times f(\varepsilon_1 | \boldsymbol{\theta}), \\ &= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{ -\frac{\varepsilon_t^2}{2\sigma_t^2} \right\} f(\varepsilon_1 | \boldsymbol{\theta}), \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1)'$ is a vector of unknown parameters and $f(\varepsilon_1|\boldsymbol{\theta})$ is a probability density function of ε_1. However, the exact form of $f(\varepsilon_1|\boldsymbol{\theta})$ is complicated. It is generally removed from the prior likelihood function, especially when the sample size is sufficiently large. This allows us to use the conditional-likelihood function which can be written as

$$f(\varepsilon_2, \ldots, \varepsilon_T | \boldsymbol{\theta}, \varepsilon_1) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\},$$

where σ_t^2 can be evaluated recursively. The model parameter estimates are obtained by maximizing the conditional likelihood under the assumption of normality. Maximizing the conditional-likelihood function is equivalent to maximizing its logarithm, which is easier to handle. The conditional log-likelihood is given by

$$\begin{aligned} l(\varepsilon_2, \ldots, \varepsilon_T | \boldsymbol{\theta}, \varepsilon_1) &= -\frac{1}{2}\sum_{t=2}^{T}\left[\ln(2\pi\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right], \\ &= \sum_{t=2}^{T}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}\right]. \end{aligned}$$

Since the first term $\ln(2\pi)$ does not involve any parameters in it, the log-likelihood estimating function becomes

$$l(\varepsilon_2 \ldots, \varepsilon_T | \boldsymbol{\theta}, \varepsilon_1) \propto -\frac{1}{2}\sum_{t=2}^{T}[\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}], \tag{2.5}$$

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

The maximization of equation (2.5) with respect to $\boldsymbol{\theta}$ is a non-linear optimization problem, which can be solved numerically [Franke et al., 2008]. The conditional estimator of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1)'$. Note that alternatively the log-likelihood function as

$$l(\varepsilon_2 \ldots, \varepsilon_T | \boldsymbol{\theta}, \varepsilon_1) = -\sum_{t=2}^{T} l_t,$$

where $l_t = \frac{1}{2}\left[\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right]$ and where T is the sample size. In order to find the estimates, we differentiate with respect to parameters α_0 and α_1 and equate the derivatives to zero. More generally, the partial derivative of l with respect to $\boldsymbol{\theta} = (\alpha_0, \alpha_1)'$

$$\begin{aligned}\frac{\partial l}{\partial \boldsymbol{\theta}} &= \sum_{t=2}^{T} \frac{\partial l_t}{\partial \sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}},\\ &= -\frac{1}{2}\sum_{t=2}^{T}\left\{\frac{1}{\sigma_t^2} - \frac{\varepsilon_t^2}{\sigma_t^4}\right\}\begin{pmatrix}1\\ \varepsilon_{t-1}^2\end{pmatrix},\end{aligned}$$

where $\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} = (1, \varepsilon_{t-1}^2)'$. Moreover $\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = 0$, then the Hessian matrix is given by

$$\begin{aligned}\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} &= \sum_{t=2}^{T} \frac{\partial^2 l_t}{\partial \sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'},\\ &= -\frac{1}{2}\sum_{t=2}^{T}\left\{\frac{\varepsilon_t^2}{(\sigma_t^2)^3} + \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^4}\right\}\begin{bmatrix}1 & \varepsilon_{t-1}^2\\ \varepsilon_{t-1}^2 & \varepsilon_{t-1}^4\end{bmatrix}.\end{aligned}$$

The Fisher information which is denoted by $I(\boldsymbol{\theta})$ is defined to be the negative of the expected value of the Hessian, that is

$$I(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right].$$

Since

$$E[(\frac{\varepsilon_t^2}{\sigma_t^2} - 1)\frac{1}{\sigma_t^4}\begin{bmatrix}1 & \varepsilon_{t-1}^2\\ \varepsilon_{t-1}^2 & \varepsilon_{t-1}^4\end{bmatrix}|\boldsymbol{\Phi}_{t-1}] = 0$$

and

$$E\left[\frac{\varepsilon_t^2}{(\sigma_t^2)^3}|\boldsymbol{\Phi}_{t-1}\right] = \frac{\sigma_t^2}{\sigma_t^6} = \frac{1}{\sigma_t^4}$$

it follows that

$$I(\boldsymbol{\theta}) = \frac{1}{2}\sum_{t=2}^{T}\left(\frac{1}{\sigma_t^4}\right)\begin{bmatrix}1 & \varepsilon_{t-1}^2,\\ \varepsilon_{t-1}^2 & \varepsilon_{t-1}^4\end{bmatrix}$$

as in Engle [1982]. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ cannot be obtained analytically. In order to mitigate this difficulty we require iterative optimizations. A particular optimization routine that is often employed to estimate the model in ARCH models is BHHH algorithm named after Berndt et al. [1974]. To introduce this algorithm we employ a vectorial notation $\boldsymbol{\theta} = (\alpha_0, \alpha_1)'$ and $\frac{\partial l}{\partial \boldsymbol{\theta}} = (\frac{\partial l}{\partial \alpha_0}, \frac{\partial l}{\partial \alpha_1})'$

where

$$\begin{aligned} l &= \sum_{t=2}^{T} l_t, \\ &= -\frac{1}{2}\sum_{t=2}^{T}\left[\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right]. \end{aligned}$$

According to this algorithm the i^{th} estimator is obtained as

$$\hat{\boldsymbol{\theta}}_i = \hat{\boldsymbol{\theta}}_{i-1} + \phi\left(\sum_{t=2}^{T}\frac{\partial l_t}{\partial\boldsymbol{\theta}}\frac{\partial l_t}{\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{i-1}}\right)^{-1}\sum_{t=2}^{T}\frac{\partial l_t}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{i-1}}, \tag{2.6}$$

where $\phi > 0$ is used to modify the step length. This method is a modification of the Newton-Raphson method. Furthermore the algorithm is very sensitive to the initial values. Note that the computations should make sure that $\alpha_0 + \alpha_1 < 1$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1)'$ is asymptotically normal, that is, $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N(0, I(\boldsymbol{\theta})^{-1})$ and $I(\boldsymbol{\theta})$ is approximated.

Forecasting with the ARCH (1) model

Forecasting is one of the primary goals of time series modeling. We consider the series $\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_T$. The $l-$step ahead forecast for $l = 1, 2, \ldots$ at the forecast origin T, denoted by $\varepsilon_T(l)$, is assumed to be the minimum square error predictor. The mean squared prediction error is given by

$$MSE = E[\varepsilon_{T+l} - f(\varepsilon)]^2,$$

where $f(\varepsilon)$ is a function of observations, then

$$\varepsilon_T(l) = E[\varepsilon_{T+l}|\varepsilon_1, \ldots, \varepsilon_T].$$

For the simple ARCH (1) model we have

$$\varepsilon_T(l) = E[\varepsilon_{T+l}|\varepsilon_1, \ldots, \varepsilon_T] = 0.$$

The forecasts for the series ε_t are not useful and it is therefore essential to consider the squared mean-corrected returns ε_t^2. That is,

$$\varepsilon_T^2(l) = E[\varepsilon_{T+l}^2 | \varepsilon_1^2, \ldots, \varepsilon_T^2].$$

The one step ahead forecast is given by

$$\varepsilon_T^2(1) = \hat{\alpha}_0 + \hat{\alpha}_1 \varepsilon_T^2,$$

which is equivalent to

$$\sigma_T^2(1) = E[\sigma_{T+1}^2 | \mathbf{\Phi}_T] = \hat{\alpha}_0 + \hat{\alpha}_1 \varepsilon_T^2$$

where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are conditional maximum likelihood estimates of the model parameters [Tsay, 2010]. Analogously, the two-step ahead forecast of ε_t^2 follows from the law of iterated expectations,

$$\begin{aligned}
\varepsilon_T^2(2) &= E[\varepsilon_{T+2}^2 | \mathbf{\Phi}_T], \\
&= \sigma_T^2(2), \\
&= E[\sigma_{T+2}^2 | \mathbf{\Phi}_T], \\
&= \hat{\alpha}_0^2 + \hat{\alpha}_1 E[\varepsilon_{T+1}^2 | \mathbf{\Phi}_T], \\
&= \hat{\alpha}_0^2 + \hat{\alpha}_1 (\hat{\alpha}_0^2 + \hat{\alpha}_1 \varepsilon_T^2), \\
&= \hat{\alpha}_0 (1 + \hat{\alpha}_1) + \hat{\alpha}_1 \varepsilon_T^2.
\end{aligned}$$

A generic expression for a l-step ahead forecast can be formulate by repeatedly substitution and is given by

$$\varepsilon_T^2(l) = E[\varepsilon_{T+l}^2 | \mathbf{\Phi}_T] = \sigma_T^2(l) = \sum_{i=0}^{l-1} \hat{\alpha}_0 \hat{\alpha}_1^i + \hat{\alpha}_1^l \varepsilon_T^2.$$

This result has been derived elsewhere in the literature [Tsay, 2010].

2.2.2 ARCH (p) model

Model description

The family of ARCH models was introduced by Engle [1982] to accommodate the the dynamics of conditional heteroscedasticity [Gourieroux and Jasiak, 2001]. Its advantages are simplicity of formulation and ease of estimation [Gourieroux and Jasiak, 2001]. The ARCH models have been extensively used in financial empirical research and have bee extended in various respects. Let r_t denote the stochastic process of returns, $E[r_t|\mathbf{\Phi}_{t-1}] = \mu_t$ be mean of returns, ε_t represent a discrete time stochastic process of mean-corrected returns or shocks with conditional mean and variance parameterized by finite dimensional vector, and let $\mathbf{\Phi}_{t-1}$ represent the available information set at time $t-1$.

Definition 2. *A stochastic process $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t\}$ follows an ARCH model of order p if*

$$r_t = E[r_t|\mathbf{\Phi}_{t-1}] + \varepsilon_t = \mu_t + \varepsilon_t,$$

where $E[\varepsilon_t|\mathbf{\Phi}_{t-1}] = 0$ and the conditional variance

$$\begin{aligned} Var[\varepsilon_t|\mathbf{\Phi}_{t-1}] = \sigma_t^2 &= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2, \\ &= \alpha_0 + \alpha(L)_p\varepsilon_t^2, \end{aligned}$$

where L is the lag operator such that $L^k\varepsilon_t = \varepsilon_{t-k}$ and $\alpha(L)_p$ is a polynomial in the lag operator given by

$$\begin{aligned} \alpha(L)_p &= \sum_{i=1}^{p} \alpha_i L^i, \\ &= \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_p L^p. \end{aligned}$$

Alternative specification of the ARCH (p) model is

$$\varepsilon_t = \sigma_t z_t, \ z_t \sim iid(0,1),$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2,$$
$$= \alpha_0 + \alpha(L)_p \varepsilon_t^2.$$

To ensure that the conditional variance is positive the parameters have to satisfy the constraints $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, \ldots, p$. The random variable z_t is not necessarily to be normally distributed. It can follow a leptokurtic distribution. The stochastic process $\{\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_T\}$ is a martingale difference sequence with conditionally heteroscedastic errors. The conditional mean and variance are then $E[\varepsilon_t|\mathbf{\Phi}_{t-1}] = 0$ and

$$Var[\varepsilon_t|\mathbf{\Phi}_{t-1}] = E[\varepsilon_t^2|\mathbf{\Phi}_{t-1}] - (E[\varepsilon_t|\mathbf{\Phi}_{t-1}])^2 = E[\varepsilon_t^2|\mathbf{\Phi}_{t-1}] = \sigma_t^2$$

We also see that $E[\varepsilon_t^k|\mathbf{\Phi}_{t-1}] = 0$ if k is odd.

Theorem 4. *Suppose that the process $\{\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_t\}$ is an ARCH (p) process with $Var[\varepsilon_t] = \sigma^2 < \infty$. Then*

1. *$\nu_t = \sigma_t^2(z_t^2 - 1)$ is a white noise process and*

2. *ε_t^2 is an AR(p) with*

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \nu_t.$$

Proof.

1. We first prove that $\nu_t = \sigma_t^2(z_t^2 - 1)$ is a white noise process.

   (a) The expected value of ν_t is

   $$E[\nu_t] = E[\sigma_t^2(z_t^2 - 1)] = E[\sigma_t^2]E[z_t^2 - 1] = 0.$$

   (b) The variance of ν_t is given by

   $$Var[\nu_t] = E[\nu_t^2] - (E[\nu_t])^2 = E[\nu_t^2] = E[\sigma_t^4(z_t^2 - 1)^2].$$

This expression further simplifies to

$$\begin{aligned} Var[\nu_t] &= E[\sigma_t^4]E[z_t^4 - 2z_t^2 + 1], \\ &= E[\sigma_t^4](E[z_t^4] - 2E[z_t^2] + 1), \\ &= E[\sigma_t^4](3 - 2 + 1), \\ &= 2E[\sigma_t^4], \\ &= 2E[(\sigma_t^2)^2], \\ &= 2E[(\alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2)^2], \end{aligned}$$

which is a constant independent of t.

(c) The covariance between ν_t and ν_{t+s} is given by

$$\begin{aligned} Cov[\nu_t, \nu_{t+s}] &= E[\sigma_t^2(z_t^2 - 1)\sigma_{t+s}^2(z_{t+s}^2 - 1)])^2, \\ &= E[\sigma_t^2(z_t^2 - 1)\sigma_{t+s}^2]E[z_{t+s}^2 - 1], \\ &= 0 \end{aligned}$$

for $s \neq 0$.

2. The desired result follows from:

$$\begin{aligned} \varepsilon_t^2 &= \sigma_t^2 z_t^2, \\ &= \sigma_t^2 + \sigma_t^2(z_t^2 - 1), \\ &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \nu_t. \end{aligned}$$

□

Since the stochastic process ν_t is a martingale difference sequence, it implies that ε_t is an uncorrelated process. Moreover the error term ε_t is stationary with mean zero and constant unconditional variance.

Theorem 5. *Suppose that the process $\{\varepsilon_1, \ldots, \varepsilon_t\}$ is an $ARCH(p)$ process with*

$Var[\varepsilon_t] = \sigma^2 < \infty$. Then

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

with $\alpha_1 + \alpha_2 + \ldots + \alpha_p < 1$.

Proof. The variance of

$$Var[\varepsilon_t] = E[\varepsilon_t^2] - (E[\varepsilon_t])^2 = E[\varepsilon_t^2] = E[E[\varepsilon_t^2|\mathbf{\Phi}_{t-1}]] = E[E[z_t^2\sigma_t^2|\mathbf{\Phi}_{t-1}]],$$

which simplifies to

$$\begin{aligned} Var[\varepsilon_t] &= E[\sigma_t^2 E[z_t^2|\mathbf{\Phi}_{t-1}]], \\ &= E[\sigma_t^2], \\ &= \sigma^2. \end{aligned}$$

Assuming second-order stationarity of ε_t, then the variance of ε_t is obtained as

$$\begin{aligned} E[\sigma_t^2] &= E[\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2], \\ &= \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] + \alpha_2 E[\varepsilon_{t-2}^2] + \ldots + \alpha_p E[\varepsilon_{t-p}^2], \\ &= \alpha_0 + \alpha_1 E[\sigma_t^2] + \alpha_2 E[\sigma_t^2] + \ldots + \alpha_p E[\sigma_t^2], \end{aligned}$$

which implies that

$$\begin{aligned} E[\sigma_t^2] &= \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p}, \\ &= \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}, \\ &= \sigma^2. \end{aligned}$$

$\square$

In order for second-order stationarity of ε_t to hold then the constraint $\sum_{i=1}^{p} \alpha_i < 1$ has to be satisfied. If instead $\sum_{i=1}^{p} \alpha_i \geq 1$, then the unconditional variance does not exist and the process is not covariance-stationary. It is intricate with the $ARCH(p)$ model that in some applications a larger order p must be used, since larger lags only loose their influence on the volatility slowly. The disadvantage of larger order is that many parameters have to be estimated under restrictions.

Estimation of parameters of the ARCH (p) model

There are several likelihood functions which are frequently employed in the ARCH estimation, depending on the distributional assumption of ε_t^2 [Tsay, 2010]. Consider a time series of mean-corrected returns $\{\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_T\}$. Under the normality assumption, the likelihood function of an ARCH (p) model is defined as

$$\begin{aligned} L(\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_T|\boldsymbol{\theta}) &= f(\varepsilon_T|\boldsymbol{\Phi}_{T-1}) f(\varepsilon_{T-1}|\boldsymbol{\Phi}_{T-2}) \ldots f(\varepsilon_{p+1}|\boldsymbol{\Phi}_p) f(\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_p|\boldsymbol{\theta}), \\ &= \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\} f(\varepsilon_1|\boldsymbol{\theta}) f(\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_p|\boldsymbol{\theta}), \end{aligned}$$

with $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ the vector of unknown parameters and $f(\varepsilon_1, \ldots, \varepsilon_p|\boldsymbol{\theta})$ is a joint probability density function of $\{\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_p\}$. Since the exact form of $f(\varepsilon_1, \ldots, \varepsilon_p|\boldsymbol{\theta})$ is unknown, it is commonly dropped from the conditional likelihood function, especial when the sample size T is sufficiently large [Tsay, 2010]. Therefore, the conditional likelihood used is given by

$$L(\varepsilon_{p+1}, \ldots, \varepsilon_T|\boldsymbol{\theta}, \varepsilon_1, \ldots, \varepsilon_p) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\},$$

where σ_t^2 can be evaluated recursively. The maximum likelihood estimates are obtained by maximizing this expression, or, equivalently the log-likelihood function

$$l(\varepsilon_{p+1}, \ldots, \varepsilon_T|\boldsymbol{\theta}, \varepsilon_1, \ldots, \varepsilon_p) = \sum_{t=p+1}^{T} l_t,$$

where $l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}$ is the log-likelihood of observation at time t. The methods of optimization used are the same as those employed in the $ARCH(1)$ model.

Forecasting with the ARCH (p) model

The procedure of forecasting using an ARCH (p) model is very similar as that of an ARCH (1) model. If we have a time series of mean-corrected returns $\{\varepsilon_1, \ldots, \varepsilon_T\}$, the $l-$step ahead forecast, represented by $\varepsilon_T(l)$ is the minimum mean squared error predictor that minimizes $E[\varepsilon_{T+l} - f(\varepsilon)]^2$ where $f(\varepsilon)$ is a function of observations, ε [Talke, 2003]. Therefore, since $E[\varepsilon_T(l)] = 0$, this predictor is not instrumental ε_t [Talke, 2003]. Therefore, we consider the squared mean-corrected returns ε_t^2. The forecasts of the ARCH (p) model are obtained recursively.

The singe-step ahead forecast for σ^2_{T+1} is given by

$$\begin{aligned}\sigma^2_T(1) &= \hat{\alpha}_0 + \hat{\alpha}_1\varepsilon^2_T + \ldots + \hat{\alpha}_p\varepsilon^2_{T+1-p},\\ &= \hat{\alpha}_0 + \sum_{i=1}^{p}\varepsilon^2_{T+1-i},\end{aligned}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_p)'$ is the vector of conditional estimates [Tsay, 2010].

For the $2 - step$ ahead forecast for σ^2_{T+2}, we need the forecast of ε^2_{T+1} which is given by $\sigma^2_T(1)$. We, therefore, have

$$\sigma^2_T(2) = \hat{\alpha}_0 + \hat{\alpha}_1\sigma^2_T(1) + \hat{\alpha}_1\varepsilon^2_T \ldots + \hat{\alpha}_p\varepsilon^2_{T+2-p}.$$

The $l - step$ ahead forecast for σ^2_{T+k} is given by

$$\begin{aligned}\sigma^2_T(l) &= \hat{\alpha}_0 + \hat{\alpha}_1\sigma^2_T(l-1) + \ldots + \hat{\alpha}_p\sigma^2_T(l-p),\\ &= \hat{\alpha}_i + \sum_{i=1}^{p}\hat{\alpha}_0\sigma^2_T(l-i),\end{aligned}$$

for $i = 1, 2, \ldots$, where $\sigma^2_T(l-i) = \varepsilon^2_{T+l-i}$ if $l \leq i$.

2.2.3 ARCH models with non-Gaussian error distributions

In spite of the strengths of the assumption that the mean-corrected returns or errors ε_t are conditionally normal, ARCH models can be specified and estimated using alternative distributional assumptions. The consideration for implementing distributions different from the normal can enhance the model. A more suitable choice of the conditional distribution of the standardized returns may enhance the precision of the volatility process parameter estimates, in the case of maximum likelihood estimation, the estimates will be efficacious. There are three distributions among the many that have been employed to estimate the parameters of the ARCH process. The first distribution is a standardized student's t distribution with given degrees of freedom say v [Bollerslev, 1987]. The distribution of ε_t follows a Student-t distribution if its probability density function is given by

$$f(\varepsilon_t, v, \sigma_t^2) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{\pi(v-2)}} \frac{1}{\sigma_t} [1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2}]^{-\frac{(v+1)}{2}}, \tag{2.7}$$

where $\Gamma(.)$ is the gamma function. That is,

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy.$$

This distribution is only well defined if $v > 2$.

Thus we may express the conditional likelihood of ε_t as

$$f(\varepsilon_{p+1}, \ldots, \varepsilon_T | \boldsymbol{\theta}, \boldsymbol{\Phi}_p) = \prod_{t=p+1}^{T} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{\pi(v-2)}} \frac{1}{\sigma_t} [1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2}]^{-\frac{(v+1)}{2}}, \tag{2.8}$$

where $v > 2$ and $\boldsymbol{\Phi}_p = (\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_p)$ [Tsay, 2010, Hamilton, 1994]. We refer to the estimates that maximizes the conditional likelihood function as the maximum likelihood estimates under t distribution [Tsay, 2010]. A value of degrees of freedom between 4 and 8 is often used if it is pre-specified [Tsay, 2010]. Thus if the degrees of freedom v of the $Student - t$ distribution is pre-specified, then the conditional log-likelihood function is given by

$$l(\varepsilon_{p+1}, \ldots, \varepsilon_T | \boldsymbol{\theta}, \boldsymbol{\Phi}_p) = -\sum_{t=p+1}^{T} [\frac{v+1}{2} \ln(1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2}) + \frac{1}{2} \ln(\sigma_t^2)]. \tag{2.9}$$

Nevertheless, if the degrees of freedom parameter v is to be estimated by maximum likelihood estimation, then the log-likelihood function is modified into

$$l(\varepsilon_{p+1},\ldots,\varepsilon_T|\boldsymbol{\theta},v,\boldsymbol{\Phi}_p) = (T-p)\left\{\ln[\Gamma(\frac{v+1}{2})] - \ln(\Gamma[\frac{v}{2}]) - \frac{1}{2}\ln[(v-2)\pi]\right\} + l(\varepsilon_{p+1},\ldots,\varepsilon_T|\boldsymbol{\theta},\boldsymbol{\Phi}_p)$$

that incorporates additional terms.

The second distribution suggested in the literature is the generalized error distribution [Nelson, 1991]. A random variable ε_t with shape parameter v, a mean of zero, and variance σ_t^2 belongs to the generalized error distribution if its has a probability density function given by

$$f(\varepsilon_t|\boldsymbol{\theta},v) = \frac{v\exp(-\frac{1}{2}|\frac{\varepsilon_t}{\lambda\sigma_t}|^v)}{2^{\frac{v+1}{v}}\lambda\Gamma(\frac{1}{v})}, \tag{2.10}$$

where $\lambda = \frac{2^{-\frac{2}{v}}\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}$.

When the shape parameter $v = 2$ the generalized error distribution becomes a standard normal distribution. The GED is fat-tailed when $v < 2$ and thin-tailed when $v > 2$. In order for this distribution to be employed for forecasting ARCH parameters, it is necessary that $v \geq 1$ since the variance is not finite when $v < 1$. The maximum likelihood estimates can be obtained by maximizing the log-likelihood function, and using BHHH algorithm in the R-numerical optimization routines.

The third useful distribution introduced by Hansen [1994] extends the standardized $student-t$ distribution to accommodate skewness of returns. Thus the probability density function of ε_t is given by

$$f(\varepsilon_t|v,\lambda,\boldsymbol{\theta}) = \begin{cases} bc(1+\frac{1}{(v-2)}(\frac{b\varepsilon_t+a\sigma_t}{\sigma_t(1-\lambda)})^2)^{-\frac{v+1}{2}} & \text{for} \quad \varepsilon_t < -\frac{a}{b}, \\ bc(1+\frac{1}{(v-2)}(\frac{b\varepsilon_t+a\sigma_t}{\sigma_t(1+\lambda)})^2)^{-\frac{v+1}{2}} & \text{for} \quad \varepsilon_t \geq -\frac{a}{b}, \end{cases} \tag{2.11}$$

where $a = 4\lambda c(\frac{v-2}{v-1}), b = 1+3\lambda^2 - a^2$ and $c = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi(v-2)}\Gamma(\frac{v}{2})}$.

The parameters v and λ in this distribution control the kurtosis and skewness respectively. This distribution may be a better approximation to the true distribution since they allow kurtosis which is greater than that of the normal distribution.

2.3 GARCH (p, q) model

In this section, we discuss the GARCH (p, q) model. We first describe the $GARCH(1, 1)$ model and then examine statistical consideration for estimation and forecasting approaches that exist in the literature.

2.3.1 GARCH $(1, 1)$ model

Despite the fact that the ARCH model is simple, it often requires many parameters to fit the data. To circumvent this difficulty, Bollerslev [1986] proposed a useful extension of ARCH model known as the generalized ARCH (GARCH) model. Their results suggested that the GARCH model better captures the volatility in a return series than the ARCH model. Let $\varepsilon_t = r_t - \mu_t$ be the mean-corrected return at time t.

Definition 3. *A stochastic process* $\{\varepsilon_1, \ldots, \varepsilon_t\}$ *follows a GARCH* $(1, 1)$ *model if*

$$\varepsilon_t = \sigma_t z_t,$$

and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2.12}$$

where $\{z_t\}$ *is a sequence of independent and identically distributed random variables with zero mean and variance one.*

An imperative condition for the variance to be positive is that parameters are constrained such that $\alpha_0 > 0$, $\alpha_1 \geq 0$, and $\beta_1 \geq 0$. Equation (2.12) exhibits that large past mean-corrected squared returns ε_{t-1}^2 or past conditional variances σ_{t-1}^2 or both give rise to large σ_t^2. This means that a large ε_{t-1}^2 tends to be followed by another large ε_t^2, generating, again, the well-known behavior of volatility clustering in financial time series Tsay [2010].

The stochastic process $\{\varepsilon_t\}$ is a martingale difference sequence with conditionally heteroscedasticity errors, since $E[\varepsilon_t|\mathbf{\Phi}_t] = 0$. As in the ARCH (1) model, the GARCH $(1,1)$ model can be expressed as

$$\begin{aligned}
\varepsilon_t^2 &= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2, \\
&= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 - \beta_1\varepsilon_{t-1}^2 + \beta_1\varepsilon_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2, \\
&= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\varepsilon_{t-1}^2 + \beta_1(\sigma_{t-1}^2 - \varepsilon_{t-1}^2) + \varepsilon_t^2 - \sigma_t^2, \\
&= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\varepsilon_{t-1}^2 - \beta_1\nu_{t-1} + \nu_t, \\
&= \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1\nu_{t-1} + \nu_t,
\end{aligned}$$

where $\nu_t = \varepsilon_t^2 - \sigma_t^2$.

As in the case of ARCH (1) model, the conditional mean of ν_t is zero, that is, $E[\nu_t|\mathbf{\Phi}_{t-1}] = 0$, and ν_t is martingale difference sequence. The information history is given by $\mathbf{\Phi}_{t-1} = \left\{\varepsilon_1, \sigma_1^2, \ldots, \varepsilon_{t-1}, \sigma_{t-1}^2\right\}$. However, unlike ARCH (1) process which can be transformed into AR(1), the GARCH $(1,1)$ process is transformed into ARMA(1,1). By implementing the same methods that were employed to the ARCH (1) model for deriving the unconditional variances of ε_t, the conditional variance for the GARCH $(1,1)$ return process is given by

$$\begin{aligned}
Var[\varepsilon_t] &= E[\varepsilon_t^2] - (E[\varepsilon_t])^2, \\
&= E[\varepsilon_t^2], \\
&= E[\alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1\nu_{t-1} + \nu_t], \\
&= \alpha_0 + E[(\alpha_1 + \beta_1)\varepsilon_{t-1}^2], \\
&= \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)},
\end{aligned}$$

since $E[\varepsilon_t^2] = E[\varepsilon_{t-1}^2]$. The requirements for stationarity are such that $1 - \alpha_1 - \beta_1 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_0 > 0$.

Theorem 6. *Suppose that the process $\{\varepsilon_1, \ldots, \varepsilon_t\}$ is a GARCH $(1,1)$ process with $Var[\varepsilon_t] = \sigma^2 < \infty$ and z_t is normally distributed with mean zero and variance one. Then $E[\varepsilon_t^4] < \infty$ holds if and only if $2\alpha_1\beta_1 + 3\alpha_1^2 + \beta_1^2 < 1$. The kurtosis is given* *as*

$$Kurt[\varepsilon_t] = 3 + \frac{6\alpha_1^2}{1 - 2\alpha_1\beta_1 - 3\alpha_1^2 - \beta_1^2}.$$

Proof. If we assume that z_t is normally distributed with mean zero and variance one and fourth-order stationarity holds, then, $E[z_t^4] = 3$ and henceforth

$$\begin{aligned} E[\varepsilon_t^4] &= E[E[\varepsilon_t^4 | \mathbf{\Phi}_{t-1}]], \\ &= E[E[\sigma_t^4 z_t^4 | \mathbf{\Phi}_{t-1}]], \\ &= E[\sigma_t^4 E[z_t^4 | \mathbf{\Phi}_{t-1}]], \\ &= E[3\sigma_t^4], \\ &= 3E[\sigma_t^4]. \end{aligned}$$

The expression for $E[\sigma_t^4]$ is given by

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

The expression for $E[\sigma_t^4]$ can be simplified further as

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

and finally, upon substituting $\sigma^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$ for σ^2 in this expression we get

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

The kurtosis of a GARCH (1,1) model is then given by

$$\begin{aligned} Kurt[\varepsilon_t] &= \frac{E[\varepsilon_t^4]}{(Var[\varepsilon_t])^2}, \\ &= \frac{3(1+\alpha_1+\beta_1)(1-\alpha_1-\beta_1)}{1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2} > 3, \\ &= 3 + \frac{6\alpha_1^2}{1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2}. \end{aligned}$$

□

Therefore, the required condition for the kurtosis to exist is such that

$$1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2 > 0.$$

and the GARCH $(1,1)$ model has fatter tails than those of normal distribution. In practical applications it is frequently shown that models with smaller order sufficiently describe the data. In most cases GARCH $(1,1)$ is sufficient.

Estimation of parameters of the GARCH $(1,1)$ model

The estimation of parameters of the GARCH $(1,1)$ model is carried in a similar fashion as that of the ARCH (1) model. For parameter estimation, Bollerslev [1986] suggested that the unconditional variance for ε_t should be considered as a starting value for variance. That is,

$$E[\varepsilon_t^2] = \frac{\alpha_0}{1-\alpha_1-\beta_1}.$$

Under the normality assumption, the likelihood function of the GARCH $(1,1)$ model is expressed as

$$\begin{aligned} f(\varepsilon_1,\dots,\varepsilon_T,\sigma_1^2,\dots,\sigma_t^2) &= f(\varepsilon_T,\sigma_T^2|\mathbf{\Phi}_{T-1})f(\varepsilon_{T-1},\sigma_{T-1}^2|\mathbf{\Phi}_{T-2})\dots f(\varepsilon_2,\sigma_2^2|\mathbf{\Phi}_1)f(\varepsilon_1,\sigma_1^2|\boldsymbol{\theta}), \\ &= \prod_{t=2}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\} f(\varepsilon_1,\sigma_1^2|\boldsymbol{\theta}), \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1)'$ is the vector of unknown parameters and $f(\varepsilon_1, \sigma_1^2|\boldsymbol{\theta})$ is a probability density function. Since $f(\varepsilon_1, \sigma_1^2|\boldsymbol{\theta})$ is complicated, it is commonly dropped from the conditional likelihood function, especially when the sample size is sufficiently large.

Henceforth, we consider the conditional likelihood function

$$f(\varepsilon_2, \ldots, \varepsilon_T, \sigma_2^2, \ldots, \sigma_T^2|\boldsymbol{\theta}, \varepsilon_1^2, \sigma_1^2) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\}$$

to estimate $\boldsymbol{\theta}$.

The maximum likelihood estimates are obtained by directly maximizing this expression or, equivalently the log-likelihood function given by

$$l(\boldsymbol{\theta}|\varepsilon_1, \sigma_1^2) = \sum_{t=2}^{T} l_t(\boldsymbol{\theta}),$$

where

$$l_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2},$$

is the log-likelihood function of observation t.

The algorithm for optimization employed to find the conditional maximum likelihood estimates is similar to that discussed for the ARCH (1) model.

Forecasting with the GARCH $(1, 1)$ model

Forecasts for the GARCH $(1, 1)$ model are obtained recursively in a similar way as for the ARCH (1) model. Let T to be the forecast origin. Then the single$-$*step* ahead forecast for ε_{T+1}^2 is

$$\varepsilon_T^2(1) = \hat{\alpha}_0 + \hat{\alpha}_1\varepsilon_T^2 + \hat{\beta}_1\varepsilon_T^2.$$

Since $\varepsilon_T^2 = \sigma_T^2 z_T^2$, the GARCH $(1,1)$ model can be re-written as

$$\begin{aligned}\sigma_T^2 &= \alpha_0 + \alpha_1 \varepsilon_{T-1}^2 + \beta_1 \sigma_{T-1}^2, \\ &= \alpha_0 + (\alpha_1 + \beta_1)\sigma_{T-1}^2 + \alpha_1 \sigma_{T-1}^2 (z_{T-1}^2 - 1),\end{aligned}$$

so that at time $T+2$, we have

$$\sigma_{T+2}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{T+1}^2 + \alpha_1 \sigma_{T-1}^2 (z_{T+1}^2 - 1)$$

with $E[(z_{T+1}^2 - 1)|\mathbf{\Phi}_T] = 0$. Therefore the $2-step$ ahead forecast is

$$\sigma_T^2(2) = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_T^2(1). \tag{2.13}$$

In general, the $l-step$ ahead forecast for σ_{T+l}^2 is

$$\sigma_T^2(l) = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_T^2(l-1) \tag{2.14}$$

for $l > 1$.

By repeated substitution of (2.14) the $l-step$ ahead forecast for GARCH $(1,1)$ can be expressed as

$$\sigma_T^2(l) = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1)^{l-1}(\sigma_T^2(1) - \hat{\sigma}^2)$$

for $l > 1$,where $\hat{\sigma}^2 = \frac{\hat{\alpha}_0}{1-\hat{\alpha}_1-\hat{\beta}_1}$.

The expression shows that $\sigma_T^2(l) \to \hat{\sigma}^2$ as $l \to \infty$, provided that $(\hat{\alpha}_1 + \hat{\beta}_1) < 1$. The multi-step ahead forecast of a GARCH $(1,1)$ model converges to the unconditional variance of ε_t as the forecast horizon goes to infinity provided that the variance of ε_t exists [Tsay, 2010].

2.3.2 GARCH (p,q) model

The GARCH (p,q) model is a generalization of the GARCH $(1,1)$ model which was discussed in the previous section. For a log return series $\{r_t\}$, let $\varepsilon_t = r_t - \mu_t$

be the mean corrected return at time t [Tsay, 2010].

Definition 4. *A stochastic process $\{\varepsilon_1, \ldots, \varepsilon_t\}$ follows a GARCH (p, q) model if*

$$\begin{aligned}\varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,\end{aligned}$$

where $\{z_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance one, $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, 2, \ldots, p)$, $\beta_j \geq 0$ $(j = 1, 2, \ldots, q)$, and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$.

Theorem 7. *Suppose that the process $\{\varepsilon_1, \ldots, \varepsilon_t\}$ is GARCH (p, q) process with $E[\varepsilon_t^4] = c < \infty$ and z_t normally distributed with mean zero and variance one. It follows that*

1. *$\nu_t = \sigma_t^2(z_t^2 - 1)$ is white noise.*

2. *ε_t^2 is an Autoregressive Moving Average(m,p) process with*

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} \gamma_i \varepsilon_{t-i}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-j} + \nu_t \tag{2.15}$$

   with $m = max(p, q)$, $\gamma_i = \alpha_i + \beta_i$, $\alpha_i = 0$ when $i > p$, and $\beta_j = 0$ when $j > q$.

Proof.

1. In order to show that $\nu_t = \sigma_t^2(z_t^2 - 1)$ is white noise, we find the expected value of ν_t and the variance of ν_t.

   (a) The expected value of ν_t is

   $$E[\nu_t] = E[\sigma_t^2(z_t^2 - 1)] = E[\sigma_t^2]E[z_t^2 - 1] = 0$$

   (b) The variance of ν_t is

   $$Var[\nu_t] = E[\nu_t^2] - (E[\nu_t])^2 = E[\nu_t^2] = E[\sigma_t^4(z_t^2 - 1)^2] = E[\sigma_t^4]E[z_t^4 - 2z_t^2 + 1]$$

which simplifies to

$$Var[\nu_t] = E[\sigma_t^4](E[z_t^4]-2E[z_t^2]+1) = E[\sigma_t^4](3-2+1) = 2E[\sigma_t^4] = 2E[(\sigma_t^2)^2]$$

and finally leads

$$Var[\nu_t] = 2E[(\alpha_0 + \sum_{i=1}^{p} \varepsilon_{t-i}^2)^2] = constant,$$

which is a constant independent of t.

(c) For $k \neq 0$, the covariance between ν_t and ν_{t+k} is

$$\begin{aligned} Cov[\nu_t, \nu_{t+k}] &= E[\sigma_t^2(z_t^2 - 1)\sigma_{t+s}^2(z_{t+k}^2 - 1)])^2 \\ &= E[\sigma_t^2(z_t^2 - 1)\sigma_{t+k}^2]E[z_{t+k}^2 - 1], \\ &= 0 \end{aligned}$$

2. We can rewrite ε_t^2 as follows

$$\varepsilon_t^2 = \sigma_t^2 z_t^2 = \sigma_t^2 + \sigma_t^2 z_t^2 - \sigma_t^2 = \sigma_t^2 + \sigma_t^2(z_t^2 - 1).$$

Given that

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

it then follows that

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \nu_t.$$

But from $\nu_{t-j} = \sigma_{t-j}^2(z_{t-j}^2 - 1)$ we obtain

$$\sigma_{t-j}^2 = \sigma_{t-j}^2 z_{t-j}^2 - \nu_{t-j} = \varepsilon_{t-j}^2 - \nu_{t-j}$$

so that

$$\begin{aligned}
\varepsilon_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j (\varepsilon_{t-j}^2 - \nu_{t-j}) + \nu_t, \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-j} + \nu_t, \\
&= \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-j} + \nu_t, \\
&= \alpha_0 + \sum_{i=1}^{m} \gamma_i \varepsilon_{t-i}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-j} + \nu_t.
\end{aligned}$$

□

The series $\{\nu_t\}$ is a martingale difference sequence, that is, $E[\nu_t] = 0$ and $Cov(\nu_t, \nu_{t-j}) = 0$ for $j \geq 1$. Nonetheless, the series $\{\nu_t\}$ in general is not an independent and identical distributed sequence. The equation (2.15) is expressed as an ARMA form for the mean-corrected squared series ε_t^2. The fundamental and adequate conditions for the positivity of the conditional variance in higher-order GARCH models are formidable than the sufficient conditions stipulated and have been provided in [Nelson and Cao, 1992].

Theorem 8. *Suppose that the process $\{\varepsilon_1, \ldots, \varepsilon_t\}$ is a GARCH (p, q) process with $Var[\varepsilon_t] = \sigma^2 < \infty$. Then*

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}.$$

with $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$

Proof. The variance of ε_t is given by

$$Var[\varepsilon_t] = E[\varepsilon_t^2] - (E[\varepsilon_t])^2 = E[\varepsilon_t^2] = E[E[\varepsilon_t^2 | \mathbf{\Phi}_{t-1}]] = E[\sigma_t^2] = \sigma^2$$

Then

$$
\begin{aligned}
Var[\varepsilon_t] &= E[\alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2], \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i E[\varepsilon_{t-i}^2] + \sum_{j=1}^{q} \beta_j E[\sigma_{t-j}^2], \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma^2 + \sum_{j=1}^{q} \beta_j \sigma^2
\end{aligned}
$$

Therefore, we have

$$
E[\varepsilon_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j},
$$

given that the denominator of the function is positive [Tsay, 2010]. $\square$

However, there are several disadvantages of the GARCH models. Firstly, they do not allow leverage effect which is a significant feature in stock returns. Secondly, we need to impose restrictions on the parameters to ensure the positivity of the unconditional variance of ε_t^2 which complicate the estimation procedure. Furthermore, the interpretation of the persistence in the GARCH models is indeterminate.

Estimation of parameters of the GARCH (p, q) model

In general, to estimate the parameters of the GARCH (p, q) model, we employ the maximum likelihood estimation. If we consider a time series of returns $\{r_1, \ldots, r_T\}$, and the denote $m = max(p, q)$ the number of observations lost for initializing the process. Under normality assumption, the likelihood function of a GARCH (p, q) model is

$$
\begin{aligned}
f(\varepsilon_1, \ldots, \varepsilon_T | \boldsymbol{\theta}) &= f(\varepsilon_T | \boldsymbol{\Phi}_{T-1}) f(\varepsilon_{T-1} | \boldsymbol{\Phi}_{T-2}) \ldots f(\varepsilon_{m+1} | \boldsymbol{\Phi}_m) f(\varepsilon_1, \ldots, \varepsilon_m | \boldsymbol{\theta}), \\
&= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{ -\frac{\varepsilon_t^2}{2\sigma_t^2} \right\} f(\varepsilon_{m+1} | \boldsymbol{\Phi}_m) f(\varepsilon_1, \ldots, \varepsilon_m | \boldsymbol{\theta}),
\end{aligned}
$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ is a vector of unknown parameters and $f(\varepsilon_1, \ldots, \varepsilon_m | \boldsymbol{\theta})$ is a joint probability density function of $\varepsilon_1, \ldots, \varepsilon_m$.

The exact form of $f(\varepsilon_1, \ldots, \varepsilon_m|\boldsymbol{\theta})$ is complicated, therefore it is dropped from the conditional likelihood function, especially when the sample size T is sufficiently large. Thus we focus on the conditional likelihood function

$$f(\varepsilon_{m+1}, \ldots, \varepsilon_T|\boldsymbol{\theta}, \varepsilon_1, \ldots, \varepsilon_m) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\} \tag{2.16}$$

Tsay [2010]. The maximum likelihood estimates are obtained by maximizing equation (2.16) with respect to $\boldsymbol{\theta}$, or, equivalently the conditional likelihood function

$$l(\boldsymbol{\theta}|\varepsilon_t, t = 1, 2, \ldots, T) = \sum_{t=m+1}^{T} l_t(\boldsymbol{\theta})$$

where $l_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}$ is the log-likelihood of observation t with $\varepsilon_t = r_t - \mu_t$. We employ the same methods of optimization as those discussed in the ARCH (1) model. Further, we can apply generalized error distribution, student t distribution and standardized student's t distribution to enhance the accuracy of the parameter estimation.

Forecasting with the GARCH (p, q) model

Forecasting volatility of a return series $\{r_t\}$ with the GARCH (p, q) is done in the same way as that of a GARCH $(1, 1)$ model. Let T be the starting date for forecasting. Then the $1 - step$ ahead forecast for σ_{T+1}^2 is given by

$$\sigma_T^2(1) = \hat{\alpha}_0 + \sum_{i=1}^{m}(\hat{\alpha}_i + \hat{\beta}_i)E[\varepsilon_{T+1-i}^2|\boldsymbol{\Phi}_T] - \sum_{i=1}^{q}\hat{\beta}_i E[\nu_{T+1-i}|\boldsymbol{\Phi}_T],$$

where $(\varepsilon_T^2, \ldots, \varepsilon_{T+l-m})$ and $(\sigma_T^2, \ldots, \sigma_{T-q+1}^2)$ are known at time T.

By extension, the $l - step$ ahead forecast of the conditional variance in a GARCH

(p, q) model is given by

$$\begin{aligned}\sigma_T^2(l) &= E[\varepsilon_{T+l}^2|\mathbf{\Phi}_T], \\ &= \hat{\alpha}_0 + \sum_{i=1}^{m}(\hat{\alpha}_i + \hat{\beta}_i)E[\varepsilon_{T+l-1}^2|\mathbf{\Phi}_T] - \sum_{i=1}^{q}\hat{\beta}_i E[\nu_{T+l-i}|\mathbf{\Phi}_T]\end{aligned} \tag{2.17}$$

where $E[\varepsilon_{T+l}^2|\mathbf{\Phi}_T]$ is given recursively in equation (2.17).

2.4 Exploratory and model diagnostic techniques

2.4.1 Investigating Stationarity

One of the fundamental concepts underpinning time series analysis is stationarity. A time series $\{r_t\}$ is said to be strictly stationary if the joint distribution $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+\tau}, \ldots, r_{t_k+\tau})$ for all τ, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers [Tsay, 2010].

A weaker classification of stationarity is often assumed. A time series $\{r_t\}$ is weakly stationary if both the mean of r_t and covariance between r_t and r_{t-l} are unchanging with time, where l is an arbitrary integer. Mathematically, a series $\{r_t\}$ is weakly stationary if

- $E[r_t] = \mu$, which is a constant, and
- $Cov(r_t, r_{t-l}) = \gamma_l$, which only depends on l [Tsay, 2010].

Suppose we have observed a time series data points $\{r_1, \ldots, r_T\}$, the weak stationarity suggests that the time plot of the data would portray that the T values of the series fluctuate with a constant variation around a fixed level. The weak stationarity allows us to make inference about future observations [Tsay, 2010].

Autocorrelation

Suppose we have a weakly stationary time series $\{r_t\}$. In order to assess the linear dependence between r_t and r_{t-l}, we focus on the correlation. The correlation coefficient between r_t and r_{t-l} is called the lag-l autocorrelation of r_t and is denoted by ρ_l, which under weak stationarity assumption is a function of l only. Basically, we define

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{Cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}$$

where the property $Var(r_t) = Var(r_{t-l})$ for a weakly stationary time series is used. Autocorrelations take values in the interval $[-1, 1]$. That is, $-1 \leq \rho_l \leq 1$. Furthermore, a weakly stationary time series $\{r_t\}$ is not serially uncorrelated if and only if $\rho_l = 0$ for all $l > 0$.

Portmanteau test and Box-Ljung test for autocorrelations

In financial time series modelling it is essential to test jointly that several autocorrelations of r_t are zero. Box and Pierce [1970] proposed the Portmanteau statistic which is given by

$$Q^*(m) = T \sum_{l=1}^{m} \hat{\rho}_l^2,$$

where T is the sample size of returns, m is the number of lags and $\hat{\rho}$ is the estimate of the l^{th} autocorrelation of returns given by

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}, 0 \leq l < T - 1$$

with

$$\bar{r} = \sum_{t=1}^{T} \frac{r_t}{T}.$$

The aim is to the test a null hypothesis $H_0 : \rho_1 = \ldots = \rho_m = 0$ against the alternative hypothesis $H_a : \rho \neq 0$ for $i = 1, 2, \ldots, m$. If underlying assumption is that $\{r_t\}$ is an independent and identically distributed sequence with certain moment conditions, then $Q^*(m)$ is asymptotically a chi-squared random variable with m degrees of freedom.

Ljung and Box [1978] revised the $Q^*(m)$ statistic to enhance the power of the test in finite samples. The Box-Ljung statistic is given by

$$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{T-l},$$

The decision rule is to reject the null hypothesis if $Q(m)$ is greater than χ^2_α, where χ^2_α denotes the $100(1-\alpha)^{th}$ percentile of a chi-squared distribution with m degrees of freedom.

Jarque-Bera test

Jarque-Bera is a test statistic that is employed for testing whether the series of T observations is normally distributed. The Jarque-Bera (JB) statistic is calculated as

$$JB = \frac{T}{6}(S^2 + \frac{1}{4}(K-3)^2)$$

where S is the sample skewness, K is the sample kurtosis and T is the sample size. Under normality assumption S and $K-3$ are distributed asymptotically as normal with zero mean and variances $\frac{6}{T}$ and $\frac{24}{T}$, respectively. We reject the null hypothesis at $\alpha\%$ significance level if $JB > \chi^2_{1-\frac{\alpha}{2}}$, where $\chi^2_{1-\frac{\alpha}{2}}$ is the critical value of the chi-square distribution with 2 degrees of freedom [Tsay, 2010].

2.4.2 Testing for ARCH effect

It is intuitive to check if there are ARCH effects in the residuals before we fit the GARCH model. Let $\varepsilon_t = r_t - \mu_t$ be the residuals of the mean equation. The time series $\{\varepsilon_t^2\}$ is used to detect conditional heteroscedasticity and two possible test procedures can be used for this purpose. The first test that we can implement is the usual Box-ljung test $Q(m)$ to for auto-correlations in the series $\{\varepsilon_t^2\}$ [McLeod and Li, 1983]. The second test which can be used to detect the ARCH effect in a time series of returns is the Lagrange Multiplier (LM) test of [Engle, 1983]. The null hypothesis is that the first m lags of the ACF of the series $\{\varepsilon_t^2\}$ are zero. The lagrange multiplier test for testing the null hypothesis corresponds to the normal $F - statistic$ in the linear regression

$$\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_m\varepsilon_{t-m}^2 + e_t, t = m+1, \ldots, T, \tag{2.18}$$

where e_t represents the error term, m is the number of lags, and T is the sample size. Let $SSR_0 = \sum_{t=m+1}^{T}(\varepsilon_t^2 - \hat{\mu})^2$, where $\hat{\mu} = (\frac{1}{T})\sum_{t=1}^{T}\varepsilon_t^2$ is the sample mean of $\{\varepsilon_t^2\}$, and $SSR_1 = \sum_{t=m+1}^{T}\hat{e}_t^2$, where $\hat{e}_t$ is the least squares residuals of the prior linear regression. Then we can define the statistic

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

which is asymptotically distributed as a chi-square distribution with m degrees of freedom under the null hypothesis. Furthermore, the decision rule stipulates that we reject the null hypothesis if F is greater than $\chi_m^2(\alpha)$, where the $\chi_m^2(\alpha)$ is the upper $100(1-\alpha)th$ percentile of χ_m^2 [Tsay, 2010].

2.4.3 Diagnostic Checking

When a GARCH model has been fitted to the data, we need to explore the adequacy of the fitted model by using several graphical and statistical diagnostics. For a properly specified GARCH model, the standardized residuals given by

$$\tilde{\varepsilon_t} = \frac{\hat{\varepsilon_t}}{\hat{\sigma_t}} \tag{2.19}$$

form a sequence of independent and identical distributed random variables. We explore the goodness of fit of the model by examining the series of estimated standardized residuals. If the GARCH model is correctly specified the residuals should portray no serial correlation, conditional heteroscedasticity or any type of non-linear dependence. Further, the distribution of the standardized residuals should match the specified error distribution used in the estimation. To detect the ARCH effects, we can plot the ACF of the squared standardized residuals. Statistically, the modified Box-Ljung statistics can be used to test the null hypothesis of no autocorrelation up to a specified lag and Engle's ARCH can be employed to test the ARCH effect. If it is assumed that errors are normally distributed, then a Quantile-Quantile plot should look roughly linear, and the Jarque-Bera statistic should not be too much larger than six [Zivot, 2009].

2.5 Model Selection

In time series modelling it often essential to identify the model that best fits the data from a set of candidate models. We now consider several selection procedures that have been proposed in the time series literature for selecting among various possible choices for GARCH models.

The most common selection criterion employed are the Akaike information criterion (AIC) and the Schwartz Bayesian information criterion (BIC). If the possible models possess the nested structure in terms of the possible parameters, we can simply compare their empirical likelihoods. But that is not always the case where the models differ in the number of parameters. In such cases, we can apply the Akaike information criterion which make improvements to the likelihood function to account for the number of parameters in the model.

If the number of parameters in the model is denoted as p, then the AIC is defined by

$$AIC(p) = -2\ln(ML) + 2p,$$

where ML is the maximum likelihood.

The second model criterion selection criteria is given by

$$BIC(p) = -2\ln(ML) + p\ln(T)$$

where ML is the maximum likelihood and T is the number of observations. For a large data set, the BIC has a heavier penalty for the number of parameters in the model, therefore it will select a more parsimonious model than AIC does [Zivot, 2009].

2.6 Data Analysis

2.6.1 Data characteristics

This section presents a preliminary description of the data set used and provides an exploratory analysis of the data prior to the complete analysis. The time series data employed in this study consist of daily closing prices of Johannesburg Stock Exchange (JSE) market index over the period from June 30, 1995 to June 6, 2012, constituting 4320 observations during. The data for the JSE closing prices were obtained with permission from the McGregorBFA website (http://research.mcgregorbfa.com/Default.aspx). To prepare the data for analysis, we employ the continuously compounded daily returns which are given by

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2.20}$$

where P_t and P_{t-1} are the daily closing market index of the JSE at time t and $t-1$, respectively. The return series consist of 4319 observations because one observation is lost due to differencing the daily closing price series.

Figure 2.1 presents the time series plot of the JSE daily closing prices. It is evident from the figure that the time series of JSE daily closing prices is non-stationary due to the non-constant mean. With the purpose of getting stationary financial time series, we transformed the prices into a natural logarithmic returns, which are displayed in Figure 2.2. The series of returns has constant mean except clearly

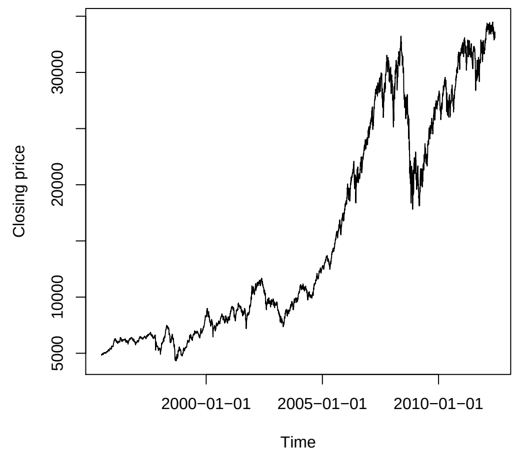

Figure 2.1. The daily closing price of the JSE All share index for the period 30 June 1995 to 6 June 2012.

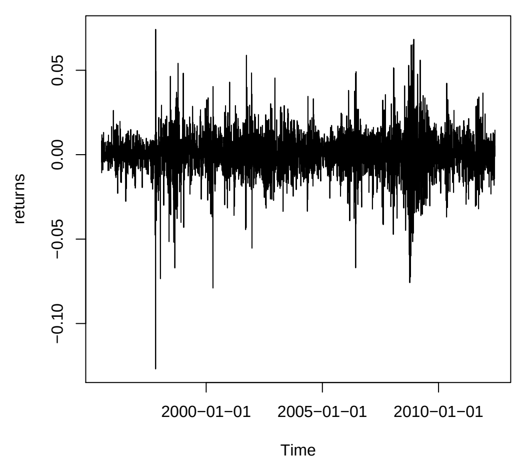

Figure 2.2. The daily returns for the JSE All share index for the period 30 June 1995 to 6 June 2012.

non-constant variance.

Figure 2.3 is a plot of the normal distribution with same mean and standard deviation as for the returns data, whose statistics appears in Table 2.1, with a histogram of the same daily returns also embedded in the figure.

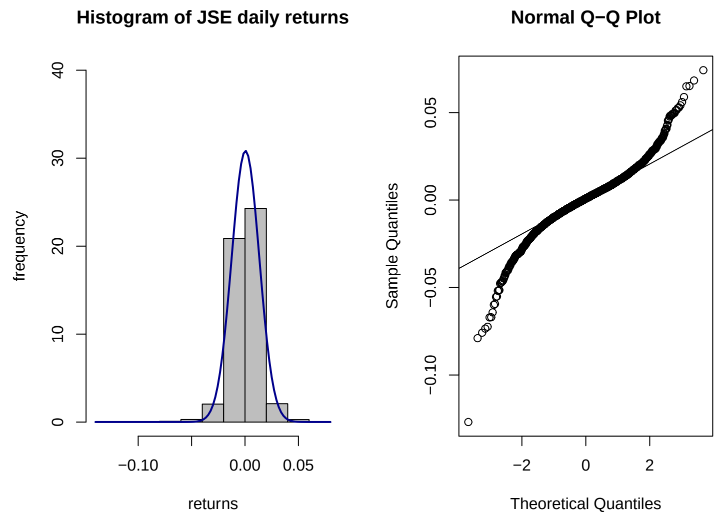

Figure 2.3. A density histogram and the QQ plot for the JSE All Share Index returns data for the period 30 June 1995 to 6 June 2012.

In the Quantile-Quantile plot, the curvature implies that the data does not come from the normal distribution. Evidently from the table the kurtosis for the JSE daily index returns is 8.936735 which is higher than the value of normal distribution. The high value of the kurtosis confirms that the time series of returns possesses the fat-tail characteristic. This characteristic is frequently known to exhibit itself in data from financial markets. The returns of the JSE index are left skewed since the value of the coefficient is -0.466593. Moreover, the Jarque-Bera test statistic for the JSE daily returns is 6363.881. The large value for the Jarque-Bera test statistic suggests the underlying non-normality in the return series.

2.6.2 Results

This section presents the Autocorrelation, Heteroscedasticity by employing Engle's ARCH test. Furthermore the estimation of parameters of the GARCH model and its extensions, model selection, diagnostics of the models and forecasting the volatility are part of the results discussed. All of the tests and parameter estimations of the models are processed using the R software package.

Table 2.1. Descriptive Statistics for JSE daily returns

Statistics	values
Standard deviation	0.01293513
Mean	0.00045608
Minimum	-0.1268996
Maximum	0.07423013
Skewness	-0.4665953
Kurtosis	8.936735
Jarque-Bera	6363.881
p-value	0.000

2.6.2.1 Autocorrelation

Apparently in Figure 2.4, ACF for the JSE index daily returns becomes significant at lags 1, 3, 4, 5, 7, 8, 18, 26, 34, 35 and 36 at 95% confidence level. Thus, the correlation among JSE index returns is significant. To test joint significance for the first 36 lags of ACF, Box-Ljung test is employed. As we can see in Table 2.2, the value of the Box-Ljung test statistic is 107.0798 and its corresponding p-value is less than 0.001. Therefore, the null hypothesis (that there is no autocorrelation) is rejected at significance level 0.001. We accept an alternative hypothesis that there is autocorrelation.

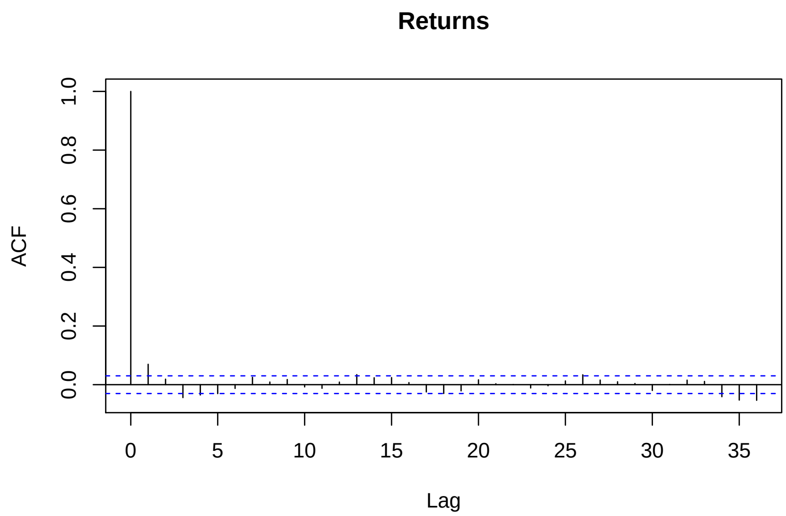

Figure 2.4. The ACF for the JSE All Share Index returns

Table 2.2. Box-Ljung Q-Statistic for Autocorrelation

Lag	Critical-value	Statistic	p-value
10	23.209	44.5817	0.0000
15	30.578	55.3774	0.0000
20	37.566	65.0827	0.0000
36	58.619	107.0798	0.0000

2.6.2.2 Heteroscedasticity

In order to detect heteroscedasticity, we plot the ACF for the square returns. The ACF of the squared returns in Figure 2.5 exhibits a higher length of serial autocorrelation through to the 36^{th} lag. Furthermore the ARCH effect is investigated using Engle's ARCH test for the mean corrected returns which is given below in Table 2.3. From Table 2.3 it can be seen that all values of the Engle's (LM) ARCH statistics are greater than their corresponding critical values and the null hypothesis is rejected at significance level 0.001. Thus, we reject the null hypothesis and conclude that there is heteroscedasticity in the returns.

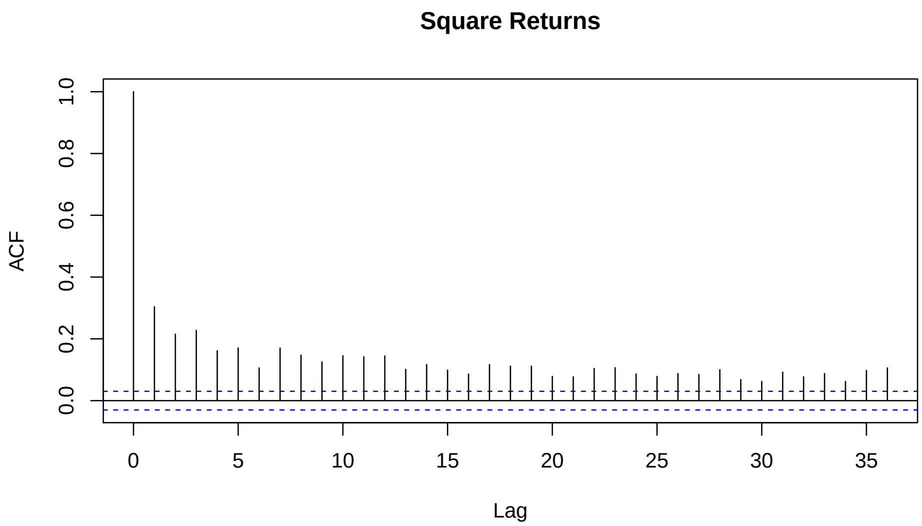

Figure 2.5. The ACF for the JSE All Share Index square returns

Table 2.3. Engle's ARCH test for Heteroscedasticity

Lag	Critical-value	Statistic	p-value
10	23.209	627.1386	0.0000
15	30.578	643.8851	0.0000
20	37.566	654.0117	0.0000
36	58.619	679.7896	0.0000

2.6.2.3 Model Selection

An essential task in modelling volatility using GARCH models is the determination of the ARCH order p and GARCH order (p, q) for a particular series. Considering GARCH models is similarly done as ARMA models using residuals and classical model selection criteria such as the Akaike information criterion (AIC) and the Schwartz-Bayesian information criterion (BIC) can be employed for choosing models. The R software package is employed to compute the AIC and BIC to determine best fitting model. Generally, GARCH models with $p, q \leq 2$ are typically selected by AIC and BIC. In the Table 2.4 below smaller AIC and BIC is an indication of an appropriate model. Therefore, since GARCH $(1, 1)$ has the smallest AIC and BIC among the contesting models we consider it as the appropriate model.

Table 2.4. Model selection for the estimated GARCH (p, q) models assuming Normal distribution

Model	AIC	BIC	Log-Likelihood
GARCH(1,0)	-5.9562	-5.9517	12597.48
GARCH $(1, 1)$	**-6.1621**	**-6.1560**	13033.66
GARCH(1,2)	-6.1615	-6.1540	13033.57
GARCH(2,1)	-6.1618	-6.1543	13034.07
GARCH(2,2)	-6.1615	-6.1525	13034.48

2.6.2.4 Estimation of the GARCH $(1, 1)$ Model

In Table 2.5 below all the estimated parameters were obtained by employing the maximum likelihood estimation process which is performed using the R software package. As shown in Table 2.5, all the estimated parameters are statistically significant and their correspond standard errors are to small which is an indication of a good fit. Furthermore, the sum of the estimated parameters $\hat{\alpha}_1 + \hat{\beta}_1$ is equal to 0.991, which indicates that volatility shocks are persistent.

Table 2.5. Parameter estimates for GARCH $(1, 1)$

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
μ	0.0008231	0.001484	5.545	< 0.001
α_0	0.000002	< 0.001	4.357	< 0.001
α_1	0.112	0.0106	10.593	< 0.001
β_1	0.879	0.011	79.728	< 0.001

2.6.2.5 Diagnostic Checking of the GARCH $(1, 1)$ Model

After the specification of the GARCH model, it is imperative to investigate its adequacy. To explore the relationship between the residuals obtained from the fitted model, the corresponding conditional standard deviations, and the observed returns are studied. We can perceive that both residuals and returns in the Figure 2.6 portray volatility clustering. Nevertheless if we plot the time series of standardized residuals as shown in Figure 2.7, it can be perceived that they become generally stable with little clustering. In order to assess normality for the residuals we can plot the quantile-quantile plot. If the standardized residuals come from

Table 2.6. Box-Ljung Q-statistic test for squared standardized residuals, Engle's ARCH test, and Jarque-Bera test for normality

Statistic	value	p-value
Q2(10)	20.1	0.028
Q2(15)	22.1	0.106
Q2(20)	27.5	0.121
Engle's ARCH test	27.3	0.128
Jarque-Bera test	575.74	< 0.001

the gaussian distribution the plot should be a straight line. Apparently in Figure 2.9 the quantile-quantile plot does not form a straight line which implies that the standardized residuals do not come from the normal distribution. Moreover the value do Jarque-Bera statistics is 575.7491 which is too large, imply that the normality assumption of the standardized residuals is rejected. The ACF of the squared standardized residuals in Figure 3.8 portrays no autocorrelation except for the two critical values that appear on lag 1 and lag 2. To corroborate this conclusion we employed the Box-Ljung test for standardized residuals. The value of the Box-Ljung test statistic Q(20) is equal to 27.53516 and less than the critical value 37.566 at 0.001 significance level. Therefore, there is no heteroscedasticity left in the fitted model. In addition the correlation of the standardized residuals is tested using Engle's ARCH test. The Engle's (LM) ARCH test which is shown in Table 2.6 indicate that heteroscedasticity has been removed.

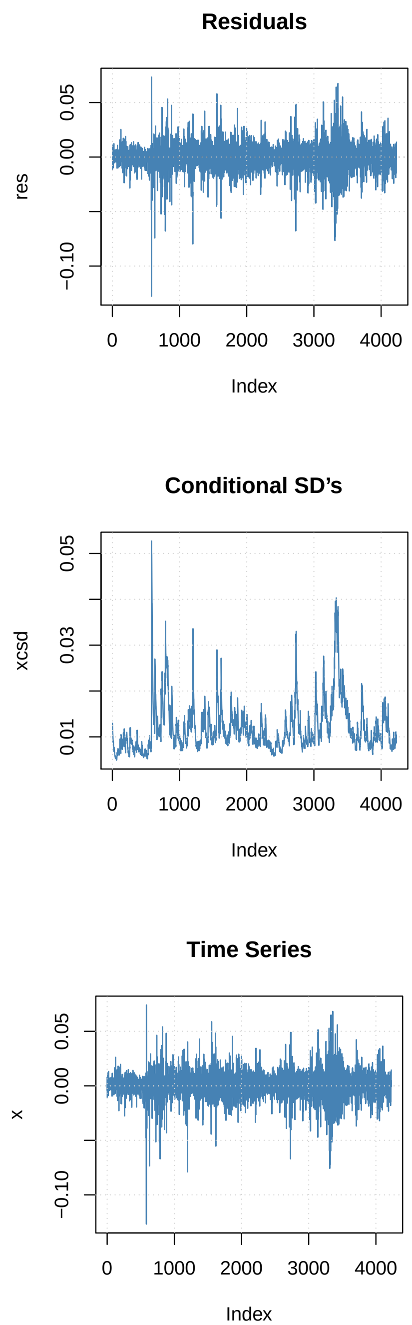

Figure 2.6. The plot of residuals, estimated conditional standard deviations and returns

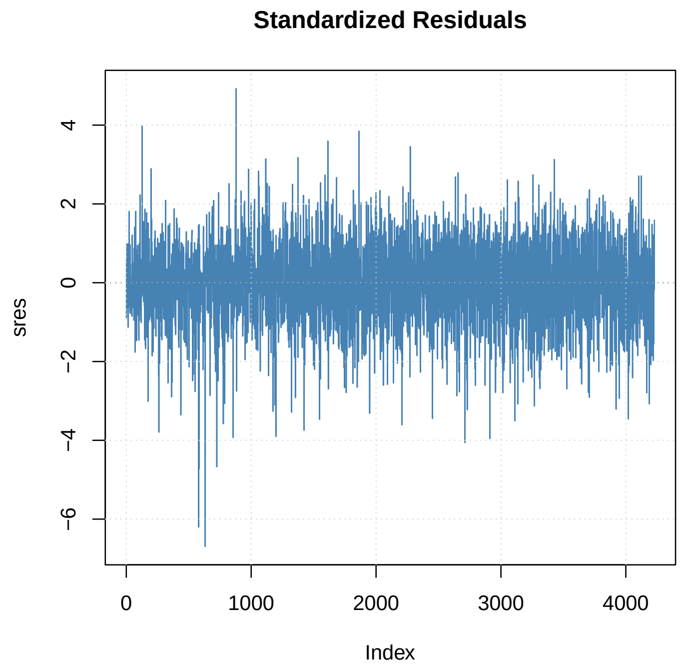

Figure 2.7. Standardized residuals of GARCH $(1, 1)$

Before applying the GARCH $(1, 1)$ model to the data, both the Engle's ARCH test and Box-Ljung test illustrated rejection of their respective null hypothesis showing overwhelming evidence in support of ARCH effects. In the post estimation applying standardized residuals based on the estimated GARCH $(1, 1)$ model, the corresponding test results is an affirmation of their respective null hypothesis. The results justify effectiveness of the GARCH $(1, 1)$ model.

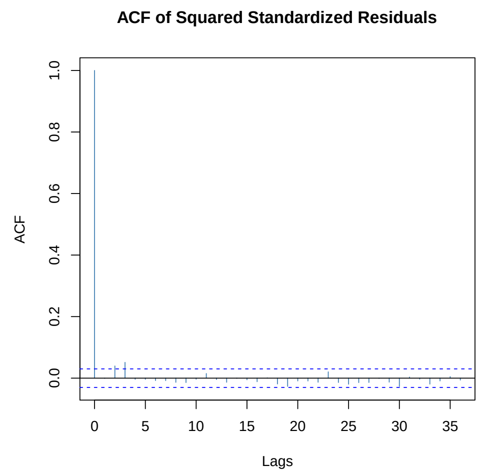

Figure 2.8. ACF for the squared standardized residuals

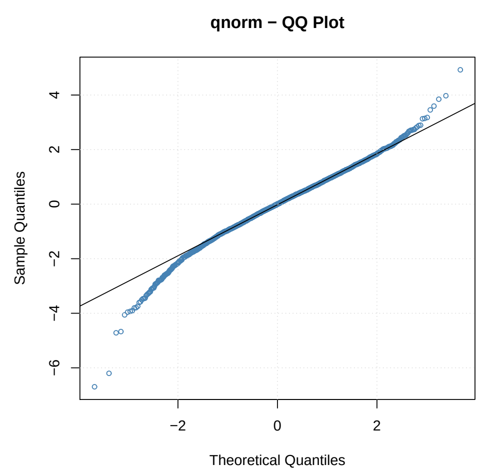

Figure 2.9. QQ-plot for the standardized residuals

2.6.3 Forecasting with the GARCH $(1, 1)$ model

In this section we forecast the volatility of the JSE all share index for the next 10 trading days (during the period 7-20 June 2012). The forecast for the volatility for the next day (t+1) is generated by the equation

$$\hat{\sigma}_{t+1}^2 = 2.208e - 06(\pm 5.068e - 07) + 0.112(\pm 0.01057)\varepsilon_t^2 + 0.879(\pm 0.01103)\sigma_t^2,$$

where σ_t and ε_t are known at time. For the following days $i = 2, \ldots, 10$, the volatility can be forecasted as:

$$\hat{\sigma}^2_{t+i} = \alpha_0 + (\hat{\alpha_1} + \hat{\beta}_1)\sigma^2_{t+i-1},$$

Table 2.7. Ten day forecasts of JSE all share index from GARCH $(1, 1)$

Day	mean-forecast	forecasted standard deviation	observed volatility
1	0.0008283	0.009439	0.0129439
2	0.0008283	0.009515	0.0129424
3	0.0008283	0.009589	0.0129410
4	0.0008283	0.009662	0.0129403
5	0.0008283	0.009734	0.0129407
6	0.0008283	0.009804	0.0129394
7	0.0008283	0.009874	0.0129374
8	0.0008283	0.009943	0.0129364
9	0.0008283	0.010010	0.0129348
10	0.0008283	0.010077	0.0129351

In order to assess the forecasting ability of the estimated GARCH $(1, 1)$ model we have estimated within sample forecasts. If the size of the difference between forecasted volatility and observed volatility is small then the model has good forecasting ability. In Table 2.7, it is evident that GARCH $(1, 1)$ produces good results.

Chapter 3

Extensions of GARCH models

3.1 Introduction

In the basic GARCH model, we consider the squared residuals ε_t^2, in the equation of the conditional variance, therefore the sign of the residuals or mean corrected returns have no impact on conditional volatility. However, a stylized fact of financial volatility is that bad news (negative residuals) tends to have a larger influence on the volatility than good news (positive residuals) of the same magnitude [Black, 1976]. The asymmetric impact on volatility is mainly referred to as the leverage effect. We now provide a brief outline of the three models that are generally used to accommodate the leverage effect.

3.2 Asymmetric GARCH Models

3.2.1 Exponential GARCH model

The leverage effect can be incorporated into a GARCH model in various ways. The first model that takes into account the leverage effect was proposed by Nelson [1991], and is given by

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

where the value of $g(z_t)$ rely upon several elements. To accommodate the asymmetric relation between stock returns and volatility changes, the value of $g(z_t)$ must be a function of both the magnitude and the sign of z_t. Thus the following expression for g(z_t) is used namely:

$$g(z_t) = \psi z_t + \gamma[|z_t| - E[|z_t|]] \tag{3.1}$$

An advantage of the exponential GARCH over the fundamental GARCH model is that the conditional variance σ_t^2 is guaranteed to be positive regardless of the values of the coefficient in equation (3.1), because the logarithm of σ_t^2 instead of σ_t^2 itself is modelled Zivot [2009].

3.2.2 GJR-GARCH models

Another extension to the traditional GARCH model used to model the leverage effect is the GJR-GARCH model, named after Glosten et al. [1993] who developed the model. In this extension conditional variance can be expressed as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i} \varepsilon_{t-i}^2) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.2}$$

where

$$S_{t-i} = \begin{cases} 1 & \text{for} \quad \epsilon_{t-i} < 0 \\ 0 & \text{for} \quad \epsilon_{t-i} \geq 0 \end{cases} \tag{3.3}$$

The TGARCH model of Zakoian [1994] is very similar to the GJR model but models the standard deviation of the conditional variance.

3.2.3 APARCH model

Ding et al. [1993] introduced the Asymmetric Power ARCH (APARCH) model to allow for the leverage effect. The APARCH (p, q) model can be expressed as

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j-1}^{q} \beta_j \sigma_{t-j}^\delta, \tag{3.4}$$

where $\alpha_0 > 0$, $\delta \geq 0$, β_j $(j = 1, \cdots, q)$,α_i $(i = 1, \cdots, p)$, and $|\gamma_i| < 1$ $(i = 1, \cdots, p)$. The exponent δ may also be estimated as an additional parameter which enhances the flexibility of the model. Furthermore the family APARCH models includes other seven GARCH extensions as special cases:

- ARCH model of Engle [1982] when $\delta = 2$, $\gamma_i = 0$ $(i = 1, \cdots, p)$ and $\beta_j = 0$ $(j = 1, \cdots, q)$
- GARCH model of Bollerslev [1986] when $\delta = 2$, and $\gamma_i = 0$ $(i = 1, \cdots, p)$
- TS-GARCH of Taylor [1986] and Schwert [1990] when $\delta = 1$ and γ_i $(i = 1, \cdots, p)$
- GJR of Glosten et al. [1993] when $\delta = 2$
- TARCH of Zakoian [1994] when $\delta = 1$
- NARCH of Higgins and Bera [1992] when γ_i $(i = 1, \cdots, p)$ and $\beta_j = 0$ $(j = 1, \cdots, q)$
- The log-ARCH of Geweke [1986] and Pantula [1986] when $\delta \to 0$.

3.3 Parameter Estimation

Consider a stochastic process $\{\varepsilon_1, \varepsilon_2, \cdots, \epsilon_T\}$ of residuals that are conditionally independent. Let $\mathbf{\Phi}_{t-1}$ represent the history information available up to time t-1. Then the likelihood function for the residuals series is given by

$$f(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T|\mathbf{\Phi}_{T-1}) = \prod_{t=1}^{T} f(\varepsilon_t|\mathbf{\Phi}_{t-1}), \tag{3.5}$$

where f is the general notation for a density function. Under the normal assumption, the conditional density function of ε_t is

$$f(\varepsilon_t|\mathbf{\Phi}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\}, \tag{3.6}$$

where $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i(|\epsilon_{t-i}| - \gamma_i\epsilon_{t-i})^2 + \sum_{j=1}^{q} \beta_j\sigma_{t-j}^2$. We employ the maximum likelihood method to estimate the unknown parameters in the APARCH (p, q) model. To make the model less difficult to work with, we define some vector parameters. We define the vector $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \cdots, \alpha_p, \beta_1, \beta_2, \cdots, \beta_q, \gamma_1, \gamma_2, \cdots, \gamma_p)$, which contains unknown parameters. In order to estimate the parameters we differentiate the log-likelihood function with respect to $\boldsymbol{\theta}$ and equate to zero. However, the maximum likelihood for $\hat{\boldsymbol{\theta}}$ is nonlinear, thus its maximization must be performed using appropriate nonlinear optimization routine. To solve the problem of nonlinearity in the estimation, we use BHHH iterative algorithm to estimate the unknown parameters. Similarly for the EGARCH model parameters can be estimated by employing the maximum likelihood method in conjuction with BHHH algorithm.

3.4 Forecasting with the Asymmetric GARCH Models

The conditional variance can be forecasted independently from the mean. For an APARCH (p, q) process, the distribution of residuals may have an impact on the forecast, the h-step ahead forecast for the conditional variance is computed recursively from

$$\begin{aligned}
\hat{\sigma}^{\hat{\delta}}_{T+h|T} &= E[\sigma^{\delta}_{T+h}|\Phi_T] \\
&= \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i E[(|\varepsilon_{T+h-i}| - \hat{\gamma}_i \varepsilon_{T+h-i})^{\hat{\delta}}|\Phi_T] + \sum_{j=1}^{q} \hat{\beta}_j \sigma^{\hat{\delta}}_{T+h-j|T}
\end{aligned}$$

[Pasha et al., 2007].

For an EGARCH $(1, 1)$ process, the h-step ahead forecast for the conditional variance is computed recursively from

$$\begin{aligned}
\hat{\sigma}^2_{T+h|T} = \sigma^{2\hat{\beta}_1}_{T+h-1} \exp\left(\hat{\alpha}_0 - \hat{\alpha}_1\sqrt{\frac{2}{\pi}}\right) \left\{ \exp\left(\frac{(\hat{\gamma}+\hat{\alpha}_1)^2}{2}\right) N(\hat{\gamma}+\hat{\alpha}_1) \right. \\
\left. + \exp\left(\frac{(\hat{\gamma}-\hat{\alpha}_1)^2}{2}\right) N(\hat{\gamma}-\hat{\alpha}_1) \right\},
\end{aligned}$$

where N(.) is the standard Normal cumulative distribution [Tsay, 2010].

3.5 Data Analysis

3.5.1 Estimation Results

Tables 3.1, Table 3.2, and Table 3.3 present the parameter estimates for GARCH $(1, 1)$, GJR-GARCH $(1, 1)$, EGARCH $(1, 1)$ and APARCH $(1, 1)$ assuming normal, student-t and skewed-t distribution respectively. The values in the parentheses are standard errors of corresponding parameter estimates. The numbers in square brackets are p-values of test statistics. Log(L) represents the value of the maximized log likelihood function. The values Q[10], Q[15] and Q[20] are Box-Ljung Q-statistic calculated on standardized residuals of order 10, 15 and 20 respectively. The values Q2[10], Q2[15] and Q2[20] are Box-Ljung Q-statistic calculated on squared standardized residuals of order 10, 15 and 20 respectively. Furthermore, JBStat represents Jarque-Bera test statistic for normality. The estimates of the coefficient α_1 that accommodates the effect of the new shocks on volatility are statistically significant for all estimated four models. The estimate is positive for the GARCH $(1, 1)$, GJR-GARCH $(1, 1)$ and APARCH $(1, 1)$ but negative in the case of EGARCH $(1, 1)$ model. However the superiority of EGARCH $(1, 1)$ model is that it allows asymmetric response to past positive or negative returns and uses logarithmic volatility to mitigate the parameter constraints. The parameter estimate β_1 that measures persistence of volatility shocks, is positive and statistically significant at 1% significance level. The asymmetry coefficient γ is positive and statistically significant at 1% significance level in all four models with different distributional assumptions. This strongly supports the assertion that residuals have asymmetric influence on the volatility of JSE index closing prices, especially, the positive sign indicates that negative residuals increase volatility more than positive residuals of the same magnitude. The coefficient of the the APARCH $(1, 1)$ model δ assuming normal, student-t and skewed-t distribution is positive and statistically significant at 1% significance level. In Table 3.2 and Table 3.3 the estimates of the shape parameter are statistically significant for the GARCH models. Therefore this shows that JSE index return series is leptokurtic. In Table 3.3 the skewness estimate is approximately equal to $0.9(\pm 0.02)$ and is statistically significant for all the estimated GARCH models. The results show that the density of the JSE index return is skewed to the right in all the estimated models. The Jarque-Bera test indicates that the residuals that resulted from the estimated GARCH models do not conform to normal distribution. Furthermore, because of the presence of leverage effect, the asymmetric models apparently perform better than the symmetric GARCH models. The EGARCH $(1, 1)$ model has both smallest AIC and BIC under skewed-t distributional assumption. Thus, the EGARCH $(1, 1)$ model assuming skewed-t distribution is more likely to be the best fit of JSE index returns.

Table 3.1. Estimation results from GARCH $(1,1)$, GJR-GARCH $(1,1)$, EGARCH $(1,1)$ and APARCH $(1,1)$ under normal distribution assumption

	JSE							
	GARCH		GJR-GARCH		EGARCH		APARCH	
Coefficients								
μ	8.231e-04	(< 0.001)	5.859e-04	(< 0.001)	5.840e-04	(< 0.001)	5.803e-04	(< 0.001)
α_0	2.208e-06	(< 0.001)	2.434e-06	(< 0.001)	-0.213	(0.037)	3.639e-05	(< 0.001)
α_1	0.112	(0.011)	0.0924	(0.011)	-0.077	(0.008)	0.104	(0.010)
β_1	0.879	(0.011)	0.887	(0.0109)	0.976	(0.004)	0.892	(0.010)
γ	-	-	0.260	(0.038)	0.205	(0.017)	0.333	(0.049)
δ	-	-	-	-	-	-	1.421	(0.153)
Log(L)	13033.66		13064.37		13074.94		13065.13	
Residuals test statistics								
Q[10]	42.0	[< 0.001]	41.9	[< 0.001]	42.2	[< 0.001]	40.0	[< 0.001]
Q[15]	44.8	[< 0.001]	45.2	[< 0.001]	45.2	[< 0.001]	43.4	[< 0.001]
Q[20]	53.1	[< 0.001]	53.8	[< 0.001]	53.3	[< 0.001]	51.4	[< 0.001]
JBStat	575.7	[< 0.001]	408.8	[< 0.001]	420.1	[< 0.001]	466.3	[< 0.001]
Q2[10]	20.2	[0.028]	16.7	[0.082]	28.6	[0.001]	31.1	[0.001]
Q2[15]	22.1	[0.105]	20.4	[0.156]	32.4	[0.006]	34.7	[0.003]
Q2[20]	27.6	[0.119]	25.9	[0.171]	38.5	[0.008]	41.0	[0.004]
Information Criteria								
AIC	-6.1621		-6.1761		-6.1811		-6.1760	
BIC	-6.1560		-6.1686		-6.1736		-6.1670	

Table 3.2. Estimation results from GARCH $(1,1)$, GJR-GARCH $(1,1)$, EGARCH $(1,1)$ and APARCH $(1,1)$ under student-t distribution assumption

	JSE							
	GARCH		GJR-GARCH		EGARCH		APARCH	
Coefficients								
μ	8.605e-04	(<0.001)	7.040e-04	(<0.001)	6.870e-04	(<0.001)	6.933e-04	(<0.001)
α_0	2.080e-06	(<0.001)	2.270e-06	(<0.001)	-0.176	(0.036)	1.025e-04	(<0.001)
α_1	0.106	(0.011)	0.090	(0.011)	-0.068	(0.009)	0.091	(0.010)
β_1	0.885	(0.012)	0.890	(0.011)	0.981	(0.004)	0.907	(0.010)
γ	-	-	0.235	(0.045)	0.180	(0.018)	0.372	(0.069)
δ	-	-	-	-	-	-	1.156	(0.168)
shape	8.663	(1.000)	9.498	(1.196)	9.532	(1.196)	9.332	(1.143)
Log(L)	13099.76		13117.24		13128.67		13112.30	
Residuals test statistics								
Q[10]	41.9	$[<0.001]$	41.4	$[<0.001]$	41.9	$[<0.001$	32.2	$[<0.001]$
Q[15]	44.7	$[<0.001]$	44.7	$[<0.001]$	44.8	$[<0.001]$	36.1	[0.002]
Q[20]	53.0	$[<0.001]$	53.3	$[<0.001]$	52.8	$[<0.001]$	42.8	[0.002]
JBStat	574.1	$[<0.001]$	414.2	$[<0.001]$	466.8	$[<0.001]$	15038.1	$[<0.001]$
Q2[10]	23.3	[0.010]	18.1	[0.054]	46.8	$[<0.001]$	14.5	[0.153]
Q2[15]	25.3	[0.046]	21.7	[0.114]	50.8	$[<0.001]$	15.5	[0.414]
Q2[20]	30.8	[0.057]	27.3	[0.127]	57.4	$[<0.001]$	16.8	[0.666]
Information Criteria								
AIC	-6.1928		-6.2006		-6.2060		-6.1978	
BIC	-6.1853		-6.1916		-6.1970		-6.1873	

Table 3.3. Estimation results from GARCH (1,1), GJR-GARCH (1,1), EGARCH (1,1) and APARCH (1,1) under skewed student-t distribution assumption

	JSE							
	GARCH		GJR-GARCH		EGARCH		APARCH	
Coefficients								
μ	6.980e-04	(< 0.001)	5.616e-04	(< 0.001)	5.620e-04	(< 0.001)	5.615e-04	(< 0.001)
α_0	1.925e-06	(< 0.001)	2.146e-06	(< 0.001)	-0.168	(0.034)	8.874e-05	(< 0.001)
α_1	0.102	(0.011)	0.088	(0.010)	-0.064	(0.009)	0.088	(0.010)
β_1	0.889	(0.011)	0.895	(0.019)	0.981	(0.004)	0.909	(0.010)
γ	-	-	0.225	(0.044)	0.176	(0.018)	0.356	(0.068)
δ	-	-	-	-	-	-	1.179	(0.169)
shape	9.124	(1.109)	9.840	(1.287)	9.851	(1.265)	9.662	(1.230)
skewness	0.901	(0.021)	0.906	(0.021)	0.907	(0.021)	0.907	(0.021)
Log(L)	13110.05		13126.53		**13137.52**		13121.92	
Residuals test statistics								
Q[10]	42.2	[< 0.001]	42.1	[< 0.001]	42.6	[< 0.001]	33.6	[< 0.001]
Q[15]	45.1	[< 0.001]	45.31	[< 0.001]	45.5	[< 0.001]	37.4	[0.001]
Q[20]	53.2	[< 0.001]	53.8	[< 0.001]	53.4	[< 0.001	44.2	[0.001]
JBStat	583.6	[< 0.001]	420.1	[< 0.001]	484.2	[< 0.001]	7996.4	[< 0.001]
Q2[10]	25.3	[0.005]	19.6	[0.033]	51.2	[< 0.001]	20.9	[0.021]
Q2[15]	27.3	[0.027]	23.3	[0.079]	54.9	[< 0.001]	22.4	[0.097]
Q2[20]	33.0	[0.034]	28.9	[0.090]	61.7	[< 0.001]	24.4	[0.225]
Information Criteria								
AIC	-6.1972		-6.2046		**-6.2098**		-6.2019	
BIC	-6.1882		-6.1940		**-6.1992**		-6.1899	

3.5.2 Diagnostic Checking of the GJR-GARCH $(1,1)$ Model

According to the AIC and BIC, the EGARCH $(1,1)$ model with skewed-t distribution in the errors perform better than all models considered in this study since it has both the least AIC and BIC values. On exploring the EGARCH $(1,1)$ model further we noted that heteroscedasticity was not completely discarded in the residuals. The GJR-GARCH $(1,1)$ model with student-t errors seems to be the best fit since it removes the heteroscedasticity in the residuals. Figure 3.1 shows that the fitted residuals portray volatility clustering. Moreover if we plot the standardized residuals as shown in Figure 3.2, they appear generally stable with little clustering. This ACF of the standardized residuals plot in Figure 3.3 exhibits no autocorrelation except for the two values that appear on lag 2 and lag 3.

In order to justify this outcome we use the Box-Ljung for squared standardized residuals. The value of the Box-Ljung statistic Q(20) is equal to 27.45 and smaller than the critical value 37.57 at 0.001 significance level. Thus, we can conclude that there is no heteroscedasticity in the fitted model. Furthermore the correlation of the standardized residuals is tested using Engle's ARCH test. The p-value of the Engle's ARCH statistic at lag 12 which is shown in Table 3.4 is greater than 0.05. Therefore we accept the null hypothesis that there is no heteroscedasticity. The value of the Jarque-Bera statistic as presented in Table 3.4 is 414.2105 which is quite a big number, therefore the assumption of the normal standardized residuals is rejected. Furthermore the empirical density and qq-plot substantiates the same result that were suggested by the Jarque-Bera statistic. In the pre-estimation both Engle's ARCH test and Box-Ljung statistic supported the rejection of their respective null hypothesis. After the GJR-GARCH (1,1) model has been fitted, the same tests signify acceptance of their respective null hypothesis. These results justify the efficiency of the GJR-GARCH model.

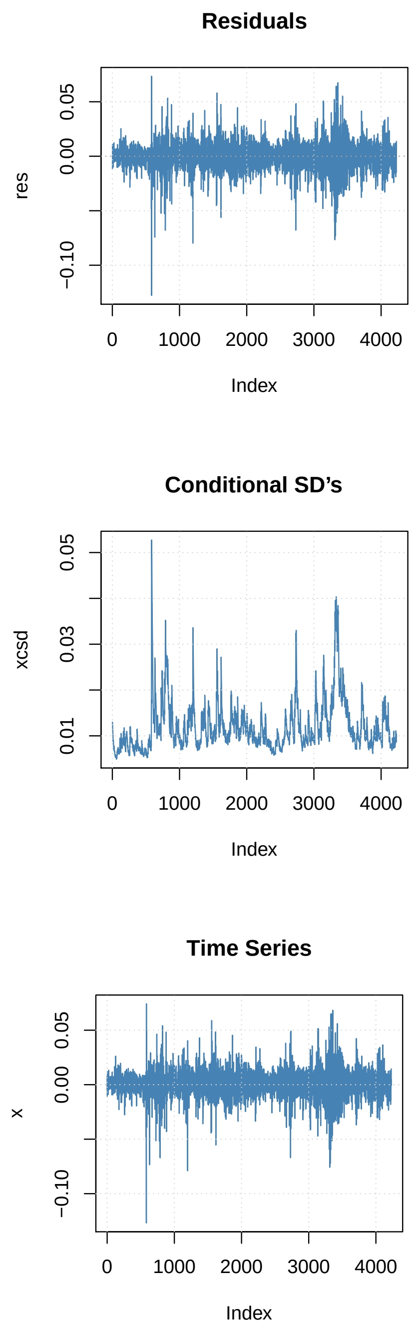

Figure 3.1. Comparison of residuals, estimated conditional standard deviations and returns

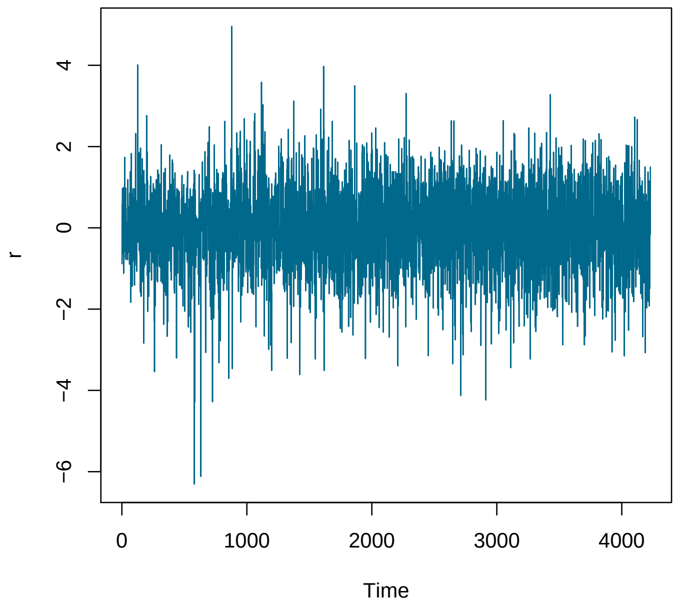

Figure 3.2. Standardized residuals of GJR-GARCH (1,1)

Table 3.4. Box-Ljung Q-Statistic test for squared standardized residuals, Engle's ARCH test for standardized residuals, and Jarque-Bera test for normality

Statistic	value	p-value
Q2(10)	18.2	0.052
Q2(15)	21.9	0.111
Q2(20)	27.5	0.123
Engle's ARCH test	20.8	0.053
Jarque-Bera test	414.2	< 0.001

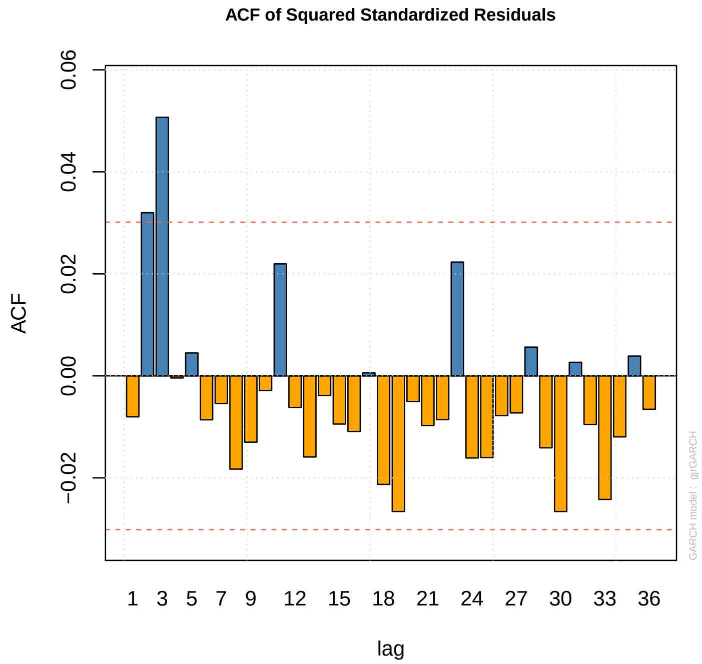

Figure 3.3. ACF plot of squared standardized residuals for GJR-GARCH (1,1)

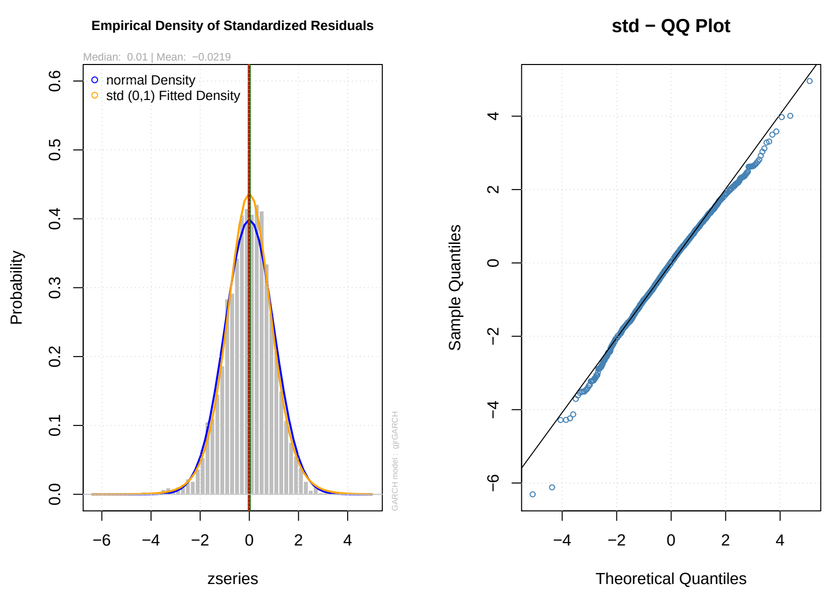

Figure 3.4. Empirical density and QQ-plot of standardized residuals

Chapter 4

Multivariate GARCH Process

4.1 Introduction

Modelling volatility in financial markets has been a key issue in econometrics and time series analysis following the landmark contributions by Engle [1982], Bollerslev [1986], and Taylor [1986] who introduced the GARCH and the stochastic volatility models. Since modelling volatility of returns in financial markets has been the subject of extensive research, an understanding of the co-movements of returns among different markets globally is of great significance. Multivariate volatilities play a crucial role in portfolio selection and asset allocation and can also be implemented to calculate the value at risk of a financial position comprising multiple assets. Rather than a limited discussion concerning volatility in a univariate return series greater value can be found in extending the ideas concerning GARCH models to a multivariate setting. For instance, asset pricing and risk management significantly rely upon the conditional covariance structure of the assets of a portfolio. Modelling a covariance is a non-trivial issue because of the likely high dimensionality of the problem and the constraint that compels the positive definiteness of the covariance matrix. The most essential part in modelling the multivariate volatility is to render a realistic but parsimonious specification of the covariance matrix ensuring its positivity. A major drawback in the multivariate approach is that the number of parameters in the GARCH model increases rapidly leading to complexities associated with parameter estimation and inference. This dictates the analyst or modeller to tend to confine or limit the number of assets (financial markets) to be incorporated into the model for easy of estimation and interpretation. Numerous specifications of multivariate GARCH have been introduced in the literature. For the purpose and scope of this study, we will limit our attention to the Dynamic Conditional Correlation (DCC) model of Engle [2002].

4.2 Multivariate GARCH Models

The volatilities of financial indices are known to move synchronously across different markets or slightly delayed. Essentially, detecting how financial markets are inter-related is of paramount importance. For an investor holding a variety of assets, the dynamic relationship between returns on the assets play a crucial role in decision making.

In modelling the multivariate volatility, we consider an N-dimensional vector series of returns $\mathbf{r}_t = (r_{1t}, \ldots, r_{Nt})'$. The vector $\mathbf{r}_t$ has an $N-$dimensional conditional mean vector $\boldsymbol{\mu}_t$ and a $(N \times N)$ conditional covariance matrix $\mathbf{H}_t$.

Letting $\boldsymbol{\Phi}_{t-1}$ represents the information history generated by the observed time series $\mathbf{r}_t$ up to and including $t-1$ and $\boldsymbol{\theta}$ be a finite vector of parameters, $\mathbf{r}_t$ is conditionally heteroscedastic in the following manner:

$$\mathbf{r_t} = \mu_\mathbf{t}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon_t} \tag{4.1}$$

where $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ is the conditional mean vector with respect to the information set $\boldsymbol{\Phi}_{\mathbf{t-1}}$ and

$$\varepsilon_t = \mathbf{H}_t^{\frac{1}{2}}(\boldsymbol{\theta})\mathbf{z_t} \tag{4.2}$$

where $\mathbf{H}_t^{\frac{1}{2}}(\theta)$ is a positive definite matrix of order $(N \times N)$.

Further, we assume that the N-dimensional vector $\mathbf{z}_t$ is an independent and identical distributed random vector that satisfies $E[\mathbf{z_t}] = \mathbf{0}$ and $E[\mathbf{z_t z_t'}] = \mathbf{I_N}$, where $\mathbf{I_N}$ is an identity matrix of order N.

Definition 5 (Second-order stationarity)**.** *The process* $\mathbf{r_t}$ *is said to be second-order stationary if:*

1. $E[\mathbf{r_t}] = \boldsymbol{\mu} = constant$

2. $E[r_{it}^2] < \infty$, *for* $i = 1, 2, \ldots, N$

3.

$$\begin{aligned}
\mathbf{\Gamma}(k) &= E[(\mathbf{r_t} - \boldsymbol{\mu})(\mathbf{r_{t-k}} - \boldsymbol{\mu})\prime] \\
&= \begin{bmatrix} \gamma_1 & \gamma_{12}(k) & \ldots & \gamma_{1N} \\ \gamma_{21}(k) & \gamma_2 & \ldots & \gamma_{2N}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1}(k) & \gamma_{N2}(k & \ldots & \gamma_N \end{bmatrix}
\end{aligned}$$

where the expectation is taken element by element over the the joint distribution of $\mathbf{r_t}$*. The mean is N-dimensional vector consisting of unconditional expectations of the components of* $\mathbf{r_t}$*. The covariance matrix* $\mathbf{\Gamma}(k)$ *is an* $(N \times N)$ *matrix where* γ_{it} *is the* i^{th} *diagonal element and where the* $(i,j)^{th}$ *element is the covariance between* r_{it} *and* r_{jt}*, and it is a function of* k*.*

The main goal is to model the conditional covariance matrix $\mathbf{H}_t$ which is an $(N \times N)$ positive definite matrix.

4.2.1 The CCC-GARCH (p, q) Model

Bollerslev[1990] proposed a multivariate specification in which all conditional conditional correlations are time invariant and thus the conditional variances are modelled by univariate GARCH models. This specification is called Constant Conditional Correlation (CCC). This specification abridges the number of parameters to be estimated and thus alleviates the estimation. Then, the conditional covariance matrix $\mathbf{H}_t$, may be expressed as:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t \quad \text{where} \quad h_{ijt} = \rho_{ij}\sqrt{h_{iit}h_{jjt}}, \forall i \neq j,$$
$$\mathbf{D}_t = \text{diag}(\sqrt{h_{11t}}, \sqrt{h_{22t}}, \cdots \sqrt{h_{NNt}}),$$

where h_{iit} can be defined as a univariate $GARCH$ model, R is an $(N \times N)$ correlation matrix. The elements of the correlation matrix are given by

$$\rho_{ijt} = \frac{Cov(\varepsilon_{it}, \varepsilon_{jt}|\mathbf{\Phi}_{t-1})}{\sqrt{Var(\varepsilon_{it}|\mathbf{\Phi}_{t-1})Var(\varepsilon_{jt}|\mathbf{\Phi}_{t-1})}} \tag{4.3}$$

with $\rho_{ij} = 1, \forall i = 1, \cdots, N$. $\mathbf{D}_t$ represents the $(N \times N)$ diagonal matrix consisting of the conditional standard deviations of element ε_t. The CCC-GARCH specification assumes that the conditional correlations are time invariant $\rho_{ijt} = \rho_{ij}$, so that the temporal variation in $\mathbf{H}_t$ is only determined by the time varying conditional correlations for each in ε_t. As long as each conditional variances are positive, the CCC guarantees that the resulting conditional matrices are positive definite.

4.2.2 The DCC-GARCH (p, q) Model

Despite the simplicity of the Constant Conditional Correlation, the major drawback is that the correlation tends to vary over time in real application. The assumption of constant correlations may be seen unrealistic in many empirical applications. Thus, Engle [2002] and Tse and Tsui [2002] propose a generalization of the conditional constant correlation model by making the conditional correlation matrix time-dependent. The generalized model is the called a Dynamic Conditional Correlation (DCC) model. The DCC model of Engle [2002] and Tse and Tsui [2002] are genuinely multivariate and are instrumental when modelling high dimensional data sets.

Definition 6. *The DCC model of Tse and Tsui [2002] is defined as:*

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{4.4}$$

where $\mathbf{D}_t = diag(\sqrt{h_{11t}}, \ldots, \sqrt{h_{NNt}})$ *is an* $(N \times N)$ *diagonal matrix,* h_{iit} *can be obtained as any univariate GARCH model and*

$$\mathbf{R}_t = (1 - \theta_1 - \theta_2)\mathbf{R} + \theta_1 \mathbf{\Psi}_{t-1} + \theta_2 \boldsymbol{R}_{t-1}. \tag{4.5}$$

In equation (4.5), θ_1 *and* θ_2 *are nonnegative scalar parameters satisfying* $\theta_1 + \theta_2 < 1$, $\mathbf{R}$ *is a symmetric* $(N \times N)$ *positive definite matrix with* $\rho_{ij} = 1$ *and* $\mathbf{\Psi}_{t-1}$ *is the* $(N \times N)$ *correlation matrix of* u_τ *for* $\tau = t - M, t - M + 1, \ldots, t - 1$. *Its* $(i, j)^{th}$ *element is given by*

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^{M} u_{i,t-m} u_{i,t-m}}{\sqrt{(\sum_{m=1}^{M} u_{i,t-m}^2)(\sum_{m=1}^{M} u_{j,t-m}^2)}} \tag{4.6}$$

where $u_{it} = \frac{\varepsilon_{it}}{\sqrt{h_{iit}}}$

A necessary condition to ensure positivity of $\mathbf{\Phi}_{t-1}$, and therefore also of $\mathbf{R}_t$, is that $N \leq M$.

Definition 7. *The DCC model Engle [2002] is defined as:*

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{4.7}$$

with

$$\mathbf{R}_t = (\mathbf{Q}_t^*)^{-1} \mathbf{Q}_t (\mathbf{Q}_t^*)^{-1} \tag{4.8}$$

where $\mathbf{Q}_t^* = diag(\mathbf{Q}_t)^{\frac{1}{2}}$, $\mathbf{Q}_t$ *is the* $(N \times N)$ *symmetric positive definite matrix given by*

$$\mathbf{Q}_t = (1 - \theta_1 - \theta_2)\bar{\mathbf{Q}} + \theta_1 u_{t-1} u_{t-1}' + \theta_2 \mathbf{Q}_{t-1}. \tag{4.9}$$

θ_1 *and* θ_2 *are nonnegative scalar parameters satisfying* $\theta_1 + \theta_2 < 1$. $\bar{\mathbf{Q}}$ *is the* $(N \times N)$ *unconditional covariance matrix consisting of standardized residuals resulting from the first-step estimation, where* $\mathbf{u_t} = (u_{1t}, \ldots, u_{Nt})'$ *is the standardized residuals vector (*$u_{it} = \frac{\varepsilon_{it}}{\sqrt{h_{iit}}}$, *for* $i = 1, 2, \ldots, N$*)*

$$(\mathbf{Q}_t^*)^{-1} = diag(\frac{1}{\sqrt{q_{11}}}, \ldots, \frac{1}{\sqrt{q_{NN}}})$$

The typical element $\mathbf{R}_t$ *will be of the form*

$$\rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{ii} q_{jj}}}$$

A disadvantage of the DCC models is that θ_1 and θ_2 are scalars, so that all the conditional correlations obey the same dynamics. This is necessary to ensure that $\mathbf{R}_t$ is positive through sufficient conditions on the parameters.

4.2.3 Estimation of DCC-GARCH Model

In order to estimate the unknown parameters of the DCC-GARCH model, we employ the method of conditional maximum likelihood estimation.

Under the assumption of normally distributed errors, parameters can be estimated by maximizing the log-Likelihood function given by

$$
\begin{aligned}
l(\theta) &= -\frac{1}{2}\sum_{t=1}^{T}\left[N\ln(2\pi)+\ln(|\mathbf{H}_t|)+\boldsymbol{\varepsilon}'\mathbf{H}_t^{-1}\boldsymbol{\varepsilon}_t\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left[N\ln(2\pi)+\ln(|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t|)+\mathbf{u}_t'\mathbf{D}_t\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\mathbf{D}_t\mathbf{u}_t\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left[N\ln(2\pi)+2\ln(|\mathbf{D}_t|)+\ln(|\mathbf{R}_t|)+\mathbf{u}_t'\mathbf{R}_t^{-1}\mathbf{u}_t\right]
\end{aligned}
$$

where T represents the number of observations, N is the number of stock markets and θ denotes the number of parameters to be estimated. However, the log-likelihood function is non-linear, therefore it is maximized by an iterative numerical algorithm called BHHH algorithm which in turn estimates the desired parameters.

4.2.4 Diagnostics of DCC-GARCH model

An exploratory diagnostic analysis in require to ascertain the adequacy of DCC-GARCH model. Such a procedure confirms whether an estimated DCC-GARCH model can render the desired estimates and statistical inference. In order to check the overall significance of the residual correlation, we consider the Box-Ljung portmanteau statistic. This test was proposed by Ljung and Box [1978] as one of the diagnostics in Autoregressive Moving Average (ARMA) time series modelling. Let $\hat{\mathbf{z}}_t = \hat{\mathbf{H}}_t^{-\frac{1}{2}}\hat{\boldsymbol{\varepsilon}}_t$ represent the N-dimensional vector of standardized residuals. Following Hosking [1980], a multivariate version of the Box-Ljung statistic is given by

$$Q_m(k) = T^2\sum_{j=1}^{k}(T-j)^{-1}tr\{C_{\mathbf{Z}_t}^{-1}(0)C_{\mathbf{Z}_t}(j)C_{\mathbf{Z}_t}^{-1}(0)C_{\mathbf{Z}_t}'(j)\}, \tag{4.10}$$

where tr denotes the trace of a matrix. $\mathbf{Z}_t = vech(\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t')$, $C_{\mathbf{Z}_t}(j)$ is the sample autocovariance matrix of order j and $vech$ is the operator that stacks a matrix as column vector. Under the null hypothesis (No ARCH effect), $Q_m(k)$ is asymptotically as χ^2 with N^2k degrees of freedom. Duchesne and Lalancette [2003] generalized this statistic using a spectral approach and obtained higher asymptotic power by using different kernel than the truncated uniform kernel used in $Q_m(k)$. This test is also employed to detect the misidentification in the conditional matrix $\mathbf{H}_t$ [Bauwens et al., 2006].

4.3 Empirical Results

4.3.1 Summary Analysis

The data set contains weekly prices for the period June 1995 to June 2012 for three selected stock market, namely the JSE All Share Index, FTSE-100 and NASDAQ-100.

Table 4.1 presents the summary statistics of the log weekly return for the three stock market indices. Both FTSE-100 and NASDAQ-100 show the highest kurtosis. The Jarque-Bera test rejects the normality assumption of the returns.

One of the key aims of this study was to investigate the nature of the relationship between South African market and key players in global markets. Table 4.2 presents the correlation matrix computed for the three stock markets. The results confirm the coexistence of correlation among markets. Correlation between all the markets is positive, suggesting that there is a common factor driving markets in the same direction. In particular, correlation between JSE and both FTSE and NASDAQ indices is low, specifically less than 0.5. The autocorrelation Function (ACF) and Cross-correlation Function (CCF) in Figure 4.2 also suggest a decline in correlation between the JSE stock market index and the two other indices with increasing lag. The results suggest that it will be appropriate to model the conditional variance of returns of the stock markets.

The Box-Ljung statistics for lag 12 implemented on squared returns (denoted by Q2[12]) suggest significant time dependence of the second moments.

Furthermore, the null hypothesis of no ARCH effects in each of the returns is rejected at 1% level of significance.

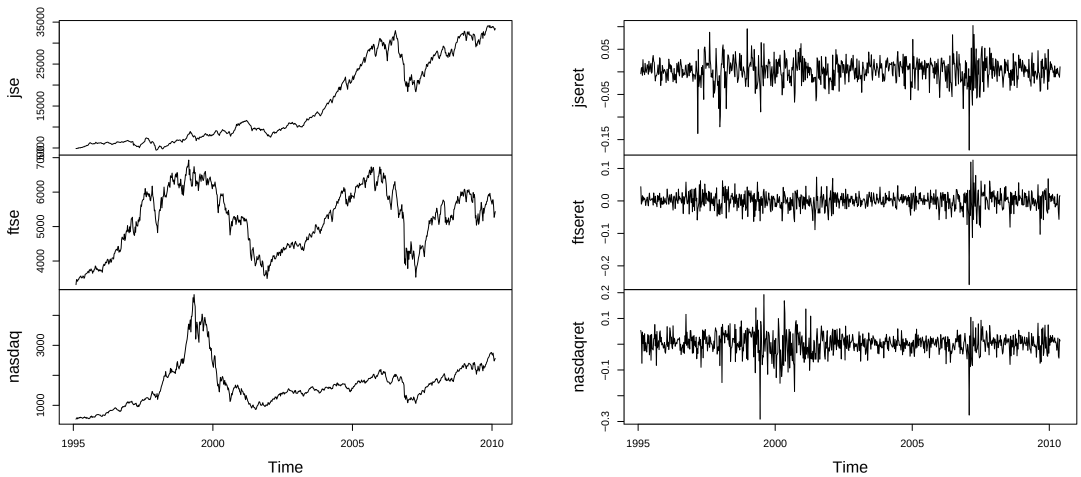

Figure 4.1. Time series plots of the price (left) and return (right) series for the JSE, FTSE and NASDAQ indices

Table 4.1. Descriptive Statistics of JSE All Share Index, FTSE-100 and NASDAQ-100 returns

Summary statistics	JSE returns	FTSE-100	NASDAQ-100
number of observations	781	781	781
Standard deviation	0.02652	0.0266	0.04264
Mean	0.002465	0.0006332	0.001997
Minimum	-0.1732	-0.2578	-0.291
Maximum	0.1026	0.1258	0.1926
Skewness	-0.7839	-1.279	-0.8298
Kurtosis	7.487	15.73	9.191
Q2[12]	93.29	118.4	111.4
p-value	(< 0.001)	(< 0.001)	(< 0.001)
Engle's ARCH test[12]	54.14	82.38	64.77
p-value	(< 0.001)	(< 0.001)	(< 0.001)
Jarque-Bera	735.1	5485	1337
p-value	(< 0.001)	(< 0.001)	(< 0.001)

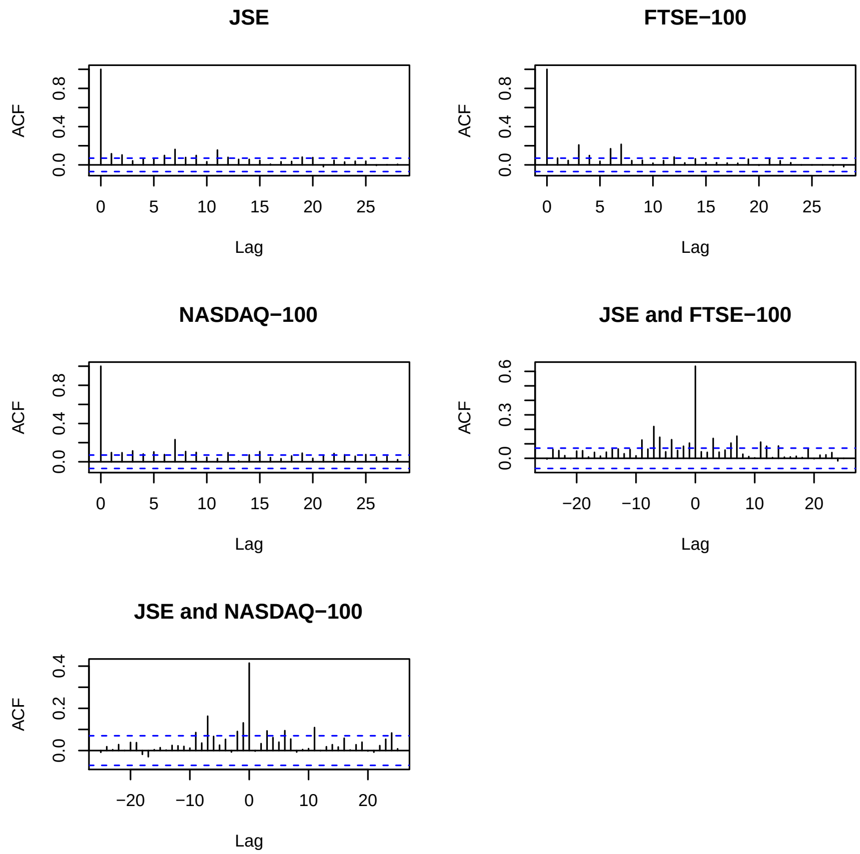

Figure 4.2. ACF and CCF for the JSE All Share Index, FTSE-100 and NASDAQ-100 squared returns

Table 4.2. Constant Correlation Estimates

	JSE returns	FTSE-100 returns	NASDAQ-100
JSE returns	1.0000	0.4604	0.3406
FTSE-100 returns	0.4604	1.0000	0.6206
NASDAQ-100returns	0.3406	0.6206	1.0000

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

The univariate GARCH $(1,1)$ models for each stock market in Table 4.3 are representing the diagonal elements of $\mathbf{D}_t$ in Definition 6 and 7. The coefficient α_0 is highly significant at 1% level in all the three stock markets. The significant α_1 for all three stock markets are indicating persistence of volatility and a large coefficient of asymmetric volatility for all the three stock markets, which could suggest possible transmissions in volatility from other stock markets that was taken in by other stock markets. The coefficient β_1 is also significant in all three markets and indicates a large asymmetric impact implying that the stock markets are reacting to different sources of information from different markets and adapt their portfolio consequently. The DCC-GARCH $(1,1)$ parameters θ_1 and θ_2 are also presented in Table 4.3, both parameters are significant implying the correlations are dynamic, Thus, this enables us to interpret the results in terms of time varying correlations for South Africa and other stock markets. The sum of θ_1 and θ_2 is very close to 1 which implies that the conditional covariance is highly persistent.

Table 4.3. DCC-GARCH(1,1) Estimates for JSE, FTSE100 and NASDAQ100 Indices

Parameter	Estimate	Standard Error
GARCH parameters		
$\alpha_{0,JSE}$	0.00004	(< 0.001)
$\alpha_{1,JSE}$	0.102	0.027
$\beta_{1,JSE}$	0.847	0.034
$\alpha_{0,FTSE100}$	0.00003	(< 0.001)
$\alpha_{1,FTSE100}$	0.176	0.072
$\beta_{1,FTSE100}$	0.803	0.0506
$\alpha_{0,NASDAQ100}$	0.00006	(< 0.001)
$\alpha_{1,NASDAQ100}$	0.137	0.040
$\beta_{1,NASDAQ100}$	0.840	0.042
Correlation parameters		
θ_1	0.018	0.003
θ_2	0.981	0.005

4.3.3 Diagnostic Checking for the DCC-GARCH $(1, 1)$

Once the model has been fitted, it is essential to assess the adequacy of the specification of the model. In order to check the adequacy of the fitted models we use Box-Ljung statistics on the standardized squared residuals for univariate GARCH models and we employ multivariate version of Box-Ljung for the estimated squared standardized residuals which are provided presented in Table 4.4 and 4.5 respectively. For the univariate GARCH models which estimates the diagonal elements of the matrix $\mathbf{D}_t$, the diagnostic tests of residuals are all insignificant. Thus, the result suggest that heteroscedasticity has been removed in the residuals. Similarly for the multivariate version of Box-Ljung statistics, all diagnostic statistics are all insignificant. Therefore there is no heteroscedasticity left in in the estimated DCC-GARCH $(1, 1)$ model.

Table 4.4. Box-Ljung statistics for squared standardized residuals

	JSE	FTSE-100	NASDAQ-100
Q2[15]	7.6(0.939)	13.4(0.576)	3.5(0.999)

Table 4.5. Multivariate Box-Ljung Q-statistic test for squared standardized residuals

Statistic	value	p-value
Q2(5)	46.6	0.406
Q2(10)	100.8	0.204
Q2(15)	136.8	0.442
Q2(20)	152.6	0.932

Chapter 5

Conclusion

Volatility of stock returns is a fundamental measurement to several financial decision making models, hence it has been the subject of many empirical studies around the world. Furthermore, the volatilities of financial markets are known to move synchronously across different markets or slightly delayed. Thus, it is essential to examine how financial markets are inter-related. There has been little amount of studies conducted on the emerging stock markets especially in South Africa. The main aim of this study has been to review models to describe the volatility in the Johannesburg Stock Exchange (JSE) index. In this thesis, we focused on modelling volatility and correlation dynamics between JSE index, NASDAQ 100 and FTSE 100 indices. In modelling volatility of the JSE index, we employed daily JSE all-share index closing prices during the sampled period 30 June 1995 to 6 June 2012.

In the preliminary analysis in Chapter 2, daily closing price series for the JSE index was found to be non stationary. We then transformed JSE index closing prices into continuously compounded returns to make it stationary. The time series of returns exhibited volatility clustering, which implies that variance of returns was not constant during the sampled period. The assumption of normality in the JSE returns was rejected. The QQ-plot in Figure 2.3 also revealed the same results that the returns do not conform to normal distribution. The distribution of the returns was negatively skewed, as having a value of -0.4666 which implies that the distribution has a long left tail. In addition, the distribution of the returns was found to be fat-tailed, since the value of the kurtosis was equal to 8.9367 which is greater than that of the normal distribution. The large value of the kurtosis implies thatr large price changes occurred often during the sample period.

For modelling the volatility of the JSE index returns, we first checked the autocorrelation and heteroscedasticity effect. The null hypothesis that there was no autocorrelation in the returns was rejected. Moreover, the null hypothesis that there was no heteroscedastic effect in the returns was rejected. We thus concluded that there was heteroscedasticity present in the returns. We fitted both symmetric GARCH models to accommodate some financial time series characteristics assuming the normal distribution in the errors. Based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), GARCH $(1,1)$ performed better than all models that were investigated in the JSE index for symmetric GARCH models. The parameters of the GARCH $(1,1)$ were all statistically significant. The sum of the estimated parameters $\hat{\alpha}_1$ and $\hat{\beta}_1$ is equal to 0.991 which implies that the volatility of the residuals was persistent.

In Chapter 3, we fitted GARCH $(1,1)$, EGARCH $(1,1)$, GJR-GARCH $(1,1)$ and APARCH $(1,1)$ models with the normal, student-t and skewed student-t distribution to accommodate some financial time series characteristics. The GARCH $(1,1)$ was employed and other three GARCH models, namely, GJR-GARCH $(1,1)$, EGARCH $(1,1)$ and APARCH $(1,1)$ for capturing leverage effect which was noted by Black [1976]. The parameter estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$ are statistically significant in both symmetric and asymmetric GARCH models and their sum is less than one. These results suggested volatility of residuals was persistent. The leverage parameter $\hat{\gamma}$ was statistically significant and positive indicating the leverage effect was present in the JSE index return series. The results revealed that negative residuals increase volatility more than positive residuals of the same magnitude. The EGARCH $(1,1)$ model with the assumption of the skewed student-t distribution on the error terms was selected by both AIC and BIC as the best model. However the heteroscedasticity was not removed in the model, so we decided to select the second best model which was GJR-GARCH $(1,1)$ with the assumption of the student-t distribution in the errors. The results are in line with Cifter [2012].

In Chapter 4, we studied the correlation dynamics between JSE index, NASDAQ 100 and FTSE 100 index. We employed weekly returns during the sampled period 30 June 1995 to 6 June 2012. In all three time series returns that were employed in the study, the assumption of normality was rejected. In addition, the heteroscedasticity was found present in the returns of all the indices. The correlation between JSE index and FTSE 100 index is equal to 0.4604 and the correlation between JSE index and NASDAQ 100 index is 0.3406. This implies that the condition for international portfolio diversification (that is the correlation among the markets should be low) is satisfied. We then fitted the DCC-GARCH $(1, 1)$ model to study the correlation dynamics between JSE, FTSE 100 and NASDAQ 100 index returns. The DCC-GARCH $(1, 1)$ model parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ are statistical significant at 1% significance level indicating that the correlation in this study are dynamic. Moreover, the sum of the parameter estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ was less than one, which implies that the correlation of the returns was persistent.

For future research, in modelling the volatility of the JSE index returns we can employ artificial neural networks to account for the nonlinearity patterns of the returns. Moreover, we can apply the BEKK-GARCH $(1, 1)$ and GO-GARCH $(1, 1)$ models with different distributions on the errors to study further the dynamics of the correlation between financial markets. In addition, we can apply Multivariate GARCH models to study the returns from more than one sector of the JSE market to describe relationships between the sectors. Even further, study how the JSE impacts local microeconomic variables.

Appendix A

R Code for Univariate and Multivariate GARCH Models

A.1 R Code for Asymmetric GARCH Models

```
##################################################
########### Asymmetric power arch ###############
#######   Assuming Normal distribution###########
##################################################

aparch=garchFit(~aparch(1,1),data=returns,trace=FALSE
,cond.dist="norm")
summary(aparch)

gjr=garchFit(~aparch(1,1),data=returns,trace=FALSE,delta=2,
include.delta=FALSE,cond.dist="norm")
summary(gjr)
###################################################
library(rugarch)
returns= data$returns[-1]
egarch11.spec = ugarchspec(variance.model=list(model=
"eGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0))
,distribution.model="norm")
p1=ugarchfit(egarch11.spec,returns)
p1
```

```
plot(p1)

################################################
########### Asymmetric power arch ###############
#######   Assuming skew-t distribution###########
################################################

aparchtsk=garchFit(~aparch(1,1),data=returns,
trace=FALSE,cond.dist="sstd")
summary(aparchtsk)

gjrtsk=garchFit(~aparch(1,1),data=returns,
trace=FALSE,delta=2,
include.delta=FALSE,cond.dist="sstd")
summary(gjrtsk)
#################################################
library(rugarch)
returns= data$returns[-1]
egarch11.spec = ugarchspec(variance.model=
list(model="eGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="sstd")
p1=ugarchfit(egarch11.spec,returns)
p1
plot(p1)

################################################
############ Asymmetric power arch ###############
#######   Assuming t distribution    ############
################################################

apt=garchFit(~aparch(1,1),data=returns,trace=FALSE,
cond.dist="std")
summary(apt)

tgt=garchFit(~aparch(1,1),data=returns,
trace=FALSE,delta=1,
include.delta=FALSE,cond.dist="std")
summary(tgt)

gjrt=garchFit(~aparch(1,1),data=returns,
```

```
 trace=FALSE,delta=2,
 include.delta=FALSE,cond.dist="std")
 summary(gjrt)
 plot(gjrt)
##################################################
library(rugarch)
returns= data$returns[-1]
egarch11.spec = ugarchspec(variance.model=
list(model="eGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="std")
p1=ugarchfit(egarch11.spec,returns)
p1
plot(p1)
jarque.bera.test(returns)
plot(p1)
```

A.2 R Code for Symmetric GARCH Models

```
############################################
#############GARCH Models#################
############################################
library(fGarch)
m1=garchFit(~garch(1,1),data=
returns,trace=FALSE)
summary(m1)

 m2=garchFit(~garch(1,2),data=
 returns,trace=FALSE)
 summary(m2)

 m3=garchFit(~garch(2,1),data=returns,
 trace=FALSE)
 summary(m3)

 m4=garchFit(~garch(2,2),data=returns,
 trace=FALSE)
 summary(m4)
```

```
m5=garchFit(~garch(1,0),data=
returns,trace=FALSE)
summary(m5)
```

A.3 R Code for DCC-GARCH $(1, 1)$ GARCH Models

```
library(rugarch)
jseret= data$jseret[-1]
garch11jse.spec = ugarchspec(variance.model=
list(model="sGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="norm")
p1=ugarchfit(garch11jse.spec,jseret)
p1
plot(p1)

ftseret= data$ftseret[-1]
garch11ftse.spec = ugarchspec(variance.model=
list(model="sGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="norm")
q1=ugarchfit(garch11ftse.spec,ftseret)
q1
plot(q1)

nasdaqret= data$nasdaqret[-1]
garch11nasdaq.spec = ugarchspec(variance.model=
list(model="sGARCH",
garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="norm")
r1=ugarchfit(garch11nasdaq.spec,nasdaqret)
r1
plot(r1)
```

```
 library(ccgarch)
 library(fGarch)
 y=cbind(jseret,ftseret,nasdaqret)
 a = c(0.000039,0.000027,0.000057)
 A = diag(c(0.101569,0.176424,0.136712))
 B = diag(c(0.846575 ,0.802763,0.839890))

dccpara = c(0.2,0.1)
dccresults = dcc.estimation(inia=a,
iniA=A, iniB=B,
ini.dcc=dccpara,dvar=x, model="diagonal")
dccresults$out
DCCrho = dccresults$DCC[,2]
matplot(DCCrho, type='l')
```

Bibliography

P. Alagidede and T. Panagiotidis. Modelling stock returns in africa's emerging equity markets. *International Review of Financial Analysis*, 18(1):1–11, 2009.

N.K. Anaraki. The european stock market impulse to the u.s. financial crisis. *Journal of International Business and Cultural Studies*, 22, 2011.

R.T. Baillie and T. Bollerslev. The message in daily exchange rates: A conditional-variance tale. *Journal of Business & Economic Statistics*, 7(3), 1989.

L. Bauwens, S. Laurent, and J.V.K. Rombouts. Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109, 2006.

A. Bera, E. Bubnys, and H. Park. Conditional heteroscedasticity in the market model and efficient estimates of betas. *The Financial Review*, 23(2):201–214, 1988.

E.K. Berndt, B.H. Hall, R.E. Hall, and J.A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3 (4):653–665, 1974.

F. Black. Studies in stock price volatility change. In *Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Selection, American Statistical Association*, pages 177–181, 1976.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547, 1987.

T. Bollerslev, R.Y. Chou, and K.F. Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1):5–59, 1992.

G.E.P. Box and D.A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970.

Z. Chinzara and M.J. Aziakpono. Dynamic returns linkages and volatility transmission between south african and world major stock markets. *Economic Research Southern Africa Working Papers*, 2009.

A.A. Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432, 1982.

A. Cifter. Volatility forecasting with asymmetric normal mixture garch model: Evidence from south africa. *Journal for Economic Forecasting*, (2):127–142, 2012.

R.A. Connolly. An examination of the robustness of the weekend effect. *Journal of Financial and Quantitative Analysis*, 24(02):133–169, 1989.

Z. Ding, C.W.J. Granger, and R.F. Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

Pierre Duchesne and Simon Lalancette. On testing for multivariate arch effects in vector time series models. *Canadian Journal of Statistics*, 31(3):275–292, 2003.

R. Engle. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

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

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

E.F. Fama. The behavior of stock-market prices. *Journal of business*, pages 34–105, 1965.

R. Ferreira and J.D. Krige. The application of fundamental indexing to the south african equity market for the period 1996 to 2009. *Investment Analysts Journal*, 73:1–12, 2011.

J. Franke, W. Härdle, and C.M. Hafner. *Statistics of financial markets: an introduction*. Springer, 2008.

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

L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, pages 1779–1801, 1993.

C. Gourieroux and J. Jasiak. *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press, 2001.

C.W.J. Granger and A.P. Andersen. *An introduction to bilinear time series models*. Vandenhoeck und Ruprecht, 1978.

R. Gupta and A. Mollik. Volatility, time varying correlation and international portfolio diversification: an empirical study of australia and emerging markets. *International Research Journal of Finance and Economics*, 18, 2008.

J.D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

B.E. Hansen. Autoregressive conditional density estimation. *International Economic Review*, pages 705–730, 1994.

M.L. Higgins and A.K. Bera. A class of nonlinear arch models. *International Economic Review*, pages 137–158, 1992.

W.J. Horng, T.C. Hu, and J.L. Tsai. Dynamic relatedness analysis of two stock market returns volatility: An empirical study on the south korean and japanese stock markets. *Asian Journal of Management and Humanity Sciences*, 4(1): 1–15, 2009.

J.R.M. Hosking. The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608, 1980.

K. Jefferis, C. Okeahalam, and T. Matome. *International stock market linkages in Southern Africa*. Number 17. Botswana Institute for Development Policy Analysis Gaborone, 1999.

S. Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2):231–254, 1988.

Hung-Chun Liu and Jui-Cheng Hung. Forecasting s&p-100 stock index volatility: The role of volatility asymmetry and distributional assumption in garch models. *Expert Systems with Applications*, 37(7):4928–4934, 2010.

G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

M.R. Makhwiting, M. Lesaoana, and C. Sigauke. Modelling volatility and financial market risk of shares on the johannesburg stock exchange. *African Journal of Business Management*, 6(27):8065–8070, 2012.

B. Mandelbrot. The variation of certain speculative prices. *Journal of business*, pages 394–419, 1963.

R. Mangani. Modelling return volatility on the jse securities exchange of south africa. *The African Finance Journal*, 10(1):55–71, 2008.

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

J.H. Mittelman. *The globalization syndrome: transformation and resistance.* Princeton University Press, 2000.

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

D.B. Nelson and C.Q. Cao. Inequality constraints in the univariate garch model. *Journal of Business & Economic Statistics*, 10(2):229–235, 1992.

R.A. Olowe. Modelling naira/dollar exchange rate volatility: application of garch and assymetric models. *International Review of Business Research Papers*, 5(3):377–398, 2009.

T. Olweny and K. Omondi. The effect of macro-economic factors on stock return volatility in the nairobi stock exchange, kenya. *Economics and Finance Review*, 1(10):34–48, 2011.

C.E. Onwukwe, B.E.E. Bassey, and I.O. Isaac. On modeling the volatility of nigerian stock returns using garch models. *Journal of Mathematics Research*, 3 (4):31–43, 2011.

S.G. Pantula. Modelling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):71–74, 1986.

G.R. Pasha, T. Qasim, and M. Aslam. Estimating and forecasting volatility of financial time series in pakistan with garch-type models. *Lahore Journal of Economics*, 12(2):115–149, 2007.

N. Sariannidis, G. Konteos, and E. Drimbetas. Volatility linkages among india, hong kong and singapore stock markets. *International Research Journal of Finance and Economics*, 58:141–149, 2010.

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

G.W. Schwert and P.J. Seguin. Heteroskedasticity in stock returns. *The Journal of Finance*, 45(4):1129–1155, 1990.

I.S. Talke. Modelling volatility in time series data. Master's thesis, University of Kwa-Zulu Natal, 2003.

S. Taylor. *Modelling financial time series*. John Wiley & Sons, 1986.

R.S. Tsay. *Analysis of Financial Time Series, Third Edition.* John Wiley & Sons, Inc., 2010.

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

J.M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994.

E. Zivot. Practical issues in the analysis of univariate garch models. *Handbook of Financial Time Series*, pages 113–155, 2009.