A comparative analysis of ARCH and GARCH type models before and after the removal of outliers

L.D Sepato

orcid.org/0000-0002-0269-3758

Dissertation submitted in fulfilment of the requirements for the degree *Master of Commerce* in Statistics at the North-West University

Supervisor: Prof N.D Moroke

Co-supervisor: Dr F Matarise

LIBRARY
MAFIKENG CAMPUS
CALL NO.:
2021 -02- 01
ACC.NO.:
NORTH-WEST UNIVERSITY

Graduation Ceremony: October 2017

Student number: 20840373

http://dspace.nwu.ac.za/

DECLARATION

I declare that the study titled "*A comparative analysis of ARCH and GARCH type models before and after the removal of outliers*" towards the award of the M.Com degree is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full names: Lesego Dorothy Sepato Date: 26/10/2017

Signed: Sepato

Signature.. Date.................................
Supervisor

ACKNOWLEDGEMENTS

Firstly, I thank God for giving me the strength and knowledge to carry out this dissertation, it was not an easy journey but I managed. Secondly, I would like to express my deep gratitude to Prof N.D Moroke, my supervisor, for her patience, guidance, encouragement and useful criticisms towards this dissertation. I appreciate that despite everything you still did not give up on me, and for that, I will be forever grateful. Thirdly, I thank Dr. Matarise, who took the time to supervise me despite the long distance barrier between us and the University of Zimbabwe.

I also thank my beloved father for his advice and assistance in keeping my progress on schedule, without him this dissertation would not be a success. Lastly, I thank my family for their support and encouragement throughout my studies.

DEDICATION

This dissertation is dedicated to my late mother Ditsamai Sepato, you are truly missed.

ABSTRACT

Forecasting volatility of the JSE financial data in literature is extremely popular in South Africa (SA). The procedure involves modeling, assessing, and predicting time-varying return volatility. Numerous volatility estimators and models have been recommended to quantify unpredictability of stock returns which adds to upgraded asset pricing, portfolio management, and risk management. However, financial data often contain observations caused by unexpected events, called interventions, and such extreme returns are often found to disturb volatility less than a standard time series model would forecast.

The main purpose of this study is to assess the performance of the ARCH and GARCH type family models before and after the outlier(s) have been removed from a set of data. Investigating this subject proves to be vital since no attention has been paid to outliers in volatile financial time series data. The study explored through the guidance of the ACFs and PACFs the ARMA (2, 2), ARCH (2), together with the ARMA enhanced GARCH models such as the ARMA (2, 2)-GARCH (1, 1), ARMA (2, 2)-EGARCH (1, 1), ARMA (2, 2)-TGARCH (1, 1), ARMA (2, 2)-APARCH (1, 1), ARMA (2, 2)-GJR-GARCH (1, 1), ARMA (2, 2) -CGARCH (1, 1), together with ARMA (0, 2), ARCH (2), ARMA (0, 2)-GARCH (1, 1), ARMA (0, 2)-EGARCH (1, 1), ARMA (0, 2)-TGARCH (1, 1), ARMA (0, 2)-APARCH (1, 1), ARMA (2, 2) -CGARCH (1, 1) and ARMA (0, 2)-GJR-GARCH (1, 1) models to assess the volatility in stock returns data. These models were applied to outlier contaminated and outlier free data. Daily time series data from 03 January 2011 until 21 April 2016 was sourced from the JSE database.

SAS 9.4, E-Views 9 and Gretl version were used to obtain results. Preliminary data analysis was conducted to check the variable description before the estimation of the models. The stock return data conceded the diagnostics such as independence, unit root except for normality. Seven models selected according to minimum information criteria were exposed to model diagnostics testing. The ARMA (2, 2)-EGARCH (1, 1) t-distribution model was confirmed to be adequate and stable for the data for original return series and ARMA (0, 2)-EGARCH (1, 1) was appropriate in the outlier free data. These models were recommended for further analysis and were later used for producing forecasts of JSE stock returns.

The ARMA (2, 2)-EGARCH (1, 1) model proved to be consistent, reasonable and effective for forecasting volatility. The model further delivered a high unpredictability contrasted to other

models. In view of these findings, the study suggested the utilisation of this model for further analyses. These forecasts might be utilised when setting out new strategies concerning JSE trading in South Africa. A subsequent study was suggested where other GARCH family models ought to be assessed and the outcomes contrasted with those acquired in this study.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Appendices

List of Acronyms

ADF	Augmented Dickey-Fuller
ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ALO	Additive Level Outlier
AO	Additive Outliers
APARCH	Asymmetric Power Generalized Autoregressive Conditional Heteroskedastic
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedastic
ARMA	Autoregressive moving average
AVO	Additive Volatility Outlier
BIC	Bayesian information criterion
BM	Bounded-M
CCC	Conditional Constant Correlation
CGARCH	Component Generalized Autoregressive Conditional Heteroskedastic
D-BEKK	diagonal Baba-Engel-Kraft-Kroner
DCC	Dynamic Conditional Correlation
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedastic
FIGARCH	Fractional Integrated Generalized Autoregressive Conditional Heteroskedastic
FTSE	Financial Times Stock Exchange
GARCH	Generalized Autoregressive Conditional Heteroskedastic
GARCH-M	Generalized Autoregressive Conditional Heteroskedastic- in-mean
GARCH-t	Generalized Autoregressive Conditional Heteroskedastic- t distributed errors
GJR-GARCH	Glosten, Jagannathan, and Runkle-Generalized Autoregressive Conditional Heteroskedastic
IGARCH	Integrated Generalized Autoregressive Conditional Heteroskedastic

MAFE	Mean Absolute Forecasting Error
JSE	Johannesburg Stock Exchange
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
LB-Q	Ljung-Box
LM	Lagrange Multiplier
LS	Level Shift Outliers
MA	Moving average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimate
NA-GARCH	Nonlinear Asymmetric Generalized Autoregressive Conditional Heteroskedastic
PACF	Partial Autocorrelation Function
QML-t	Quasi Maximum Likelihood t-distribution
RMSE	Root Mean Square Error
SA	South Africa
S & P	Standard & Poor's
TC	Transitory Change Outliers
TGARCH	Threshold Generalized Autoregressive Conditional Heteroskedastic
TS-GARCH	Taylor/Schwert Generalized Autoregressive Conditional Heteroskedastic
VGARCH	Vector Generalized Autoregressive Conditional Heteroskedastic or (Multivariate Generalized Autoregressive Conditional Heteroskedastic)

List of special symbols

σ_t^2 = variance with respect to time

ε_t = error term

μ_t = mean

r_t=return

CHAPTER ONE
ORIENTATION OF THE STUDY

1.1. Introduction

The South African stock markets have been growing as far back as the dispatch of the Johannesburg Stock Exchange (JSE) in 1887, amid the main gold rush in South Africa (SA). The historical backdrop of the JSE goes back to 1947. This occurred in the wake of taking after the principal enactment covering in money related markets where the JSE joined the World Federation Exchange in 1963 and redesigned its exchanging frameworks in the mid-1990's (Michie, 2006). Predominantly, with the present boom in South African's economy; South Africa's stock markets have been drawing a boundless measure of consideration according to (Câmara & Manuel, 2011). The country has become more and more interrelated in its globalisation and has profited in numerous establishments. This ranges from general economies, strategy makers, investors, savers and scholastics (Hou, 2013). According to Investopedia the share trading system is very much characterised as a standout amongst the most vital parts of a free market economy. This division gives organisations with access to resources in return for giving shareholders a rate of proprietorship in the business (www.investopedia.com).

Stock markets are subject to irregular growth and decline. This is due to market crashes which are difficult to comprehend and these unexpected fluctuations affect the dynamics of the data temporarily or permanently (Foley, 2014). In principle, stock markets are volatile and behave differently from other markets. As a result in the previous years, financial markets around the world have experienced high volatility and unpredicted decreasing returns making investors particularly profound to guard their portfolios (Storch et al., 2013). Volatility is defined as the "*variability in prices or returns that can be measured by the standard deviation*" (Heracleous, 2003:1). Subsequently, volatility gives an important part in the option pricing decision making practice and is a vital unknown variable in option pricing and assessments (Ederington & Guan, 2006, Song & Xiu, 2016).

Literature published on forecasting volatility is ample and is still growing. This literature helps in understanding the nature and temporal behaviour of the stock return forecastability (Rapach & Zhou, 2013). There is also a vast theory and literature behind volatility that further aids to enhance modeling problems in financial time series such as forecasting investment, option pricing, and risk management, etc. (Poon & Granger, 2005). Furthermore, this high frequency data allows a better understanding of dynamic properties of highly persistent volatility as it allows spread of announcements in markets that activate shocks in volatility (Storch et al., 2013). The number of crashes (i.e. long periods of rather normal price movements often disturbed by short periods of somewhat large and abrupt price movements) and the extent of their impacts have compelled many researchers to read into the stationarity of volatility with respect to time. Thus resulting to move their attention towards improvement change of econometric models that provide reliable and reasonable forecasts, producing fewer forecast errors about such swings done by returns' instability (Matei, 2009).

The application of volatility is vital to anyone analysing financial data. In general, volatility has been connected to risks, falsely priced securities or even failing of the entire market. Therefore, for that reason modeling and forecasting stock market volatility is still to date the main obstacle that academics and practitioners are faced with in terms of formulating a generally accepted model that could capture all the aspects of volatility. The stock returns can be characterised by models that are able to capture the core stylised features existing in financial time series, such as (i) time-varying volatility; (ii) financial series that is leptokurtic (fat tails); (iii) significant serial correlation between volatility (i.e. volatility clustering) and (iv) the leverage effect (Makhwiting et al., 2012; Samouilhan & Shannon, 2008).

Though there are many models being used in a univariate series, this study focused on the Autoregressive conditional heteroscedastic (ARCH) models proposed by Engle (1982), the generalised ARCH (GARCH) of Bollerslev (1986), the Glosten, Jagannathan, and Runkle-GARCH (GJR-GARCH) model proposed by Glosten et al. (1993), the Asymmetric Power ARCH (APARCH) by Ding et al. (1993), the Threshold GARCH (TGARCH) of Glosten et al. (1993), Component GARCH (CGARCH) of Ding and Granger (1996); Engel and Lee (1991) and the Exponential GARCH (EGARCH) model of Nelson (1991). These models help in characterising the dynamic evolution of conditional variances. These models are most referenced, most popular in financial time series literature and are simple historical models. Furthermore, just like in linear models, outliers affect model estimation and identification in GARCH family models. These models have the ability to capture the behaviour of stock volatility based on past standard deviations, hence the model is most favoured in the financial literature (Kgosietsile, 2015).

In summary, ARCH parameter captures the short run persistence of shocks, whereas the GARCH parameter captures the impact of shocks to long run persistence (Verhoeven & McAleer, 2004). Meanwhile, both the GJR-GARCH and EGARCH models attempt to address volatility clustering in an innovation process, the APARCH model can well express the fat tails, excess kurtosis and leverage effects and the TGARCH may be useful in determining asymmetry in the financial data. The CGARCH model is helpful as it captures the long range dependence in volatility. Despite the popularity of ARCH and GARCH type models, little attention was paid to the analysis of data containing outliers. Studies by Andersen et al. (2007); Boudt et al. (2013); Laurent et al. (2016), and Verhoeven & McAleer (2004) caution that the consequence of neglecting jumps (outliers) in GARCH models usually overestimate the volatility during several days, if not weeks, after the occurrence of these jumps. Though these models have enjoyed the success in studying the evolution of financial time series modeling, they are unable to explain the high frequency and size of extreme jumps commonly occurring in practice, which may wrongly suggest conditional heteroscedasticity (Carnero et al., 2007).

The remainder of this chapter is organised as follows: the problem statement is given in Section 1.2 Section 1.3 outlines the purpose of the study and Section 1.4 discusses the objectives of the study. Section 1.5 explains the significance of the study, Section 1.6 highlights the preliminary construction of the entire study and the chapter summary is given last in Section 1.7.

1.2. Problem statement

Over the past decades, the ARCH and GARCH type models have remained without uncertainty the most generally utilised models for volatility estimation as indicated in the literature. Then again, forecasts of market volatility across financial markets have additionally been the center stage for academics and experts for several years (Pandey, 2003). This has prompted researchers to the improvement of predominant forecasting techniques. However, there is still no reasonable conclusion as to which of these predicting techniques are best suited to predict financial market volatility with a high degree of accuracy (Wang & Zou, 2010). Literature around this area is very fragmented, hence, this study explores the applicability of the above mentioned models to provide sound conclusions.

Like many other indices, the JSE index is volatile due to its nature. Therefore, customary volatility models, such as the GARCH-family methodologies depend persistently on volatility specification and known distributions of returns. These models are used in this study to determine the volatility of the South African market. Although GARCH type models with fat tails seem to perform adequately in forecasting basic volatility, they, however, perform poorly in predicting extreme observations (Verhoeven & McAleer, 2004). There is a dearth of literature on studies that detect outliers in time series prior to modeling and producing forecasts. Therefore, in studying the problem of outliers, the results may be used, among others, as a diagnostic tool to test the strength and weakness of the model and also make inferences about the parameter. This may help improve the model's goodness-of-fit. Furthermore, this study looks at the influence of outliers, irrespective of whether this influence is known or unknown to the researcher.

1.3. Purpose of the study

In principle, the study intends to explore the efficiency of GARCH type models in outlier free and outlier contaminated data. The researcher sought to determine which of the models when applied to these data types would give better results. The predictive power of the models was assessed with least forecast error generated.

1.4. Objectives of the study

The main objective of this study is to assess the performance of the symmetric GARCH and asymmetric GARCH type models in outlier free and outlier contaminated data. It is assumed that the data capture "stylised facts" of financial time series. Specific objectives of this study are to:

- estimate standard time series predictive ARCH and GARCH type models for South African top 40 indices,
- explore the type of outliers present in the proposed data,
- investigate the effects of outliers in the diagnostics of ARCH and GARCH type models,
- determine the predictive performance of ARCH and GARCH type models on stock performance, for series which is affected and unaffected by outliers,
- use the findings of the study in formulating suggestions with regards to policy formulation and future studies.

1.5. Significance of the study

This study contributes to the stock market data modelling and prediction by comparing the ARCH, GARCH (p, q), GJR-GARCH (p, q), TGARCH (p, q), APARCH (p, q), CGARCH (p, q) and EGARCH (p, q) models in the presence and absence of outliers. This study is anticipated to give awareness and foundation for future studies in the field or other related fields. The study intends to contribute to policy through estimating a forecasting model which policymakers could refer to. To the best of the researcher's knowledge, there are no studies that have compared the performance of ARCH and GARCH type models with or without outliers in the context of South Africa. Therefore, the findings of this study may also contribute to the literature on the subject and may help in popularising outlier detection methods such as wavelength outlier detection, adaptive outlier detection, robust outlier detection and empirical likelihood for outlier detection among other things.

1.6. Research outline

This study proceeds as follows: Chapter two gives a review of theory and literature in relation to the study. This entails a critical evaluation of previous research methods to be used in the study.

Chapter three presents a concise description of the data and research methodology including the research design and procedure to be used for data analysis.

Chapter four presents and discusses the results by providing statistical analysis and interpretation of results.

Chapter five gives potential options, conclusions and recommendations for future studies and policy adjustments.

Appendices follow after Chapter five, with a list of references as cited in the study.

1.7. Chapter summary

This chapter displayed the prologue to the study. The research problem was clearly stated and used to generate the study objectives. The purpose and significance of this study were highlighted.

The chapter that follows provides a review of the literature.

CHAPTER TWO
LITERATURE REVIEW

2.1. Introduction

This chapter reviews the theoretical literature on univariate ARCH and GARCH type modeling for modeling and forecasting stock return volatility. The study firstly explores the empirical literature in Section 2.2; Section 2.3 focuses on the theoretical literature. Section 2.4 highlights the review on outliers; thereafter Section 2.5 entailing the overview outliers and Section 2.6 entailing gap identified follows respectively thereafter.

2.2. Empirical Literature

This section aims to provide a brief overview of the vital developments relevant to the GARCH-family models.

2.2.1. Brief overview of estimating GARCH volatility in the presence of outliers

This study evaluates and compares empirical performance of one of the different unconditional volatility estimators and conditional volatility models called the ARCH and GARCH type models with respect to outliers. The performance of these models is shown on the JSE top 40 index, a value-weighted index of 40 stocks traded on the JSE as recommended by Pandey (2003). Attention is given to three types of outliers mentioned before. A brief overview of the types of outliers is that an additive outlier (AO) affects only one observation, an innovation outlier (IO) acts as an addition to the noise term at a specific series point, and in stationary series, an IO affects numerous observations after its occurrence. For nonstationary series, IO may affect every single observation starting at a particular series point. A level shift (LS) is an outlier that shifts all observations by a constant, starting at a particular series point. An LS could result from a change in policy. The reason for outlier identification may be to find those unexpected data, whose conduct is very exceptional when contrasted with the rest of the dataset.

A key characteristic of time series data is the noticeable large (small) absolute returns that tend to be followed by large (small) absolute returns. This means that there are eras which show high (low) volatility, thus, classifying volatility as a measure of risk (Charles & Darné, 2005). A paramount characteristic of a GARCH model is that it might be fitted to data which have excess kurtosis as stated. Regardless of the qualities that the GARCH model possess, the empirical distributions are more peaked about the mean and have fatter tails than a normal distribution (Franses & Ghijsels, 1999). The authors further opined that the accuracy of the model and confidence limits are affected. One conceivable foundation to this result is that some observations in the returns are alleged to be perceived as AOs which cannot be captured by a standard GARCH model (Carnero et al., 2007). This means that the estimated residuals from GARCH models still registering excess kurtosis are observations on returns that cannot be fitted by a Gaussian GARCH model. Not even a t-distributed GARCH model (Student t-distribution) may be used to fit such residuals. These observations are considered to be influential since they can affect the estimation of parameters, the conditional homoscedasticity tests and the out-of-sample volatility forecasts. Some authors denote such observations as outliers (Doornik & Ooms, 2005).

2.2.2. Modelling returns using GARCH models with outliers

Catalán & Trivez (2007) conducted a study on forecasting volatility in GARCH models with AOs. The general purpose of the study was to uncover threefold mistakes researchers make by paying limited attention to outliers. Firstly, the study investigated the effects of outliers on partial autocorrelation function to identify the ARCH model. Secondly, the study determined the impact of outliers on the size and power of the Lagrange multipliers. The last step investigated the effects of outliers on the estimation of parameters of the model. Mainly, the study measured the effects of two types of outliers: the additive level outlier (ALO) and additive volatility outlier (AVO) on the results. The data used was generated using Monte Carlo simulation and the results showed that the unfamiliarity of AVO and ALO has dire consequences in forecasting volatility. The size of the outlier was found to have an effect on the relative increase in the mean absolute forecast error (MAFE) (i.e. the results showed that as the MAFE decreases the size of the time horizon of the forecast increases). These findings implied that if the outliers are not treated, they can lead to model mis-specification and poor forecasts. Therefore, ignorance of such outliers could also lead to an increase in the mean absolute forecast error, where the error is increased by bias produced in the estimation of the coefficients of the model.

Louw (2008) conducted a study on the evidence of volatility clustering on the FTSE/JSE top 40 index. The study investigated whether the presence of volatility clustering exists on the FTSE/JSE top 40 index daily closing prices. The study used five different return interval size to seek the distributional characteristics and to check if the normality assumption is violated by analysing each of the return interval sizes. Moreover, for every interval sizes a one-step-ahead volatility was used for linear regression, exponential smoothing, GARCH (1, 1) and EGARCH (1, 1) models to determine the most powerful model. The results showed that the FTSE/JSE top 40 index has the volatility clustering, and was not normally distributed. The data proved to be leptokurtic with a very high kurtosis with negative skewness.

Charles (2008) conducted a study on forecasting volatility with outliers in GARCH models. The study aimed at detecting and correcting irregular returns in 17 French stock returns and the French index CAC40 during the period from 6 January 1997 to 4 April 2002. The data consisted of 1435 observations from the closing prices. The study also sought to get a breakthrough from AO detection method in GARCH models. Specifically, Charles (2008) study was similar to studies conducted by (Franses & Ghijsels, 1999) and (Charles & Darné, 2005). The findings showed the parameters of the equation of AR (*p*)-GARCH (1, 1) (i.e. the AR-GARCH) to be an appropriate model to predict volatility. Changing aspects showed biasness when the exclusion of outliers was not taken into account. Volatility forecasts revealed were more accurate as opposed to a series contaminated with outliers. This was so even if Gaussian innovation processes like the GARCH-t process was taken into account. The findings further showed that outliers lead to model mis-specification and forecasts. These findings underline the fact that outliers can lead to a mis-specification of the model when ignored.

Carnero et al. (2012) conducted a study estimating GARCH volatility in the presence of outliers. The aim of the study was to develop GARCH models when the residual has excess kurtosis. The authors benchmarked on previous studies such as Sakata and White, (1999); Carnero et al. (2007) and Muler & Yohai, (2008) which concentrated on the effects of outliers on the Gaussian Maximum Likelihood (ML) estimator of GARCH parameters. The analysis was done on the daily return of the S&P 500 from January 2, 1987 till February 19, 2008. The methods used were the Quasi Maximum Likelihood t-distribution (QML-t) estimates, the bounded – M (BM) estimator and the Gaussian ML estimator. The study proved that the model with outliers gave biased GARCH parameter estimates. The study also found that GARCH provided more robust BM estimator.

Ilango et al. (2013) conducted a study on outlier detection and influential point observation in linear regression using clustering techniques in financial time series data. The study was aimed at discovering a new method which is built on clustering techniques for the outlier. The

Expectation Maximization cluster (EM-Cluster) algorithm was used to find the "optimal" parameters of the distributions that maximise the likelihood function. Regression centered outlier technique was used to detect influence point. The study further investigated outliers, volatility clustering and risk-return trade-off in the Indian stock markets NSE Nifty and BSE SENSEX. Engle's ARCH Test and $AR\ (1) - EGARCH\ (p, q) - in - Mean$ model were used to conduct the investigation. The findings showed that the stock market return had evidence of volatility clustering/ARCH effects judging from the significant ARCH-LM test statistics. The empirical results of $AR\ (1) - EGARCH\ (p, q) - in - Mean$ model revealed that volatility is persistent and there is leverage effect supporting the Indian stock markets. Moreover, the study revealed positive but insignificant relationship between stock return and risk for NSE Nifty and BSE SENSEX stock markets. Gaussian mixture model with EM algorithm is a powerful approach for volatility clustering.

Veiga et al. (2014) analysed outliers in multivariate GARCH models and found that moderate amount of outliers have a huge impact on stock prices and returns affecting both estimations of parameters and volatilities when fitting a GARCH type model. The paper aimed at uncovering the impact of AOs (isolated, patches and volatility outliers) on the estimation of correlations. The study made use of the diagonal Baba-Engle-Kraft-Kroner (D-BEKK) by Engle and Kroner (1991), the Conditional Constant Correlation (CCC) model by Bollerslev (1990) and the dynamic conditional correlation (DCC) model by Engle (2002) with simulated data. The results showed that through Monte Carlo simulation, outliers influence the evaluated correlations. Also, this impact will be stronger for the restrictive correspondence models CCC and DCC. The Monte Carlo simulation proved to be accurate because it detects the rate of right detections and this means it also detects the right amount of false AOs. The procedure can also be interpreted as a model mis-specification test since it is grounded on residual diagnostics. The simulation studies showed that correlations are extremely affected by the inclusion of outliers and that the new method is both effective and reliable.

2.2.3. Models for evaluating volatility

Financial time series modeling has been an increasingly important aspect in financial literature over about 20 decades ago, and it has been a widely researched topic since the global financial crises of 2008. This led practitioners and academics to review the adequacy of many financial models (Brownlees et al., 2011). It is still an important factor in the financial sector to date and there are a number of models that have been tried and tested to forecast/ model volatility.

However, research has surprisingly produced conflicting conclusions on which model best describes volatility forecasting. This made volatility prediction an acute task in asset evaluation and risk management for investors and financial arbitrators.

Numerous financial return models with changing volatilities have been developed by researchers and are currently very popular in the recent literature describing how the stock market moves, how the stock index behaves and the patterns of currency returns (Heynan & Kat, 1994). In 1971, Box and Jenkins promoted the ARMA model used to identify the linear process generated by a given time series on condition that it is stationary (Montgomery and Weatherby, 1980). Thus, this model gained more attention as it captured the volatility movement, but a disadvantage of the ARMA model was that it broke the non-negativity constraint by not overcoming the Random Walk (RW) benchmark model (Einarsen, 2014). This lead to the development of one model that stands out and is frequently used, the ARCH (p) model proposed by Engle (1982). In this model, the conditional variance is written as a linear function of p past squared innovations. This model proved to be a breakthrough due to its capacity to give detail the non-linear changing aspects of financial data in an enhanced way. Later Bollerselev (1986) generalised the ARCH model by permitting conditional variances as well as the linear GARCH (p, q) model. A disadvantage of the GARCH model is that it is symmetric (i.e. leverage effect- alludes to the usually negative correlation between an asset return and its fluctuations of volatility). Moreover, the effect leads to the asymmetric GARCH-model (Einarsen, 2014). Nelson (1991) introduced the EGARCH (p, q) to model the return variance which does not only depend on the magnitude but by the sign of the past errors. This model has been proven to be quite successful in modeling the stock returns (Heynen & Kat, 1994).

Thereafter, studies regarding time series volatility modelling and forecasting that could influence its raw form to model the asymmetric effects observed in financial time series data and long memory in variance have been proposed in many studies to obtain specifications for σ_t, that includes the IGARCH model, the Taylor(1986)/Schwert (1989)(TS-GARCH) pattern, the A-GARCH^2, the NA-GARCH and the V-GARCH design introduce to by Engle and Ng (1993), the threshold GARCH dummy (Threshold.-GARCH) by Zakoian (1994), the GJR-GARCH shape of Glosten et al. (1993) to mention a few (Hansen & Lunde, 2005).

In most studies, the purpose is for modeling, forecasting and identifying the best GARCH specification that captures the symmetric and asymmetric effects of the data. ARCH, GARCH and GARCH-M are usually effective at capturing the leptokurtic (fat tails) and also the volatility clustering. Conversely, asymmetric effects such as leptokurtosis as well as volatility clustering, leverage effects and volatility persistence are captured better with the EGARCH, GJR-GARCH, APGARCH, FIGARCH, CGARCH (Kgosietsile, 2015).

2.2.4. Applications of volatility measurement

Volatility is a measure of the improbability of the return for a particular asset (i.e. something that varies or changes such as the rate or revenues of equity over time). Large values of volatility indicate that returns fluctuate in a wide range – statistically in terms of the standard deviation which is a measure that offers an indication of the dispersion or spread of the data (www.jse.co.za). Since it was first proposed by pioneers Black-Scholes-Merton in 1976, volatility has played an important part of the Black-Scholes-Merton option pricing model. Beginning from the 'efficient markets' or 'random walk' model, asset price movements can be described by the following equation:

$$r_t = \frac{S_t - S_{t-1}}{S_t} = \mu_t + \varepsilon_t, \tag{2.1}$$

where $E[\varepsilon_t] = 0$ and $\text{var}[\varepsilon_t] = \sigma_t^2$.

"The return at time t, r_t, is the percentage change in the asset price S, over period from t to t-1. This is equal to μ, a non-random mean return for period t, and a zero mean random disturbance e_t" (Kambouroudis, 2012:1).

Even though the model (Black Scholes) treats the volatility as a constant variable over time, the three pioneers found that volatility changes over time. Black & Scholes (1973) defined *volatility "as the standard deviation because it measures the variability in the returns of the underlying asset."* Therefore it can be said that volatility is statistically persistent, that is, the present state may remain similar in the near future. This is also known as volatility clustering; a typical characteristic found in financial time series data.

Furthermore, financial stability gets affected by volatility during an international financial crisis. Financial stability is of vital importance to any country concerned and also to other countries that have relations with South Africa. The volatility estimation used in this study was calculated from return series only. To measure the historical volatility, the close-to-close market share was used to measure the implied volatility, which is the simplest and most common type of calculation used from reliable closing stock prices. Volatility is important in the study of asset returns because returns are not serially correlated by they are non-linear dependent (Figlewski & Wang, 2001).

In order to review the empirical research to analyse and draw conclusions, there are several models such as symmetric GARCH, asymmetric GARCH type and parametric models developed to measure volatility.

2.2.5. Empirical consistencies of stock return

According to a study conducted by Kambouroudis (2011), following the studies of Mandelbrot (1963) and Fama (1965) who highlighted that "experimental distribution of stock returns is significantly non-normal, caused by large kurtosis (i.e. stock returns are usually leptokurtic) and also if the returns are skewed to the left or right with a constant variance over time can display volatility clustering." Discussed below are the stylised facts regarding financial price data.

- **Volatility clustering**

Volatility clustering is characterised by extensive changes which are trailed by substantial changes or small changes in a time series data. It is recommended that for time samples appear to be uncorrelated across time, they are actually dependent on time (Babu & Reddy, 2014). Numerous financial time series share similar properties, e.g., relative price changes (returns) showing the recurrent occurrence of large eruptions (volatility clustering). This results in the power-law scaling of tails in their probability distributions, as well as of the autocorrelation function of their absolute values (Krawiecki et al., 2002). Volatility clustering can be measured by the autocorrelation plot of absolute returns data, or either using the autocorrelation plot of squared returns (Babu & Reddy, 2014).

According to a study by Mandelbrot (1963:418), "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." A quantitative appearance is that, while returns themselves are uncorrelated, absolute returns or their squares show a positive, significant and slowly decaying autocorrelation function $corr(|r_t|, |r_{t+\tau}|) > 0 \quad for\, t$ ranging from a few minutes to several weeks (Cont, 2007).

- **Thick tails**

Asset returns are leptokurtic. This implies that the tails of the distribution comprise of too many extreme points to fit the normal distribution, where these extreme observations tend to cluster together over time (Franses, 1998). The presence of excess kurtosis can possibly be explained by the presence of outliers (Charles, 2008).

- **Leverage effects**

Leverage can be defined as "*the ratio of a company's loan capital (debt) to the value of its ordinary shares (equity)*"; or "*the use of borrowed capital for (an investment), expecting the profits made to be greater than the interest payable*" (Bach, 2003:108). The leverage effect illustrated by Black (1976) refers to the negative correlation between stock returns and hence cause changes in stock volatility. The negative relation between leverage and stock returns indicates how leverage should be priced and taken into account whilst evaluating risk in the asset pricing models (Adami et al, 2010). In a nutshell, this basically means that a company with debt and unresolved equity normally becomes highly leveraged when the value of the company falls thus rising equity returns volatility (Ouso, 2012). A negative relation between stock returns and leverage suggests that leverage is priced by the market. According to Black (1976), it can be proved that fixed costs (i.e. financial and leverage) give an explanation of the leverage effect in stock prices (Kambouroudis, 2012).

- **Co-movements in volatilities**

According to Black (1976) "*generally it is fair to say that when stock volatilities change, they tend to change in the same direction.*" Therefore, there are also many studies that back up this argument of the existence of joint factors clarifying volatility activities in stock data. Bollerslev et al. (1994) who mentioned the fact that "*volatilities that move together should be encouraging to model builders since it indicates that few common factors may explain much of the temporal variation in the conditional variances and covariance of asset returns, which is the basis of the ARCH modeling*".

2.3. Theoretical Framework

This section entails different models that have been developed to forecast volatility. Some basic model description on some of the financial econometric models that have been previously proposed to model volatility, specifically the models that were used in this study, is given. Thus, this study explored the following volatility models: i. ARMA (p, q), ii. ARCH (p), iii.

GARCH (p, q), iv. GJR-GARCH (p, q) and v. EGARCH (p, q) models applied to the JSE stock index.

2.3.1. ARMA model

The ARMA model (p, q) in a stationary series X_t for every t is expressed as:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i X_{t-i} \tag{2.2}$$

where $\{\varepsilon_t\}$ is *independently and identically distributed (iid)* with $N(\mu, \sigma^2)$, c a constant, ϕ_i and θ_i are parameters to be estimated and polynomials $(1-\phi_1\varepsilon-\ldots-\phi_p\varepsilon^p)$ and $(1+\theta_1\varepsilon+\ldots+\theta_p\varepsilon^q)$ have no common factors.

The process $\{X_t\}$ is said to be an ARMA (p, q) process with mean μ if $\{X_t-\mu\}$, variance σ^2 and c a constant. The time series $\{X_t\}$ is said to be an autoregressive process (AR) of order p, and a moving-average (MA) process of order q (Näsström, 2003). Thus the number of AR and MA terms selected are based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.

2.3.2. ARCH (p) model

The ARCH model, proposed by Engle (1982) has been widely used and generalised by many authors including Bollerslev (1986) and Gourieroux et al. (1997). The ARCH model has contributed a lot to the empirical work and theoretical results of the analysis of data on the exchange rate, stock prices and inflation rate to list a few. The ARCH model is used for predicting time series data for non-linear models which address the heteroscedasticity in the data and does not assume a constant variance (Li, 2013). Volatility clustering or volatility pooling is a motivating factor for using ARCH models to forecast volatility. Mandelbrot (1963) and Fama (1965) both claim that financial markets display volatility clustering. The ARCH model which addresses the conditional variance, which depends on past information is defined by the resulting formulae as presented by Ladokhin (2009):

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} z_t \tag{2.3}$$

$$h_t = \varpi + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \tag{2.4}$$

where r_t is the return at time t, μ is the mean of return, ε_t are the residuals ϖ is the intercept and α_j are parameters of the model, h_t is the conditional variance with $z_t \sim iidN(0,1)$ a normally distributed random variable. The process z_t is scaled by h_t which follows an autoregressive process. In order to ensure that variance h_t is positive, ϖ>0 and α_j ≥0, for *j=1, 2,…, q* must be satisfied (Ladhokhin, 2009). Heteroscedasticity in financial time series present meaning that the variance is not constant over time, where the ARCH-family models are able to capture the magnitude which might appear in clusters (Einarsen, 2014).

2.3.3. GARCH (*p, q*) model

The GARCH model is used to estimate the volatility of an asset. The model was proposed by Bollerslev (1986) as a new form of ARCH model and named it the GARCH model. This model allowed long memory and more flexible lag structure. One of the disadvantages of modeling with an ARCH model is that there may be a requirement for a large value of the lag *q*, hence a large number of parameters. This may result in a model with a large number of parameters, violating the principle of parsimony. This can present difficulties when using the model to effectively describe the data. A GARCH model contains fewer parameters as compared to an ARCH model, and thus a GARCH model may be preferred to an ARCH model.

The GARCH model with the assumption of normality with zero mean can be expressed as:

$$\begin{aligned} y_t &= \varepsilon_t = e_t\sqrt{h_t} \\ y_t|\psi_{t-1} &\sim N(0,h_t) \end{aligned} \tag{2.5}$$

where ψ_{t-1} denotes the past information through time t-1, and the conditional variance h_t can be written as:

$$h_t = \varpi + \sum_{j=1}^{m} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{s} \beta_i h_{t-i} \tag{2.6}$$

where $p \geq 0; q > 0$

$\omega > 0, \alpha_j > 0, j = 1,2,3,\dots,q$

$\beta_i \geq 0, i = 1,2,3,\dots,p$

and q is the order of the autoregressive ARCH term, p is the order of the moving average terms of GARCH model, α and β are the vectors of unknown parameters (Li, 2013). It is noted that ARCH (1) can also be defined as GARCH (0, 1).

2.3.4. GJR-GARCH (*p*, *q*) model

The GJR model is another non-linear extension of the standard GARCH model and was first introduced by Glosten, Jaganathan and Runkle (1993) and have developed the GJR-GARCH model which estimates effects of good news and bad news in financial stock markets. Therefore, to take into account this kind of effect, a dummy variable is introduced into the symmetric GARCH model. This model covers asymmetric or leverage effect confidently along with long memory. GJR-GARCH can be explored as:

$$\sigma_t^2 = \gamma_0 + \sum_{j=1}^{q} \left(\gamma_i \varepsilon_{t-j}^2 + \beta_j d_{t-j} \varepsilon_{t-j}^2\right) + \sum_{i=1}^{p} \omega_i \sigma_{t-1}^2 \qquad (2.7)$$

where d_t is a dummy variable that takes the value "1" when the error term ε_t is negative and "0" and when the error term is positive (Raza et al., 2015).

2.3.5. EGARCH (*p*, *q*) model

The ARCH and the GARCH model are the two models that are able to model the persistence of volatility and volatility clustering but the models both assume that positive and negative shocks have a similar bearing on volatility. In financial asset, volatility innovations have an asymmetric impact. To be able to model this behaviour and overcome the flaws of the GARCH model, Nelson (1991) proposed the first extension to the GARCH model, called the Exponential GARCH (EGARCH) model, which was able to allow for asymmetric effects of positive and negative asset returns (Wennström, 2014). The EGARCH model is designed to capture the leverage effect noted in Black (1976) (Narayan & Narayan, 2007). A simple variance specification of EGARCH (p, q) is given by:

$$\log \sigma_t^2 = \omega + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2 + \sum_{i=1}^{p} \alpha \left(\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \rho_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right) \qquad (2.8)$$

where $p \geq 0$; $q > 0$; $\omega > 0, \alpha_i > 0$ for $i = 1, 2, 3\ldots q$ and $\beta_j \geq 0$ for $j = 1, 2, 3\ldots, p$.

where ε_t a white noise process with zero mean, σ_t^2 is the conditional variance, q and p are the order of the ARCH and GARCH terms respectively. The ρ parameter is the measurement of the asymmetry effect. If $\rho = 0$, then the impact to conditional variance is symmetric and the leverage effect is asymmetric if the parameter $\rho \neq 0$ (Li, 2013).

2.3.6. TGARCH (*p, q*) model

The TGARCH (*p, q*) model was introduced by Glosten et al. (1993) and later developed by Zakoian (1994). This model defines the conditional variance as a piecewise function and captures the asymmetric effect (Zhang & Zhang, 2016). Thus the TGARCH (p, q) model specification for conditional variance is given by:

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j}^2 d_{t-j} + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2.9}$$

where $d_t = 1$ if $\varepsilon_t < 0$ and $d_t = 0$ otherwise. Hill et al. (2007) cited by Makhwiting et al. (2014) explains that in equation (2.9) good news is defined by $\varepsilon_t > 0$ and bad news defined by $\varepsilon_t <$ 0, which have distinct effects on the conditional variance. With that said good news have a significant impact on α, while the bad news has an impact in $\alpha + \gamma$ and therefore leverage effects will exist if $\gamma > 0$ and if $\gamma \neq 0$ the bad news increases volatility thus these news impacts are asymmetric.

2.3.7. APARCH (*p, q*) model

The APARCH model is a conventional GARCH model that regress the variance over its residuals and lagged values of variance. Thus, modeling standard deviation of innovations of mean specification is known as Asymmetric Power ARCH model. This model estimates the power of standard deviation instead of forced standard deviation. Therefore the APARCH (*p, q*) model is a nonlinear extension of GARCH (*p, q*) model. APARCH model permits to take into account of both asymmetry and (likely) long memory property. The APARCH specifications are given as follows:

$$\sigma_t^\delta = \gamma_0 + \sum_{j=1}^{q} \gamma_j \left(|\varepsilon_{t-j}| - \beta \varepsilon_{t-j} \right)^\delta + \sum_{i=1}^{p} \omega_i \sigma_{t-i}^\delta \tag{2.10}$$

where $p \geq 0, q > 0$, $\gamma_0 > 0, \gamma_i \geq 0$, $|\beta_j| \leq 1$ for all $j = 1, 2, \ldots, q$ & $\omega_i \geq 0$, $\forall$, $j = 1, 2, 3, \ldots, p$ and $\sigma > 0$ (Raza, et al. 2015).

2.3.8. CGARCH (*p, q*) model

The CGARCH model is a type of GARCH model that considers long-run and short-run volatility effects in a manner similar to that of the Beveridge-Nelson (1981) decomposition of conditional mean ARMA models for an econometric time-series. The GARCH model evidence mean reversion in volatility to $\frac{\omega}{(1-\alpha-\beta)}$, which converges asymptotically on the unconditional variance, while the CGARCH model allows mean reversion to a time-varying long-run volatility level. The CGARCH model specification is as follows:

$$\sigma_t^2 = q_t + \alpha\left(\varepsilon_{t-1}^2 - q_{t-1}\right) + \beta\left(h_{t-1} - q_{t-1}\right)$$

$$q_t = \omega + \rho q_{t-1} + \phi\left(\varepsilon_{t-1}^2 - h_{t-1}\right)$$

in which q_t represents the long-run volatility, the past forecast error $\left(\varepsilon_{t-1}^2 - h_{t-1}\right)$ provides the driving force for the time-dependent movement of q_t, and the difference between the conditional variance and its long-run volatility $\left(h_{t-1} - q_{t-1}\right)$ defines the transitory component of volatility. The transitory component then converges to zero with powers of $(\alpha + \beta)$, whereas the permanent component converges on $\omega/(1-\rho)$ with powers of ρ. It is assumed $0 < \alpha + \beta < \rho < 1$ for the long-run volatility component to be more persistent than the short-run volatility component (Kang et al, 2009).

2.4. A review on outliers

Outlier detection is, in essence, a broad field and has thoroughly been studied in the perspective of a large number of application areas. Ismail (2009), Tesfamicael (2007), Carnero et al. (2007) and Verhoeven (2000) provide an extensively detailed outline of outlier detection techniques. Outlier detection has been studied in a wide range of data fields including high-dimensional data, uncertain data, streaming data, network data and time series data (Gupta et al., 2014). There are many methodologies used for detecting outliers with the aim to improve model efficiency and adequacy of statistical analyses. There are recent methods in time series analysis which comprise of an iterative process to identify the locations and types of outliers. This removes the effect the outliers have on the data and models the data until the estimated model is uncontaminated and "outlier-free" (Ismail, 2009).

In the past, not much focus was aimed at outlier detection, but rather consistency was made solely on the assumption of *iid* observation for detecting outliers. Fox (1972) introduced a study of outliers in time series data and was the first to define two types of outliers in an autoregressive model, where Fox (1972) was able to prove that the power of the test performed better than that of the *iid* assumption thus many researchers such as (e.g., Chen and Liu, 1993; Davies & Gather, 1993; Hadi & Simonoff, 1993; Ljung, 1993; Rocke & Woodruff, 1996; Penny, 1996) followed suit to formulate ARMA model to detect and model in the presence of outliers in a time series data. This made outlier discovery essential for accomplishing the source of factual statistical inference and has been a fascinating theme of various studies. Later, after the good breakthrough, many researchers extended their studies to other models such as ARIMA, ARCH, GARCH and other models (Zainol, et al., 2010).

Recently different studies have been proposed apart from the standard financial time series models. As a result, several models have been proposed in outlier detection of non-linear models.

Franses & Ghijsels (1999) demonstrated the AO by adapting the comparison of GARCH(1, 1) model as equal to that of the ARMA model for ε_t^2, with the idea proposed by Chen & Liu (1993) on the study of outlier detection in ARMA time series. The work of Franses & Ghijsels (1999) was later extended by Charles & Darne (2005) to detect the presence of the AO and IO in the GARCH models. Recently there are many types of outliers that have been examined in time series data, such as the AO and IO. These were the first to be studied. Later other types of outliers in time series were examined, namely temporary change (TC) or sometimes called as transient change and level change (LC) or be referred to as level shift (LS) which then became famous, and were thoroughly investigated by researchers. These types of outliers were classified by Tsay (1988) who has also introduced another type of outlier, known as variance change (VC). Later, Wu et al. (1993) defined a new type of outlier known as reallocation outlier (RO) (Zanoil et al., 2009).

On the other hand, little consideration has been given to remote perceptions in the standard GARCH. Among a couple of others, Van Dijk et al. (1999) and Franses & Ghijsels (1999) found that dismissing AOs dampens the tests and estimates of ARCH impacts. Before surveying the impacts of outliers on the ARCH and GARCH type models, the study characterise what is meant by anomalies in the time arrangement models.

Two noteworthy sorts of anomalies have been characterised by Fox (1972), one is known as the added substance of AO model; the other is referred as the innovation inconsistency IO model. An IO speaks to an unprecedented upset at time t impacting $y_t, y_{t+1}, \ldots$ through the dynamic framework depicted by the mathematical statement. In the IO model, irregular developments have bigger change than the greater part and along these lines, can show up as anomalies. In the AO model, the disconnected anomaly has an added substance transient character that is negligible to the time arrangement model (Carenero, 2007). In this manner the AO is likewise called a gross slip, subsequent to just the level of t^{th} perception is influenced. Truth be told, IO-type anomalies transmit their impact through to later perceptions while AO-type exceptions do not. Likewise taking note of the IO model will make a substantial tailed appropriation and ARCH model is overwhelming tailed. ARCH models have the capacity to catch IOs by development.

2.5. Overview of outliers

The subtleties and improvement of financial data are better when analysed utilising stochastic modeling approach that captures the time dependent structure altered in the time series market data (Lux, 2008). Various volatility estimators and models have been proposed in the literature to measure the volatility of stock returns. However, time series often contain observations caused by unexpected events, called interventions (observations are named with different names such as "contaminants", "outliers" and "extreme values") (Murek, 2014). One special characteristic of financial time series data is that there are observable high amplitude fluctuations at certain time points of the data. These relatively rare but vital occasions are caused by for example, stock market crash, political conflicts, news announcements, recession, decrease in equity prices, and changes in policy regimes which might have critical impacts on macroeconomic execution (Baker & Wurgler, 2000), and such extreme returns are often found to disturb volatility less than a standard GARCH model would forecast (Boudt et al., 2013).

Thus, according to Reider (2009), volatility not only spikes up during a crisis, but it eventually drops back to approximately the same level of volatility as before the crisis. It may be a single spike or a collection of spikes at a time interval depending on the coefficient of the time series process as observed in Figure 2.1 (Ismail, 2009). Therefore, in such instances, the utilization for standard GARCH models leads to an overestimation of the volatility for the times (days) following the spikes. Therefore, the unconditional volatility estimation will be upward biased and volatility forecasts might have a tendency to be bigger than it would be. A related argument applies to correlation estimates, where if only one of the stocks exhibit a large jump in prices, that biases the correlation estimates towards zero (Boudt et al., 2013).

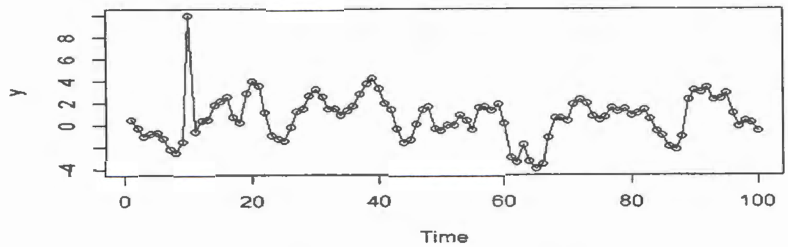


Figure 2.1: Outlier representation

Previously, researchers have concentrated a lot of studies on the issues of outliers and they often relied on the assumption of independent and identically distributed *(iid)* to address their analysis. This assumption was somehow not correct and reliable when analysing time series data. The assumption implied that the data have the same distribution (Tolvi, 2000). Those not well-matched observations are considered outliers (Zainol, 2010). However, if the time of and reasons for their occurrence are known, methods of intervention analysis can be applied to them. Thus, the occurrence of outliers can be verified by comparing the estimate of the parameter ω around $\hat{\omega}$, to its standard error. If the time of outlier occurrence is unknown, it is important to identify outliers and clean the time series from them as recommended by Murek (2014).

2.5.1. Types of outliers

Outlier detection has then become a vital measure of time series analysis as it influences modeling, testing and inference. Outliers can lead to model mis-specification, biased parameter estimation, poor forecasts and inappropriate decomposition of the series (Kaiser and Maravall, 1999). Literature behind outlier detection in time series was first familiarized by Fox (1972) with two types of outliers, namely the additive and innovation outliers. When the timing of outliers is known, they can be detected using tests of hypothesis. This means that outliers are detected by hypothesis testing on the residuals by portraying an outlier which is big at the starting point. The hypothesis testing proficiently identifies only data points that are dissimilar or varying with time series data. According to (Akouemo and Povinelli, 2014), as the outliers are removed from the data, parameter estimates are improved and consequently improves the results. Chang (1982) proposed an iterative procedure to detect multiple outliers when the timing of outliers is unknown. This procedure is described in Box and Tiao (1975) and in Ljung (1993). Tsay (1988) generalised this procedure to detect four types of outliers namely the additive, innovational, level and temporary change (Bui & Jun, 2012). There are several types of outliers being studied in financial time series namely: innovation outlier (IO), additive outlier AO, level shift (LS), temporary change (TC), local trend (LT) and seasonal level shift (SLS). This study, despite the different types of outliers, focused on four famous and commonly used types of outliers in time series namely the AO, IO, TC and LS outliers discussed in Fox (1972), Chen and Liu (1993) and Tsay (1988). The SLS, is a type of outlier that exhibits seasonal variation, and the LT outlier, is a type of outlier that yields a drift in the series caused by a pattern in the outliers after the onset of the initial outlier these two types are not applicable due to the nature of the financial time series data used.

Four basic main types of models have resulted in the theory of outliers, describing their most frequently occurring types. Namely, they are:

- Additive outlier
- Level Shift outlier
- Innovation outlier
- Transitory change

These types of outliers affect financial time series in different ways, therefore it can be noted that not all outliers will be present in the data since the effect of an AO, an LS or a TC on an observed financial series is independent of the ARMA, ARCH and GARCH model.

Thus, the effect of an IO on an observed series comprises on a preliminary shock that spreads in the observations that follow with the weights of the ARIMA, ARCH and GARCH type model (Kaiser and Maravall, 1999). Therefore, it can be seen in studies such as Balke and Fomby (1994) that the most occurring outlier in high frequency data is the IO (Charles, 2008). Furthermore, AO and TC outliers might be related with outliers influencing those unpredictable components (irregular trend), LS outliers might be connected with those trend-cycle components and, finally, IOs are associated with simultaneously influencing the trend-cycle and the seasonal segments (Kaiser and Maravall, 1999).

2.5.1.1. Additive Outlier (AO)

The AO is the least difficult and the most generally over researched in time. It is the added constituent inconsistency which is otherwise called Type I outlier (Fox, 1972). An AO just influences a solitary observation (Figure 2.2), which is either littler or bigger in worth contrasted with the normal qualities in the information. After this unsettling influence, the arrangement comes back to its ordinary way as though nothing has happened.

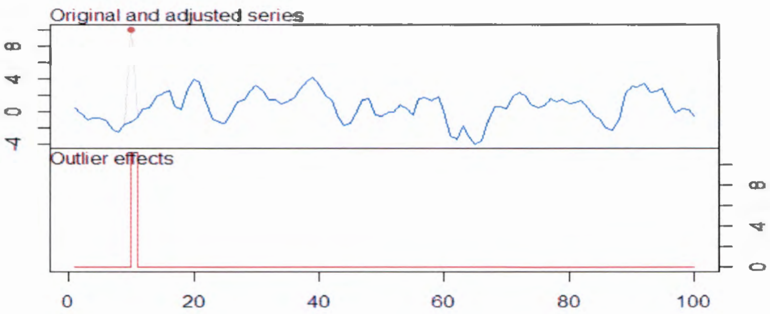


Figure 2.2: Additive outlier

Added substance variances are the most troublesome exclusions due to the fact that they expand two sequential residuals, one preceding the AO and the other after the AO. An added substance exclusion compares to an outside mistake or exogenous change of the watched estimation of the time arrangement at a specific time point. In a study conducted by (Tsay, 1988; Chen and Liu, 1993) the standard variance impact brought on by AO at once $t = T$, with the size of the impact is meant by ω is given by:

$$Z_t = X_t + \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)} \tag{2.11}$$

where $I_t^{(T)} = 1, t = T$ and $I_t^{(T)} = 0, t \neq T$. The true rule suggests that the shock caused by an AO affects the original observation at $t = T$ only with the magnitude of ω and the rest remained unaffected. AO affect the estimated residuals and the estimates of the parameters, it can be seen that large AOs push all the autocorrelation coefficients towards zero (Gounder et al., 2007).

2.5.1.2. Level Shift (LS) Outliers

LS outlier is a "step function" and permanent, that is, it can sometimes be called structural breaks. LS characterizes sudden movements in all observation at an observed period after approximately some time point t_0 (Figure 2.3). If one level shifts occur at time t_0, the observed series is $Z_t^* = Z_t + c$ for all $t \geq t_0$ with c being a constant. The model for a LS at time t can be defined as:

$$Z_t = X_t + \omega \frac{1}{1-B} I_t^{(T)} \tag{2.12}$$

where $I_t^{(T)} = 1, t = T$ and $I_t^{(T)} = 0, t \neq T$ (Tsay, 1988).

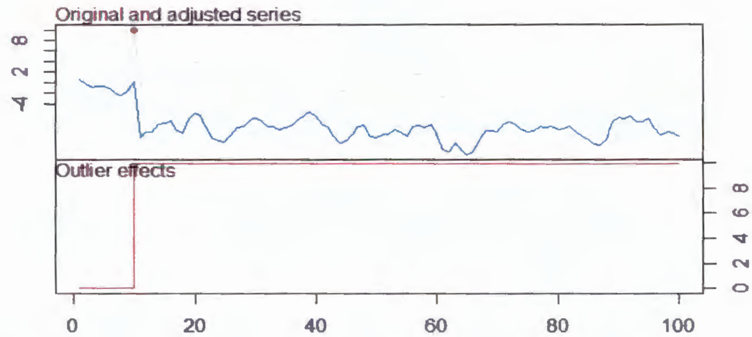


Figure 2.3: Level shift outlier

2.5.1.3.Innovation Outlier (IO)

Contrary to the AO, the IO also known as Type II outlier (Fox, 1972) disturbs numerous observations. An AO affects only one residual at the date of the outlier, and thus the effect of the IO on a detected series consists of a preliminary shock that spreads to the following observations with the weights of the *moving average* (MA) representation of the ARIMA model (Figure 2.4). Thus these weights are often volatile, where the influence of the IO may increase as the date of its event becomes more and more detached in the past, leading to a rather undesirable future.

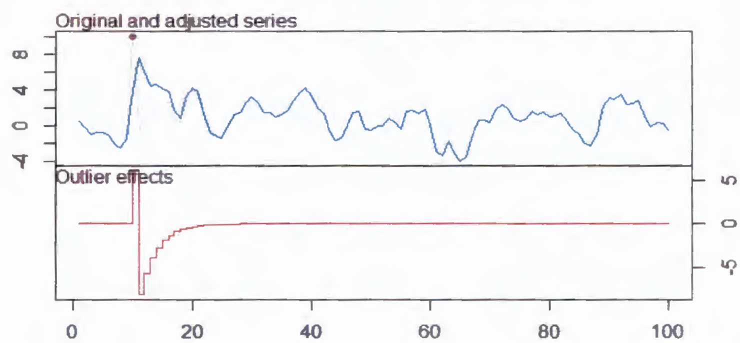


Figure 2.4: Innovation outlier

The effect of the IO depends on the particular model of the series, and for series with stationary change. Let Z_t be the observed series and X_t be an outlier free series. Assume $\{X_t\}$ follows ARMA (p,q) model:

$$\phi(B)X_t = \theta(B)a_t \tag{2.13}$$

where $\phi(B) = 1 - \phi_1 B - .... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - .... - \theta_q B^q$ are stationary and invertible operators. The IO is defined as:

$$Z_t = X_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} (a_t + \omega I_t^{(T)}) \tag{2.14}$$

It can be noted that an additive outlier affects only the T^{th} observation, that is, z_T, whereas an innovation outlier affects all observations z_T, z_{T+1}......., beyond time T through the recollection of the structure described by $\frac{\theta(B)}{\phi(B)}$. Generally, a time series containing, k outliers of different types, the general model is defined as:

$$Z_t = \sum_{j=1}^{k} \omega_j v_j(B) I_t^{(T)} + X_t \tag{2.15}$$

where $X_t = \frac{\theta(B)}{\phi(B)} a_t$, $v_j(B) = 1$, for AO and $v_j(B) = \frac{\theta(B)}{\phi(B)}$ for IO at $t = T_j$ (Grounder, 2007).

2.5.1.4. Transitory Change (TC) Outliers

The TC outlier is similar to the LS outlier and AOs, and it has been defined by Tsay (1986:1) as the "spike that takes few times to decrease exponentially over successive observations". The TC similar to AO causes a temporary preliminary impact on observations but the impact is passed on to the next observation (Grounder, 2007).

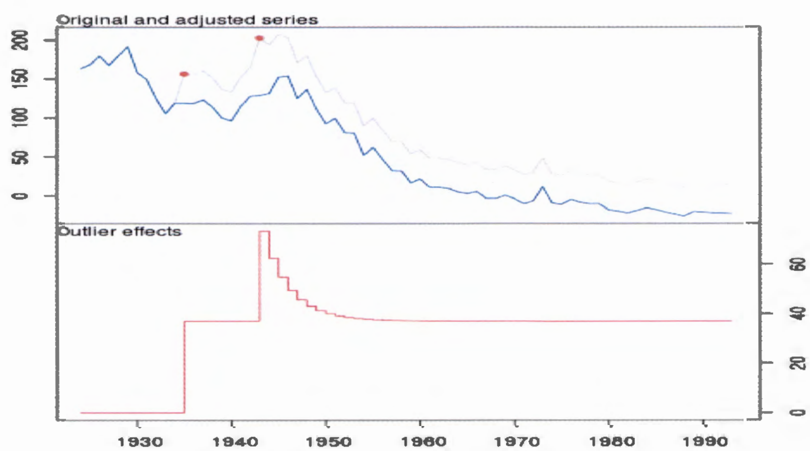


Figure 2.5: Transitory outlier

2.5.2. ARCH models and outliers

The ARCH model is a basic, yet helpful technique for displaying heteroscedasticity in time series. Initially proposed by Engle (1982), the model is well known in both theoretical and empirical literature, particularly in the financial markets. Literature about the subject can be found, for instance, in Bollerslev (1986) and Engle & Nelson (1994). Consider a stationary AR (p) model such that,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots \phi_\pi Y_{t-p} + \varepsilon_t \tag{2.16}$$

$$\varepsilon_t | \Omega_{t-1} \sim F(0, \sigma_t)$$

$$\sigma_t = \mu + \alpha_1 \varepsilon^2_{t-1} + \alpha_2 \varepsilon^2_{t-2} + \ldots \alpha_\tau \varepsilon_{t-r} + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \ldots + \beta_s h_{t-s} \tag{2.17}$$

where ε_t is an *iid* Gaussian error, and follows a GARCH (p, q) process. The conditional error distribution F is usually assumed to be normal. The least difficult test for ARCH effects in a series (residuals) is the Lagrange multiplier test of Engle (1982). This is probably the most broadly utilised test for ARCH, probably for the most part on account of its effortlessness and simplicity of utilisation Engle and Nelson (1994). ARCH models can represent some excess kurtosis. It has been noted, even subsequent to fitting an ARCH model to the sometime arrangement, the residuals are still not good with the suspicion of typically appropriated blunders, yet actually still have abundance kurtosis. This perception has prompted numerous proposed arrangements, one of which is to accept a heavier-tailed appropriation, for example, Student's t, for the mistakes. Another probability is to accept that exceptions are the purpose behind this overabundance kurtosis.

A level outlier is a simple AO in the level of the series that has no effect on the conditional variance but will affect the estimated model by inflating the estimated volatility. A volatility outlier, on the other hand, is passed on to the following observations and has a decaying effect on the volatility of the series through the conditional variance equation (2.13).

2.5.3. The nature of outliers in ARCH and GARCH type (p, q) model

In order to fully comprehend the development of GARCH model, the empirical properties have to be taken into account with respect to the price data. This section derives the formulation of two types of outliers (i.e. the AO and IO) by looking at the effects of outliers on observations, residuals, autocorrelation and on conditional heteroscedasticity.

2.5.3.1. Effects of outliers on observations

Suppose $Y^{*}_{t,AO,IO}$ are the observed values from GARCH (1, 1) process with an AO occurs at time point *t=d* with the magnitude ω and an IO occurs at time *t<d*, therefore $Y^{*}_{t,AO,IO} = Y_t$, $t = 1,\ldots,T$. Suppose that Y_t is the observation at time *t* attained if there were no outliers in the series would be referred to as "original observation". For *t<d*, $Y^{*}_{t,AO,IO} = Y_t$, and for $t \geq d$ the values will be given by the following formulation for AO:

$$Y^{*}_{t,AO} = \begin{cases} y_t + w & if \quad t = d \\ y_t & \quad t \neq d \end{cases}$$

The above equation proposes that shocks caused by AO disturb the original observation at *t=d* only with an extent ω and the rest remains unaffected. Therefore, the formulation of IO effects on observations is derived as follows:

For *t=d:*

$$h_t = \varpi + \sum_{j=1}^{q} \alpha_j \varepsilon^2_{t-j} + \sum_{i=1}^{p} \beta_i h_{t-i}$$

$$Y^{*}_{t,IO} = \sum_{i=1}^{q} \alpha_j e_{d-j} + e_d + \omega + \sum_{k=p+1}^{p} \beta_i Y^{*}_{d-k,IO} \tag{2.18}$$

$$= Y_d + \omega$$

$$= Y_d + \omega A_{0,IO} \text{ (Ismail, 2009).}$$

2.5.3.2.Effects of outliers on the sample autocorrelations of squared observations

This unit derives analytically the outcome of large AOs on the sample autocorrelations of squared observations caused by stationary processes that are either homoscedastic or heteroscedastic.

The series of interest, $y_t, t = 1,\ldots,T$, is a stationary series with a finite fourth-order moment which is "contaminated" from time τ forward by *k* consecutive outliers of the same size ω. Therefore the observed series is given by:

$$z_t = \begin{cases} y_t + w & if \quad t = \tau, \tau+1, \ldots \tau+k-1 \\ y_t & otherwise \end{cases} \tag{2.19}$$

The main focus of this study is based on additive outliers because the type of data of interest is used in the analysis is daily financial returns of JSE which are often considered to be uncorrelated. In this context, the traditional distinction between additive and innovative outliers is not relevant. On the other hand, it is important to distinguish whether an outlier affects future conditional variances or not. It is assumed that the AOs defined in (2.15) are LOs in which that they affect the level of the series but not the evolution of the underlying volatility (Hotta & Tsay, 1998). The autocorrelation of order $h, h \geq 1$, of the squared observations of the contaminated series in (2.18) is estimated by:

$$r(h) = \frac{\sum_{t=h+1}^{T} z_t^2 z_{t-h}^2 - \frac{T-h}{T^2}\left(\sum_{t=1}^{T} z_t^2\right)^2}{\sum_{t=1}^{T} z_t^4 - T^{-1}\left(\sum_{t=1}^{T} z_t^2\right)^2} \tag{2.20}$$

However, if the sample size, T, is large relative to the order of the estimated autocorrelation, *h*, the numerator of *r(h)* can be written as follows:

$$\sum_{t\in T(h)} y_t^2 y_{t-h}^2 + \sum_{i=0}^{h-1} (y_{\tau+i} + \omega)^2 y_{\tau+i}^2 - h + \sum_{i=h}^{k-1} (y_{\tau+i} + \omega)^2 (y_{\tau+i-h} + \omega)^2$$
$$+ \sum_{i=k} y_{\tau+i}^2 (y_{\tau+i-h} + \omega)^2 - T^{-1}\left[\sum_{t\in T(0)} y_t^2 + \sum_{i=0}^{k-1} (y_{\tau+i} + \omega)^2\right]^2 \tag{2.21}$$

where $T(s) = \{s+1, \ldots, \tau-1, \tau+k+j, \ldots, T\}$

When the order of the autocorrelation is less than the number of successive outliers, that is $h <$ k, then the third summation in (2.21) contains h terms which depend on ω^4. Therefore, equation (2.17) is equal to $(k-h-(k^2/T))\omega^4 + o(\omega^4)$. Moreover, if h $\leq$k, then the third summation in (2.17) disappears and the numerator of *r(h)* is equal to $-(k^2/T))\omega^4 + o(\omega^4)$. Then,

$$\lim_{\omega\to\infty} r(h) = \begin{cases} 1 - \frac{h}{k(1-\frac{k}{T})} & if \quad h < k \\ \frac{k}{k-T} & if \quad h \geq k \end{cases} \tag{2.22}$$

Therefore, if there is one single large outlier (k= 1) it will bias towards zero *r(h)* for all lags, but a set of *k* large successive outliers generates a positive *r(h)* for $h<k$ and zero for the others. Furthermore, for $h<k$ and large T, the autocorrelations will then follow a linear decay for large outliers (i.e. if two large consecutive outliers produce an autocorrelation of the squares of order one approximately equal to 0.5, all the others being close to zero). Thus, if a heteroscedastic series is contaminated by a huge single outlier, the detection of honest heteroscedasticity will become difficult. On the other hand, when a homoscedastic series is contaminated by several large consecutive outliers, the positive autocorrelations of squares generated by the outliers can be confused with conditional heteroscedasticity. Thus, it can be noted that the limits in (2.18) are valid regardless of whether the return y_t is a homoscedastic or heteroscedastic procedure (Carnero, 2007).

2.6. Gap identified and conclusion

From the literature collected above, it is clear and evident that in finance and financial economics, modeling volatility of returns is a vital factor in risk management, portfolio selection, policy making…etc. Since volatility is a measure of risk due to the large (small) absolute returns followed by large (small) absolute returns, there are times which show high (low) volatility. GARCH is the most commonly used model to forecast volatility and has been used by different authors. It is evident that the estimated residuals from the GARCH model are effective in revealing excess kurtosis. Therefore, one possible reason could be the inclusion of outliers. The study identified commonly used GARCH model for the purpose of identifying statistical methods and the variables used.

Many studies showed that the GARCH-t is the commonly used method for forecasting volatility. Few studies show that EGARCH is a suitable model for forecasting stock returns. Not many studies take into account volatility clustering where the high and low shocks give rise to unusual observations which may be due to external factors like strikes, policy shifts etc. Methods for dealing with returns that are subjected to such exogenous forces are sometimes ignored by researchers and hence few studies that analysed JSE returns in the absence of outliers are done in South Africa. Most of the studies reviewed analysed simulated data as opposed to real life stock returns. Therefore, it is evident that the application of the GARCH model on outlier free data has not been exhausted in financial time series in a South African context.

It is vital to understand the effects of outliers when forecasting volatility, the reason being that outliers affect the identification and estimation of GARCH models by hiding true conditional heteroscedasticity. Since the presence of outliers affects the accuracy of parameter estimation and volatility, many studies made use of the Monte Carlo simulation method to detect outliers. The point of interest for this technique has many advantages namely: it can be used in both multivariate and univariate instability, secondly it is great at fitting for identifying disengaged single/multiple outliers and also patches of outliers; Thirdly, the technique is not difficult and it is fast, which makes it an engaging device for utilization of practitioners; fourthly, it has a chance to be connected to a number of series. Also finally, it is dependable, since it detects not many false outliers.

Outlier detection methods are used in many countries for detecting different stock returns. This study reviewed relevant studies over the years that have used outlier detection methods on different stock returns, with the main aim of identifying the different statistical methods used on stock returns and also the different convenient and simple methods to detect outliers (i.e. the Monte Carlo simulation, the wavelet-based detection, regression centred outlier technique, robust outlier detection, the bootstrap-based procedure, the Diebold and Mariano (1995) tests etc.). However, even though GARCH is the most used technique, many techniques can be used. This is a clear indication that applications and theories about the methods and application of stock returns with the inclusion of outliers are still not exhausted.

It was shown that many studies since its introduction by Fox (1972) have tried to improve the methods of detecting outliers in time series data to obtain better forecasts. It is clear that the properties that financial time series exhibit (i.e. implied volatility, co-movements, volatility clustering, leverage effects…etc.) have a huge impact on modeling the series for future forecasts, thus neglecting outliers, to some extent have a huge disadvantage. Four basic and commonly used types of outliers are discussed, namely the AO, LO TC and IO. Furthermore, this section aimed at deriving the formulation of two types of outliers' effect on observations, residuals, sample autocorrelations and conditional heteroscedasticity from GARCH (1, 1) model.

The next chapter describes the methodology used to conduct the study.

CHAPTER THREE
METHODOLOGY

3.1. Introduction

Financial data has received significant attention in literature over the past two decades. Several models have been proposed for capturing special features of financial data. Most of these models are in possession of the conditional variance (or the conditional scaling) dependent on the past (Ngailo, 2011). One of the commonly known and most often used is the ARCH process which was introduced by Engle (1982). This section, introduces the theory behind ARCH (p), GARCH (p, q), TGARCH (p, q), GJR-GARCH (p, q) and APARCH (p, q) model. The study discusses the symmetric ARCH (q) and GARCH (p, q) and asymmetric TGARCH (p, q), GJR-GARCH (p, q), CGARCH (p, q) and APARCH (p, q) models contexts to model and estimate the South African stock exchange volatility. It compares the performance of the parametric models with outliers and without outliers.

The study makes use of a univariate GARCH type modeling approach which follows a straightforward comparison of volatility ARCH and GARCH type models with and without outliers. The issue of outliers has been thoroughly researched since it was proposed by Fox (1972) and has often produced conflicting results. A model testing approach for modeling volatility and financial market risk of JSE stock prices is used in this study.

This chapter consists of the following sections: Section 3.2 discusses ethical considerations, Section 3.3 entails the data used in this study and Section 3.4 discusses issues governing the ARCH and GARCH type models. Section 3.5 gives an overview of the proposed method. Diagnostic tests used on the selected model are discussed in Section 3.6. Section 3.7 provides the forecast error metrics used for selecting the best performing model. The last section gives a summary of the chapter.

3.2. Ethical considerations

The study did not have contact with either animal or human beings, as a result, ethical issues were not considered a challenge. The study followed guidelines as prescribed by the university which is outlined in the postgraduate manual. Ethical clearance to conduct the study was sought from the university through the supervisor(s).

3.3. Sampling Technique, Data Description and Source

Data used in this study consist of the daily closing prices of the primary South African indices the FTSE/JSE index obtained from the Johannesburg Stock Exchange (JSE) from January 3, 2011, to April 29, 2016, totaling 1330 observations. The data will be used to explore the recent change in the stock market since its upgraded trading systems in the 1900's. The daily prices are downloaded from http://www.jse.com. Hence, all data are converted to their daily natural log returns and are presented in percentages denoted as (r). Returns(r) are defined as the natural logarithm of price relatives multiplied by 100 to generate percentage changes in price. Multiplying will assist to reduce errors since the raw log returns denoted as r_t seem to be very small, hence the percentage log returns are calculated as:

$$\begin{aligned} r_t &= 100\nabla \ln X_t \\ &= 100(1-B)\ln X_t \\ &= 100(\ln X_t - \log X_{t-1}), \\ &= 100 * \ln\left(\frac{X_t}{X_{t-1}}\right) \end{aligned} \tag{3.1}$$

where X_t is the daily closing price of JSE index stock market at time t, and X_{t-1} denotes lagged stock price on day $t-1$, this is used to forecast ARCH and GARCH (p, q) type model.

This study presents time series plots and tables about stock prices. The tests conducted to check whether squared correlation exist on stock prices and also check for the presence of heteroscedasticity in stock prices with and without outliers. The tests will be used to apply ARCH and GARCH type models for forecasting volatility in a univariate setting.

When modeling a GARCH model; firstly the number of lags (p, q) for the symmetric GARCH with a normal distribution is chosen, secondly the study tests for particularly the GARCH (p, q) type models. Finally, when the best model is chosen, the GARCH model is tested and a validation process is done to check if correlation and heteroscedasticity have been removed by using a system validation process introduced by Ljung in 1993 (Nasstrom, 2003). As pointed out by Bera & Higgins (1993), most of the applied financial works show that GARCH (1, 1) provides a flexible and parsimonious approximation to the uncertain variance dynamics and is capable of representing the majority of financial series (Hou, 2013). Therefore, rejection of the unit root hypothesis is necessary to support stationarity.

3.4. Preliminary data analysis

Making inferences about a population is based on information within the sample and measuring the goodness of the inference in well-defined quantities used to summarise the data within the sample. These quantities are called numerical descriptive measures of a set of data. Two types of descriptive measures are measures of central tendency and measures of dispersion or variation (Wackerly et al., 2008:9). The mean is used to test the measure of the trend (description) as an important measure of volatility and the variance is used to check for the variability of the returns.

Moreover, skewness and kurtosis are used to check the symmetry of the returns. The descriptive statistics established an association in the interpretation of inspecting the characteristics of stock returns and portraying the significance behind the formation of predicting volatility (Makan et al., 2012). The measures of central tendency assist in summarising the returns in a single measurement and the measure of dispersion aids in checking the variability of the returns (Rice, 2007:401).

3.4.1. Testing for stationarity

Many theories of time series require stationarity, therefore it is imperative to determine whether a time series is stationary or not. Though many stationarity tests have been proposed in the literature, the Dickey-Fuller test by Dickey and Fuller (1979) and the KPSS test proposed by Kwiatkowski et al. (1992) are the most recommended (Forni, 2002). The purpose of stationarity is governed by the assumption that the stochastic properties of a series do not change with time. This basically implies that there must be a statistical balance. A series X_t is said to be stationary if the joint distribution of $X_{t1}, X_{t2}, \ldots, X_{tn}$ is the identical as the joint distribution of $X_{t1-k}, X_{t2-k}, \ldots, X_{tn-k}$ for all selections of time points $t_1, t_2, \ldots, t_n$ and all selections of time lag k. Thus, the series is second order stationary if the mean is constant for all t and if the covariance of t and k between X_t and X_{t+k} depends on the lag difference of k (Sigauke et al., 2014). The next subsections provide discussions for the ADF and KPSS tests.

3.4.2. Augmented Dickey-Fuller test (ADF)

In time-series regression, non-stationary variables on each other lead to possibly false inferences about the estimated parameters and the point of association. To test for stationarity the Augmented Dickey-Fuller (ADF) test was used. This is an augmented version of the

Dickey-Fuller test for data which is considered large and complicated for time series models. The testing for ADF is similar to the Dickey-Fuller and can be applied to the regression model:

$$\Delta x_i = \alpha_0 + \alpha x_{t-1} + \sum_{i=1}^{n} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.2}$$

where Δ is the difference operator, α_0 is the constant, α is the coefficient on a time trend and ϕ is the lag of the autoregressive process. The null hypothesis that the series is non-stationary is formulated as:

$H_0 : \phi = 1$ (Data is non-stationary or is a unit root)

$H_a : \phi < 1$ (Data is stationary or there is no unit root)

The ADF test statistic is given as:

$$\hat{\tau}_{ADF} = \frac{\hat{\phi}_1 - 1}{se(\hat{\phi}_1)} \tag{3.3}$$

The null hypothesis of a unit root $H_o : \phi_1 = 1$: is rejected if $\hat{\tau}_{ADF}$ is less than the appropriate critical value, at some level of significance, or if the observed probability of the test statistic is in excess of some conventional level of significance.

3.4.3. Kwiatkowski–Phillips–Schmidt–Shin (KPSS)

To avoid the drawback that ADF test always has a low power, Kwiatkowski et al. (1992) proposed an alternative unit root test where y_t is supposed to be (trend) stationary under the null hypothesis. The KPSS test is a Lagrange multiplier test and the test statistic can be calculated by initially regressing the response variable y_t on a constant and a time trend t based on the residuals from the Ordinary Least Squares (OLS) regression of y_t on the exogenous variables x_t. Furthermore, the test statistic is given by:

$$KPSS = \sum_{t=1}^{T} S_t^2 \Big/ s_\varepsilon^2 \tag{3.4}$$

where S_t is the partial sum of the error terms defined as $S_t = \sum_{j=1}^{t} \varepsilon_j$ and s_ε^2 is the estimated error variance from the regression. The null hypothesis that the series is non-stationary is formulated as:

$H_0 : \sigma^2 = 0$ (Data is non-stationary or is a unit root)

$H_a: \sigma^2 > 0$ (Data is stationary or there is no unit root)

For the assumption to be robust, using the unit root test and the stationary test jointly was found to be beneficial. The results of these two tests can be matched to see if the same conclusion is achieved. If there are contradictive results based on both ADF and KPSS tests, KPSS test is preferred due to the drawbacks of ADF tests.

3.4.4. Testing for Normality

Testing for normality is a skill full habit always examined to check if indeed the data is from a normal distribution or other distributions. There are several statistical methods that can be used to see if the data is from a Gaussian distribution. The Jarque – Bera (JB) is a common statistical test used to test for normality in time series.

3.4.4.1. Skewness and kurtosis

Skewness, which is known as the third 'moment' of a return distribution, measures the degree of asymmetry of a distribution around its mean. It is a measurement of the lack of symmetry, where positive skewness indicates a distribution with an asymmetric tail extending toward more positive values (long right tails). Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values (long left tails) (Li, 2013). Skewness is defined as

$$\frac{E(X-\mu)^3}{Var(X)^{\frac{3}{2}}} \tag{3.5}$$

where μ and $\sqrt{Var\ (X)}$ are the mean and standard deviation of X.

Kurtosis is known as the fourth moment of the return distribution which measures the degree to which a distribution is more or less peaked than a normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution. A normal distribution has a kurtosis of 3. Therefore, stock characterised by high kurtosis has 'fat tails' at extreme negative and positive ends of the distribution curve. Kurtosis is interpreted as follows:

- Kurtosis > 3 - Leptokurtic distribution, sharper than a normal distribution, with values concentrated around the mean and thicker tails. This means high probability for extreme values.

- Kurtosis < 3 - Platykurtic distribution, flatter than a normal distribution with a wider peak. The probability for extreme values is less than for a normal distribution, and the values are wider spread around the mean.
- Kurtosis = 3 - Mesokurtic distribution - normal distribution for example (Nasstrom, 2003). Kurtosis is described by the equation:

$$\frac{E(X-\mu)^4}{Var(X)^2} \tag{3.6}$$

3.4.4.2.Jarque-Bera statistics

This assumption refers to the degree to which data follows a normal distribution. The Jarque-Bera (JB) statistic is used to test whether a set of closing prices are normally distributed which is a type of Lagrange Multiplier test to test for normality, heteroscedasticity and serial autocorrelation or autocorrelation of regression residuals (Jarque and Bera, 1980). The JB test for normality comes from a Chi-square distribution which is calculated using the skewness and kurtosis with 2 degrees of freedom given as:

$H_0 : E(\varepsilon^2)=0$ (Residuals are normally distributed)

$H_a : E(\varepsilon^2)\neq 0$ (Non-normal distribution)

$$JB = N * \left[\frac{skewness^2}{6} + \frac{(kurtosis-3)^2}{24}\right] \sim \chi^2 \tag{3.7}$$

where N is the number of observations of the closing prices. The rejection region is calculated if the JB-test statistics is greater than the critical value the null hypothesis of the normality is rejected (Li, 2013).

The feeling about this test is that the larger the *JB*-value is, the lower the probability that the given series is drawn from a normal distribution. The test statistic of the Jarque-Bera test is χ^2-distributed with 2 degrees of freedom under the null hypothesis that the series is normally distributed (Hamann 2001).

3.5. Estimation of ARMA (p, q) model

The $ARMA\ (p,q)$ refers to the model with p autoregressive terms and q moving average terms. The model contains $AR\ (p)$ and $MA\ (q)$ models noted in equation (2.2), where c is a constant and ε_t is a sequence of independent random variables with mean 0 and variance σ^2

3.6. Estimation of ARCH (1) model

The ARCH model clearly recognises the type of temporal dependence in stock returns (i.e. volatility clustering). It assumes that the conditional error distribution is normally distributed, but with conditional variance equal to a linear function of historical squared errors. Therefore, it is evident that the stock returns exhibit extreme values that are followed by other extreme values, but with unpredictable signs making the unconditional error distribution in the ARCH model to be leptokurtic. This accounts for the observed unconditional kurtosis in the data. One of the most essential issues before applying the ARCH methodology is to first examine the residuals for evidence of heteroscedasticity. The ARCH model constructs the conditional variance σ_t^2 of asset returns with maximum likelihood method (Ngailo, 2011).

The time varying volatility is taken by letting volatility depend on the lagged values of the innovation terms ε_t and q selected such that the residuals of the variance equation are white noise (Makhwiting et al., 2014). If the ARCH model is based on the assumption of normality made on the residual ε_t by z_t the model makes use of the Maximum Likelihood Estimation (MLE). Let $r_1, r_2, .., r_t$ be a recognition from ARCH (1) process, the likelihood estimate of the data is the product of the conditional joint probability written as:

$$\begin{aligned} Likelihood\ f(r_1 \ldots, r_t|\theta) &= \prod_{t=2}^{T} f(r_t|r_{t-1}) f(r_1|\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-r_T-\mu}{2\sigma^2}\right)^2 * \ldots * \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-r_T-\mu}{2\sigma^2}\right)^2 \end{aligned} \tag{3.8}$$

where $\theta = (\varpi, \alpha_j)'$.

Since the ln L function is monotonically increasing function of L, we can maximise the log of the likelihood function,

$$\ln L(\alpha_0,\alpha_1,\beta_1,\mu|r_1\ldots r_T)=\frac{T}{2}\ln 2\pi-\frac{1}{2}\sum_{i=1}^{T}\ln\sigma_i^2-\frac{1}{2}\sum_{i=1}^{T}\left(\frac{-(r_T-\mu)^2}{\sigma^2}\right)$$
$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-r_T-\mu}{2\sigma^2}\right)^2*\ldots*\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-r_T-\mu}{2\sigma^2}\right)^2 \tag{3.9}$$

3.7. Estimation of GARCH (1, 1) model

Similar to the ARCH estimation the MLE is applicable in estimating parameters for the GARCH model. In GARCH (1, 1) model estimation, the initial values of the squared returns and past conditional variances are needed to estimate the parameters of the model. It can be seen that the most commonly used model in practice is the GARCH (1, 1) model. This model is considered the best because it is parsimonious and is proven to be adequate because it captures the volatility clustering in the data without the requirement of higher order models. The GARCH (1, 1) is specified as:

$$\sigma_t^2=\omega+\alpha_1\varepsilon_{t-1}^2+\beta_1\sigma_{t-1}^2 \tag{3.10}$$

(Ngailo, 2011).

3.8. Estimation of a GJR-GARCH (1, 1) model

The GJR (1, 1) model is a GARCH formulation that includes an additional term that captures asymmetry. The GJR (1, 1) model can be represented as:

$$\sigma_t^2=\omega+\alpha\varepsilon_{t-1}^2+\beta\sigma_{t-1}^2+\gamma\varepsilon_{t-1}^2\Gamma_{t-1} \tag{3.11}$$

where $\gamma,\alpha,\ \beta$ and ω are parameters. The coefficients of GARCH (1, 1), EGARCH (1, 1) and GJR (1, 1) models are estimated by the ML procedure, since $r_t \sim N(0,1)$ then the conditional distribution of y_t is normal:

$$f(\psi_{t-1}|r_1\ldots r_T)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-r_T-\mu}{2\sigma^2}\right)^2 \tag{3.12}$$

$$L(\vartheta|)=\frac{T}{2}\ln 2\pi-\frac{1}{2}\sum_{i=1}^{T}\ln\sigma_i^2-\frac{1}{2}\sum_{i=1}^{T}\left(\frac{-(r_T-\mu)^2}{\sigma^2}\right)$$ (Brooks & Chong, 2001).

3.9. Estimation of an EGARCH (1, 1) model

The conditional variance equation is introduced to the EGARCH model $\log\sigma_t^2$ Therefore an EGARCH (1, 1) model can be represented as:

$$\log\sigma_t^2 = \omega + \beta\log\sigma_{t-1}^2 + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \rho_i\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) \tag{3.13}$$

Where the coefficient γ captures the asymmetry or the leverage effect. When $\gamma = 0$ the model is symmetric. When $\gamma < 0$ negative shocks have a greater impact on volatility than negative shocks. When $\gamma > 0$ then positive innovations are more weakening than negative innovations. Since the conditional variance is in logarithmic form implies that the leverage effect is exponential. Thus if the conditional variance is negative the exponent in the EGARCH model will always make the conditional variance positive even if the coefficients are negative. β parameter is the persistence in captured conditional volatility (Makhwiting et al., 2014).

3.10. Estimation of a TGARCH (1, 1) model

The TGARCH (1, 1) model conditional variance specification is given as:

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta_1\sigma_{t-1}^2 \tag{3.14}$$

Thus to express a TGARCH (1, 1) model, positive news has an impact of γ_1, negative news has an impact of $\alpha_1 + \gamma$, with negative (positive) news having a greater effect on volatility if $\gamma > 0$ ($\gamma < 0$). Therefore, the coefficient γ is known as the asymmetric term, therefore when γ = 0 the model becomes a standard GARCH model (Makhwiting et al., 2014).

3.11. Estimation of an APARCH (1, 1) model

APARCH represents a general class of models that include both ARCH and GARCH models. When modeling apart from leptokurtic returns, the APARCH model, as the GARCH model, captures other stylized facts in financial time series, like volatility clustering. The volatility is more probable to be high at time *t* if it was also high at time *t-1*. An alternative way of seeing this is noting that a shock at time *t-1* also influences the variance at time *t*. As stated, the APARCH model also provides the long-memory property of returns discussed in Ding, Granger, and Engle (1993). The APARCH model, as the GJR-GARCH model, additionally captures asymmetry in return volatility. That is, volatility tends to rise more when returns are negative, as compared to positive returns of the same magnitude.The APARCH (1, 1) model is specified as:

$$\sigma_t^{\delta} = \omega + \alpha\left(\left|\varepsilon_{t-1}\right| - \gamma\varepsilon_{t-1}\right)^{\delta} + \beta\sigma_{t-1}^{\delta} \tag{3.15}$$

3.12. Estimation of a CGARCH (1, 1) model

The CGARCH model allows both a long-run component of conditional variance, q_t, which is slowly mean regressive and a short-run component, h_t–q_t that is further volatile. When modeling volatility the long-run component is a newly important determinant of the conditional variance than the short run component (Guo and Neely, 2008).

The CGARCH (1, 1) model of Engle and Lee's (1999) specification is given as:

$$\begin{aligned} \sigma_t - q_t &= \alpha_1\left(\varepsilon_{t-1}^2 - q_{t-1}\right) + \beta_1\left(h_{t-1} - q_{t-1}\right) \\ q_t &= \alpha_0 + \rho q_{t-1} + \phi\left(\varepsilon_{t-1} - h_{t-1}\right) \end{aligned} \tag{3.16}$$

Where q_t denotes the long run volatility, $\varepsilon_{t-1} - h_{t-1}$ provides for the time dependent movement of q_t, the transitory component of volatility is given by $h_{t-1} - q_{t-1}$ which is the difference between the conditional variance and its long run volatility. The CGARCH model must adjust to the stationary non negativity conditions when $\rho < 1$ and $(\alpha + \beta) < 1$, at this time the conditional variance in a long run will exist (Samuelson Hong:143, 2013).

3.13. Innovation of residuals

This study presents an extension of the ARCH and GARCH type model to allow for conditionally t-distributed errors and normally distributed (Gaussian errors) (Bollerslev, 1986). Thus, deliberation for applying the innovations differently from the actual model is to improve the accuracy of the parameter estimates of the model.

3.13.1. Normal distribution

The normal distribution is the commonly used distribution because of its simplicity. It assumes that r_t comes from a normal distribution $r_t \sim N(0,1)$. The returns r_t are said to be standardised (normally distributed) if the probability density function (pdf) takes the form:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{z^2/2} \qquad -\infty < z < \infty. \tag{3.16}$$

Under a Gaussian standard normality assumption, equation (3.16) becomes $z_t|\Omega_{t-1} \sim N(0,1)$, the Gaussian GARCH model considers volatility clustering, where the number of very highs and lows of returns observed suggests a flatter- tailed distribution that characterise the error process for South African stock market returns (Su, 2010).

3.13.2. Students t-distribution

The students t-distribution is another test used when the model exhibits flatter tails (leptokurtic). The pdf of the students'- distribution $(z_t \sim ST(0,1,\nu))$ is defined as:

$$f(z,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{z^2}{\nu}\right)^{-\left(\frac{\nu+i}{2}\right)} \qquad -\infty < z < \infty, \tag{3.17}$$

where $\nu > 0$ is the degree of freedom, $2 < \nu \leq \infty$ and $\Gamma(\cdot)$ is the gamma function which copies the Cauchy distribution. When $\nu \rightarrow \infty$ it indicates that the student t-distribution almost equal the normal distribution. Thus, the lower the v, the heavier the tails, so the student t-distribution reflects the fat-tail of the volatility index more exactly. Therefore this distribution may be an improved approximation to the exact distribution since they agree to excess kurtosis which is superior to that of a normal distribution.

3.13.3. GED distribution

The Generalized Error Distribution (GED) is a 3 parameter distribution belonging to the exponential family with conditional density given by,

$$f(x) = \frac{\kappa e^{-0.5\left|\frac{x-\alpha}{\beta}\right|^{\kappa}}}{2^{1+\kappa^{-1}}\beta\Gamma(\kappa^{-1})} \tag{3.18}$$

with α, β and κ representing the location, scale and shape parameters. Since the distribution is symmetric and unimodal the location parameter is also the mode, median and mean of the distribution (i.e. μ). By symmetry, all odd moments beyond the mean are zero. The variance and kurtosis are given by,

$$\mathrm{var}(x) = \beta^2 2^{2/\kappa} \frac{\Gamma(3\kappa^{-1})}{\Gamma(\kappa^{-1})}$$
$$Ku(x) = \frac{\Gamma(5\kappa^{-1})\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})\Gamma(3\kappa^{-1})} \tag{3.19}$$

The log likelihood function for GARCH with GED residuals is expressed as:

$$L = T\{\log(v/\lambda) - (1+v^{-1})\log(2) - \log(\Gamma(1/v))\}$$
$$-(1/2)\sum_{t=1}^{T}\left|\left(y_t - x_t\beta\right)\Big/\lambda\sqrt{h_t}\right|^{v} - \left(\tfrac{1}{2}\right)\sum_{t=1}^{T}\log(h_t) \tag{3.20}$$

where T is the sample size, $\Gamma(.)$ is the gamma function, λ is a constant given by:

$$\lambda = \left\{\frac{2^{(-2/v)}\Gamma\left(1/v\right)}{\Gamma\left(3/v\right)}\right\}^{1/2} \tag{3.21}$$

and v is a positive parameter governing the thickness of the tails of the distribution. Note that for $v = 2$, constant $\lambda = 1$, and the GED is the standard normal distribution (Ghalanos, 2015).

3.13.4. Skewed normal

Let X be a skew normal random variable. Its probability density function is given by:

$$f(x) = 2\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right) \tag{3.22}$$

where μ, σ and λ are the locations, scale and shape parameters respectively. Specifically, when $\mu = 0$ and $\sigma = 1$ it becomes a standard skew normally distributed variable. The skew normal distribution family is well known for modeling and analysing skewed data. It is a distribution family that extends the normal distribution family by adding a shape parameter to regulate the skewness, which has the higher flexibility in fitting a real data where there is some skewness existing (Ngunkeng, 2013).

3.13.5. Skewed student –t

The skewed student-t distribution captures the Skewness and kurtosis in the stock returns and is important in financial applications in many respects (in asset pricing models, portfolio selection, option pricing theory or Value-at-Risk among others).

For a Standardized (zero mean and unit variance) Skewed Student, the log-likelihood is:

$$L_T = \ln\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \ln\left(\Gamma\frac{v}{2}\right) - 0.5\ln[\pi(v-2)] + \ln\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \ln(s)$$

$$-0.5\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (1+v)\ln\left(1+\frac{az_t+m}{v-2}\xi^{-I_t}\right)\right] \tag{3.23}$$

where ξ is the asymmetry parameter, v the degree of freedom of the distribution, $\Gamma(.)$ the gamma function, $I^t = \begin{cases} 1 \text{ if } z_t \geq -\frac{m}{a} \\ -1 \text{ if } z_t < -\frac{m}{a} \end{cases}$, $m = \frac{\Gamma\left(\frac{v+1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right)$ and

$a = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}$ (Peters, 2001).

3.14. Model Selection Criteria

In order to use the time series models to forecast volatility, the first step is to fit the data into the time series model, which is done by estimating the unknown parameters of the model. This study evaluates the unknown parameters for the form of GARCH model which was best chosen by the information criteria. The method is represented briefly as: $y_t = \varepsilon_t = e_t\sqrt{h_t}$ and $y_t|\psi_{t-1} \sim N(0,h_t)$ this implies that the error term e_t is independent and *iid* with standard normal distribution (i.e. $\varepsilon_t|\psi_{t-1} \sim N(0,h_t)$, so the conditional function of ε_t is given by:

$$f(\varepsilon_t|\psi_{t-1}) = \frac{1}{\sqrt{2\pi h_t}}\exp\left[\frac{-(\varepsilon_t)^2}{2h_t}\right]. \tag{3.24}$$

$$\textit{Likelihood } f(\varepsilon_1 \ldots, \varepsilon_n|\psi_{t-1}) = \prod_{t=1}^{n} f(\varepsilon_n|\psi_{t-1})$$

$$= \frac{1}{(2\pi)^n}\exp\left(\frac{-(\varepsilon_t)^2}{2h_t}\right) * \frac{1}{\sqrt{h_t}} * \ldots * \frac{1}{(2\pi)^n}\exp\left(\frac{-(\varepsilon_t)^2}{2h_t}\right) * \frac{1}{\sqrt{h_t}},$$

where h_t is the conditional variance and is the function of the unknown parameters. Taking log on both sides will yield the following equation:

$$l = -\frac{1}{2}\sum_{i=0}^{n}\left[\log 2\pi + \log h_t + \frac{\varepsilon_i^2}{h_t}\right]. \tag{3.25}$$

The log likelihood is then maximised to estimate the unknown parameters involved (Li, 2013).

The strategy in choosing the appropriate model from competing models is based on the Akaike information criterion (AIC); the Bayesian information criterion (BIC); standard error (SE) and on the significance tests. The idea is to have a parsimonious model that captures as much difference in the data as possible. Usually, the simple GARCH model captures most of the variability in most stabilised series having small lags for p and q. The criteria used in selecting the best model for this study is, "the smaller the AIC and SBC the better the model". According to Bollerslev et al. (1992) and Edward (2011), $GARCH\ (2,1), GARCH\ (1,2), GARCH\ (1,1)$ are the best models. All the models are fitted as an affirmative measure to ensure that the best model is chosen, but the most parsimonious model that can capture conditional variance is chosen.

3.14.1. Akaike Information Criterion (AIC)

Order selection is very important when using ARMA process. As stated before, the higher the order in the model, the smaller estimated errors result, but the higher the order model the more complex the model is (Yang & Yang, 2007). The AIC makes adjustments to the likelihood function to account for the number of parameters (P). The AIC is given by:

$$AIC\ (P) = \ 2ln\ (maximum\ likelihood) - \ 2P, \tag{3.26}$$

where P is the number of parameters in the fitted model and L is the value of the maximised likelihood-function for the model. Assuming the model errors have the same variance, then

$$AIC(P) = 2P + N\left[\log(2\pi\sum_{i=1}^{N}\frac{\varepsilon_i^2}{(N-1)}\right]. \tag{3.27}$$

Applying the model into forecasting, the mean squared error of the forecasts will be large, which depends on errors from the estimation of the parameters of the fitted model.

3.14.2. The Bayesian information criterion (BIC)

The BIC or SBC is another powerful technique for selecting the best model and is also related to the Bayes factor. The BIC model is defined as:

$$BIC\ (P) = -2ln\ (likelihood) + (k + k\ ln\ N, \tag{3.28}$$

where k denotes the number of parameters and N denotes the number of observations. What makes the BIC superior is that it penalises more complex models with many parameters relative to simpler models, thus allowing many models to be compared simultaneously. The model with the highest subsequent probability is the one that minimises BIC, a desirable model is one that minimises the AIC or the BIC (Ngailo, 2011).

3.14.3. Hannan-Quinn Information Criterion (HQIC)

A criterion, which was also designed for strong consistency, the Hannan-Quinn Information Criterion (Hannan and Quinn, 1979), was specifically developed in the context of order selection for autoregressive models. It can be calculated as:

$$HQIC = n \ln \hat{\sigma}^2 + 2p \ln \ln n \tag{3.29}$$

HQIC (Hannan and Quin, 1979) and SIC (Schwarz, 1978) statistics perform the same function. For the Information Criteria statistics, smaller values are looking forward to because they indicate relatively less lost information and, hence, better models (Bierens, 2004).

3.15. Model Robustness

Testing the goodness of fit of the ARCH-GARCH model is formulated on the residuals and also by means of the standardised residuals. The residuals are assumed to be *iid* indicating that they either follow a normal or standardised t-distribution (Tsay, 2002). The use for evaluation of model robustness are discussed in subsequent sections.

3.15.1. Testing for statistical independence

The assumption of returns being identically distributed confirms that the mean and variance of returns do not vary over time, which is a consistent assumption in time series data. The independence assumption ensures that projected value changes are dissimilar to each other at any point in time (Abraham et al, 2002).

3.15.1.1. Runs test

The runs test and turning point test are two widely used tests for independence (Cromwell et al. 1994). This study tests independence using Runs test for testing independence of the volatility of stock price. Runs test is a non-parametric statistical technique used to determine whether or not successive price changes are independent. Contrasting parametric techniques like the serial correlation independence test which is equivalent to the runs test has traditionally been used to test independence but the runs test does not require returns to be normally distributed (SAS support, 2013). A run is a sequence of consecutive price changes with the same sign. Furthermore, if the return series show larger tendency of change in one direction, the average run will be longer and the number of runs will be fewer than that generated by a random process.

The runs test can also be designed to count the direction of change from any base; for instance, a positive change could be one in which the return is greater than the sample mean, a negative change one in which the return is less than the mean, and zero change representing a change equal to the mean. The actual runs (R) are then counted and compared to the expected number of runs (m) under the assumption of independence with null hypothesis given as follows:

H_0 :The r_t is independent

H_a :The r_t is not independent

$$m = \frac{\left[N(N+1) - \sum_{i=1}^{3} n_i^2 \right]}{N}, \tag{3.30}$$

where N is the overall number of return observations and n_i is a sum total of price change in every classification. For a large amount of observations$(N > 30)$, m just about compares to a normal distribution with a standard error (σ_m) of runs specified as:

$$\sigma_m = \left[\sum_{i=1}^{3} n_i^2 \left\{ \sum_{i=1}^{3} n_i^2 + N(N+1) \right\} - 2N \sum_{i=1}^{3} n_i^3 - N^3 \right]^{1/2}. \tag{3.31}$$

The standard normal Z-statistic (Z= (R−m)/ σ_m) can be used to test whether the actual number of runs is consistent with the independence hypothesis. When the actual number of runs exceeds the expected runs, a positive Z value is found. A positive Z value indicates negative serial correlation in the return series (Abraham et al, 2002).

3.15.2. Testing for normality histogram

To test for normality, plots of the residuals are taken into account by means of histograms, the normal probability plot and the time plot of the residuals. A model that perfectly fits the data highlights a histogram of residuals which is approximately symmetrical and the normal probability plot should be a straight line, approximately 45°, while the time plot should exhibit random variation with no visible pattern.

3.15.3. Test for autocorrelation and heteroscedasticity

A standout amongst the most key issues before applying the ARCH methodology is to first analyse the residuals for evidence of heteroscedasticity. To test for the occurrence of heteroscedasticity in residuals of JSE index return series, the Lagrange Multiplier (LM) test for ARCH effects proposed by Engle (1982) is applied.

3.15.4. Ljung-Box q test for autocorrelation

This study will make use of the Engle's ARCH test using the LM in order to evaluate the significance of the ARCH effects. Therefore, the conditional heteroscedasticity in a variance is equal to the autocorrelation in the squared innovation process, where the residual series is given by: $e_t = r_t - \hat{\mu}_t$, where $\hat{\mu}_t$ is the conditional mean of the process and e_t is the innovation under the null hypothesis for autocorrelation of squared residuals given by the regression is:

$$H_0 : \alpha_0 = \alpha_1 = \ldots = \alpha_m = 0$$

$H_0 : \alpha_k \neq 0$ for at least one k

A formal test for white noise is found in the Ljung-Box q (LBQ) (1978), this test is performed to test whether or not a series has significant autocorrelation. The LBQ-value is given by:

$$Q_k = T(T+2)\sum_{i=1}^{k} \frac{r_i^2}{T-i} \sim \chi_{\alpha}^2 (m-j) \tag{3.32}$$

where T is the number of samples, k is the number of lags and r_i the i^{th} autocorrelation, Qk is 2 with k degrees of freedom. A large Q_k indicates that the probability process has uncorrelated data decreases (Nasstrom, 2003).

The LB statistic, or Q statistic, is used to determine whether the joint hypothesis (of a group of autocorrelation) is simultaneously significantly different from zero. The statistics can also be used in an informative manner to determine whether the residuals of an ARCH (*p, q*) behave as a white noise, or not. The general idea is that if the LBQ statistic(s) is not significantly different from zero, then this is a sign that the estimated model may "fit" the data well (Enders, 2010), as there may not be any more information in the series to model (Villegas, 2012). The LBQ statistic is given by:

$$\hat{\varepsilon}_t^2 = \rho_0 + \sum_{i=0}^{m} \rho_i \varepsilon_{t-i}^2 + \phi_t \tag{3.33}$$

where $\rho_0, \rho_1, \ldots, \rho_m$ are parameters and ϕ_t is *iid* random variable with mean 0 and variance σ_ε^2. In summary, the test procedure is performed by first obtaining the residuals e_t from the ordinary least squares regression of the conditional mean equation

The null hypothesis that there is no ARCH effect up to order q can be formulated as:

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_q = 0$$
$$H_a : \rho_i > 0 \quad \text{for at least one } i = 1,2,\ldots,q$$

The test statistic for the joint significance of the q-lagged squared residuals is the number of observations times the R-squared from the regression. The total R-square is evaluated against the Q- distribution. This is an asymptotically locally most powerful test. Firstly the ARMA and ARCH model is employed, which is the conditional mean in the return series as an initial regression. Lastly, test the null hypothesis that there are no ARCH effects in the residual series from lag 1 up to lag 12. The ARCH-LM test provides solid confirmation for rejecting the null hypothesis for all lags included. Rejecting H_0 is an indication of the presence of ARCH effects in the residuals series and therefore the variance of the return series of JSE index will be non-constant for all periods indicated.

3.16. Model performance evaluation

This study uses four evaluation measures to evaluate the forecast accuracy of JSE top 40 model for the proposed model ARCH and GARCH applied, namely the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) statistics. To estimate the forecasting performance of the methods used, one should define error functions using equations:

NWU

$$\mathrm{RMSE}=\sqrt{\frac{1}{I}\sum_{i=1}^{I}(\hat{\sigma}_i^2-\sigma_i^2)^2}\,, \tag{3.34}$$

$$\mathrm{MAE}=\frac{1}{I-(I_1-1)}\sum_{i=1}^{I}\left|\sigma_i^2-\hat{\sigma}_i^2\right|, \tag{3.35}$$

$$\mathrm{MAPE}=\frac{100}{I-(I_1-1)}\sum_{i=1}^{I}\left|\frac{\sigma_i^2-\hat{\sigma}_i^2}{\sigma_i^2}\right|, \tag{3.36}$$

where σ_t is the observed volatility (absolute value of returns) on a day t, $\hat{\sigma}_t$ is a forecast of the volatility and I is the number of days in a given dataset (both in-sample and out-of-sample) (Ladokhin, 2009).

The RMSE is the most common measure to test the forecasting power of the model, and together with the MAE measure how far the forecast is from the actual data. Therefore, the forecasting measure with the smallest RMSE and MAE is the one that gives best results/ forecasting power (Brooks, 2008). The error of both the MAE and RMSE are assumed to be unbiased and to follow a normal distribution. An advantage of RMSE is that it is sensitive to large errors which are different from the three other criteria. Thus it can be a helpful criterion when there are large estimate errors that could lead to serious problems. On the contrary, the MAE criterion severely deals with large errors less than the RMSE due to the fact that it does not square the errors. Moreover, one disadvantage of RMSE is that it can hold-up if there are large errors that could not lead to serious problems (Kgosietsile, 2015).

3.17. Outliers

An outlier is known as an extreme observation. Time series data consists of three types of observations a) ordinary observations which are realised the most frequently; b) extreme observations which are observed by the tails of the distribution usually occur frequently and clustered and c) outlying observations (Verhoeven, 2000). In a study conducted by Peters (1996) as cited in Storch et al., (2013), the variance was found only to be stable and finite for normal distributions. This finding led to the question whether or not the dataset under review is under the influence of extreme values. A study conducted by Karunanayake et al. (2012) showed that a crisis might significantly increase market volatilities and shifts to the level of a time series that cannot be explained are referred to as outliers. These observations are inconsistent with the remainder of the series and can dramatically influence the analysis, and consequently, affect the forecasting ability of the time series model. The next sections provide discussions about methods used for detecting outliers in a dataset.

3.18. General outlier detection in ARCH and GARCH models

Balke & Fomby (1994) discussed outlier detection in time series data by applying the following steps: namely outlier detection, model estimation, model diagnostics, and forecasting. If the date and type of the outlier are known, one can model the irregular component in the form of an intervention model (Box & Tiao, 1975) in which estimates of ω can be obtained from the coefficients on the intervention dummies. In this study, once possible outliers or level shifts were identified and, an intervention model was used to estimate ω. Before estimation of this intervention model, the timing of outliers was determined. To determine the existence of outliers, the study uses the outlier detection method described by Tsay (1988). Define $y_t = \left(\phi(L)/\theta(L)\right)Y_t$ which is equivalent to the ARMA residuals under the null hypothesis of no outliers and also define:

$$\pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \ldots = \phi(L)/\theta(L) \text{ and} \tag{3.37}$$

$$\eta(L) = 1 - \pi_1 L - \pi_2 L^2 - \ldots = \pi(L)/(1-L). \tag{3.38}$$

In cases where $\pi(L)/(1-L)$ and σ_n are known the following outliers can be detected:

$$\text{AO: } y_t = z_t + \sigma x_t, \tag{3.39}$$

where y_t is the observed value, σ is the magnitude of outlier and $x_t = \begin{cases} 1 & t = T \\ 0 & otherwise \end{cases}$, with residual $e_t = \omega\pi(L)X_t^i + a_t$

$$\text{IO: } y_t = \frac{\theta(L)}{\phi(L)}(e_t + \sigma x_t), \text{ with residual } e_t = \omega(L)X_t^i + a_t \tag{3.40}$$

The AO action may denominate a gross error model since only the level of the Tth notice is affected. On the other hand, an IO represents a prodigious concussion at time point T influencing $z_T, z_{T+1}, \ldots$ through procedure described by $\psi(L) = \phi(L)/\phi(L)$ (Chang et al., 1988).

$$\text{LS: } y_t = z_t + \frac{\omega}{1-L} X_t^i \text{, with residual } e_t = \omega \frac{\pi(L)}{1-L} X_t^i + a_t \tag{3.41}$$

Thus the LS indicates sudden permanent shifts by ω in the series. To allow for outliers, assume that instead of y_t one can observe an outlier contaminated series z_t given by:

$$z_t = \omega_i V_i(L) I_t^{(T_i)} + y_t, \tag{3.42}$$

where ω_i is the outlier size and $I_t^{T_i}$ denotes an impulse variable that takes the value 1 $if\ t = T_i$ and 0 otherwise and $V_i(L)$ defines the outlier type. Furthermore, when $V_i(L) = 1/\pi(L)$ this indicates an IO, $V_i(L) = 1$,

which implies an AO, $V_i(L) = 1/1-(L)$ which implies a LS outlier. Then $e_t = \pi(L) z_t \Rightarrow e_t = \omega_i x_t + a_t$, where the IO $\omega_i = \omega_I$ and $x_t = I^{(T)}$. AO $\omega_t = \omega_A$ and $x_t = \pi(L) I^{(T)}$, LS $\omega_i = \omega_L$ and $x_t = \pi(L)(1-L)^{-1} I^{(T)}$. This will test the null hypothesis for an outlier $H_0 : \omega_A = \omega_I = \omega_L = 0$

$$H_A : \omega_A \neq 0$$

$$H_I : \omega_I \neq 0$$

$$H_L : \omega_L \neq 0$$

This tests the likelihood ratio test statistics for testing H_0 vs. H_A, H_I, and H_L respectively, $\lambda_{iT} = \hat{\omega}_i / \sigma_i$ for i = I, A, and L where σ_i is the standard deviation of the estimate. Under the null hypothesis of no outliers, these statistics are asymptotically distributed as N (0, 1) (Jesús Sánchez & Pena, 2003). The next procedure is to compute the effect of outliers on residuals following (Tsay, 1986) where the actual parameters π and σ^2 are known in the modeling procedure estimated by any consistent estimator by employing:

$$\text{AO: } \lambda_A = \max_{\{T:1 \leq T \leq n\}} \left| \lambda_{A,T} \right| \tag{3.43}$$

$$\text{IO: } \lambda_I = \max_{\{T:1\le T\le n\}}\left|\lambda_{I,T}\right| \tag{3.44}$$

$$\text{LS: } \lambda_L = \max_{\{T:1\le T\le n\}}\left|\lambda_{L,T}\right|, \tag{3.45}$$

where these are used as testing criterion for outlier detection. Comparing the test statistics with critical value C the existence of outliers can be detected, where the time points at which the above maxima occur are timings of the corresponding outliers.

According to Tsay (1986) "the odds estimate (likelihood ratio test) derived by Chang and Tiao (1983) to distinguish an AO, IO and LS indicated that by testing the statistic of $\lambda_A, \lambda_I, \lambda_L$ with respect to (3.43), (3.44) and (3.45) respectively one may specify the type of the outlier without requiring any additional computation. More precisely if $\lambda_I > \lambda_A$ an IO is found; otherwise an AO is identified." The final procedure is to determine whether the observations are outliers and by removing each outlier from the series by deducting the value of the effect of ω, then apply the ARCH and GARCH modeling procedure to obtain an adequate model and use it for forecasting future values of the series.

3.19. Additive outliers in ARCH models

This section briefly explains the effects of outliers on ARMA model cited by Franses (1999) discussed in Chen and Liu (1993) outlier detection, which provides joint rather than sequential outlier detection of multiple outliers and estimation of parameters. The procedure provides an iterative system that will consist of specification-estimation-detection-removal cycles to independently lodge more significant outliers (Verhoeven, 2000). The procedure will be as follows (i) estimate the ARCH model, (ii) investigate the impact of outliers on the residual of the estimated model and (iii) the ARMA series will be adjusted on the basis of the largest outlier, and thus the model will be re-estimated with the removed outliers and the 3 steps will be repeated until there is no trace of large outliers. The general model of a time series data y_t which has influential outliers is represented as: $y_t = y_t^* + f(t)$, (3.46)

where $\{y_t^*\}$ is an unobserved clean series following an ARMA process:

$$y_t = \frac{\theta(L)}{\phi(L)}\varepsilon_t^* \tag{3.47}$$

Suppose a univariate time series y_t*, t=1, 2,…,n, described by the ARMA(p, q) model $\varphi_p(L)y_t^* = \theta(L)\varepsilon_t$ i.e., equation $\left(1-\theta_1 L - \ldots - \theta_p L^p\right)y_t^* = \left(1+\theta_1 L + \ldots + \theta_q L^q\right)\varepsilon_t$ (3.48)

where L is the lag operator defined by $L^j_{x_t} \equiv x_{t-j}$ and ε_t is a standard zero-mean white noise process with variance σ^2. Assuming that the roots of $\varphi_p(z)$ and $\theta_q(z)$ are outside the unit circle. An additive outlier model can be represented by equation:

$$y_t = y_t^* + \omega I_t(\tau) \tag{3.49}$$

where $I_t(\tau)=1 \quad t=\tau, \; I_t(\tau)=0 \quad t \neq \tau$.

The analysis will be done on a single AO, although the empirical method below can deal with multiple outliers in iterative steps. In case an ARMA (p, q) model is fitted to y_t in (3.47), one obtains the estimated residuals equation $\hat{\varepsilon}_t = \varphi(L)/\theta(1-\pi_1 L - \pi_2 L - ..)$. Assume that the values of the parameters, and the orders p and q, are known. For the AO in (3.39), the equation in (3.48) amounts to equation (3.49) $\hat{\varepsilon}_t = \varepsilon_t + \omega\pi(L)I_t(\tau)$. The expression in (3.49) can be viewed as a regression model for $\hat{\varepsilon}_t$, with

$$x_t = 0 \qquad t < \tau$$

$$x_t = 1 \qquad t = \tau$$

$$x_t + k = -\pi_k \qquad t > \tau \quad and \quad k = 1,2,\ldots \tag{3.50}$$

The impact ω of the AO at time t=τ can then be estimated as:

$$\hat{\omega}(\tau) = \sum_{t=\tau} \hat{\varepsilon}_t x_t \Big/ \sum_{t=\tau} x_r^2 \tag{3.51}$$

Chen and Liu (1993) suggest three ways to estimate a robust error variance, who introduced the "omit-one" method for computational convenience. The method calculates the error variance from the sample where the observation at t=τ has been deleted. Therefore, denoting the estimated error variance, based on the `omit-one' method, as $\hat{\sigma}_a$, the standardised statistic equation:

$$\hat{\tau} = \frac{\left[\hat{\omega}(\tau) / \hat{\sigma}_a\right]}{\left[\sum_{t=\tau}^{n} x_t^2\right]^{1/2}} \tag{3.52}$$

When $\hat{\tau}$, which is asymptotically standard normal exceeds some value k, the impact of the AO will be significant, where k will be fixed at the usual level. In case $\hat{\tau}$ is large and in excess of k, one can adjust the observation y_t at $t = \tau$ to obtain the AO-corrected y_t^* obtained in (3.49). However, if more than one AO can be observed, one can repeat this procedure until any $\hat{\tau}$ statistic becomes insignificant. In a final step, one can re-estimate the parameters for all observations, where some of these have been corrected for AOs (Franses & Ghijsels, 1999).

3.20. Additive outliers in ARCH models

In econometrics or time series analysis, it is stipulated to note that the problem of autocorrelation is a shape of time series data and heteroscedasticity is a shape of cross sectional data, but sometimes heteroscedasticity rising in time series data. To crop this situation, Engel (1982) developed autoregressive conditional heteroscedasticity ARCH model in forecasting financial time series, such as stock prices, inflation rates, foreign exchange rate etc. Generally an ARCH (p) model arises when the variance depends on the last p squared values of the derived time series (Gounder et al., 2007).

This study is aimed to model financial time series data allowing for the sensitivity accounted for on the conditional heteroscedasticity to the inclusion of AO's, where the focus is based on the Lagrange Multiplier (LM) as it is the most popular test to test for ARCH effects (Van Dijk, 1999). This study follows the procedure derived by Van Dijk et al (1999), in which the author aimed to detect outliers on ARCH model using robust test when testing for linearity of the conditional mean. The study focuses on the LM test for the ARCH biased by the contamination of outliers. The study shows the effects of an outlier robust estimator, under the hypothesis of estimating the model using outlier robust estimator can also be used to robustify the LM for ARCH model in (3.9).

3.21. Additive outliers in GARCH models

The reference point of GARCH (p, q) regression model with normally distributed errors is defined as:

$$y_t = x_t'\xi + \varepsilon_t, \qquad \varepsilon_t|\Gamma_{t-1} \sim N(0, h_t)$$
$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \qquad t = 1, \ldots, T \tag{3.53}$$

Γ_{t-1} is the division up to time t. In practice x_t consists of a constant term. The log-likelihood of (3.31) is given by:

$$\ell(\theta) = \sum \ell_t(\theta) = c - \frac{1}{2}\sum_{t=1}^{T}\left[\log(h_t) + \frac{\varepsilon_t^2}{\sigma_t}\right] \tag{3.54}$$

For which a GARCH (1, 1) model having $0 \le \beta_1 < 1$ can be written as:

$$\sigma_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3.55}$$

$$\sigma_t = \alpha_0^* + \alpha_1 \sum_{j=1}^{t} \beta_1^{j-1} \varepsilon_{t-j}^2 \tag{3.56}$$

Given ε_0 and h_0, where $\alpha_0^* = \alpha_0 (1 - \beta_1^t) / (1 - \beta_1) + \beta_1^t h_0$ (Doornik & Ooms, 2005).

An outlier detection procedure used by Fox (1972) follows the following steps to detect outliers in a time series data:

1. Model a GARCH (1, 1) model based on the original data is estimated under the assumption that there is no outliers in the data.
2. The estimated residuals are obtained and therefore, the outlier effect $\varpi(\tau)$ and the residual variance σ_v is computed
3. Using (2), the test statistic $\tau_i(\tau)$ is calculated for all possible $\tau = 1, \ldots n$, where i denotes outlier type.
4. The maximum of the absolute value of these test statistics $\tau_{\max} = Max_{t=1,..n}|\hat{\tau}_i(\tau)|$ is computed if the value of the test statistic exceeds the pre-specified critical value, C (significant) then an outlier is detected. Thus, the point t where $\tau_{\max}$ occurs is the point detected as having the outlier.

3.22. Summary

This chapter outlined the methodology of the study. The study focused on the quantitative approach of GARCH modeling with the inclusion and exclusion of outliers. The ARCH and GARCH models were reviewed together with their model selection criterion. The performance of the models will be judged based on the three error matrices discussed. A modified model after removal of outliers will also be subjected to similar model diagnostics and its performance just same way as outlier contaminated model.

Chapter four which outlines data analysis and presentation of results follows.

CHAPTER FOUR
DATA ANALYSIS AND RESULTS

4.1. Introduction

This chapter discusses the analysis of the results, and it starts with the examination of the data specifically, by analysing the return distribution and descriptive statistics. Data analysis was done using SAS, Gretl and E-views packages. The rest of the chapter is structured as follows: Section 4.2 presents the behaviour of the data and stationarity test results. Section 4.2 presents the preliminary data analysis results such as the descriptive statistics. Section 4.3 presents the results of the main analyses which include the outlier free and outlier contaminated ARCH and ARMA-GARCH type models. The last section gives a chapter summary.

4.2. Preliminary analysis

The purpose of this section was to provide and discuss the results that quantify, summarise and check the distributional properties of the financial time series. The results are presented as time series plots and descriptive statistics before applying formal econometric tests. According to Harrilall (2014:44), if the returns are normally distributed and exhibit stable moments, then the probability of a specified loss does not vary across days. Normality is assumed in this study since the number of observations is 1330 in excess of the minimum 50 according to the central limit theorem.

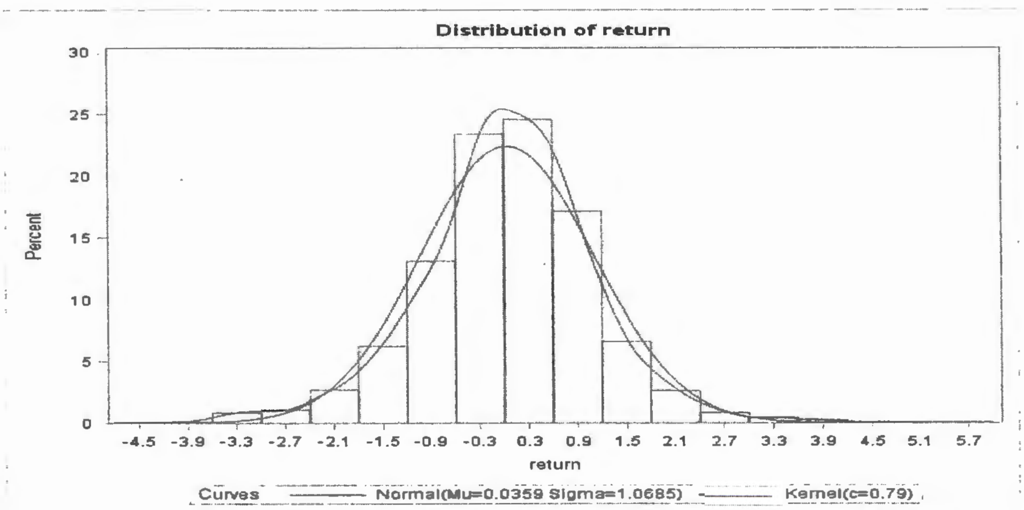


Figure 4.1: plot of the kernel density plot

The histogram of the returns visually indicates that the distribution of the returns <u>is</u> not peculiarly different to a normal bell-shaped normal curve with the red line indicating residuals of the daily log returns. The blue line indicates the normal distribution in terms of the sample mean and standard deviation. Since the log returns exhibit excess kurtosis and normality showing a taller and slimmer body (Harrilall, 2014).

4.2.1. Time series plots

This study employed both graphical (informal) methods and formal statistical tests to verify stationarity of the series. Most applied time series data are non-stationary, that is, their variance changes with time. The results are summarised in Figure 4.2 and Table 4.1.

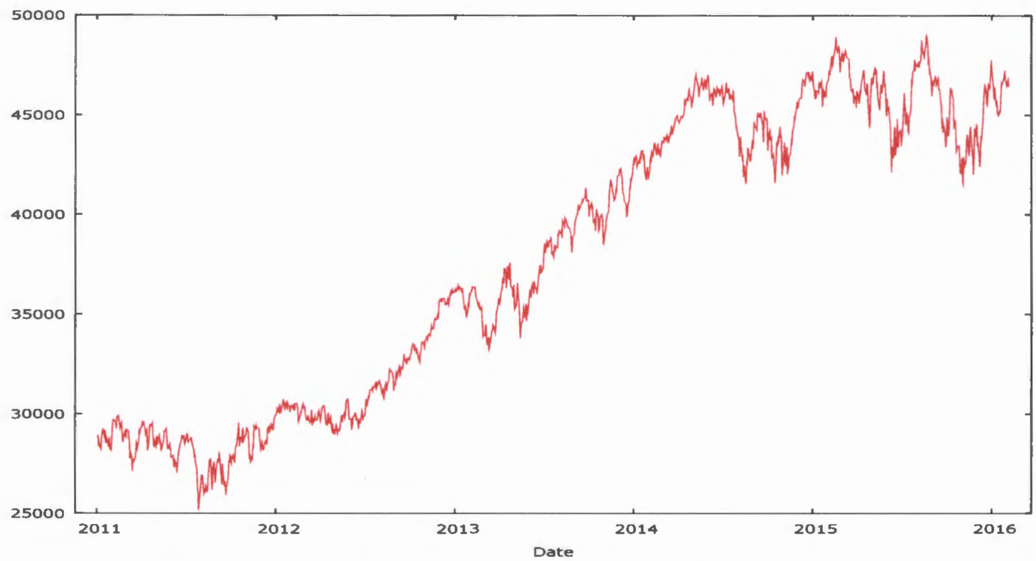


Figure 4.2: plot of the daily returns for JSE top 40 index (2011 to 2016)

Figure 4.2 shows that the variances of the original data for JSE top 40 index are changing with time. This is a sign that the series is not stationary (has unit root) and as a result, the values are converted by logarithmic transformations. The JSE top 40 index also shows a nonlinear upward stochastic trend. This could also be an indication that the returns follow a random walk process with a drift. The Figure 4.2 suggest that the estimated model should include a trend with no seasonal variations. In the absence of a drift, the series is considered a non-stationary process, with both its mean and variance increasing over time. Several peaks are also evident during each period, and this marks the period during which the prices of stocks were increasing or decreasing.

Table 4.1: ADF and KPSS test of the JSE top 40 index

		t-Statistic	Prob.*			LM-stat	Prob.*
ADF	Test statistic	-1.0463	0.7384	KPSS	Test statistic	4.2159	0.0000
Test critical values:	1% level	-3.4350		Asymptotic critical values:	1% level	0.7390	
	5% level	-2.8635			5% level	0.4630	
	10% level	-2.5678			10% level	0.3470	

Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively.

The ADF test was used to check for a unit root in the original series using the ADF formal test, shown in Table 4.1. Since the test statistic of -1.046364 with a probability value of 0.7384 is much greater than the critical values at 1% and 5%, but lower at 10%, fail to reject H_0 at a significance level <1%. It was concluded that the time series has a unit root supporting Figure 4.1. The KPSS test was also used to confirm stationarity for the JSE top 40 index. The test statistic is 4.215939 and a corresponding probability value of 0.000 is less than the critical values at 1%, 5% and 10%. Therefore fail to reject the null hypothesis indicating that the JSE top 40 index is non-stationary.

4.2.2. Stationarity test results

Due to the results presented in Figure 4.2 and Table 4.1, the JSE top 40 index was transformed to induce stationarity, and the results are summarised in Figure 4.3 and Table 4.2.

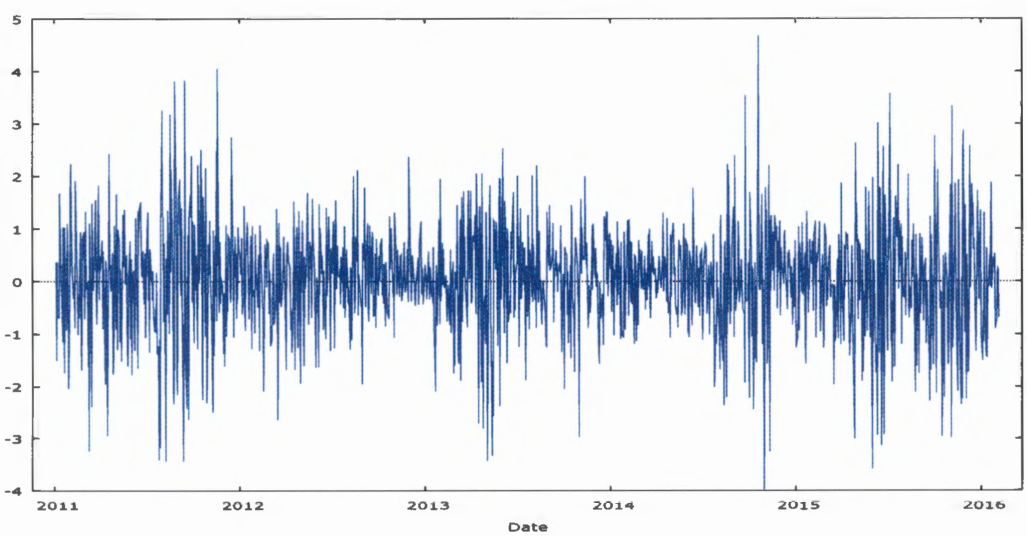


Figure 4.3: A plot of the transformed daily return series

Figure 4.3 shows the plot of the stationary JSE top 40 index. The variance and mean of this series are now fluctuating around a common location with no indication of structural breaks. Thus the log transformation has produced a stationary process. Although the plot depicts a fairly constant mean across sub samples of the data, which is consistent with other studies of financial returns, where the mean is often found to be stationary, the series is volatile. This is due to abnormalities roughly above and below the mean. Figure 4.3 also shows that there are days of increased volatility, followed by more soothing periods, which is consistent with the concept of volatility clustering. To confirm the stationarity condition of the series, the results on ADF and KPSS unit root testing are reported in the next section. Summarised in Table 4.2 is the ADF and KPSS unit root test results.

Table 4.2: ADF and KPSS Test of the JSE returns

		t-Statistic	Prob.*			LM-stat	Prob.*
ADF	Test statistic	-28.31780	0.0000	*KPSS*	Test statistic	0.06508	0.2206
Test critical values:	1% level	-3.435067		*Asymptotic critical values:*	1% level	0.73900	
	5% level	-2.863511			5% level	0.46300	
	10% level	-2.567868			10% level	0.34700	

Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively.

The results presented in Table 4.2 suggest that the null hypothesis for ADF test is rejected implying that the return series does not have a unit root at a 99% level of significance. The test statistic of -28.31780 is much lower than all the critical values leading to the rejection of H_0 at all significance levels. Thus, it can be concluded that the time series has no unit root. The KPSS test results were used to confirm whether or not the return series is stationary, thus allowing the application of the ARMA, ARCH and GARCH in this study. The results of KPSS test show that the returns are stationary since LM-stat is greater than 0.10 at all asymptotic critical values.

4.2.3. Stylised facts of the data results

This section presents the results quantifying the descriptive measures for the JSE top 40 index and returns series. A summary of the results is given in Table 4.3.

Table 4.3: showing measures of central tendency and dispersion

Measure	JSE top 40 index	Returns
N	1330	1329
Mean	37794.3015	0.03592003
Minimum	25181.00	-3.9847955
Maximum	49081.00	4.6777744
Std Deviation	7178.64679	1.06851441
Variance	51532969.8	1.14172304
Skewness	-0.0947732	-0.1293821
Kurtosis	-1.5613544	1.29215392
Jarque-Bera	136.8483(< .0001)	94.8555 (<0.0001)

Table 4.3 shows the summary statistics for the South African JSE top 40 index and return series y_t. The mean 0.0359 of the daily return series is close to zero and the mean for the JSE top 40 index is 37794.302. The statistics are both positive signifying average rate at which stock prices are increasing. The average time interval of the price increment is estimated at about 1.06851%, and this is indicated as a standard deviation. The estimated skewness measure is -0.12938 for the return series and -0.0947732 for JSE top 40 index and this is an indication of a negative skewness. This implies that most of the values of the series are concentrated on the left side of the mean. This could also mean that there are heavy tails with extreme values on the left in contrast to the symmetric normal distribution. Furthermore, kurtosis is equal to 1.292154 for the returns, which is an indication that the probability for extreme values is less for a normal distribution. On the contrary, the kurtosis for the JSE top 40 index is -1.5613544. One could cautiously conclude that the values are widely spread around the mean. As a result, the distribution of the series is platykurtic due to negative excess kurtosis.

Inferences concerning non-normality are maintained by the Jarque-Bera test statistics for the JSE top 40 index and returns which show that the null hypothesis is rejected at 5% level of significance. Thus it can be concluded that the JSE top 40 index have a non-normal distribution which is a common occurrence in stock markets. As suggested by Sigauke et al. (2014), lack of non-normality of the distribution is due to volatility clustering. Furthermore, the presence of outliers in the data pushes for lack of normality. In any case, as explained earlier, this assumption is protected according to the central limit theorem. The findings from the descriptive analysis confirm the stylise facts about financial data and are in accordance with those reported by many authors.

4.3. Primary data analysis results

This section entails a detailed analysis of the univariate ARMA (*p, q*), ARCH (*p*) and GARCH (*p, q*) models for the JSE top 40 return. In order to address the study objectives, this study relies heavily on time series analysis using traditional ARCH and GARCH type models suggested, by predicting ARCH and GARCH type model for JSE top 40 index.

4.3.1. ARMA (2, 2) model specification

Autocorrelation and partial autocorrelation plots are used to determine the order of the ARMA (*p, q*), ARCH (*p*) and GARCH (*p, q*) models. The ACF and PACF plots shown in Figure 4.4 and Figure 4.5 respectively, give a clearer indication of the nature of the model. This procedure assisted in choosing the correct number of lags for ARCH and GARCH type analysis.

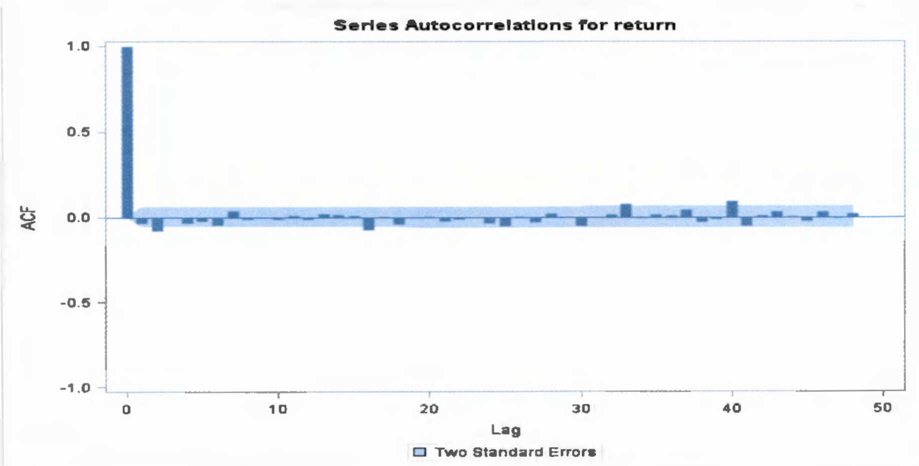


Figure 4.4: Plots of ACF results

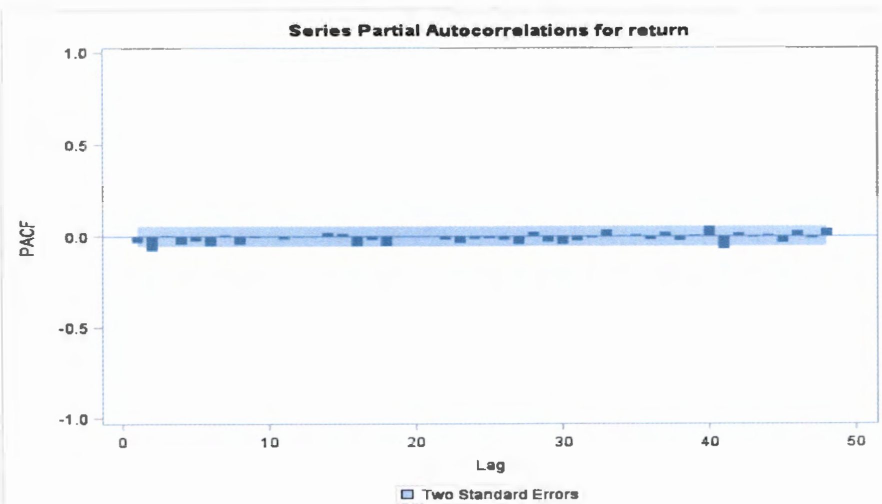


Figure 4.5: Plots of PACF results

Table 4.4: ARMA (2, 2) model summary

Statistics	Estimation
Constant	0.03607 (0.0301)**
AR (2)	0.74442 (<0.0001)***
MA (2)	0.81850 (<0.0001)***
Durbin-Watson	2.075892
AIC	3937.567
SBC	3963.528
Log likelihood	-1965.136

*Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively. Values in the parentheses are the probabilities of the test statistics of the estimated parameters.*

The results in Table 4.4 shows that the returns can be written as:

$$(1-0.03607-0.74442(L)-0.81850(L^2)(1-L)X_t = \varepsilon_t \qquad (4.1)$$

where L is the lag operator $X_t = return_t$ and $\varepsilon_t \sim iid$ (0, 1.128729). The results reveal the observed Durbin-Watson statistic as 2 indicating that there is no autocorrelation problem associated with the returns. The selected model was then tested for robustness and the results are presented in subsequent sections.

4.3.4. ARMA (2, 2) Model Robustness

This section presents and discusses the results of the tests used in evaluating the selected model for adequacy.

4.3.4.1. Test for independence

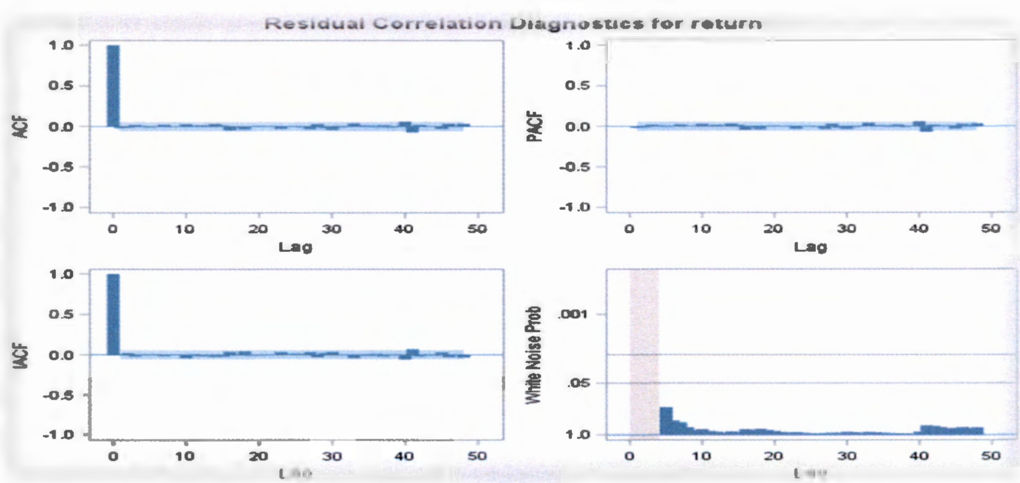


Figure 4.6: Autocorrelations of residuals

The residual autocorrelation plot in Figure 4.6 shows that most of the residual autocorrelations fall around the zero band. With reference to the runs test for independence, the observed statistic is 0.0693 with a p-value of 0.9447. Hence, the null hypothesis that the residuals follow a white noise process cannot be rejected. Therefore, the selected ARMA (2, 2) model for returns is deemed adequate.

4.3.4.2. Normality

This section tests the consistency of the inferences concerning normality and constant moments of the return series. To test for normality, it is vital to analyse the summary statistics of the return series, which gives a suggestion of where and how the data is spread, giving insight into whether or not the series is normally distributed.

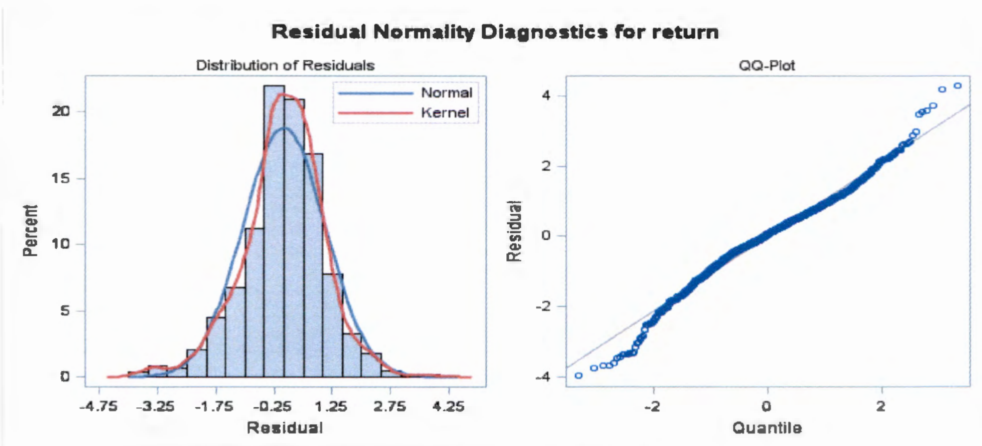


Figure 4.7 Histogram of residuals

Figure 4.7 displays the histogram and Jarque-Bera test statistic respectively, which jointly examine the skewness and kurtosis to formulate a joint test statistic with a value of 102.8008. The null hypothesis of normality was rejected at a 99% confidence level, as the p-value is less than 0.000001. Thus the alternative was favoured indicating that the returns are not normally distributed as expected. Therefore, the conclusion that return series observed in the JSE market have non-normality distribution is confirmed, which is a practical and common occurrence in developing stock markets (Makhwiting et al., 2013).

4.3.4.3. Constant variance test

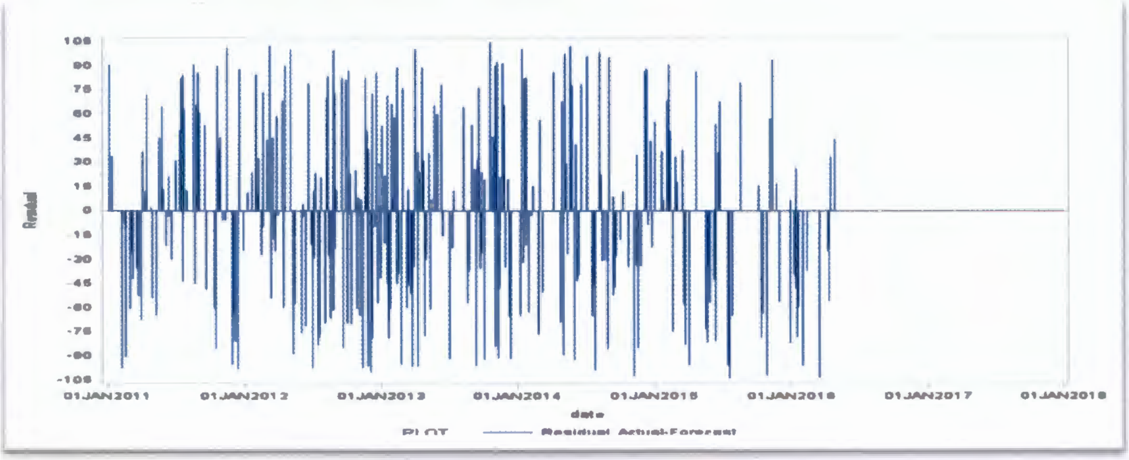


Figure 4.8: Residual plot

The residual plot in Figure 4.8 does not show any clear pattern, further confirming that the ARMA (2, 2) model residuals followed a white noise process. The Residual plot in Figure 4.8 shows that there is no trend, thus the conclusion that ARMA (2, 2) is an adequate model.

4.4. ARMA (2, 2) out-of-sample forecasts

Once the model has been confirmed for appropriateness, Box and Jenkins (1979) recommend that forecasts be produced using it. Figure 4.9a and Figure 4.9b below are used to present the out-of-sample forecasts using ARMA (2, 2) model.

Figure 4.9a: Overlay plots of estimated index and return plot

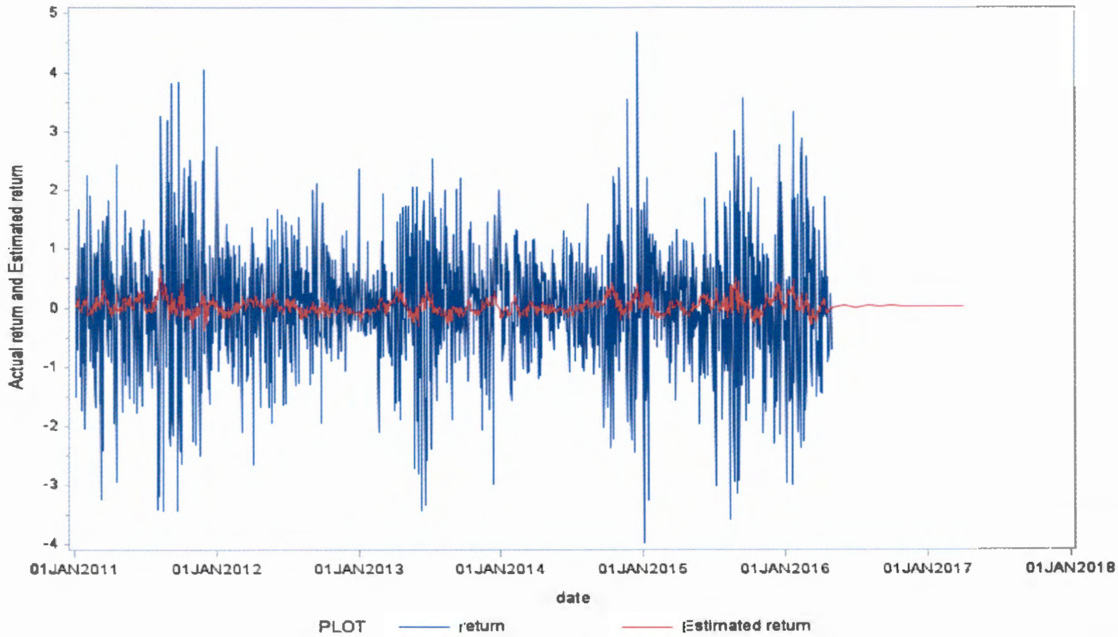


Figure 4.9b: Overlay plots of estimated index and return plot

The overlay plot of the actual index and estimated index in Figure 4.9 (a and b) above is indicative that ARMA (2, 2) model is appropriate for the data. The estimated and actual values of the data are closely related and moving along the same wavelength providing more evidence that the model is appropriate.

4.5. ARCH and GARCH type model identification results

The ARCH model is used in parametric nonlinear time series model, introduced by Engle (1982). This model allows the conditional variances to change over time as a function of squared past errors leaving the unconditional variance constant. The ARCH model together with the ARMA model specification gives better condition of the forecast error variance. Presented in this section are the results of the ARCH (2) and ARMA (2, 2)-GARCH (1, 1), ARMA (2, 2)-GJR-GARCH (1, 1), ARMA (2, 2)-EGARCH (1, 1), ARMA (2, 2)-TGARCH (1, 1), ARMA (2, 2)-APARCH (1, 1), ARMA (2, 2)-CGARCH (1, 1), ARMA (2, 2)-GARCH (1, 1) with student t distribution, ARMA (2, 2)-GARCH (1, 1) with GED, ARMA (2, 2)-GARCH (1, 1) with skewed normal, and ARMA (2, 2)-GARCH (1, 1) with skewed-t errors with constant parameters included in the model. The desire was to determine the suitability of these parameters in both the proposed ARCH model and also to determine the most suitable GARCH type model for the data. According to Bollerslev et al. (1994), models with lower orders such as $GARCH\ (1,1)$ and $GARCH\ (2,1)$, are adequate for modelling volatilities even over long sample periods. Therefore, the study considered the GARCH (1, 1) together with the

ARMA (2, 2) models. The two information criteria were used to make a choice between these models and the results are summarised in Table 4.5 and Table 4.6.

Table 4.5: ARCH and GARCH Outlier Contaminated Model estimation summary

Parameters	arch (2)	ARMA(2, 2)-GARCH (1, 1) Model t-distributed errors	ARMA(2, 2)-GARCH (1, 1)	ARMA(2, 2)-GARCH (1, 1) GED distributed errors	ARMA(2, 2)-GARCH (1, 1) skewed t-distributed errors	ARMA(2, 2)-GARCH (1, 1) skewed normal distributed errors
μ	0.03976 (0.1349)*	9.21e-06 (0.0001)***	0.0514 (0.0351)**	0.06848 (0.0063)***	0.04904 (0.0469)**	0.050334 (0.0457)**
θ_1	0.8104 (<0.0001)***	0.997254 (0.0000)***	0.0271 (0.3804)*	-0.0424 (0.1311)*	-0.0486 (0.0644)*	-0.05422 (0.0492)**
θ_2	0.0703 (0.0140)**	-0.10021 (0.0014)***	0.0744 (0.0058)***	-0.08062 (0.0038)***	-0.08699 (0.0030)***	-0.0887 (0.0021)***
ω	0.2262 (<0.0001)***	0.155771 (0.000)***	0.0197 (<0.0001)***	0.018534 (0.0172)**	0.01579 (0.0287)**	0.017238 (0.0182)**
β		0.804538 (0.0000)***	0.9112 (<0.0001)***	0.9116 (0.000)***	0.91276 (0.000)***	0.91267 (0.000)***
α				0.07263 (0.003)***	0.074448 (0.000)***	0.07239 (0.000)***
df		38.74475 (0.3398)*				
N				1.57672 (0.000)***	10.1366 (0.000)***	1.57459 (0.000)***
λ					-0.1515 (0.0002)***	-0.14142 (0.0006)***

*Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively. Values in the parentheses are the probabilities of the test statistics of the estimated parameters.*

Table 4.5 above shows the parameter estimates for the proposed three models with constants. ARCH (2) is significant with probability values below all the levels of significance. The intercept of the ARCH (2) model is however insignificant. Since the ARCH (2) model is significant according to the results, this is an indication that this mean equation could be fit to the GARCH variance equation. The parameter estimates (μ, ω, θ_1, and θ_2) of ARCH (2) model are 0.0396, 0.8104, 0.0703, and 0.2262, respectively. Furthermore, according to the findings, ARMA (2, 2)-GARCH (1, 1) proves to be sufficient for the data.

The parameters of this model are all significant except for α_1. The parameter estimates (μ, ω, θ_1, θ_2 and β_1) of ARMA (2, 2)-GARCH (1, 1) model are 0.0514, 0.0271, 0.0744, 0.0197 and 0.9112, respectively. Volatility shocks are persistent since the sum of the ARCH and GARCH

$(\alpha + \beta)$ coefficients are close to one. The coefficient for the conditional variance estimate, on the other hand, is also highly significant. For the evaluation of innovation of errors, this study used the ARMA (2, 2)-GARCH (1, 1) volatility model to estimate the conditional variance at the same time. The researcher did so by choosing the normal distribution and the student t-distribution as the error term's distribution may assist in determining the best model that can accommodate excess kurtosis identified in previous discussions.

Estimation results for ARMA (2, 2)-GARCH (1, 1) with student t distribution is also shown in Table 4.5. The results show all GARCH terms being significant at the 1%, 5% and 10% levels of significance, suggesting that the model may be adequate. Using the student t-distribution should give some sort of advantage in terms of accuracy given the heavy tails that were evident in the data. The parameter estimates (μ, ω, θ_1, θ_2 β_1 and *df*) of ARMA (2, 2)-GARCH (1, 1) with t-distributed innovations model are 0.00000921, 0.997254, -0.10021, 0.155771, 0.804538 and 38.744475, respectively.

Table 4.6 above also highlights the skewed versions distribution, the Skewed GED, can fit the data better than other two skewed models. Nevertheless, there is not a huge dissimilarity between the Skewed Student-t Distribution and Skewed GED, one possible cause is that both of them have the shape and skew parameters.

Table 4.6: GARCH type models Outlier Contaminated Model estimation summary

Parameters	ARMA(2, 2)-EGARCH (1, 1)	ARMA(2, 2)-APARCH (1, 1)	ARMA(2, 2)-TGARCH (1, 1)	ARMA(2, 2)-GJR (1, 1)	ARMA(2, 2)-CGARCH (1, 1)
μ	0.0138 (0.5482)*	0.0164 (0.4724)*	0.0106 (0.6486)*	0.0186236 (0.4519)*	0.0894 (0.000)***
θ_1	0.00913 (0.7600)*	0.0252 (0.4094)*	0.0123 (06812)*	0.01632 (0.5412)*	0.0159 (0.9258)*
θ_2	0.0804 (0.0035)***	0.0773 (0.0055)**	0.0920 (0.0010)***	0.08547 (0.0024)***	-0.10158 (0.5434)*
ω	0.00042 (0.9062)*	0.0179 (<0.0001)***	0.0170 (<0.0001)***	0.01867 (0.0111)**	1.0465 (0.000)***
α	0.0535 (0.0004)***	0.0587 (0.0003)***	-0.0288 (0.0001)***	0.03195 (<0.0001)***	0.97412 (0.000)***
β	0.9822 (<0.0001)***	1.000 (0.0002)***	0.1472 (<0.0001)***	0.99931 (0.000)***	0.0415 (0.0003)***
γ	-2.4784 (0.0006)***	0.9375 (<0.0001)***	0.9395 (<0.0001)***	0.91882 (0.000)***	-0.0797 (0.000)***
δ		0.498 (<0.0001)***			-0.0695 (0.0007)***
N					0.03458 (0.2622)*
λ					-0.47664 (0.1473)*

*Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively. Values in the parentheses are the probabilities of the test statistics of the estimated parameters.*

Table 4.6 shows the estimation for GARCH type models. The results indicate for ARMA (2, 2)-EGARCH (1, 1) model the leverage effect term $\gamma$is negative which shows that there is an existence of the leverage effect in future returns. When, $\gamma \neq 0$ indicates the news impact is asymmetric, supporting the use of skewed Student-t distribution for returns. Since β (persistence in conditional volatility) is 0.9822 which is close to 1, implies that volatility will take long to die in the JSE stock market.

The estimated ARMA (2, 2)-TGARCH (1, 1) model allows for asymmetric effect; thus treating positive and negative news differently. Similar to the GARCH model, the TGARCH model provides information as to whether the news on volatility is significant or whether it influences volatility. The ARMA (2, 2)-TGARCH (1, 1) model estimates show that the news impact is asymmetric since $\gamma \neq 0$ the γ parameter is significantly positive indicating that the leverage effect exists. The results also indicate that the estimated coefficients of ARMA in the mean equation and GARCH terms are both significant within the margin of error. Volatility shocks are persistent since the sum of the ARCH and GARCH $(\alpha + \beta)$ coefficients are close to one (Makhwiting et al., 2014).

The ARMA (2, 2)-GJR-GARCH (1, 1) model is similar to the TGARCH (1, 1) model but the volatility is specified in terms of σ_t^2 instead of σ_t. The chosen distribution for the innovation of the ARMA (2, 2) APARCH (1, 1) is the traditional normal distribution. Estimating the parameters appearing in an ARMA-GARCH structure is straightforward.

The ARMA (2, 2) CGARCH (1, 1) model indicate that the estimated $\alpha + \beta$ is bigger than that of the normal GARCH (1, 1), thereby indicating that the long run volatility component is strong. The long-run volatility component (δ) is estimated as -0.0695, thereby implying that the permanent component of the conditional variance evidence has a negatively low degree of persistence.

4.6. Model selection

In the SA stock return ARCH and GARCH type model compare the best model was selected according to the information criteria given in Table 4.7 below.

Table 4.7: ARCH and GARCH Outlier Contaminated Model selection summary

Criterion	ARCH (2)	ARMA(2, 2)-GARCH (1, 1)	ARMA(2, 2)-GARCH (1, 1) Model t-distributed	ARMA(2, 2)-TGARCH (1, 1)	ARMA(2, 2)-APARCH (1, 1)	ARMA(2, 2)-**EGARCH(1, 1)**	ARMA(2, 2)-GJR-GARCH(1, 1)	ARMA(2, 2)-GARCH (1, 1) GED distributed	ARMA(2, 2)-GARCH (1, 1) Skewed t-distributed	ARMA(2, 2)-GARCH (1, 1) skewed GED distributed
MSE	1.14088	1.1401	1.13413	1.13408	1.13307	**1.13386**	1.13558	1.13523	1.14012	1.14253
SBC	3906.995	3806.0446	3795.5774	3757.383	3761.115	**3746.342**	3757.445	3791.944	3782.2943	3784.0637
AIC	3886.227	3780.0837	3764.8243	3721.038	3719.578	**3709.997**	3721.11	3755.6094	3740.7689	37442.538
MAE	0.80554	0.805377	0.8023959	0.802223	0.80164	**0.802221**	0.802398	0.803598	0.802498	0.80359
HQC	3897,957	3784.2119	3766,8044	3795,0951	3755,456	**3717.2672**	3734,7293	3769.2838	3756.3334	3758.103
MAPE	106.821	108.6961	107.58122	109.8685	108.3038	**107.9927**	107.9899	107.699	107.5856	107.689
Log-Likelihood	-1939.113	-1885.042	-1876.412	-1853.52	-1851.79	**-1848**	-1853.56	-1870.805	-1862.384	-1863.269
Rank	(10)	(9)	(5)	(2)	(4)	(1)	(3)	(8)	(6)	(7)

The models presented in Table 4.7 were ranked according to their performance. According to the AIC and SBC, the ARMA (2, 2)-EGARCH (1, 1) model is the best model. This model also proves the best because $\alpha + \beta$ is close to 1 and it accommodates the excess kurtosis that is exhibited by the returns. Table 4.8 also illustrates which model has the smallest MAE, MAPE and RMSE for JSE returns. This model was chosen as the best and appropriate one to produce a precise forecast for the conditional variance. For the stock return, ARMA (2, 2) -EGARCH (1, 1) model has the smallest MAPE and RMSE but has a contradicting result for MAE that has a small marginal error of 0.0005 of the APARCH model. Therefore, EGARCH model is expected to forecast accurately the future conditional variance better than other models. Thus, this model was used to make forecasts in subsequent sections. The methodological specifications suggest the ARMA (2, 2) - EGARCH (1, 1) model with the mean equation as:

$$r_t = 0.0138 + \varepsilon_t$$

$$\ln \sigma_t^2 = 0.09131 + 0.0804\varepsilon_{t-1} + 0.000417\varepsilon_{t-2}^2 + 0.0535h_{t-1} + 0.9822h_{t-1}^2 - 2.4784\gamma \qquad (4.2)$$

$$\lambda = 1.0357$$

NWU

4.7.ARMA (2, 2) –EGARCH (1, 1) Model robustness

4.7.1. Test of independence results

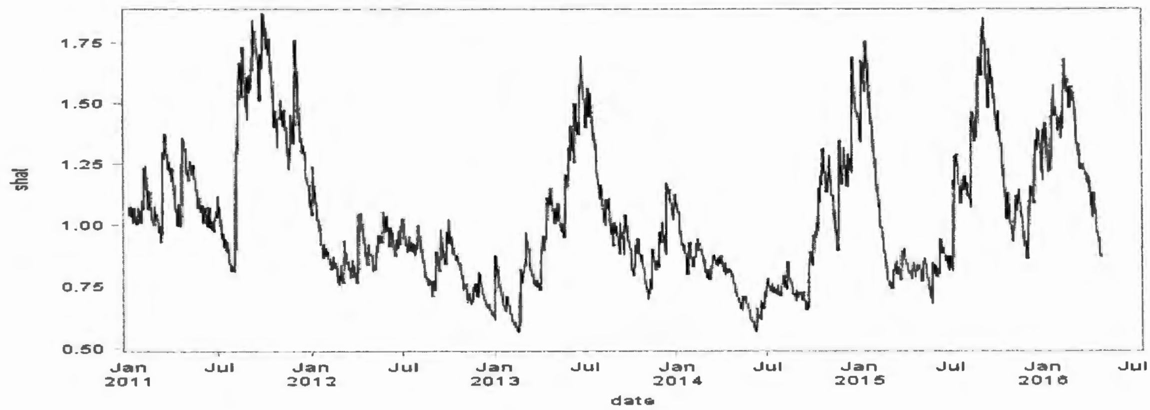


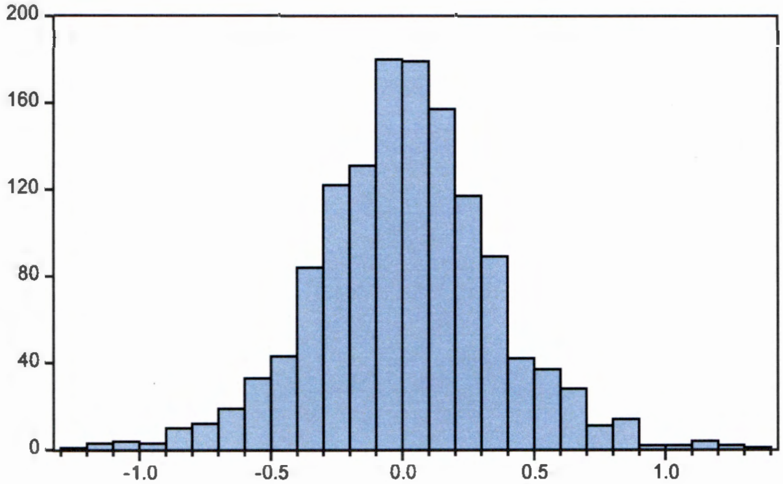
Figure 4.10: Conditional standard deviation for returns

Figure 4.10 shows the conditional standard deviation for residual of returns. Letting e_t (residual at date *t*) = return at date *t* – sample mean, thus, the squared residual e_t^2 estimates the variance of the return for t. Therefore, the plot of the squared residuals does not appear to follow a random process. Also, similar values of squared residuals come in chunks. The chunkiness of squared residuals is the result of the dependence of the variance of the return at time *t* on variances at preceding periods. Thus, by constructing the correlogram of squared residuals can confirm the variance of the squared residuals. In Table 4.11, the ACF values with the standard error, it is concluded that ACF are highly significant at all (10) lags. Q-stats are also highly significant. This dependence of the variance at time *t* on variances of preceding periods is called the conditional heteroscedasticity, which is quite common for high frequency asset return data.

Table 4.11: Correlelogram of squared residuals

Autocorrelation	Partial Correlation	Lags	AC	PAC	Q-Stat	Prob*
|** |	|** |	1	0.268	0.268	95.987	0.000
|** |	|** |	2	0.291	0.236	208.88	0.000
|* |	| |	3	0.149	0.030	238.61	0.000
|* |	|* |	4	0.182	0.089	282.91	0.000
|* |	|* |	5	0.183	0.102	327.73	0.000
|* |	| |	6	0.161	0.050	362.38	0.000
|* |	| |	7	0.128	0.016	384.36	0.000
|* |	| |	8	0.151	0.064	414.95	0.000
|* |	| |	9	0.117	0.020	433.34	0.000
|* |	| |	10	0.124	0.024	453.99	0.000

4.7.2. Jarque-Bera test for normality of residuals



Series: Standardized Residuals	
Sample 1/03/2011 4/29/2016	
Observations 1330	
Mean	0.012391
Median	0.013063
Maximum	1.399987
Minimum	-1.242720
Std. Dev.	0.344757
Skewness	0.028599
Kurtosis	4.006349
Jarque-Bera	56.30392
Probability	0.000000

Figure 4.12: Jarque-Bera test for normality

Presented in Figure 4.12 are the ARMA (2, 2)-EGARCH (1, 1) forecasts results for JSE top 40 returns. The residuals show that the errors are not normally distributed. As previously stated, financial data has a common characteristic of not being normally distributed. So this finding is not a surprise.

4.7.3. Test for heteroscedasticity

Prior to estimating the parameters of the ARCH and GARCH models, it is important to detect the presence of ARCH disturbances in time series data (Weiss, 1984).

Table 4.8 Tests for ARCH Disturbances Based on Residuals

Order	Q	Pr > Q	LM	Pr > LM
1	4.1731	0.0411	4.1942	0.0406
2	53.9189	<.0001	52.3088	<.0001
3	68.2422	<.0001	62.1362	<.0001
4	77.6676	<.0001	64.4023	<.0001
5	93.2064	<.0001	70.9407	<.0001
6	110.3101	<.0001	79.1926	<.0001
7	142.5453	<.0001	95.8121	<.0001
8	163.3515	<.0001	102.7914	<.0001
9	218.2298	<.0001	128.1818	<.0001
10	222.2179	<.0001	128.3622	<.0001
11	235.2117	<.0001	128.4654	<.0001
12	252.8319	<.0001	132.6749	<.0001

Testing for heteroscedasticity was done using the Engle's Lagrange Multiplier (LM), squared residual test. In essence, the ARCH effects were tested by scrutinising whether or not the volatility of the daily returns on the JSE index and the squares of the daily returns were serially correlated, and the results are summarised in Table 4.8 above.

The Q-statistics test for changes in the variance across time by using windows ranging from 1 through 12. The test strongly shows that there is heteroscedasticity, with p-values<0.0001 for all lag windows, except lag 1. The *p*-value of LM statistics for the lag windows fails to reject the null hypothesis of the absence of ARCH effects. The LM test further suggests a strong heteroscedasticity of errors for GARCH model. The LM tests also indicate a long memory of GARCH model to modeling time series with heteroscedastic errors (SAS support, 2013).

4.8. Out-of-sample forecasting for ARMA-EGARCH (1, 1)

A decent forecast capability of volatility models give an applied tool for stock market analysis and permit investors to give more suitable securities pricing strategies. The prediction error statistics used in the assessment of the out-of-sample predictions based on the selected model shows that the adequate model ARMA (2, 2)-EGARCH (1, 1) will be used to make forecasts.

Presented in Figure 4.13 is a forecast plot for ARMA (2, 2)-EGARCH (1, 1).

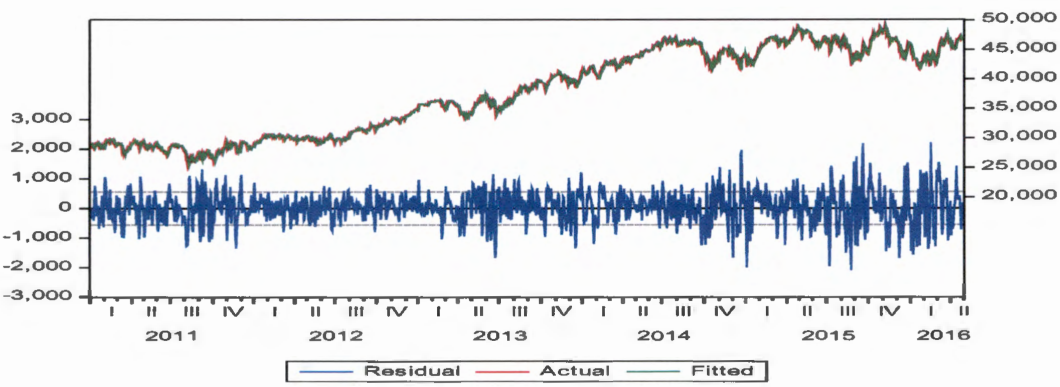


Figure 4.13: Forecast plot for ARMA (2, 2)-EGARCH (1, 1)

ARMA (2, 2)-EGARCH (1, 1) model is adequate to capture volatility. Therefore, innovation of errors was presented to check whether the flat tails of the data contributes to the significance of the forecasts. The forecasts presented in Figure 4.13 show that the ARMA (2, 2)-EGARCH (1, 1) model moves with actual and fitted series, indicating that the selected model is the best. Further analysis on the detection of outliers was done to investigate if the outliers have any influential effect on the good estimation of parameters for better forecasts.

4.9. The ARMA, ARCH and GARCH type modeling results - "outlier free"

4.9.1. General Outlier detection

This section focused on outlier detection processes to analyse the outlier-adjusted returns $\{y_t^*\}$ and to also examine time-varying volatility. The study also investigates the effect of outliers through test statistics and performance of the criteria and outlier detection procedure described in chapters 2 and 3. In detecting the presence of a single outlier, an approach is to apply the likelihood ratio criteria. The likelihood ratio criterion is derived for testing the null hypothesis against the alternative hypothesis for each type of outlier. The detected type of outlier is based on the largest measure of the test statistic (in absolute value).

Stage I: Locate outliers

In this step, outliers are detected and located by checking the significance of the three types of outliers at any possible time points. Therefore, the location of outliers was made by computing the corresponding t-statistics and selecting those that are significant given a critical value.

Table 4.9: outlier detection summary

Outlier Detection Summary	
Maximum number searched	1328
Number found	99
Significance used	0.01

Table 4.9 shows the outlier detection summary. The maximum number of outliers found is 99 at 1% level of significance under the tests of hypothesis discussed in Chapter 3. It is also evident in Table 7.2 (appendices), (summary of outliers) that there are two types of outliers detected namely the AO and LS at 0.01 level of significance. The high absolute statistics higher than the threshold shows the time point of a potential outlier (with asterisks). Therefore $\omega = 99$, with a threshold critical value of $cval = 3.5$. As stated, there are possibly two types of outliers for a given time point (e.g. $at\ t = 991\ or\ t = T = 18/12/2014, \tau_{AO} = 4.5309$) At first, in Table 7.2 (appendix), it is clear that the AO and LS were detected, but later through looping the LS transformed into a TC. Thus, the TC is a generalisation of AO and LS in the sense that it bases an initial influence like an AO but the outcome is also passed on to the following observations. The impact of a TC is not enduring however, it decays exponentially (Tolvi, 2000).

4.9.2. Locate outliers using loops

This section locates outliers using inner and outer loops in order to remove duplicates by keeping the type of outlier with the highest t-statistic in absolute value. Reported in Table 7.3 (appendices) following the Chen and Liu (1993) procedure, outliers are detected through inner and outer loops. The outlier detection process shows that there are two types of outliers found, namely the AO and TC. Iterations around the function locate outliers until no additional outliers are found or the maximum number of iterations is reached. After each iteration, the effect of the outliers on the residuals of the fitted model is removed and the t-statistics are obtained again for the modified residuals. No model selection or refit of the model is conducted within this loop. At the end of each iteration, the detected outliers are removed from the original data and a new check for the presence of outliers is carried out. Then, the time series model is fitted (or selected) again for the adjusted series, and a new search for outliers is executed. The outer loop stops when no additional outliers are detected. Table 7.3 (appendices) confirms that the outliers highlighted with asterisks in Table 7.2 (appendices) are the most significant values (i.e. $cval =$ 3.5).

4.9.3. The effects of outliers

The purpose of this section was to graphically show the effects of outliers on stock returns. In the previous tables, there are two types of outliers found in the data namely the AO and TC. This section examines the effects of the two outliers on observation values and their respective residuals. Figure 4.13 shows the data for JSE top 40 index the plot shows the measure of outlier effects, $\hat{\omega}_{TP} = TC$ (t =1... 1000). It is clear that all $\hat{\omega}_{TC_t}$ lie in the interval [-4, 4]. As it was observed in Table 7.3 (appendices) there are five TC outliers of size $\omega = 3.5$. Consequently, this indicates that an ARMA or ARCH and GARCH type model will be able to isolate time point at which TC occurs. Figure 4.9 shows the original data (grey line), the adjusted series (blue line), the location of the detected outliers (red points) and their estimated effects (red line) for the return series.

The process was intended to illustrate the behaviour of the AO and TC influencing the original series. The influence of an outlier is represented by a red dot. Note that not all outliers are represented in the diagram. Therefore, due to the nature of the outliers detected the effect of outliers is not permanent as it affects a single observation at a particular time.

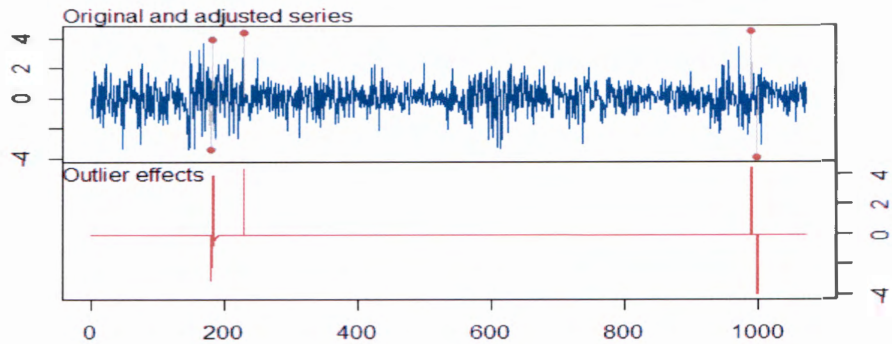


Figure 4.13: Effects of outliers AO

Stage II: Remove outliers

The detection of outliers is already made in stage I. Thus the choice of the model and the parameter estimates on the observed series may be affected by the deletion of outliers. Therefore, if any of the outliers turn to be non-significant then they are removed from the set of potential outliers. An outlier free data was used for further analysis. Thus the new ARMA (*p, q*), ARCH (*p*) and ARMA (*p, q*)-GARCH (*p, q*) type models were modeled with 1312 observations after removal of 18 significant outliers from a set of data.

Stage III: iterate stages I and II for the adjusted series (model fit)

4.9.4. ARMA estimation

After the outliers were removed the ACF and PACF were plotted and shown as in Figure 4.14a and 4.14b. According to the results, AR (2), MA (2) are seen to be significant in this section.

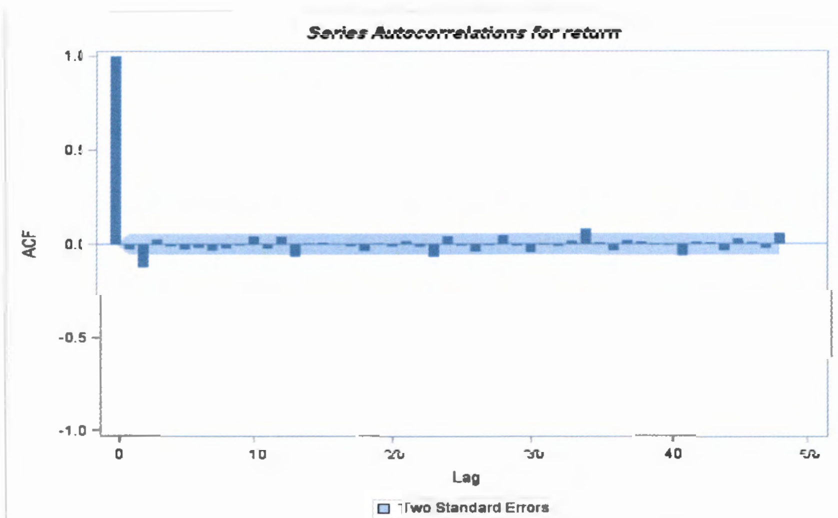


Figure 4.14a: ACF plot (outlier free)

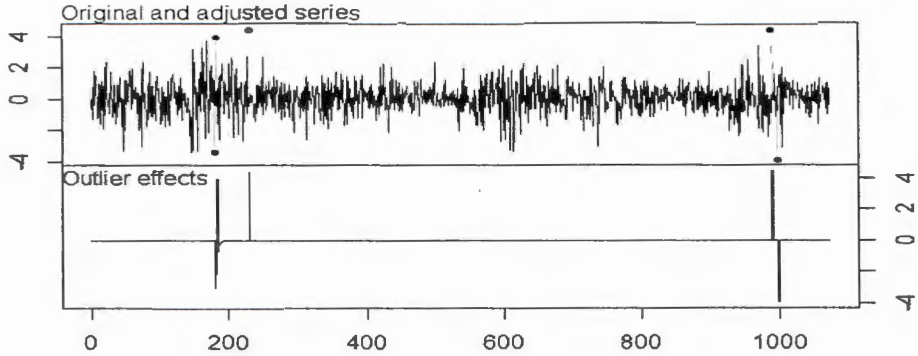


Figure 4.13: Effects of outliers AO

Stage II: Remove outliers

The detection of outliers is already made in stage I. Thus the choice of the model and the parameter estimates on the observed series may be affected by the deletion of outliers. Therefore, if any of the outliers turn to be non-significant then they are removed from the set of potential outliers. An outlier free data was used for further analysis. Thus the new ARMA (p, q), ARCH (p) and ARMA (p, q)-GARCH (p, q) type models were modeled with 1312 observations after removal of 18 significant outliers from a set of data.

Stage III: iterate stages I and II for the adjusted series (model fit)

4.9.4. ARMA estimation

After the outliers were removed the ACF and PACF were plotted and shown as in Figure 4.14a and 4.14b. According to the results, AR (2), MA (2) are seen to be significant in this section.

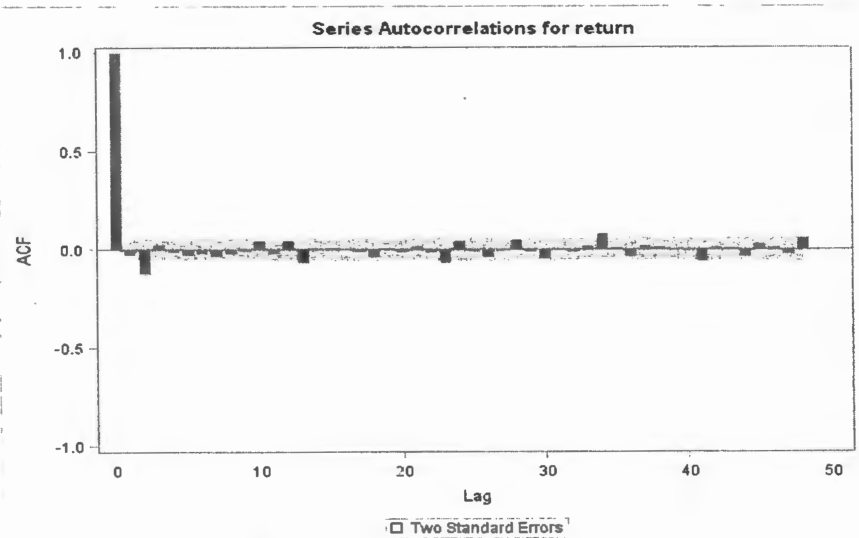


Figure 4.14a: ACF plot (outlier free)

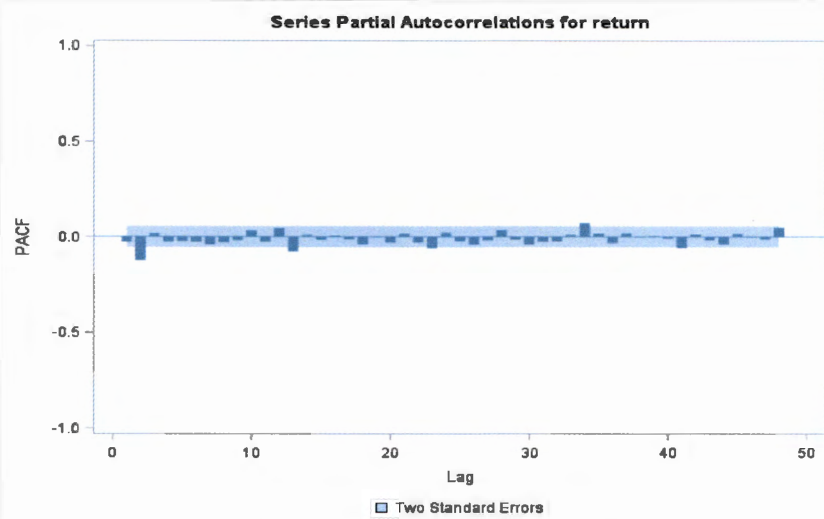


Figure 4.14b: PACF plot (outlier free)

Table 7.4 (appendices) shows the competing ARMA model when outliers are removed and ARMA (0, 2) model was found to be the best fitting model based on AIC and SBC when outliers are removed. Therefore, parameters for ARMA (0, 2) were estimated and the results are presented in Table 7.4 (appendices).

4.9.5. Model identification ARMA (0, 2)

Table 4.10: ARMA (0, 2) model summary

Statistics	Estimation
Constant	0.03657 (0.1446)*
MA (1)	0.02749 (0.3160)*
MA (2)	0.13006 (<0.0001)***
Durbin-Watson	2.075892
AIC	395.062
SBC	3938.149
Log likelihood	-1958.801

The returns can be written as:

$$(1-0.03647-0.02749(L)-0.13006(L)(1-L)X_t = \varepsilon_t \qquad (4.3)$$

where L is the lag operator $X_t = return_t$, and $\varepsilon_t \sim iid$ (0, 1.161378) as demonstrated in Table 4.10.

4.9.6. Model estimation

This section presents the results for the ARCH (2) model estimated using outlier free data shown in Table 4.11 below. ARCH (2) is chosen because the most significant lag is lag 2.

Table 4.11: ARCH and GARCH uncontaminated model summary

Parameters	ARCH (2)	ARMA(0, 2)-GARCH (1, 1)	ARMA(0, 2)-GARCH (1, 1) Model t-distributed errors	ARMA(0, 2)-GARCH (1, 1) GED distributed errors	ARMA(0, 2)-GARCH (1, 1) skewed t-distributed errors	ARMA(0, 2)-GARCH (1, 1) GED t-distributed errors
μ	0.0391 (0.1432)*	0.05823 (0.0158)**	0.062072 (0.0063)***	0.075198 (0.0053)***	0.04979 (0.0445)**	0.05057 (0.0493)**
ω	0.7543 (<.0001)***	0.1014 (0.0004)***	-0.099623 (0.0003)***	0.0206 (0.0208)**	0.015992 (0.0218)**	0.01853 (0.0185)**
ϑ_1	0.1691 (<.0001)***	0.025665 (0.0001)***	0.017276 (0.0301)**	-0.04169 (0.1420)*	-0.04698 (0.0775)*	-0.0482 (0.0869)*
ϑ_2	0.2037 (<.0001)***	0.088371 (0.0000)***	0.071488 (0.0000)***	-0.10489 (0.0005)***	-0.1069 (0.0002)***	-0.1113 (0.0001)***
β_1		0.891771 (0.0000)***	0.914441 (0.0000)***	0.90703 (0.0000)***	0.916 (0.0000)***	0.91048 (0.000)***
α				0.07619 (0.0002)***	0.07069 (0.00001)***	0.07395 (0.0000)***
df			8.764220 (0.0000)***			
N				1.4831 (0.0000)***	8.7826 (0.000)***	1.5043 (0.0000)***
Λ					-0.16055 (0.000)***	-0.1586 (0.0033)***

Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively. Values in the parentheses are the probabilities of the test statistics of the estimated parameters.

A simple ARCH (2) model is a model which is really AR's in disguise as illustrated in Table 4.12. Thus the AR (2) model is statistically significant at 1%, 5% and 10% but insignificant at 10% for the intercept indicating that the intercept adds little or no value to ARCH (2) model estimation. The parameter estimates can be represented as (μ , ω , α_1 , and α_2) of ARCH (2) model are 0.0391, 0.7543, 0.1691, and 0.2037, respectively.

The skewed distribution, show that the densities of the two Student's t-distributions (symmetric and skewed) clearly outperform the normal distribution. The parameters of the GARCH-GED (Skewed and t-GED) models are all statistically significant at the 1% level, indicating that the distributions of the returns series are right-skewed and fat-tailed. Deceptively, each GED distribution is skewed towards the right, and is thinner and more platykurtic than the normal distribution.

Table 4.12: GARCH type models with outlier free model estimation summary

Parameter	ARMA(0, 2)-EGARCH (1, 1)	ARMA(0, 2)-APARCH (1, 1)	ARMA(0, 2)-TGARCH (1, 1)	ARMA(0, 2)-GARCH-GJR (1, 1)
μ	0.0126 (0.5861)*	0.0163 (0.4870)*	0.0107 (0,6449)*	0.01964 (0,4292)*
θ_1	-0.005998 (0.8384)*	0.00681 (0.8266)*	0.001014 (0,9721)*	-0.00561 (0,8458)*
θ_2	0.0942 (0.0007)***	0.0977 (0.0006)**	0.1042 (0,0002)***	-0.0998 (0,0005)***
ω	0.000901 (0.8133)*	0.0218 (<0.0001)***	0.0182 (<0,0001)***	0.02182 (0,0226)**
α	0.0541 (<0.0001)***	0.0592 (0.7771)***	-0.0412 (0,0001)***	0.036161 (0,0009)***
β	0.9811 (<0.0001)***	1.000 (0.8486)***	0.1565 (<0,0001)***	0.99933 (0,000)***
γ	-2.6465 (0.0001)***	0.9210 (<0.0001)***	0.9454 (<0,0001)***	0.90835 (0,000)***
δ		0.6496 (<0.0001)***		

Notes: ***, **, * denotes significance at 1%, 5% and 10% levels respectively. Values in the parentheses are the probabilities of the test statistics of the estimated parameters.

Table 4.12 shows the parameter estimates of the ARMA (0, 2)-GARCH (1, 1) estimation, ARMA (0, 2)-GARCH (1, 1) t distributed errors, ARMA (0, 2)-TGARCH (1, 1), ARMA (0, 2)-EGARCH (1, 1), ARMA (0, 2)-APARCH (1, 1) and ARMA (0, 2)-GJR-GARCH (1, 1). The parameter estimates shown are reported when outliers are removed from the data. According to the results, the ARMA (0, 2)-GARCH (1, 1) is statistically significant at all levels of significance except for the intercept which is significant at 5% level of significance. The ARMA (0, 2)-GARCH (1, 1) t distributed errors is also significant at all levels of significance except for α_1 which is significant at 5%. The finding suggests that the three models are significant. The intention of this study was to use the most appropriate and efficient model as possible, hence the information criteria is used to select the best among the three models.

Like in GARCH estimation with outliers, the intercept in the mean equation adds little value as it is not significantly different from zero for all models in Table 4.12. All GARCH terms are significant only at the 10% level. There is little difference in the GARCH coefficients of the outlier-contaminated and outlier-free model estimation summary in Table 4.6 and 4.13 respectively. The estimates remain significant at the 10% level, whilst the other GARCH terms are statistically significant from zero.

4.9.7. Model selection

Table 4.13: ARCH and GARCH Outlier Uncontaminated Model selection summary

criterion	ARCH (2)	ARMA(0, 2)-GARCH (1, 1)	ARMA(0, 2)-GARCH (1, 1) Model t-distributed	ARMA(0, 2)-TGARCH	ARMA(0, 2)-EGARCH	ARMA(0, 2)-APARCH	ARMA(0, 2)-GJR-GARCH(1, 1)	ARMA(0, 2)-GARCH (1, 1) GED distributed	ARMA(0, 2)-GARCH (1, 1) skewed t-distributed	ARMA(0, 2)-GARCH (1, 1) skewed GED distributed
MSE	1.17866	1.16091	1.16107	1.16172	1.16272	1.16140	1.16283	1.658	1.1644	1.648
SBC	3868.030	3776.4873	3737.093	3704.638	3704.312	3721.042	3710.516	3742.3529	3722.108	3730.8389
AIC	3847.322	3745.4114	3700.838	3668.383	3668.057	3679.607	3674.272	3706.109	3680.686	3689.417
HQC	3858.971	3750.819	3708.097	3677.283	3675.038	-	3687.865	3719.702	3696.222	3704.952
MAE	0.80501	0.79761	0.7968157	0.79974	0.80025	0.799247	0.79998	0.796658	0.79845	0.79689
MAPE	106.782	118.7425	120.2215	111.0643	110.0026	110.6335	112.058	120.2358	111.655	110.569
Log-Likelihood	-1919.66	-1886.71	-1843.419	-1827.192	-1827.029	-1831.804	-1830.136	-1846.054	-1832.343	-1836.708
Rank	(10)	(9)	(7)	(2)	(1)	(4)	(3)	(8)	(5)	(6)

Table 4.13 shows the model selection summary for ARCH (2), ARMA (0, 2)-GARCH (1, 1) and ARMA (0, 2)-GARCH (1, 1) t-distributed errors and ARMA (0, 2)-GARCH (1, 1) type models together with their ranks. The results suggest ARMA (0, 2)-EGARCH (1, 1) as an adequate model based on the information criterion and the model performance evaluations. Therefore this model was used to make forecasts for outlier free series. The returns of ARMA (0, 2)-EGARCH (1, 1) are formulated as:

$$r_t = 0.0126 + \varepsilon_t$$

$$\ln \sigma_t^2 = -0.005998 + 0.0942\varepsilon_{t-1} + 0.000901\varepsilon_{t-2}^2 + 0.0541h_{t-1} + 0.9811h_{t-1}^2 - 2.6465\gamma \qquad (4.4)$$

$$\lambda = 1.0352$$

The next section evaluates the model robustness.

4.10. Outlier-free Model Robustness results

The results presented here are obtained from outlier-free models.

4.10.1. Test for heteroscedasticity

Table 4.15 shows the ARCH MLE free from outliers. The table shows the Q-statistics test for changes in variance across time by using windows ranging from 1 through 12. The tests strongly suggest there is heteroscedasticity, with $p - values < 0.05$ for all lag windows.

Table 4.14: Tests for ARCH Disturbances Based on Residuals

Order	Q	Pr > Q	LM	Pr > LM
1	54.6393	<.0001	54.5658	<.0001
2	87.1424	<.0001	71.4727	<.0001
3	91.5819	<.0001	71.4729	<.0001
4	98.0911	<.0001	73.5355	<.0001
5	109.2378	<.0001	79.1121	<.0001
6	116.3997	<.0001	80.1876	<.0001
7	119.9454	<.0001	80.2803	<.0001
8	130.7352	<.0001	85.4063	<.0001
9	133.8257	<.0001	85.4281	<.0001
10	150.5647	<.0001	92.7498	<.0001
11	163.3177	<.0001	95.8003	<.0001
12	179.9652	<.0001	99.2761	<.0001

The *p*-value of LM statistics for the lag windows fails to reject the null hypothesis in the absence of ARCH effects. The results for the first lag have improved from being insignificant (as seen in Table 4.8 in the presence of outliers) to being significant (outlier free). Therefore, under the methodological specifications *reject* the null hypothesis of homoscedasticity in favour of the alternative hypothesis of heteroscedasticity.

4.10.2. Test for statistical independence

Table 4.15: The runs test of independence

RUNS	Pr > |RUNS|
0.6269	0.5307

The observed Runs test statistic 0.6269 with its associated probability (0.5307) is insignificant at 1%, 5% and 10% significance levels shown in Table 4.16 above. There is evidence that the majority of the lag orders are significant resulting in the rejection of the null hypothesis. Therefore, the residuals of the ARMA (0, 2)-EGARCH (1, 1) are not independent in favour of the alternative hypothesis. The residual autocorrelations usually fall within 95% confidence band, which is located roughly around zero band at $\frac{2}{\sqrt{n}}$ as shown in Figure 4.15 below. Thus the residuals are significant since most of the autocorrelations are located roughly at $\frac{2}{\sqrt{n}} = 0.055$.

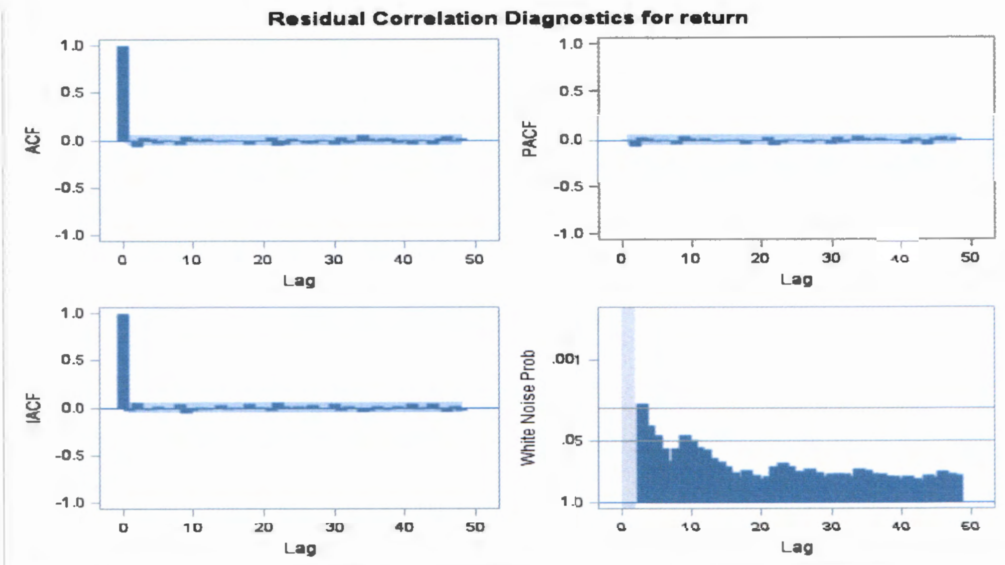


Figure 4.15: Residual autocorrelation plot (outlier free)

4.10.3. Test for normality

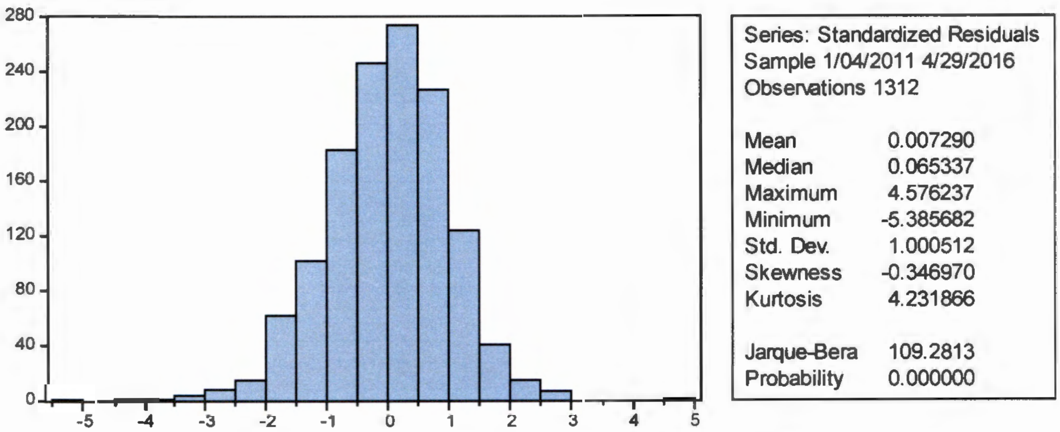


Figure 4.16: Jarque-Bera test for normality

The residuals of the ARMA (0, 2) - EGARCH (1, 1) are not normally distributed as shown in Figure 4.16. This was expected from financial data.

4.11. Out-of-sample forecasts

This section compares the volatility specification by diagnostic checking. The ARMA (0, 2)-EGARCH (1, 1) was selected as the best model to compare out-of-sample performance.

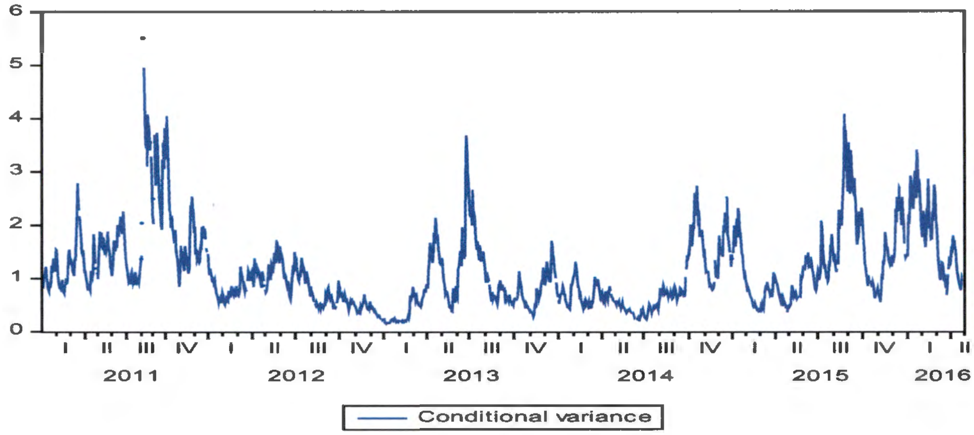


Figure 4.17: GARCH variance forecasts

Figure 4.17 shows the ARMA (0, 2) - GARCH (1, 1) conditional variance forecasts. If only daily closing prices are presented, the only volatility substitution that fully accentuates the time varying property is the daily squared returns (Wennström, 2014). Unconditional variance plot in Figure 4.17 shows that the squared returns provide a highly constant volatility for the JSE returns, indicating constant forecasting proxy which is significantly close to 0.

4.12. Out-of-sample forecasts for ARMA (0, 2) - GARCH (1, 1) t-distribution model

Outlier free ARMA (0, 2) - GARCH (1, 1) t-distribution model forecasts are presented in Figure 4.18.

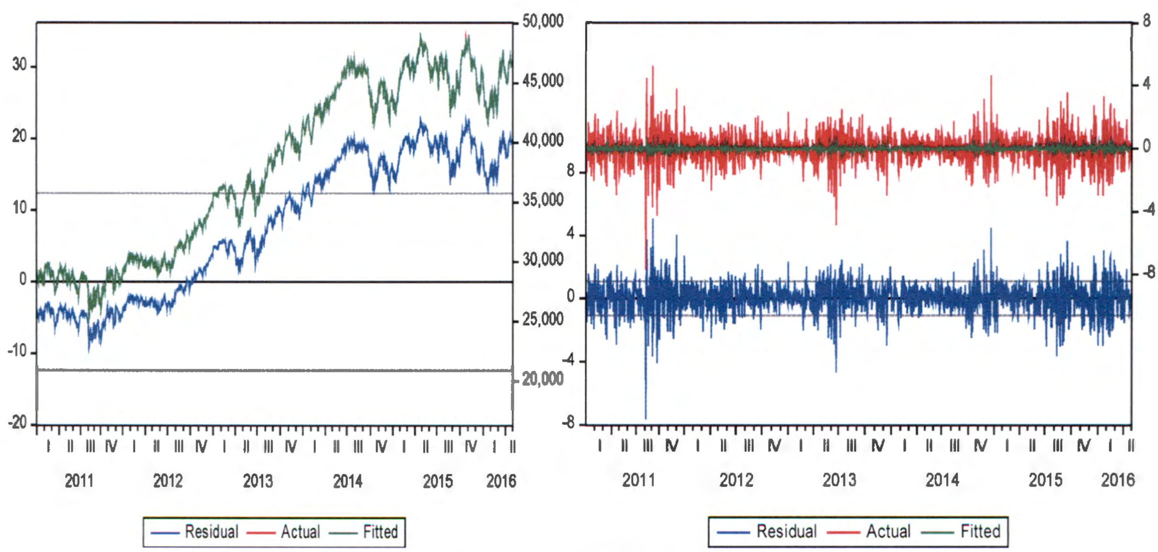


Figure 4.18: Forecast plot for ARMA (0, 2)-EGARCH (1, 1)

The forecasts in Figure 4.18 show that the actual and fitted series move very closely together. The MAE for outlier free model is low as compared to the "original series" meaning that the outlier free model gives accurate forecasts. The RSME/MSE in the "original series" is lower thus according to Brooks (2008) the forecast with the smallest RMSE and MAE provides the most accurate forecasts. MAPE shows that the original series is the best since the MAPE value is closest to 100. Therefore, the error statistics in Table 4.8 and 4.14 cannot give a clear distinction between the competing models, therefore SBC and AIC will be used to select the best model. This is a confirmation that outlier free ARMA (2, 2) - EGARCH (1, 1) model is good for the data. By comparing the criterion in Table 4.7 and 4.13, the information criterion for outlier free model is much lower than that in the "original series" in terms of both SBC and AIC, indicating that the outlier free GARCH model is good together with a small forecasting error compared to the "original series". In conclusion, the outlier free data provide good forecasts to predict volatility of JSE top 40 index.

4.13. Chapter summary

Chapter four presented the results pertaining to the volatility of the JSE. The performance of the data and stationarity tests on the data were presented. The overall results revealed the properties of financial data such as excess volatility, volatility clustering, excess kurtosis and non-normality. The out-of-sample forecasting evaluations indicated that outlier contaminated ARMA (2, 2)-EGARCH (1, 1) model and outlier free ARMA (0, 2)-EGARCH (1, 1) model estimated most precise volatility forecasts in the original series.

Presented in the next chapter are the conclusions and recommendations of the study.

CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

5.1. Introduction

The discussion presented in this chapter entails the final conclusions deduced from data analysis presented in Chapter four. The main goal of this study was to assess the performance the ARCH and GARCH type models before and after the outlier(s) have been removed from the data. This chapter is organised as follows: Section 5.2 entails the discussion with regards to the objectives outlined in Chapter one and deductions thereof; Section 5.3 outlines the findings, Section 5.4 outlines the recommendations based on the findings, Section 5.5 highlights the study limitations and Section 5.6 provides the chapter summary.

5.2. Objectives and Conclusions

This section is aimed at outlining the objectives and the following findings were observed.

5.2.1. Objective 1: To develop a standard time series predictive ARCH and GARCH type model before removing the outliers.

The GARCH (1, 1) type models were applied to the JSE top 40 index to predict future stock returns. It was found that the outlier contaminated ARMA (2, 2)-EGARCH (1, 1) produced better results than the standard ARMA (2, 2)-GARCH (1, 1) model with normally distributed errors. From this finding, it can be concluded that the EGARCH accommodated for the stylised facts contained in the stock returns. The predictive power of the chosen model was chosen by means of predictive accuracy models such as the RMSE, MAE and MAPE. The EGARCH (1, 1) model produced good out-of-sample-forecasts of the stock returns since these forecasts moved along with the actual data of the returns. Thus, even though a standard model was produced, it clearly captured the predictive accuracy to reach the study objective.

5.2.2. Objective 2: To determine the type of outliers present in the proposed data.

An outlier detection procedure proposed by Chen & Liu (1993) was used to detect the types of commonly found outliers in financial time series data namely the TC, AO, IO and LS. The outlier detection procedures aimed to identify the type of outliers present in the JSE stock returns. The study searched for the outliers and two types of outliers found in the data; namely the AO and the LS. Through looping process, it was later found that the TC and AO were present in the data. This was done by identifying outliers which were more significant using

3.5 as a standard benchmark. Therefore, by detecting the TC which is a generalisation of the AO and LS, indicated that the initial influence of an AO is also passed on to the following observations. This meant that the unusual spikes in the data were not permanent, hence the increase or decrease in stock prices affected only a single day, but not consecutive days.

5.2.3. Objective 3: To determine the effects of outliers in the ARCH and GARCH type models.

The 18 outliers that were identified and confirmed through inner and outer loop outlier identification were removed from the data and a new model was fitted with outlier free data. The new results were produced with modified data. It was found that without outliers, the ARCH model parameters improved and the LM became significant in all lags. Likewise, the ARMA (0, 2)-EGARCH (1, 1) model parameters gave more significant values tested at 1%. The ARMA (0, 2)-EGARCH (1, 1) still proved to be a better model among others. Even though outliers were removed, the unique stylised fact about financial time series called excess kurtosis was still present. The forecasts produced by this model moved parallel with the actual values but not as close as those produced with outlier contaminated data. Therefore, one could conclude that the exclusion of outliers in financial time series data has a questionable effect on the forecasts.

5.2.4. Objective 4: To use the findings of the study in formulating suggestions with regards to policy formulation and future studies.

Forecasts for JSE obtained through ARMA (2, 2)-GARCH (1, 1) approach is vital for formulating appropriate strategies. This type of analysis would go a long way to derive appropriate decisions on several issues at various times and help investors and portfolio managers decide whether or not to invest in the stock exchange. This could assist policy makers to make correct decisions from time to time and well in advance in order to meet the targets set in their organizations. All this would ultimately result in efficient management on a sound statistical basis.

5.3. Findings

The current study evaluated the JSE top 40 index volatility in SA. The analysis took into consideration the objectives as outlined in Chapter one, and thus a quantitative research approach was used to conduct the study. Secondary 5-day daily data from 03 January 2011 till 21 April 2016 was obtained from the JSE databases. The study generated the results by SAS, Gretl and E-views packages. Stationarity testing on the series was performed using the ADF test. The series was found to be non-stationary at their level but stationary once the percentage change log returns were formulated. The PACF and ACF were employed to select an optimal lag of two which proved to be helpful in the ARCH model estimation where ARMA (2, 2) and ARMA (2, 2)-EGARCH (1, 1) showed to be more significant at 1%.

As a prerequisite, the variables were further checked for normality using Jarque-Bera test for normality and statistical independence using runs test for independence. The independence assumption was the only one not violated. Due to the size of the data, the normality assumption was also protected according to the central limit theorem. For primary analyses, this study applied ARMA (2, 2), ARCH (2), ARMA (2, 2)-GARCH (1, 1), ARMA (2, 2)-GARCH (1, 1) t-distribution, ARMA (2, 2)-EGARCH (1, 1), ARMA (2, 2)-TGARCH (1, 1), ARMA (2, 2)-APARCH (1, 1) and ARMA (2, 2)-GJR-GARCH (1, 1) models to assess stock return volatility in SA. This was done to determine the model which reveals high stock return volatility. ARMA (2, 2)-EGARCH (1, 1) outperformed the other models.

The goodness-of-fit tests were used to confirm the robustness of the selected models. The models were selected using the information criteria. The selected model was used to produce forecasts for the data. The forecasts produced generated few errors according to the forecast error measures MAE, MAPE, and RMSE. The forecasts produced moved along with the original series confirming the efficiency of the model. Even though the ARMA (2, 2)-GARCH (1, 1) model was also found to be fit and stable for the data according to the diagnostic checking. An innovation analysis was conducted to check whether or not from the data conforms to a t-distribution or a Gaussian distribution.

The second set of analysis was done on outlier free GARCH models. Upon doing outlier analysis, a total of 99 outliers were identified. This was a literal indication that the JSE top 40 market data is relatively volatile. The *p*-value of ARCH model significantly changed when the data were cleaned of outlying observations. The parameter were inflated from 0.804538 to 0.91441 indicating that presence of outliers in the have an effect on the results. Four types of outliers were tested namely the AO, IO, TC and LS. Two outliers were detected; namely the AO (16) and TC (5) were found to be more significant as their critical values were greater or equal to 3.5. The results showed new improved model also with improved parameter estimates as compared to the model with outlier contaminated data.

However, the α_1 estimates were essentially bigger while the β_1 estimates were considerably smaller, demonstrating that innovations have huge ARCH effects but smaller GARCH effects. The outliers were also tested and the ARMA (0, 2)-EGARCH (1, 1) also proved to be better. The removal of outliers had pros and cons in the results. Model parameter estimates were improved but the forecasts generated were found not as good as those generated from outlier contaminated data.

5.4. Recommendations

This study is not definitive, rather it prepares the way for future research. This study utilised daily data but subsequent studies can be conceded out using weekly or monthly data given that the data is accessible on these interims. The results may be compared with the current study to check for biasness. Thus, when fitting GARCH type models to conditionally heteroscedastic series it is recommended to check if the series is affected by outliers or not. If the effects of outliers prove to influence the results badly, the researcher is advised to be cautious about removing such observations as insignificant results may be obtained from the data.

There is a bias of outlier methods, therefore is no permanent rule of thumb on what to do with outliers in financial time series data. Consequently, it is up to the analyst to decide what to do, how and why. Thus, it is best to detect for outliers in any dataset and then concentrate on the normal procedures. There are several procedures such as Wavelet-Based Detection of Outliers, Efficient Clustering-Based Outlier Detection Algorithm, Iterative outlier detection proposed in the literature to test for outliers in the presence of GARCH effects and the comparison of these procedures should be the subject of further research.

5.5. Study scope Limitations

Though different GARCH type models are available to forecast volatility in the stock market, the current study was limited to the use of traditional ARCH and GARCH (*p, q*) type models. Even though the data used was obtained from the South African database, internationally approved methods were used. Therefore the results of this study may be referred to by other markets internationally. Furthermore, this study restricted the attention to the univariate parametric GARCH (*p, q)* type model only since it is often used in practice (Teräsvirta, 2009). The use of univariate model GARCH models was also informed by the fact that modeling of stock returns does not search for the possibility of effects and interactions between the X and Y-share markets.

5.6. Summary

This chapter provided the summarised conclusions and recommendation of this study with specific reference to the objectives and research problem. The main purpose of this study was to compare the ARCH and GARCH type models with and without outliers. The results from both data types proved to almost similar though statistically outlier contaminated data models were found to be of more interest than the other. This could imply that the outliers identified do not necessarily serve as influential observations since their removal posed some pros and cons to the result.

Appendices

Table 7.1: Competing model summary with outlier contaminated data

Model	AIC	SBC
ARMA (0,1)	3948.627	3953.82
ARMA (0,2)	3942.102	3952.486
ARMA (1,0)	3948.867	3954.059
ARMA (1,1)	3940.526	**3950.911**
ARMA (1,2)	3943.591	3959.167
ARMA (2,0)	3942.776	3953.161
ARMA (2,1)	3944.338	3959.915
ARMA (2,2)	**3939.571**	3960.339

Table 7.2: Outlier details

Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
991	18-DEC-2014	Additive	4.53805	24.69	<.0001*
230	30-NOV-2011	Additive	4.16483	20.98	<.0001*
1000	05-JAN-2015	Additive	-3.87796	18.26	<.0001*
147	04-AUG-2011	Additive	-3.84758	17.98	<.0001*
1151	12-AUG-2015	Additive	-3.67482	16.46	<.0001*
1176	16-SEP-2015	Additive	3.61362	15.95	<.0001*
170	07-SEP-2011	Additive	3.49695	14.95	0.0001*
184	27-SEP-2011	Additive	3.47061	14.75	0.0001*
181	22-SEP-2011	Additive	-3.46914	14.76	0.0001*
616	20-JUN-2013	Additive	-3.46561	14.76	0.0001*
156	18-AUG-2011	Additive	-3.43613	14.55	0.0001*
49	10-MAR-2011	Additive	-3.41307	14.37	0.0002*
973	21-NOV-2014	Additive	3.38122	14.14	0.0002*
163	29-AUG-2011	Additive	3.29155	13.40	0.0003
610	11-JUN-2013	Additive	-3.28691	13.42	0.0002
1007	14-JAN-2015	Additive	-3.24039	13.09	0.0003
151	11-AUG-2011	Additive	3.19701	12.75	0.0004
149	08-AUG-2011	Additive	-3.13899	12.34	0.0004
1264	22-JAN-2016	Additive	3.09141	12.05	0.0005
1159	24-AUG-2015	Additive	-3.08710	12.24	0.0005
1250	04-JAN-2016	Additive	-3.06103	12.06	0.0005
1160	25-AUG-2015	Additive	2.97743	11.42	0.0007
1262	20-JAN-2016	Additive	-2.90812	10.90	0.0010
739	12-DEC-2013	Additive	-2.88564	10.77	0.0010
1126	07-JUL-2015	Additive	-2.87365	10.71	0.0011
1280	15-FEB-2016	Additive	2.87224	10.78	0.0010
1128	09-JUL-2015	Additive	2.79143	10.20	0.0014
1240	17-DEC-2015	Additive	2.75393	9.93	0.0016
1165	01-SEP-2015	Additive	-2.74782	9.90	0.0017
223	21-NOV-2011	Additive	-2.70904	9.62	0.0019
75	18-APR-2011	Additive	-2.70204	9.57	0.0020
604	03-JUN-2013	Additive	-2.67103	9.42	0.0021
1289	26-FEB-2016	Additive	2.64858	9.29	0.0023
250	03-JAN-2012	Additive	2.64554	9.30	0.0023
193	10-OCT-2011	Additive	2.63656	9.24	0.0024

1168	**04-SEP-2015**	Additive	-2.58813	8.99	0.0027
1167	**03-SEP-2015**	Additive	2.56281	8.83	0.0030
1276	**09-FEB-2016**	Additive	-2.53237	8.64	0.0033
985	**09-DEC-2014**	Additive	-2.51568	8.56	0.0034
631	**11-JUL-2013**	Additive	2.51407	8.58	0.0034
228	**28-NOV-2011**	Additive	2.51406	8.59	0.0034
315	**04-APR-2012**	Additive	-2.50579	8.56	0.0034
77	**20-APR-2011**	Additive	2.49192	8.55	0.0034
618	**24-JUN-2013**	Additive	-2.47676	8.51	0.0035
1232	**04-DEC-2015**	Additive	-2.46516	8.47	0.0036
958	**31-OCT-2014**	Additive	2.44487	8.40	0.0038
597	**23-MAY-2013**	Additive	-2.43931	8.36	0.0038
189	**04-OCT-2011**	Additive	-2.43888	8.48	0.0036
206	**27-OCT-2011**	Additive	2.42342	8.41	0.0037
168	**05-SEP-2011**	Additive	-2.41215	8.38	0.0038
1188	**05-OCT-2015**	Additive	2.40017	8.37	0.0038
943	**10-OCT-2014**	Additive	-2.39003	8.37	0.0038
186	**29-SEP-2011**	Additive	-2.38674	8.37	0.0038
500	**02-JAN-2013**	Additive	2.36028	8.23	0.0041
1279	**12-FEB-2016**	Additive	2.35297	8.19	0.0042
23	**02-FEB-2011**	Additive	2.34933	8.17	0.0043
627	**05-JUL-2013**	Additive	-2.34295	8.14	0.0043
1284	**19-FEB-2016**	Additive	-2.25833	7.62	0.0058
1244	**23-DEC-2015**	Additive	2.25767	7.65	0.0057
52	**15-MAR-2011**	Additive	-2.25162	7.61	0.0058
202	**21-OCT-2011**	Additive	2.23809	7.53	0.0061
948	**17-OCT-2014**	Additive	2.21870	7.42	0.0064
1278	**11-FEB-2016**	Additive	-2.21834	7.42	0.0065
427	**14-SEP-2012**	Additive	2.19003	7.30	0.0069
680	**19-SEP-2013**	Additive	2.16779	7.15	0.0075
1253	**07-JAN-2016**	Additive	-2.16674	7.17	0.0074
1212	**06-NOV-2015**	Additive	-2.15494	7.10	0.0077
383	**13-JUL-2012**	Shift	0.06359	7.21	0.0072*
979	**01-DEC-2014**	Additive	-2.22032	7.70	0.0055
946	**15-OCT-2014**	Additive	-2.19111	7.50	0.0062
209	**01-NOV-2011**	Additive	-2.14849	7.22	0.0072
930	**22-SEP-2014**	Additive	-2.14371	7.25	0.0071
222	**18-NOV-2011**	Additive	-2.12567	7.15	0.0075
536	**21-FEB-2013**	Additive	-2.12443	7.16	0.0075
215	**09-NOV-2011**	Additive	-2.09190	6.95	0.0084
213	**07-NOV-2011**	Additive	2.12944	7.22	0.0072
972	**20-NOV-2014**	Additive	-2.08369	6.92	0.0085
295	**06-MAR-2012**	Additive	-2.07503	6.89	0.0087
1180	**22-SEP-2015**	Additive	-2.06265	6.84	0.0089
1097	**26-MAY-2015**	Additive	-2.06075	6.87	0.0088
1006	**13-JAN-2015**	Additive	2.05944	6.88	0.0087
176	**15-SEP-2011**	Additive	2.05349	6.84	0.0089
718	**13-NOV-2013**	Additive	-2.04812	6.90	0.0086
652	**12-AUG-2013**	Additive	2.03811	6.83	0.0090
1241	**18-DEC-2015**	Additive	-2.03513	6.81	0.0090
1137	**22-JUL-2015**	Additive	-2.03435	6.84	0.0089
747	**27-DEC-2013**	Additive	2.02576	6.81	0.0091
434	**26-SEP-2012**	Additive	-2.01891	6.77	0.0093
167	**02-SEP-2011**	Additive	-2.01729	6.79	0.0091
165	**31-AUG-2011**	Additive	2.13454	7.64	0.0057
602	**30-MAY-2013**	Additive	2.01561	6.87	0.0088
628	**08-JUL-2013**	Additive	1.99816	6.75	0.0093
1156	**19-AUG-2015**	Additive	-1.98077	6.67	0.0098

544	**05-MAR-2013**	Additive	1.97152	6.66	0.0099
673	**10-SEP-2013**	Additive	1.97128	6.69	0.0097
562	**03-APR-2013**	Additive	-1.96545	6.68	0.0097
950	**21-OCT-2014**	Additive	1.96132	6.67	0.0098
664	**28-AUG-2013**	Additive	-1.95981	6.68	0.0097
173	**12-SEP-2011**	Additive	-1.95854	6.69	0.0097

Table 7.3: Outlier detection: inner and outer loops

	OUTER LOOP				**INNER LOOP**		
TYPE	**TIME ID**	**COEFHAT**	**TSTAT**	**TYPE**	**TIME ID**	**COEFHAT**	**TSTAT**
AO	07-Sep-11	3.870239	4.418528	AO	10-Mar-11	-3.316543	-3.786392
AO	22-Sep-11	-3.73737	-4.266836	AO	04-Aug-11	-3.37602	-3.854295
AO	27-Sep-11	3.904602	4.45776	AO	18-Aug-11	-3.185753	-3.637073
AO	30-Nov-11	4.159271	4.748508	AO	29-Aug-11	3.079603	3.515885
AO	18-Dec-14	4.32413	4.936721	AO	07-Sep-11	3.870239	4.418528
AO	05-Jan-15	-3.699429	-4.22352	AO	22-Sep-11	-3.73737	-4.266836
TC	20-Jun-13	-2.43836	-4.006183	AO	27-Sep-11	3.904602	4.45776
AO	11-Aug-11	3.463976	4.014625	AO	30-Nov-11	4.159271	4.748508
TC	02-aug-211	-2.406181	-4.021456	AO	11-Jun-13	-3.235001	-3.693297
TC	05-Aug-11	-2.433364	-4.066888	AO	21-Nov-14	3.243048	3.702485
TC	04-Aug-11	-2.549548	-4.261066	AO	05-Jan-15	-3.699429	-4.22352
TC	05-Aug-11	2.682333	4.483131	TC	02-aug-211	-2.170064	-3.565378
TC	09-Aug-11	2.733591	4.568801	TC	05-Aug-11	-2.143677	-3.522024
AO	18-Aug-11	-3.651	-4.28634	AO	11-Aug-11	3.351046	3.825783
				TC	20-Jun-13	-2.43836	-4.006183
				AO	18-Dec-14	4.32413	4.936721
				AO	14-Jan-15	-3.275653	-3.843651
				TC	04-Sep-11	-2.153391	-3.636322

Table 7.4: Competing model summary uncontaminated data

Model	AIC	SBC
ARMA (0,1)	3944.17	3946.349
ARMA (0,2)	3922.733	3933.091
ARMA (1,0)	3941.475	3946.655
ARMA (1,1)	3930.791	3941.15
ARMA (1,2)	3924.247	3939.785
ARMA (2,0)	3923.307	3933.666
ARMA (2,1)	3924.573	3940.111
ARMA (2,2)	3926.195	3946.912

References

ABRAHAM, A., SEYYED, F.J. & ALSAKRAN, S.A. 2002. Testing the random walk behaviour and efficiency of the Gulf stock markets. *Financial Review*, 37(3):469-480.

ADAMI, R., GOUGH, O., MURADOGLU, G. & SIVAPRASAD, S. 2010. The leverage effect on stock returns. *European Financial Management, Portugal.*

AKOUEMO, H.N. & POVINELLI, R.J. 2014. Time series outlier detection and imputation. (*In.* 2014 IEEE PES General Meeting| Conference & Exposition organised by: IEEE. p.1-5.

ANDERSEN, T.G., BOLLERSLEV, T. & DIEBOLD, F.X. 2007. Roughing it up: Including jump components in the measurement, modelling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701-720.

AROURI, M., UDDIN, G.S., KYOPHILAVONG, P., TEULON, F. & TIWARI, A.K. 2014. Energy Utilization and Economic Growth in France: Evidence from Asymmetric Causality Test.

BABU, C.N. & REDDY, B.E. 2014. Selected Indian stock predictions using a hybrid ARIMA-GARCH model. (*In.* Advances in Electronics, Computers and Communications (ICAECC), 2014 International Conference on organised by: IEEE. p. 1-6.

BACH, D., 2003. *1001 Financial words you need to know*. OUP USA.

BAKER, M. & WURGLER, J. 2000. The equity share in new issues and aggregate stock returns. *The Journal of Finance*, 55(5):2219-2257.

BALKE, N.S. & FOMBY, T.B. 1994. Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9(2):181-200.

BERA, A.K. & HIGGINS, M.L. 1993. ARCH models: properties, estimation and testing. *Journal of economic surveys*, 7(4):305-366.

BERBEROGLU, B. ed., 2014. *The Global Capitalist Crisis and Its Aftermath: The Causes and Consequences of the Great Recession of 2008-2009.* Ashgate Publishing, Ltd.

BIERENS, H.J., 2004. Information criteria and model selection. *Manuscript, Penn State University.*

BLACK, F. & SCHOLES, M., 1973. The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), pp.637-654.

BLACK, F. 1976. The pricing of commodity contracts. *Journal of financial economics*, 3(1):167-179.

BOLLERSLEV, T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of econometrics*, 31(3):307-327.

BOLLERSLEV, T., ENGLE, R.F. & NELSON, D.B. 1994. ARCH models. *Handbook of econometrics*, 4:2959-3038.

BOUDT, K., DANIELSSON, J. & LAURENT, S. 2013. Robust forecasting of dynamic conditional correlation GARCH models. *International Journal of Forecasting*, 29(2):244-257.

BOX, G.E. & TIAO, G.C., 1975. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association, 70*(349), pp.70-79.

BROOKS, C. & CHONG, J. 2001. The Cross-Currency Hedging Performance of Implied Versus Statistical Forecasting Models. *Journal of Futures Markets*, *21*(11), pp.1043-1069.

BROOKS, C. 2008. RATS Handbook to accompany introductory econometrics for finance. *Cambridge Books.*

BROWNLEES, C.T., ENGLE, R.F. & KELLY, B.T. 2011. A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915.*

BUI, A.T. & JUN, C.-H. 2012. An Improved Iterative Procedure for Outlier Detection in Time Series. *Journal of Korean Institute of Industrial Engineers*, 38(1):17-24.

CÂMARA, D. & MANUEL, R. 2011. The price and volatility transmission of international financial crises to the South African equity market/Ricardo Manuel da Câmara. North-West University.

CARNERO, M., PENA, D. & RUIZ, E. 2007. Effects of outliers on the identification and estimation of GARCH models. *Journal of time series analysis*, 28(4):471-497.

CARNERO, M.A., PEÑA, D. & RUIZ, E. 2012. Estimating GARCH volatility in the presence of outliers. *Economics letters*, 114(1):86-90.

CATALÁN, B. & TRÍVEZ, F.J. 2007. Forecasting volatility in GARCH models with additive outliers. *Quantitative Finance*, 7(6):591-596.

CHANG, I. 1982. Outliers in time series. *Dissertation Abstracts International Part B: Science and Engineering [DISS. ABST. INT. PT. B- SCI. & ENG.]*, 43(5):1982.

CHANG, I., TIAO, G.C. AND CHEN, C., 1988. Estimation of time series parameters in the presence of outliers. *Technometrics*, *30*(2), pp.193-204.

CHARLES, A. & DARNÉ, O. 2005. Outliers and GARCH models in financial data. *Economics letters*, 86(3):347-352.

CHARLES, A. 2008. Forecasting volatility with outliers in GARCH models. *Journal of Forecasting*, 27(7):551-565.

CHEN, C. & LIU, L.-M. 1993. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284-297.

CHINOMONA, A. 2010. Time series modelling with application to South African inflation data. MSc. Thesis. University of Kwazulu–Natal, Pietermaritzburg, South Africa.

CHONG, C.W., AHMAD, M.I. & ABDULLAH, M.Y. 1999. Performance of GARCH models in forecasting stock market volatility. *Journal of Forecasting*, 18(5):333-343.

CONT, R. 2007. Volatility clustering in financial markets: empirical facts and agent-based models. Long memory in economics. Springer. p. 289-309.

CROMWELL, J.B., LABYS, W.C. & TERRAZA, M. 1994. *Univariate tests for time series models* (Vol. 99). Sage.

D'AGOSTINO, R.B. & STEPHENS, M.A. 1986. Goodness-of-fit techniques (Statistics, a series of textbooks and monographs). *Dekker*, 68:1.

DAVIES, L. & GATHER, U. 1993. The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423):782-792.

DICKEY, D.A. & FULLER, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), p.427-431.

DING, Z., GRANGER, C.W. & ENGLE, R.F. 1993. A long memory property of stock market returns and a new model. *Journal of empirical finance*, *1*(1), p.83-106.

DOORNIK, J.A. & OOMS, M. 2005. Outlier detection in GARCH models. *Econometrica: Journal of the Econometric Society*: 347-370.

EDERINGTON, L.H. & GUAN, W. 2006. Measuring historical volatility. *Journal of Applied Finance*, 16(1).

EDWARD, N. 2011. Modelling and Forecasting Using Time Series Garch Models: An Application of Tanzania Inflation Rate Data. *Master of Science, University of Dar es-Salaam, Tanzania.*

EINARSEN, R.H. 2014. A comparative study of volatility forecasting models.

ENDERS, W. 2010 *Applied Econometric Time Series*, 3rd Edition, Wiley.

ENGLE, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*: 987-1007.

ENGLE, R.F. AND NG, V.K., 1993. Measuring and testing the impact of news on volatility. *The journal of finance*, *48*(5), pp.1749-1778.

FAMA, E.F. 1965. The behaviour of stock-market prices. *The journal of Business*, 38(1):34-105.

FIGLEWSKI, S. & WANG, X., 2001. *VIs the kLeverage Effectla Leverage Effect.* V working paper, Department of Finance, New York University.

FOLEY, S.D., 2014. The impact of regulation on market quality.

FORNI, M. 2002. DP3236 Using Stationarity Tests in Antitrust Market Definition.

FOX, A.J. 1972. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*:350-363.

FRANSES, P.H. 1998. Time series models for business and economic forecasting: Cambridge university press.

FRANSES, P.H. & GHIJSELS, H. 1999. Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, 15(1):1-9.

GLOSTEN, L.R., JAGANNATHAN, R. & RUNKLE, D.E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779-1801.

GOUNDER, M.K., SHITAN, M. & IMON, R. 2007. Detection of outliers in non-linear time series: A review. *Festschrift in honor of distinguished professor Mir Masoom Ali on the occasion of his retirement*: 18-19.

GOURIEROUX, C., MONFORT, A. & GALLO, G. 1997. Time series and dynamic models. Vol. 3: Cambridge University Press.

GRANÉ, A. & VEIGA, H. 2014. Outliers, ~~GARCH-type~~GARCH type models and risk measures: A comparison of several approaches. *Journal of Empirical Finance*, 26:26-40.

GUO, H. AND NEELY, C.J., 2008. Investigating the intertemporal risk–return relation in international stock markets with the component GARCH model. *Economics letters*, *99*(2), pp.371-374.

GUPTA, M., GAO, J., AGGARWAL, C. & HAN, J. 2014. Outlier detection for temporal data. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 5(1):1-129.

HADI, A.S. & SIMONOFF, J.S. 1993. Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, 88(424):1264-1272.

HAMANN, E. (2001), Time series models for the West German economy, Master's thesis, Technische Universit¨at Wien.

HANSEN, P.R. & LUNDE, A. 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873-889.

HARRILALL, U. 2014. Forecasting volatility in the South African stock market: A comparison of methods. University of the Witwatersrand.

HAUSLER, G. 2002. The globalization of finance. *Finance and Development*, 39(1):10-12.

HERACLEOUS, M.S. 2003. Volatility Modeling Using the Student's t Distribution.

HEYNEN, R.C. & KAT, H.M. 1994. Volatility prediction: a comparison of the stochastic volatility, Garch (1, 1) and Egarch (1, 1) models. *The Journal of Derivatives*, 2(2):50-65.

HILL, R.C., GRIFFITHS, W.E. & LIM, G.C. 2007. Principles of Econometrics, 4th Edition. John Wiley and Sons, New York.

HOU, A.J. 2013. Asymmetry effects of shocks in Chinese stock markets volatility: A generalized additive nonparametric approach. *Journal of International Financial Markets, Institutions and Money*, 23:12-32.

HSING, Y. 2011. The stock market and macroeconomic variables in a BRICS country and policy implications. *International Journal of Economics and Financial Issues*, 1(1):12.

ILANGO, V., SUBRAMANIAN, R. & VASUDEVAN, V. 2013. Outlier detection and influential point observation in linear regression using clustering techniques in financial time series data. *Journal of Theoretical & Applied Information Technology*, 49(2).

ISMAIL, M.I., 2009. *Detecting additive and innovational outliers in BL (p, 0, 1, 1) process* (Doctoral dissertation, University Malaya).

JARQUE, C.M. & BERA, A.K. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255-259.

JESÚS SÁNCHEZ, M. & PENA, D., 2003. The identification of multiple outliers in ARIMA models. *Communications in Statistics-Theory and Methods*, *32*(6), pp.1265-1287.

KAISER, R. & MARAVALL, A. 1999. Seasonal outliers in time series.

KAMBOUROUDIS, D.S. 2012. Essays on volatility forecasting. University of St Andrews.

KANG, S.H., KANG, S.M. AND YOON, S.M., 2009. Forecasting volatility of crude oil markets. *Energy Economics*, *31*(1), pp.119-125.

KANG, S.H., KANG, S.M. AND YOON, S.M., 2009. Forecasting volatility of crude oil markets. *Energy Economics*, *31*(1), pp.119-125.

KARUNANAYAKE, I., VALADKHANI, A. & O'BRIEN, M., 2012. Stock market and GDP growth volatility spillovers.

KGOSIETSILE, O. 2015. Modelling and Forecasting the volatility of JSE returns: a comparison of competing univariate GARCH models.

KRAWIECKI, A., HOŁYST, J. & HELBING, D. 2002. Volatility clustering and scaling for financial time series due to attractor bubbling. *Physical review letters*, 89(15):158701.

KWIATKOWSKI, D., PHILLIPS, P.C., SCHMIDT, P. & SHIN, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, *54*(1-3), pp.159-178.

LADOKHIN, S. 2009. Volatility modelling financial markets. Amsterdam: VU University Amsterdam. Dissertation–MSc.

LAURENT, S., LECOURT, C. & PALM, F.C., 2016. Testing for jumps in conditionally Gaussian ARMA–GARCH models, a robust approach. *Computational Statistics & Data Analysis*, *100*, pp.383-400.

LI, Y. 2013. GARCH models for forecasting volatilities of three major stock indexes: using both frequentist and Bayesian approach. BALL STATE UNIVERSITY.

LJUNG, G.M. 1993. On outlier detection in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*:559-567.

LJUNG, G.M., 1993. On outlier detection in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.559-567.

LOUW, J.P. 2008. Evidence of volatility clustering on the FTSE/JSE top 40 index. Stellenbosch: Stellenbosch University.

LÜTKEPOHL, H. 2006. Forecasting with VARMA models. *Handbook of Economic Forecasting*, 1:287-325.

LUX, T. 2008. Stochastic behavioral asset pricing models and the stylized facts: Institute for the World Economy.

MAKAN, C., AHUJA, A.K. AND CHAUHAN, S., 2012. A study of the effect of macroeconomic variables on stock market: Indian Perspective.

MAKHWITING, M.R., LESAOANA, M. & SIGAUKE, C. 2012. Modelling volatility and financial market risk of shares on the Johannesburg stock exchange. *African Journal of Business Management*, 6(27):8065.

MANDELBROT, B., 1963. The variation of certain speculative prices. *The journal of business*, *36*(4), pp.394-419.

MANDELBROT, B.B. 1997. The variation of certain speculative prices. Fractals and Scaling in Finance. Springer. p. 371-418).

MANDELBROT, B.B. 2001. Stochastic volatility, power laws and long memory.

MATEI, M. 2009. Assessing volatility forecasting models: Why GARCH models take the lead. *Romanian Journal of Economic Forecasting*, 4(1):42-65.

MCDONALD, R.L., CASSANO, M. & FAHLENBRACH, R. 2006. Derivatives markets. Vol. 2: Addison-Wesley Boston.

MICHIE, J., 2006. 5 Macroeconomic reforms and employment: what possibilities for South Africa? *Section 1 Contemporary debates in a global context*, p.86.

MONTGOMERY, D.C. AND WEATHERBY, G., 1980. Modeling and forecasting time series using transfer function and intervention methods. *AIIE Transactions*, *12*(4), pp.289-307.

MULER, N. & YOHAI, V.J. 2008. Robust estimates for GARCH models. *Journal of Statistical Planning and Inference*, 138(10):2918-2940.

NARAYAN, P.K. & NARAYAN, S., 2007. Modelling oil price volatility. *Energy Policy*, *35*(12), pp.6549-6553.

NÄSSTRÖM, J. 2003. Volatility modelling of asset prices using GARCH models.

NELSON, D.B. 1991. Conditional heteroskedasticity in asset returns: A new approach.

NGAILO E. Modelling and forecasting using time series GARCH models: An application of Tanzania inflation rate data. Unpublished Master's Thesis). University of Dar es Salaam. 2011.

NGUNKENG, G., 2013. *Statistical analysis of skew normal distribution and its applications*. Bowling Green State University.

OUSO, J.N., 2012. *The effect of leverage on share prices at the Nairobi Securities Exchange* (Doctoral dissertation).

PANDEY, A. 2003. Modelling and Forecasting Volatility in Indian Capital Markets: Indian Institute of Management.

PENNY, K.I. & JOLLIFFE, I.T. 1999. Multivariate outlier detection applied to multiply imputed laboratory data. *Statistics in medicine*, 18(14):1879-1895.

PENNY, K.I., 1996. Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance. *Applied Statistics*, pp.73-81.

PETERS, E.E., 1996. *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility* (Vol. 1). John Wiley & Sons.

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

POON, S.-H. & GRANGER, C. 2005. Practical issues in forecasting volatility. *Financial Analysts Journal*, 61(1):45-56.

RAPACH, D.E. & ZHOU, G. 2013. Forecasting stock returns. *Handbook of Economic Forecasting*, 2(Part A):328-383.

RAZA, M.A., ARSHAD, I.A., ALI, N. & MUNAWAR, S. 2015. Empirical analysis of stock returns using symmetric and asymmetric GARH models: evidence from Karachi stock exchange. *Science International*, 27(1), p.795-801.

REIDER, R. 2009. Volatility forecasting I: GARCH models. *New York.*

REIMERS, H.E. 2000. Time Series Models for Business and Economic Forecasting. *Journal of the American Statistical Association*, 95(450):686-686.

RICE, J., 2007. *Mathematical statistics and data analysis*. Nelson Education.

ROCKE, D.M. & WOODRUFF, D.L. 1996. Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91(435):1047-1061.

SAKATA, S. & WHITE, H. 1998. High breakdown point conditional dispersion estimation with application to S & P 500 daily returns volatility. *Econometrica*: 529-567.

SAMOUILHAN, N. & SHANNON, G. 2008. Forecasting volatility on the JSE. *Investment Analysts Journal*, 37(67):19-28.

SAMUELSON HONG, W.C. ed., 2013. *Modeling Applications and Theoretical Innovations in Interdisciplinary Evolutionary Computation.* IGI Global.

SAMUELSON, P.A. 1965. Proof that properly anticipated prices fluctuate randomly.

SIGAUKEA, C., MAKHWITING, R.M. & LESAOANA, M. 2014. Modelling conditional heteroskedasticity in JSE stock returns using the Generalised Pareto Distribution. *African Review of Economics and Finance*, 6(1):41-55.

SONG, Z. & XIU, D. 2016. A tale of two option markets: Pricing kernels and volatility risk. *Journal of econometrics*, 190(1):176-196.

STEPHEN, F. 2004. Forecasting Volatility. *New York University Stern School of Business Research Paper*.

STORCH, N., CINER, C., MOFFETT, C. & SCHUHMANN, P. 2013. Is Volatility a Hedge or a Safehaven? An Analysis of Stocks, Vix and Vvix. *Annals of the International Masters of Business Administration at UNC Wilmington*, 6(2).

SU, C. 2010. Application of EGARCH model to estimate financial volatility of daily returns: The empirical case of China.

TERÄSVIRTA, T. 2009. An introduction to univariate GARCH models. Handbook of Financial time series. Springer. p. 17-42).

TESFAMICAEL, M.A. 2007. Forecasting prescription of medications and cost analysis using time series. University of Louisville.

TOLVI, J. 2000. The effects of outliers on two nonlinearity tests. *Communications in Statistics-Simulation and Computation*, 29(3):897-918.

TSAY, R.S. 1988. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1):1-20.

TSAY, R.S. 2002. Analysis of Financial Time Series. Financial Econometrics, A Wiley-Interscience Publication.

VAN DIJK, D., FRANSES, P.H. & LUCAS, A. 1999. Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics*: 539-562.

VEIGA, H., MARTÍN-BARRAGÁN, B. & GRANÉ, A., 2014. *Outliers in multivariate GARCH models* (No. ws140503). Universidad Carlos III de Madrid. Departamento de Estadística.

VERHOEVEN, P. & MCALEER, M. 2004. Fat tails and asymmetry in financial volatility models. *Mathematics and Computers in Simulation*, 64(3):351-361.

VERHOEVEN, P. 2000. Modelling Outliers and Extreme Observations for ARMA-GARCH Processes. (*In.* Econometric Society World Congress 2000 Contributed Papers organised by: Econometric Society.

VILLEGAS, E. 2012. *A survey of risk and ambiguity: an application to the GARCH (1, 1) model with exchange rate data* (Master's thesis, University of Stavanger, Norway).

WACKERLY, D., MENDENHALL, W. & SCHEAFFER, R., 2008. *Mathematical statistics with applications*. Nelson Education

WANG, Y. & ZOU, J. 2010. Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics,* 943-978.

WEISS, A.A. 1984. ARMA models with ARCH errors. *Journal of time series analysis*, 5(2):129-143.

WENNSTRÖM, A. 2014. Volatility Forecasting Performance: Evaluation of GARCH type volatility models on Nordic equity indices.

YANG, C.C. & YANG, C.C., 2007. Separating latent classes by information criteria. *Journal of Classification*, *24*(2), pp.183-203.

ZAINOL, M.S., ZAHARI, S.M., IBRAHIM, K., ZAHARIM, A. & SOPIAN, K. 2010. Additive outliers (AO) and innovative outliers (IO) in GARCH (1, 1) processes. (*In.* Proceedings of the 2010 American conference on applied mathematics organised by: World Scientific and Engineering Academy and Society (WSEAS). p. 471-479).

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

ZHANG, H.K. & ZHANG, Z., 2016. Value-at-Risk estimation of metal market: A GED based TGARCH-EVT approach.