

**BACKTESTING CONDITIONAL EXPECTED
SHORTFALL**

KEBBA BAH

**MASTER OF SCIENCE
(Mathematics-Financial Option)**

**PAN AFRICAN UNIVERSITY INSTITUTE FOR
BASIC SCIENCES, TECHNOLOGY AND
INNOVATION**

2017

**BAYESIAN DISEASE MAPPING IN THE PRESENCE
OF UNDER-REPORTING**

Oti-Boateng Emmanuel

MS 300-0001/15

**A Thesis Submitted to Pan African University Institute of Basic
Sciences, Technology and Innovation in Partial Fulfillment of the
Requirement for the Award of the Degree of Master in Mathematics
(Statistics Option)**

2017

DECLARATION

This thesis is my own work and has not been presented elsewhere for a degree award.

Signature Date.....

Kebba Bah

Declaration by supervisors.

This thesis has been submitted for examination with our approval as university supervisors.

Signature Date.....

Dr. Joseph K. Mung'atu,

Jomo Kenyatta University of Agriculture and Technology

Signature Date.....

Dr. Anthony G. Waititu,

Jomo Kenyatta University of Agriculture and Technology

DEDICATION

This work is dedicated to my late father Musa Bah, my mother Fatou Bah, my wife Ansata Bah, my daughter Fatoumatta Bah, my younger brother Abdoulie M. Bah, my special good senior brother Momodou Yero Bah (News), Ousman Bah (Mampu), Bacarr, my younger sisters Hawa Bah and Lalia Bah and to all the citizens of Kusaak (Yallal-Ba Village) for their support and understanding during my stay in Kenya.

ACKNOWLEDGEMENT

This work has been possible through the help of many individuals who gave their support. I thank the Almighty ALLAH for granting me the knowledge to carry out this study. I also thank African Union for the financial support through PAUSTI during my period of study in Kenya. My Sincere thanks goes to JKUAT and AICAD the host institutions of PAUSTI for allowing me to use their facilities. I can't overlook the love and support from my family and friends who encouraged me throughout my studies.

Special thanks to my supervisors Dr. Joseph Mung'tu and Dr. Anthony Waititu for accepting to supervise this work despite their busy schedules. I do appreciate their unlimited constructive advice, suggestions, ideas and comments towards the work. I highly appreciate their moral and academic support.

Table of Contents

DECLARATION	i
DEDICATION	ii
ACKNOWLEDGEMENT	iii
List of figures	viii
List of tables	ix
DEDICATION	x
ABBREVIATION AND ACRONYMS	xi
ABSTRACT	xiii
1 INTRODUCTION	1
1.1 Background Information	1
1.2 Properties of a Good Risk Measure	2
1.3 Elicitability	3
1.4 Statement of the Problem	4
1.5 Objectives of the Study	5
1.5.1 General Objective	5
1.5.2 Specific Objectives	5
1.6 Justification	5
1.7 Scope	5
2 LITERATURE REVIEW	6
2.1 Daily Return Volatility	6
2.1.1 Stylized Facts of Financial Time Series	6
2.2 Review on GARCH-EVT	7
2.3 Review on Backtesting Expected Shortfall	11
2.4 Research Gap	12
2.5 Summary	12
3 GARCH-Extreme Value Theory	13
3.1 ARCH-Model	13

3.1.1	Properties of the ARCH Model	14
3.2	GARCH Model	14
3.2.1	Properties of the GARCH Model	15
3.3	Tests for Model Suitability	15
3.3.1	Test for Normality	15
3.3.2	Test for Stationarity	16
3.3.3	Autocorrelation	17
3.3.4	Test for ARCH Effects	17
3.4	GARCH Model Selection	18
3.5	Akaike and Schwartz information Criterion	18
3.6	Autocorrelation Function and Partial Autocorrelation Functions .	19
3.7	Estimation of Volatility	20
3.8	Quasi Maximum Likelihood Estimator	21
3.8.1	Asymptotic Consistency	23
3.8.2	Asymptotic Normality	25
3.9	Extreme Value Theory	25
3.10	Generalized Paretos Distribution	28
3.11	Maximum Likelihood Estimation	29
3.12	Threshold Selection	30
3.13	Tail Estimation	32
3.14	Estimation of the Extreme Quantiles	33
3.15	Value at Risk (VaR)	33
3.16	Conditional VaR	34
3.17	Expected Shortfall	35
3.18	Conditional Expected Shortfall	35
3.19	Backtesting VaR and Expected Shortfall	36
3.19.1	Backtesting CVaR	37
3.19.2	Unconditional Coverage Test	37

3.19.3	Independence Test	38
3.19.4	Conditional Coverage Test	40
3.19.5	Bootstrap Test for the Expected Shortfall	40
3.19.6	V-test for the Expected Shortfall	42
4	EMPIRICAL ANALYSIS AND DISCUSSIONS	44
4.1	Estimation of Volatility	44
4.1.1	Data Exploration	44
4.1.2	Daily Returns	45
4.1.3	Test for Stationarity and Normality	46
4.1.4	GARCH Models Selection	47
4.2	Conditional Volatility Estimation	48
4.3	Estimation of Extreme Quantile	49
4.3.1	Threshold selections	50
4.3.2	Estimation of GPD parameters	52
4.3.3	The estimation of Extreme quantile	53
4.3.4	Estimates of Conditional VaR and Conditional Expected Shortfall	54
4.4	Backtesting VaR and ES	57
4.4.1	Test for VaR Estimates	57
4.4.2	Backtesting ES (for both normal and student-t innovations)	59
5	CONCLUSION AND RECOMMENDATION	61
5.1	Conclusion	61
5.2	Recommendation	62
	Bibliography	64
	References	64
	APPENDIX	69

List of Figures

4.1	The daily price series of NSE 20 Share index	44
4.2	Graph of the daily log>Returns	46
4.3	ACF and PACF of Log Returns	47
4.4	ACF and PACF of Log Returns Squares	47
4.5	Daily Volatility NSE 20 Share index (with student-t distribution innovation)	49
4.6	ACF and PACT of Square Residuals(GARCH with Student-t Dis- tribution innovation)	50
4.7	Q-Q plot of the data against the Normal plot	51
4.8	Mean Excess plot of data	52
4.9	Hill Plot of data	52
4.10	Conditional $VaR_{0.99}$ (GARCH model with normal distribution innovation)	54
4.11	Conditional $VaR_{0.99}$ (GARCH model with Student-t distribution innovation)	55
4.12	Conditional $ES_{0.99}$ (GARCH model with Student-t distribution innovation)	56
4.13	Conditional $ES_{0.99}$ (GARCH model with Student-t distribution innovation)	56

List of Tables

4.1	Summary Statistics of Spot price	45
4.2	Summary Statistics of daily Returns	46
4.3	Jargue Bera Test and ADF statistics	47
4.4	GARCH Model Selection (with student-t distribution innovation)	48
4.5	GARCH(1,1) (with student-t innovation)	48
4.6	JB and LM statistics for Residuals (with student-t innovation) .	49
4.7	Jargue Bera and LM test for the Standardized Residuals(GARCH with Student-t distribution innovation)	50
4.8	Estimates of GPD parameter (For GARCH with student-t distri- bution innovation)	53
4.9	VaR and ES estimates (GARCH- with student-t distribution inno- vation)	54
4.10	VaR Violations for Normal distribution innovation and student-t distribution innovation	57
4.11	Results for Backtesting VaR (for both normal and student-t inno- vations)	58
4.12	Result on backtesting ES (for both normal and student-t innovations)	59
4.13	Result on V-test for backtesting ES (for both normal and student-t innovations)	60

DEDICATION

This work is dedicated to my late father Musa Bah, my mother Fatou Bah, my wife Ansata Bah, my daughter Fatoumatta Bah, my younger brother Abdoulie M. Bah, my special good senior brother Momodou Yero Bah (News), Ousman Bah (Mampu), Bacarr and to all the citizens (Kusaak) of Yallal-Ba Village for their support and understanding during my stay in Kenya.

ABBREVIATION AND ACRONYMS

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ARCH	AutoRegressive Conditional Heteroskedasticity
ARMA	Autoregressive moving average model
BMM	Block Maxima Method
CES	Conditional ES
CVaR	Conditional VaR
ES	Expected Shortfall
D_q	Empirical q-quantile
e_q	Extreme qauntile at probability q
EVT	Extreme Value Theory
f	Probability distribution function
F_u	Conditional distribution of the exceedances
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity
GEV	General Extreme Value
GPD	Generalized Pareto Distribution
H	Non-degenerating distribution
I_t	Indicator function
JB	Jargue Bera Test
MDA	Maximum Domain of Attraction
MLE	Maximum Likelihood Estimator
N_r	number of bootstrap samples

LR_{uc}	Likelihood Ratio Statistics for Unconditional Coverage test
LR_{ind}	Likelihood Ratio Statistics for Independent test
LR_{uc}	Likelihood Ratio Statistics for conditional Coverage test
p	probability of success
PACF	Partial Autocorrelation Function
POT	Peak over Threshold
POT^n	POT (GARCH with normal Innovation)
POT^t	POT (GARCH with student-t Innovation)
Q_{MLE}	Quasi Maximum Likelihood Estimator
SIC	Schartz Information Criterion
T_1	Total number of ones
T_0	Total number of ones
VaR	Value-at-Risk
$X_{(n)}$	Ordered statistic
σ	Standard deviation
α	Coverage level
γ	True Parameter
F_t	Past information up to time $t - 1$
ε	Shape Parameter
Θ	Vector parameter
β_u	GPD scale parameter
Θ	Compact parameter space
ε	GPD shape paramter
z_i	Standardized residuals

ABSTRACT

In recent years, the question of whether expected shortfall is backtestable has been hot topic of discussion after the findings in 2011 that expected shortfall lacks the elicibility property in mathematics. However, current literature has indicated that expected shortfall is backtestable and that it does not have to be very difficult. Several researches on risk measures have revealed that Value-at-Risk (VaR) is not a coherent risk measure while Expected Shortfall is coherent. Due to this weakness of VaR, Expected Shortfall has been popular and consequently in 2012 the Basel committee suggest that bank or financial institutions should move from VaR to ES as a measure of risk for minimum capital cover for potential loss. Models are backtest to establish whether their predictions are concurrent with the actual realized values. The backtesting of VaR is simple, direct and well establish in many financial literature. That of Expected Shortfall is not well explored and widely unknown. In this work the Extreme Value Theory and GARCH model are combined to estimate conditional quantile and conditional expected shortfall so as to estimate risk of assets more accurately. This hybrid model provides a robust risk measure for the Nairobi 20 Share index by combining two well known facts about return time series: volatility clustering, and non-normality leading to fat tailedness of the return distribution. First the GARCH model with different innovations is fitted to our return data using pseudo maximum likelihood to estimate the current volatility and then the GPD-approximation proposed by EVT to model the tail of the innovation distribution of the GARCH-model. The estimates are then backtested. In backtesting VaR, three methods are used: Unconditional coverage test, independent test and conditional coverage test whereas for Expected Shortfall two methods were used: bootstrap method and V-test.

Chapter 1

INTRODUCTION

1.1 Background Information

In recent years, both practitioners and academics from the financial community have become interested in extreme events analysis particularly concerning financial risk management. The quantification of market risk for derivative pricing, portfolio choice and market risk management has been of crucial interest to financial institutions and researchers especially during the last two decades. Since the early 1990s VaR has been the leading tool for measuring risk. Indeed, the ability to estimate extreme market movements can be particularly useful for detecting risky portfolios. Supervisors increase the control on banks to make sure they have enough capital to survive in bad markets. While risk is associated with probabilities about the future, one usually uses risk measures to estimate the total risk exposure. A risk measure summarizes the total risk of an entity into one single number. While this is beneficial in many respects, it opens up a debate regarding what risk measures are appropriate to use and how one can test their performance. The Basel Committee of Banking Supervisor uses VaR for internal control as well as in the supervision of banks. VaR quantifies the maximum loss for a portfolio under normal market condition over a given holding period with a certain confidence level. Despite been universal, conceptually simple and being easy to evaluate, VaR has been criticized for not being able to account for tail risk. It only tells us the maximum we can lose if a tail event does not occur, but if tail event occurs, we can expect to lose more than VaR. It is also criticized for its lack of subadditivity, Artzner *et al.* (1997).

Because of the above limitations of VaR, these has prompted the implementation of a more coherent risk measure, Expected Shortfall. Artzner *et al.* (1997) introduced ES to overcome the shortcoming of VaR. ES quantifies the expected value of the loss if a VaR violation occurs. The Basel committee on bank su-

pervision in 2012 raised the prospect of replacing VaR with Expected Shortfall as a risk measure. The greatest challenge confronting the implementation of ES as the leading measure of market risk is the unavailability of simple tools for back testing it. In fact, Gneiting (2011) proved that ES is not elicitable, unlike VaR. This result sparks a lot of debate, some scholars believe that since ES lack such an important mathematics property it is not backtestable. However, Szekely et al. (2014) proposed three non-parametric methods for backtesting ES without exploiting its backtestability and Zaichao Du *et al.* (2015) proposed another method, conditional backtesting ES using cumulative violation.

1.2 Properties of a Good Risk Measure

A risk measure that is used for specifying capital requirements can be thought of as the amount of cash (or capital) that must be added to a position to make it risk acceptable to regulators. Artzner, *et al.* (1999) have proposed a number of properties that such a risk measure should have. These are:

- i. Monotonicity: If a portfolio has lower returns than another portfolio for every state of the world, its risk measure should be greater. ie ρ is monotonic if all loss variable L_1 and L_2 it holds are such that

$$L_1 \leq L_2 \rightarrow \rho(L_1) \leq \rho(L_2)$$

- ii. Translation invariance: Translation invariance means if an amount of cash α is added to a portfolio, its risk measure should go down by α i.e ρ is translation invariant if for all loss variable and $\alpha \in \mathbb{R}$ it holds that

$$\rho(L + \alpha) = \rho(L) - \alpha$$

- iii. Positive Homogeneity: Changing the size of a portfolio by a factor (α) while keeping the relative amounts of different items in the portfolio the

same should result in the risk measure being multiplied by α i.e

$$\rho(\alpha L) = \alpha \rho(L)$$

$$\text{for } \alpha \geq 0$$

- iv. Sub-additivity: The risk measure for two portfolios after they have been merged should be not be greater than the sum of their risk measures before they were merged i.e ρ is subadditive if all loss variable L_1 and L_2 are such that

$$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$$

- v. Normalization: This means you have no risk in taking no position. It is given by

$$\rho(0) = 0$$

- vi. Convexity:

$$\rho(\alpha L_1 + (1 - \alpha) L_2) \leq \alpha \rho(L_1) + (1 - \alpha) \rho(L_2)$$

Convexity means that when we diversify portfolios or invest different asset, our risk should never increase but may decrease.

Beside the above properties of coherent risk measures, comonotonic additivity is another property of risk measures that is mainly of interest as a complementary property to subadditivity. A risk measure satisfying the properties of translation invariance, monotonicity, positive homogeneity and subadditivity is called a coherent risk measure.

1.3 Elicitability

Osband (1985) introduced the concept of elicibility, which was further developed by Lambert *et al.* (2008). It is a mathematical property which is important in

evaluating forecast performance. In general, law invariant risk measure takes a probability distribution and transforms it into a single value point forecast. Genieiting (2011), showed that expected shortfall is not elicitable. Depending on the type forecast made different scoring functions are used in evaluating the model performance. A statistical function (δ) of a random variable Y is elicitable if it minimizes the expected value of scoring function S , i.e

$$\delta(Y) = \arg \min_x (E(S(x, Y)))$$

If δ is elicitable and given historical points of the predictions x_t for the statistics and realizations y_t of a random variable, the natural model to perform backtesting is given by the average expected score

$$\hat{S} = \frac{1}{T} \sum_{t=1}^T S(x_t, y_t)$$

The quantile function (VaR) is elicitable with the scoring function:

$$S(x, Y) = ((x > y) - \alpha)(x - y)$$

1.4 Statement of the Problem

Over the past two decades, the finance world has relied on VaR as a risk measure. The Basel committee for bank supervision in 2012 proposed to phase out VaR as risk measure and replace it with ES which is argued has more benefits as a risk measure (Basel committee, 2012). The changes are motivated by the appealing theoretical properties of ES as a risk measure which are superior to those of VaR (VaR is not subadditive and fail to control for tail risk). The greatest challenge facing financial institutions and regulatory authorities is the unavailability of simple tools for backtesting ES.

1.5 Objectives of the Study

1.5.1 General Objective

To backtest estimates of Conditional Expected Shortfall.

1.5.2 Specific Objectives

- i. To determine Conditional Expected Shortfall using GARCH-EVT.
- ii. To backtest the estimated Conditional Expected Shortfall.

1.6 Justification

Findings in this work will be helpful to both portfolio investor, risk managers, banks and bank supervisors to be able to estimate ES using extreme value theory, which have better theoretical properties than VaR. In fact, the Basel committee in 2011 proposed the replacement of VaR as risk measure by Bank with ES. Since banks are allowed by the Basel committee to use their own models for internal risk management, the ability to backtest these models is a crucial tool for supervisory authorities in order to be able to establish whether these banks are over estimating or under estimating risk.

1.7 Scope

The daily returns of Nairobi Security Exchange for the time from 2nd January 2000 to 31st December 2015 is used in this work. It consist of 4018 daily closing prices. The data was provided by Nairobi Security Exchange. The estimates of conditional Value-at-Risk and Conditional Expected Shortfall were estimated at coverage levels 95%, 99%, 99.5% and backtested at 5%.

Chapter 2

LITERATURE REVIEW

The main purpose of this section is to have an overall view on the estimation of daily return volatility and the use of extreme value theory. This will enable us to gain insight into our research work which is risk estimation and backtesting.

2.1 Daily Return Volatility

Statistically, it is often measured as the sample standard deviation and mathematically defined as follows:

$$\hat{\sigma} = \left[\frac{1}{T-1} \sum_{t=1}^T (y_t - \bar{y})^2 \right]^{\frac{1}{2}}, \quad t = 1, 2, 3, 4, \dots, T \quad (2.1)$$

where y_t represents the daily return at time t and \bar{y} is the average returns over a period of T days. Risk is related to volatility but they are not exactly the same. Is an undesirable outcome in business. Volatility can be defined as a measure of how prices change over time. Hence daily return volatility is a measure of the fluctuations in the daily returns. A high volatility scares investors, since it will be very difficult to predict. According Smith *et al.* (1990) the volatility in prices has implications on the profits and survivals of business.

2.1.1 Stylized Facts of Financial Time Series

Financial time series are known to exhibit certain stylized facts which are vital for correct model selection, estimation and forecasting. The most common among them are Volatility Clustering and Persistence, Fattailedness and leverage effects.

- i. Volatility Clustering and Persistence: Volatility clustering means small changes are followed by small changes (occur in clusters) while large changes are followed by large changes. Engle and Manganelli (2001) show that financial time series are not normal and suffer from volatility clustering.

- ii. Fat Tail: The kurtosis which is the fourth central moment of a random variable Y measures the tail behavior and denoted by $K(y)$. It is mathematically defined as:

$$K(y) = E\left[\frac{(Y - \bar{y})^4}{\sigma_x^4}\right]$$

where \bar{y} representing the first central moment whereas σ_x^4 the square of the second moment. The Kurtosis of a normal distribution is 3. If the kurtosis of a distribution is greater than 3 then it is fat tail (leptokurtosis), whereas if it is less than 3 the distribution is short tail (platykurtic). When the Kurtosis is exactly 3 then the distribution is normal. It is a well known fact that financial time series are fat tail

- iii. Non-Linear Dependence: The non-linear property means the correlations between returns depends on the individual markets. For instance there might be lower correlation in the bull market than bear market.

2.2 Review on GARCH-EVT

Engle (1982) first introduced the ARCH model for capturing time variant variance exhibited by almost all financial and economic time series. The generalized version of ARCH model (GARCH model) which gives more parsimonious results than ARCH model was formulated by Bollerslev (1986) and Nelson (1990). The two GARCH-family model that allow for asymmetric shocks to volatility are GJR-GARCH (Glosten-Jagannathan Runkle GARCH) model introduced by Glosten *et al.* (1993) and EGARCH(exponential GARCH) model proposed by Nelson(1991).

In the same vein Extreme Value Theory, which is used to study the distribution of extreme realization of a given distribution satisfying certain assumption is well established. The foundation of this theory can be trace back to the theorems of Fisher and Tippet (1928) and Gnedenko (1943), who found that the distri-

bution of extreme values of an independent and identically distributed sample converges to only one of three distribution (Frechet , Weibull or Gumbel).

The use of Extreme Value Theory has become popular in finance, after its publication in some papers such as Embrechts *et al.* (1999), Bensalah (2000) and Brodin and Kluppelberg (2006). The results showed that Extreme Value Theory methods fit the tails of heavy-tailed financial time series better than more conventional distributional approaches. It was the best approach in estimating the tail of a loss distribution. Besides finance, several research papers have demonstrated the superiority of EVT in different fields especially in finance, insurance, agriculture and meteorology. McNeil (1997) studied the perform of EVT in finance and insurance. The study uses both the POT where the Generalized pareto distribution fitted to the excess of over a given high threshold and the block maxima method. Due to the 1990s currency crisis, stock market turbulence and credit default, several researchers like Gilli and Kelleiz (2006), Mancini and Trojani (2010) show the power of EVT approach in model VaR, ES and return level by using both Block Maxima and Peak Over Threshold. It show that the POT method is more efficient in modeling data than the block maxima Method.

The work by Mc Neil (1999) combined Extremes Value Theory (EVT) and stochastic volatility models. The first step is to filter returns volatility by fitting a GARCH model using ML (Maximum Likelihood). The second step is to apply the extreme value theory to residuals extracted from an optimal GARCH model using GPD (Generalized Pareto Distribution) or GEV (Generalized Extreme Values). Other studies that follow Mcneil procedure are: Soltane et al(2012) uses GARCH-EVT to forecast volatility in the Tunisian Financial Market. The work reveal that hybrid method provide a robust risk measure for the Tunisian stock market (with much chances of predictive abilities). The backtest of the estimates show that

GARCH-EVT model is good in predicting risk for the Tunisian Stock market. Singh *et al.* (2011) also use GARCH-EVT to model forecast VaR for ASX-ALL ordinary (Australia) and The *S&P* (USA) and backtest their estimates. The result show that the forecast Values for VaR are good estimates of risk for ASX-ALL ordinary (Australia) and The *S&P* (USA). The use of GARCH-EVT does not only stop at forecasting risk but some other people use to compare markets. Mwamba *et al.* (2014) use GARCH-EVT to compare risk on the conventional stock market and the sharia compliant stock market. The study uses data from Dow Jones Islamic market, US *S&P* 500 and *S&P* Europe. The result show that the Conventional stock markets has a fatter tail and it is also riskier. Murenzi *et al.* (2015) model the Rwanda's currency-USD exchange rate volatility using GARCH-EVT. The study uses different ARMA-GARCH-models and found out all the model work out well in modeling risk but ARMA(1,1)-APARCH(1,1,1) stand out. GARCH-EVT is not only useful in estimating risk for financial related institutions but also other fields.

Odening *et. al* (2000) uses GARCH-EVT along side Variance-Covariance and Historical simulation to model risk for German Hog Market. Comparison of there result shows that the GARCH-EVT out perform the other methods in model risk for the German Hog Markets. Their show that we can use the GARCH-EVT also to model risk for other non-finance markets. Paul (2015) forecast intraday VaR using Component GARCH-EVT with different innovation on a selected stock markets. His study highlight that component GARCH-EVT out perform GARCH-EVT. Hence according to Paul (2015) it is better to use component GARCH=EVT in forecasting Intraday VaR then GARCH=EVT. In another study by Kourouma (2010) using data sets from CAC 40, *S&P* 500 Wheat and Crude Oil during financial crisis shows that VaR and Estimates using the conditional models are more reliable in predicting market risk than the unconditional

models. Frad and Zouari (2014) use GARCH-EVT to estimate VaR in the Islam Stock Market. According their study the VaR estimates show a strong stability across thresholds which imply the accuracy and reliability of the VaR estimates. This means that the GARCH-EVT is good method in forecasting VaR for the Islamic Market. Lauridsen (2001) use various methods including GARCH-EVT to estimate VaR and compare he result. The comparison of the results indicate that the GARCCH-EVT are the best. Financial turbulence are periods in which fluctuation in stock markets is very high. Uppal (2013) *et al.* uses GARCH-EVT to forecast risk during the Pre-Global Financial Crisis (GFC) and during Global Financial Crisis period, his study shows that with exception UK and US, the GARCH-EVT explains the observed distributions well in both Pre-Global Financial Crisis (GFC) and during Global Financial Crisis period. In other study by Su E. and Knowles (2006) who use the GARCH-EVT to modeled volatility and analyze VaR for the Asian Pacific Market found out that the estimates of VaR using the conditional market are more reliable. An empirical study on the dynamic VaR on the index of Shanghai Security Market base on GARCH-EVT by Yulin *et al.* (2009) shows that the estimate of VaR using GARCH-EVT are better then GARCH-norm. This is quite true since financial time series are know to posses stylized facts. The results the study include backtesting estimates to ascertain the veracity of the results. The backtest result shows that estimates by GARCH-EVT are better than that of GARCH-norm. Another market that has recently experience high Volatility is the Oil Market. Prices of oil has been peaking to around 200 dollars per barrel and dropping to less than 30 dollars within a short space of time, hence practitioner in the Oil section need reliable estimates of their expected loss. Hence Marimoutou (2006) using various methods including GARCH-EVT and filtered Historical simulation. The result indicate that the forecast by conditional estimates and filter Historical simulation perform better than other models in estimating Risk for the oil market. Gencay *et al.* (2003)

use GARCH-EVT, GARCH, Variance-covariance and historical simulations, according to their study the adaptive GPD and the non adaptive GPD give more stable quantiles forecast.

2.3 Review on Backtesting Expected Shortfall

The backtesting of Expected Shortfall is not well exploited in contrast to the VaR which has several literature's written it. The Earliest known backtesting procedure for expected shortfall was performed by McNeil and Frey 2000. The study used the bootstrap method. Also Acerbi *et al.* (2014) propose three non parametric methods for backtesting ES. They were able to show that their test at confidence level 2.5% is equivalent to backtesting VaR at 1% and their test performs better than VaR. Zaichao *et al.* (2015) proposed two other methods for backtesting ES. They proposed an unconditional backtest which is a t-test for the $p, E[H_t|\alpha, \theta_0]$ analogue to the unconditional test for VaR by Kupiec (1995). The other method proposed by them is a conditional test which is a portmanteau Box-Pierce test on the sample of $H_t(\alpha, \theta_0)$.

Maana *et al.* (2015) used EVT to establish if the extreme volatility witnessed in the daily exchange rate of the Kenya Shilling against the U.S. dollar in the period January 1999 to December 2013 could have been predicted. They also determined if the long-term stability in the exchange rate was affected during the period. The study used the GARCH (1,1) model to estimate the volatility of the exchange rate returns of the Kenya Shilling against the U.S. dollar and found it to describe the volatility process well. The analysis of results revealed three key things for volatility of the exchange rate returns of the Kenya Shilling against the U.S. dollar in the study period. First the quasi Maximum likelihood estimates corresponding to the GED parameters of the exchange rate returns are significant and

corresponds to a distribution with a tail heavier than normal. It also show that the estimated volatility in the daily exchange rates was comparatively extreme in the period 2008 to 2010. An finally its also show that every 3 years extreme volatility is expected in the exchange returns. These can help the Central Bank of Kenya to prepare for such on undesirable circumstances.

2.4 Research Gap

In literature few methods for the backtesting Expected Shortfall has been developed. The most popular methods applied by researcher are those for the backtesting Expected Shortfall. However this methods do not estimate Expected well and hence the backtesting might also be not reliable. Hence to over come these shortcomings, this study first estimate the conditional Expected Shortfall using GARCH-EVT and finally backtest results which is expected to give more plausible results.

2.5 Summary

The section review what other researcher have done on the estimation and backtesting of conditional Expected Shortfall. It has also explain some stylist facts exhibited by financial time series.

Chapter 3

GARCH-Extreme Value Theory

The ARCH models capture the stylized facts of real return data, but in order to have a good fit to real data need a larger number of parameters which reduce the data required for estimation. The GARCH model introduced by Bollerslev (1986) added the concept for tomorrow volatility depends not only the past realizations but also depend on the errors of the volatility predicted. The GARCH model has advantage over the ARCH model since it can capture the series correlation in squared residuals using a smaller number of parameters. The GARCH models have been extremely widely used in finance since they integrate the two main characteristics of financial returns, which are unconditional normalities and volatility clustering. Several parametric and non-parametric methods have been used in the estimations of CVaR and CES in different literature, notable among them are normal distribution, Kernel functions, student-t distribution and GPD.

3.1 ARCH-Model

A stochastic process is ARCH if the time varying conditional variance is heteroscedasticity. It is in the form

$$\begin{aligned} y_t &= \sigma_t e_t \text{ where } e_t \sim N(0, 1) \\ \sigma_t &= \sqrt{\beta_0 + \sum_{i=1}^p \beta_i y_{t-i}^2} \end{aligned} \tag{3.1}$$

where e_t is independent and identically distributed with mean zero and variance one, and $\beta_0 > 0$, $0 \leq \beta_i < 1$ to ensure that conditional Variance is stationary and strictly positive for all t .

$$\text{Given } y_t = \sigma_t e_t = \sqrt{\beta_0 + \beta y_{t-1}^2} e_t$$

here

$$\begin{aligned}
E(y_t^2) &= E((\beta_0 + \beta y_{t-1}^2) e_t^2) \\
&= E(\beta_0 + \beta y_{t-1}^2) E(e_t^2) \\
&= \frac{\beta_0}{1 - \beta}
\end{aligned} \tag{3.2}$$

3.1.1 Properties of the ARCH Model

- i. $E(y_t|F_t) = 0$
- ii. $Var(y_t|F_t) = \sigma_t^2$ where $F_t = y_{t-1}, y_{t-1}, \dots, y_{t-p}$ The conditional variance is a positive non-trivial parametric function of the past information.
- iii. $e_t = \frac{y_t}{\sigma_t}$ are independent and identically distributed and also independent of the past information.

3.2 GARCH Model

The ARCH models capture the stylized facts of real return data, but in order to have a good fit to real data we need a larger number of parameters which reduce the data required for estimation. The GARCH model has advantage over the ARCH model since it can capture the series correlation in squared residuals using a smaller number of parameters. The GARCH have been extremely widely use they integrate the two main characteristics of financial returns which are unconditional normalities and volatility clustering. The general GARCH(p, q) is define as follows

$$y_t = \sigma_t e_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \tag{3.3}$$

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^p \delta_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2} \tag{3.4}$$

In the above equation the Autoregression part has an order $p \geq 0$ while the number of lagged of the variance terms is $q \geq 0$. The magnitude parameters δ_i and β_j determine the short run dynamics of the resulting volatility process. To

ensure that σ_t^2 is strictly positive, δ_i and β_j must be non-negative and for the process to be strictly then $\sum_{i=1}^p \delta_i + \sum_{j=1}^q \beta_j < 1$. If δ_i large, these implies that volatility reacts to market movement whereas if β_j is large indicates that shocks to the conditional variance take long time to die out. The most widely used GARCH models in financial data *GARCH*(1,1). The GARCH is referred as a symmetric model since the impact of sign is not taken into account.

3.2.1 Properties of the GARCH Model

The y_t follows *GARCH*(p, q) if the following properties hold:

1. $E(y_t|F_{t-1}) = 0$, where F_{t-1} is the set of past information up to $t - 1$.
2. $Var(y_t|F_{t-1}) = \sigma_t^2$
3. $\sigma_t = \sqrt{\omega + \sum_{i=1}^p \delta_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2}$, $e_t = \frac{y_t}{\sigma_t}$ are independent and identically distributed and independent of past information. And F_{t-1} contains all past information of y_t and σ_t up to t-1.

3.3 Tests for Model Suitability

3.3.1 Test for Normality

Jarque and Bera test is one of the most popular test for normality in many time series studies. It can be applied using the method of moments where the first moment is the measure of location; the second moment measures the variability of the random variable. The first and the second moments determine a normal distribution. The third and fourth moments are skewness and kurtosis respectively, which are used to determine the degree of asymmetry and fat tailedness of the distribution under study. The Jarque Bera statistic is given by the following equation

$$JB = \frac{s^2}{6/T} + \frac{(k - 3)^2}{4/T} \quad (3.5)$$

where T is the number of observation, k is the kurtosis and s is the sample skewness. If the p-value of the JB statistic is less than the required significant level, then the null hypothesis that the data is normal distributed is rejected. For a larger sample size the JB statistic asymptotically has a chi-squared distribution with two degrees of freedom.

3.3.2 Test for Stationarity

In dealing with GARCH models, one assumes that the time series is stationary. This implies a putting constrains on the maximum likelihood estimators. If the data is non-stationary, there is the presence of unit root. When non-stationary time series is regressed, significant relation is obtained for uncorrelated variables, which is called spurious regression. There are several ways to check if there exist a unit root, but the most popular way is the Augmented Dickey fuller test. It is named after the statisticians David Dickey and Wayne Fuller, who developed the test in 1979. There are two type of Dickey Fuller tests, the standard Dickey Fuller test and the Augmented Dickey test. The standard Dickey test is only able to test unit root for first order Autoregressive model. It is of the form

$$\Delta y_t = (p - 1) y_{t-1} + \varepsilon_t$$

Here the Dickey Fuller test is of the form

$$H_0 : p = 1 \text{ and } H_1 : p \leq 1$$

if the null hypothesis is accepted, this simply means there is a unit root. The t-statistic converges to the distribution function of wiener process instead of the normal distribution.

On the other hand, Augmented Dickey Fuller test builds correlation for higher

order correction by including lag difference of the time series, if the time series is correlated at higher The Augmented Dickey Fuller is of the form

$$\Delta y_t = (p - 1) y_{t-1} + \sum_{i=1}^q \delta_i \Delta y_{t-i} + \varepsilon_t$$

The order q can be chosen by using AIC or BIC statistics

3.3.3 Autocorrelation

The Ljung-Box test (1978), named after Grea M. Ljung and George E. P. Box is a type of statistical test of whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the overall randomness based on a number of lags, and is therefore a portmanteau test. In general, the Ljung-Box test is defined as

H_0 : The model does not exhibit lack of fit (no autocorrelation)

H_1 : The model exhibits lack of fit (there is autocorrelation)

The test Statistics Q of length n is defined as

$$Q = T(T - 2) \sum_{k=1}^m \frac{\hat{p}_k^2}{T-1}$$

where \hat{p}_k is the estimated autocorrelation of the series at lag k and n is the number of lags been tested. The test reject H_0 ,if $Q > X_{1-\alpha,d}^2$ (that is model exhibits lack of fit) where $X_{1-\alpha,d}^2$ is the chi-square distribution with d degree of freedom and α significant level.

3.3.4 Test for ARCH Effects

Before applying the GARCH on the data, one needs to test whether the residual exhibit ARCH effects which is a pre-requisite condition for its application. Engel (1982) proposed the Langrange Multiplier test for ARCH effects. In conducting the test, we first estimate the best fitting autoregressive model

AR(q) $y_t = b_0 + \sum_{i=1}^q y_{t-i} + \varepsilon_t$. Then next we regress the square residuals ε_t on a constant and q lags as in the follows

$$\varepsilon_t^2 = b_0 + b_1\varepsilon_{t-1}^2 + b_2\varepsilon_{t-2}^2 + \dots + b_p\varepsilon_{t-p}^2 + \gamma_t$$

where the test is defined as follows: $H_0 : b_1 = b_2 = \dots = b_p = 0$ against $H_1 : b_i \neq 0$ for at least one, $i = 1, 2, 3, \dots, p$. The test statistics is $LM = TR^2 \sim \chi_d^2$. T is the sample size and the R^2 is computed from the regression of the square residuals ε_t on a constant and p lags. If the p-value is less than the significant level, we reject H_0 , indicating the existence of ARCH effects.

3.4 GARCH Model Selection

In any statistical analysis the selection of model play an integral role. It is always a decision problem on what model to select in order to perform statistical analysis, such as policy analysis, forecasting, estimation and testing. Hence, the choice of a good model plays crucial role in the analysis of financial time series. Therefore, the main challenge is to select a model that takes into account the characteristics of data. Tools such as the Akaike information Criterion (A.I.C) or Bayesian information Criterion (BIC), F-test and Q-test. The most popular are AIC and BIC which will be applied in selecting a model for the data in use.

3.5 Akaike and Schwartz information Criterion

Akaike (1973) came up with the AIC test as an extension to the maximum likelihood principle. This test was the first model selection criterion to benefit from widespread acceptance. AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and fitted likelihood function of the model. A small AIC means a model is considerable to be closer to the truth. The selection criterion is based on the information content of the model. It is mathematically defined as

$$AIC = -2(\ln(l(\hat{\theta}|y_t + k)$$

where likelihood is the probability of the data given a model and k is the number of fitted parameters in the model. In other words, AIC can also be defined as

$$AIC = -2(\log(\text{maximize Likelihood}) + 2(\text{number of fitted parameters}))$$

The first term on the right hand side of AIC equation is a measure of the lack of fit of the chosen model while the second term on the right hand side measures the increased number of model parameters. Schwartz (1978) proposed another model selection criterion based on information theory in Bayesian context call BIC. BIC is an estimate of a function of a future probability of a model being true under a certain Bayesian setup. A lower BIC means that a model is considerably more likely to be the true model. Mathematically, BIC can be defined as follows

$$BIC = -2 * \log(\text{likelihood}) + k * \log(T),$$

where T is the number of observation and ln is log-likelihood function using the k estimated parameters. This definition allows multiple models to be compared at once; where the model with the highest future probability is the one that minimizes the BIC

3.6 Autocorrelation Function and Partial Autocorrelation Functions

ACF and PACF are measures of correlation between current and past series values and show, which past series values are most useful in predicting future values. Hence, it is used to select the order of the process in GARCH model. ACF can be defined, as a set of correlation coefficient between the series and the lags of itself over time, the lag at which the ACF cuts off is the indicated number of GARCH term or conditional variance. In the same way, PACF can also be defined as partial correlation coefficient between the series and lag of itself over time. The lag at which the PACF cuts off is the indicated number of Autoregressive or ARCH term. A positive correlation indicates that large current

values correspond with small values at the specified lag. The absolute value of a correlation is a measure of strength of the association, with large absolute values indicating stronger relationship, wang *et al.* (2005). For stationary processes, autocorrelation between any two observations only depends on the time lag h between them. Define $Cov(y_t, y_{t-h}) = \gamma_h$. Lag h autocorrelation is given by $\rho_h = Corr(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0}$, where the denominator γ_0 is the lag 0 covariance. Partial autocorrelation is the autocorrelation between y_t and y_{t-h} after removing any linear dependence on $y_1, y_2, y_3, \dots, y_{(t-h)+1}$. The partial lag- h autocorrelation is denoted by $\Phi_{h,h}$ for $h = 1, 2, 3, 4, \dots, T - 1$

3.7 Estimation of Volatility

We consider the process that describes the daily returns under the following model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t \\ \sigma_t^2 &= Var(y_t | F_{t-1}) \end{aligned}$$

In order to estimate the conditional volatility, the residuals are substituted by sample residuals. The residuals of the returns can be given as $\varepsilon_t = y_t - \mu_t$. Where e_t is i.i.d random variable with $E(e_t) = 0$ and $E(e_t^2) = 1$. The residuals may be estimated as follows

$$\begin{aligned} \hat{\varepsilon}_t &= y_t - \hat{\mu}_t \\ \hat{\sigma}_t \hat{e}_t &= y_t - \hat{\mu}_t \end{aligned}$$

where $\sigma_t = \sqrt{\omega + \sum_{i=1}^p \delta_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2}$ is GARCH(p,q) model. The GARCH(p,q) model is fitted to negative return data using Quasi-Maximum likelihood Estimation procedure in order to get the current volatility. The GARCH model is defined

as

$$\hat{\sigma}_t = \sqrt{\omega + \sum_{i=1}^p \hat{\delta}_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2}$$

where σ_t is the estimated volatility and its asymptotic consistency and asymptotic normality were investigated in the next section.

3.8 Quasi Maximum Likelihood Estimator

Quasi-Maximum Likelihood Estimate is appropriate when the estimator is derived from a normal likelihood but the disturbance in the model are not truly normally distributed. An important assumption made is that the specification of the likelihood function, in terms of the joint probability density of the variable is correct. Under the condition that the maximum likelihood estimator has the desirable properties of consistency and asymptotically normally, Straumann and Mikosch (2006). Lumsdaine (1996) investigated the Q_{MLE} for GARCH models and showed that the parameters of GARCH models are consistent and asymptotically normal. In this study, we applied Q_{MLE} to estimate parameters of GARCH(p,q) models assuming that conditional expectation of returns is negligible.

To get the Quasi-Likelihood function, we consider the situation where the true probability distribution $f_\gamma(y_t, \theta_\gamma)$ of the returns at time t and incorrect probability distribution given by $f(y_t, \theta_\gamma)$ are used to build the likelihood function.

Now the model can be reformulated by letting y_t to be a sequence with the true model giving

$$y_t = \mu_t + \varepsilon_{\gamma t}, \quad \varepsilon_{\gamma t} = \sigma_{\gamma t} e_t \tag{3.6}$$

$$\sigma_{\gamma t}^2 = Var(y_t | F_{t-1}) = E(y_t^2 | F_{t-1})$$

where $\varepsilon_{\gamma t} \sim N(\theta, \sigma_{\gamma t}^2)$ “ $E(\varepsilon_{\gamma t} | F_{t-1}) = 0$ almost sure and $\sigma_t = \sigma(\varepsilon_{\gamma t}, \varepsilon_{\gamma t-1}, \dots)$ the conditional variance can be define as $E(\varepsilon_{\gamma t}^2 | F_{t-1}) = \sigma_{\gamma t}^2$ (the subscript γ indicates the true value of the parameter). We also assume $y_t = \varepsilon_t = \sigma_t e_t$ $\varepsilon_t \sim N(0, \sigma_t^2)$ to be the model for the unknown parameters (misspecified) model. Hence the true and misspecified distribution are

$$f_\gamma(y_t) = \frac{1}{\sqrt{2\pi\sigma_{\gamma t}^2}} \exp\left(-\frac{\varepsilon_{\gamma t}^2}{2\sigma_{\gamma t}^2}\right) \tag{3.7}$$

$$f(y_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \quad (3.8)$$

Assume that the innovations follow a GARCH (1,1) process

$$\sigma_{\gamma t}^2 = \omega_\gamma (1 - \beta_\gamma) + \varepsilon_{\gamma t-1}^2 + \beta_\gamma \sigma_{\gamma t-1}^2 \quad (3.9)$$

An equivalent expression for the conditional variance can be derived as

$$\sigma_{\gamma t}^2 = \omega_\gamma + \delta_\gamma \sum_{k=0}^{\infty} \beta_\gamma^k \varepsilon_{\gamma t-1-k}^2 \text{ a.s.} \quad (3.10)$$

Again assume that the process is described with true parameters in the vector form given as

$$\theta_\gamma = [\omega_\gamma, \delta_\gamma, \beta_\gamma]' \quad (3.11)$$

and for the model with the unknown parameters

$$\sigma_t^2(\theta) = \omega (1 - \beta) + \delta \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, t = 1, 2, 3, 4, \dots, T \quad (3.12)$$

with the setup or initial conditional, $\sigma_t^2(\theta) = \omega$ this gives the convenient expression for the conditional variance process

$$\sigma_t^2 = \omega + \delta_\gamma \sum_{k=0}^{\infty} \beta_\gamma^k \varepsilon_{t-1-k}^2 \quad (3.13)$$

Finally, assume that the innovation ε_t are the model for unknown parameters

$\theta = [\omega, \delta, \beta]'$ with $|\beta| \leq 1$ Now define the compact parameter space Θ in the following way

$$\Theta = [\theta : 0 < \omega_a \leq \omega \leq \omega_b, 0 < \delta_a \leq \delta \leq \delta_b, 0 < \beta_a \leq \beta \leq \beta_b \leq 1]$$

where subscripts a and b indicate lower and upper limits respectively. We assume that the true parameter $\theta_\gamma \in \Theta$ this implies that $\delta_\gamma > 0, \beta_\gamma > 0$ which means that ε_t is strictly a GARCH process. We can also define standardized residuals $\frac{\varepsilon_t}{\sigma_t}$ by constructing $E(e_t|F_{t-1}) = 0$ and $E(e_t^2|F_{t-1}) = 1$ most of the time the estimation of GARCH model is done under assumption that $e_t \sim N(0, 1)$ so that the likelihood is easily specified. The maximum likelihood estimators of the parameter of the misspecified distribution obtained by maximizing the log-likelihood function

$$\ln L(\theta) = \sum_{t=1}^n \ln f(y_t, \theta) \quad (3.14)$$

The estimator $\hat{\theta}$ is obtained by setting the first order condition given by

$$l(\theta) = \frac{\partial \ln L}{\partial \theta} = \sum_{t=1}^n \frac{\partial \ln f(y_t, \theta)}{\partial \theta} \quad (3.15)$$

$$\sum_{t=1}^n \frac{\partial}{\partial \theta} (1) = 0 \quad (3.16)$$

Thus sufficient condition for 3.16 to hold is that the model is specified correctly. There are however, some important cases where $E_{\tau}[l(\theta)] = 0$ even when the distribution is misspecified. Lets assumed that the Gaussian likelihood is applied to form the estimator. Then, the log likelihood takes the form

$$L_n(\theta) = \frac{1}{2n} \sum_{t=1}^n l_t(\theta)$$

where $l_t(\theta) = - \left[\ln \sigma_t^2(\theta) + \frac{\varepsilon_t^2}{\sigma_t^2(\theta)} \right]$ and $l_n(\theta)$ is typically referred to as quasi-likelihood function of parameter θ , since the likelihood need not to be of the correct density. The vector of parameter value, denoted by $\hat{\theta}_n$ maximizes the likelihood $l_n(\theta)$ on the subspace Θ_1 of compact space Θ and is given as

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} l_n(\theta)$$

The asymptotic consistent and asymptotic normal properties of the quasi-maximum estimator $\hat{\theta}_n$ of the GARCH process also need to be investigated.

3.8.1 Asymptotic Consistency

An estimator $\hat{\theta}_n$ is said to be consistent to the actual parameter θ_n if when the sample size is sufficiently large the estimators $\hat{\theta}_n$ will be very likely to be close to the actual parameter value θ_n . When an estimator converges in probability to the true value as the sample size increases the estimator is asymptotically consistent. Suppose that the daily returns data $y_{-p+1}, \dots, y_0, y_1, \dots, y_n$ generated by the Model 3.6 with θ as the parameter. Suppose that the data up to y_0 are available and

the process y_0 is described with the true parameter in the form a vector. The $\hat{\sigma}_{\gamma t}^2$ define

$$\hat{\sigma}_{\gamma t}^2 = \omega (1 - \beta) + \delta \hat{\varepsilon}_{\gamma t-1}^2 + \beta \hat{\sigma}_{\gamma t-1}^2, t=1,2,3,4,\dots,T \quad (3.17)$$

Together with the initialization $\hat{\sigma}_{\gamma t}^2 \geq 0$ this means that the log-likelihood of $(y_1, \dots, y_n)'$ given $(y_t \sigma)'$ under $\varepsilon \sim N(0, 1)$ is approximately equal to

$$\hat{l}_n = -\frac{1}{2} \sum_{t=1}^n \left(\ln \hat{\sigma}_{\gamma t}^2(\theta) + \frac{y_t^2}{\hat{\sigma}_{\gamma t}^2(\theta)} \right) \quad (3.18)$$

The QMLE θ_n is the parameter value, which maximizes \hat{l}_n on the parameter space Θ_1 , since Θ_1 is an approximately chosen compact subset of the parameter space Θ . The QMLE θ_n is strongly consistent if the following conditions on the random variable ε_t are satisfied

M_1 : ε_t is a sequence of iid random variables such that $E(\varepsilon_t) = 0$

M_2 :The vector parameter θ_γ is in the interior of compact set Θ .

M_3 : For some $k > 0$ there exist a constant $a < \infty$ such that $E[\varepsilon_t^{2+k}] \leq a < \infty$

M_4 : $E[\ln(\beta_0 + \delta_0 \varepsilon_t^2)] \leq 0$

M_5 : ε_t^{2+k} is non generate

M_6 :If for some t holds $\sigma_{\gamma t}^2 = \omega_0 + \sum_{p=1}^{\infty} \omega_p \varepsilon_{t-p}^2$ and $\hat{\sigma}_{\gamma t}^2 = \omega_0^* + \sum_{p=1}^{\infty} \omega_p^* \varepsilon_{t-p}^2$, hence $\omega_i = \omega_i^*$ for $1 \leq i < \infty$

If the above conditions are satisfied then from the theorem below it can be concluded that the QMLE is consistency.

Theorem

Under the above conditions M_1 to M_6 the QMLE estimator $\hat{\theta}_n$ is strongly consistent i.e $\hat{\theta}_n \xrightarrow{a.s} \theta_\gamma$ as $n \rightarrow \infty$

3.8.2 Asymptotic Normality

The distribution of estimators is said to be asymptotically normal if when the sample size increases, the distribution of the estimators approaches a normal distribution. In showing that our estimators are asymptotically normal we need the following assumptions.

M_7 : $\sigma_{\gamma t}^2$ is continuous and twice differentiable on the Θ_1 (that is C-2 regular)

M_8 : The following moment condition hold i.e $E(\varepsilon_t^4) < \infty$ (the fourth moment is finite) and

$$E\left(\frac{(\nabla\sigma_{\gamma}^2(\theta_{\gamma}))^2}{\sigma_{\gamma}^2}\right) < \infty, E\|\nabla l_n\|_{\Theta_1} < \infty \text{ and } |\nabla^2 l_{n\Theta_1}| < \infty$$

If the estimator is asymptotically consistence and the conditions for M_7 and M_8 holds, then according to the following theorem our QMLE estimator is asymptotic normal.

Theorem

Under the conditions M_1 to M_8 , the QMLE $\hat{\theta}_n$ is strongly consistent and asymptotically normal, that is $\sqrt{n}(\hat{\theta}_n - \theta_{\gamma}) \xrightarrow{d} N(0, h_{\gamma})$ as $n \rightarrow \infty$ where h_{γ} is the asymptotic variance of the $\hat{\theta}_n$

3.9 Extreme Value Theory

The extreme value theory plays a fundamental role in modeling maxima of a random variable just like the central limit does in modeling sums of random variables. Basically, there are two ways of identifying extreme in real data. The first method is done by dividing the data in to blocks and considering the maximum in each block as extreme event. The other approach is focusing on the realizations exceeding a given high threshold. Any observation above the selected threshold is considered an extreme event. The block maxima method is the traditional method used to analyze data with seasonality, as for instance meteorological data.

According to Gilli *et al.* (2006) the threshold method uses data more efficiently for that reason, seems to become the choice of most recent applications. The EVT relates to the asymptotic behavior of the extreme observation of a random variable. It provides the fundamental for the statistical modeling of rare events and is used to compute tail related risk measures.

1. Distribution of Maxima:

Let M_n be maximums (minimums) of daily returns with n denoting the size of the block, then limiting law of the block maxima is given by the theorem by Fisher and Tippett and Gnedenko Theorem (Fisher and Tippett ,1928 , Gnedenko, 1943). Let (X_n) be a sequence of iid random variables. If there exists a constants $c_n > 0$ and $d_n \in R$ and some non-degenerate distribution function H such that $\frac{M_n - d_n}{c_n} \xrightarrow{d} H$ Then H belongs to one of the three standard extreme value distributions.

$$Frechet : \Phi_n = \begin{cases} 0, & \text{if } x \leq 0 \\ \exp((-x)^{-\alpha}) & \text{if } x > 0 \end{cases}, \alpha > 0$$

$$Weibull : \theta_n = \begin{cases} \exp(-(-x)^\alpha), & \text{if } x \leq 0 \\ 1, & \text{if } x > 0 \end{cases}, \alpha > 0$$

$$Gumbel : \nabla_n = \exp(-\exp(-x)), x \in R$$

The Frechet distribution has a polynomial decaying tail and hence suits heavy tailed distribution. The Gumbel distribution is exponential decaying tail, which is the characteristic of thin tail while Weibull distribution is the asymptotic distribution of finite endpoint distribution. Jenkinson (1955) and Mises (1954) suggested the generalized extreme value distribution. Which is given as

$$H_\varepsilon(x) = \begin{cases} \exp(-(1 + \varepsilon x)^{\frac{-1}{\varepsilon}}), & \text{if } \varepsilon \neq 0 \\ \exp(-\exp(-x)), & \text{if } \varepsilon = 0 \end{cases}$$

The GEV is obtained by letting $\varepsilon = \alpha^{-1}$ for the Frechet distribution, $\varepsilon = -\alpha^{-1}$ for the Weibull distribution and by interpreting the Gumbel distribution as the case when $\varepsilon = 0$. Usually the limiting distribution of sample maxima is not known in advance; hence, the generalization is useful when computing the maximum likelihood estimates. The standard GEV is the limiting distribution for normalized extrema. But usually in practice the distribution of the returns is not known and as such will not have idea about the normalize constants, hence the three parameter specification of GEV is used and this is limiting distribution for the unnormalized maxima. In this work will not only focus on the parameters only but also on the quantiles

$$H_{\varepsilon,\sigma,\mu}(Y) = H\left(\frac{Y - \mu}{\sigma}\right), Y \in D, \quad D = \begin{cases} (-\infty, \mu - \frac{\sigma}{\varepsilon}) & \varepsilon < 0 \\ (-\infty, \infty) & \varepsilon = 0 \\ (\mu - \frac{\sigma}{\varepsilon}, \infty) & \varepsilon > 0 \end{cases}$$

where σ and μ scale and location parameter and $\sigma > 0$ and $\mu \in R$

2. Distribution of Exceedance:

The Peak Over Threshold (POT) is another method for considering the distribution of exceedance over a chosen threshold. Since the POT method uses data more efficiently it will use in chapter 4 in selecting the extremes Let Y_n to be i.i.d random variables and u_1, u_2, \dots, u_n to be exceedance over given threshold u .

Lets assume that the exceedance over u are i.i.d with conditional distribution function F_u threshold u is less than endpoint $y_F \leq \infty$. F_u is called the conditional distribution and is defined as $F_u(z) = P(Y - u \leq z | Y > u)$

$0 \leq z \leq y_F - u$ where Y is random variable, u is the chosen threshold and $y_F < \infty$ is the right endpoint of F . The conditional distribution F_u can be expressed as

$$\begin{aligned} F_u(z) &= \frac{P(X \leq u + z) - P(Y \leq u)}{P(Y > u)} \\ &= \frac{F(u + z) - F(u)}{1 - F(u)} \\ 1 - F_u(z) &= 1 - \frac{F(u + z) - F(u)}{1 - F(u)} \end{aligned} \quad (3.19)$$

According to Todorovic and Zelenhasic (1970), POT gives the frame of estimating the distribution function of the exceedance over high threshold u , which shows the starting of the tail.

Theorem:

(Limiting distribution of $F_u(z)$. (Balkan and Haan, (1974) and Pickand (1975)). For $F \in MDA(H_\varepsilon, \varepsilon > 0)$, the generalized Pareto distribution (GPD) is the limiting distribution for the distribution of excesses, as the threshold tends to the far right endpoint y_F , i.e. $\lim_{u \rightarrow y_F} \sup_{0 \leq y < y_F - u} |F_u(y) - G_{\varepsilon, \beta(u)}(y)| = 0$

3.10 Generalized Paretos Distribution

According to Balkan (1975) and Pickand theorem (1974) for a sufficiently high threshold u the distribution function of the excess may be approximated by the GPD, which is defined as follows

$$G_{\beta, \varepsilon}(z) = \begin{cases} 1 - (1 + \frac{\varepsilon z}{\beta})^{\frac{1}{\varepsilon}} & \text{if } \varepsilon \neq 0 \\ 1 - \exp(-\frac{z}{\beta}) & \text{if } \varepsilon = 0 \end{cases} \quad (3.20)$$

But $F(u) = \frac{n-k}{n}$ where n is the total number of observation k is the number of exceedances. Hence substituting $F(u)$ with $\frac{n-k}{n}$ and $F_u(z)$ by $G_{\beta, \varepsilon}$ in 3.19 gives

$$F(y) = 1 - \frac{k}{n} \left[\left(1 - \frac{\varepsilon(x-u)}{\beta} \right)^{\frac{1}{\varepsilon}} - 1 \right] \quad (3.21)$$

where ε shape parameter, α is the tail index and β is the scale parameter .Also ε and β can be estimated using the maximum likelihood method for $Y > u$. For a given probability $q > F(u)$ the tail quantile is given by

$$xy_q = u + \frac{\beta}{\varepsilon} \left[\left(\frac{n(1-q)}{k} \right)^{-\varepsilon} - 1 \right] \quad (3.22)$$

3.11 Maximum Likelihood Estimation

The maximum likelihood method of estimation is always an important method in statistics, so that the MLEs of the parameters for the GPD are preferred in the literature. In fact, Smith (1984) showed that when $\varepsilon \leq \frac{1}{2}$, the MLEs for the GPD are consistent, asymptotically normal, and efficient. The MLEs for the GPD have been studied by many authors, including Smith (1985), Davison (1984, Choulakian and Stephens (2001), Grimshaw (1993), and Hosking and Wallis (1987). Let $Y = (Y_1, Y_2, \dots, Y_n)$ be a random vector with PDF, $f(y_1, y_2, \dots, y_n : \theta, \theta \in \Theta)$, Then the likelihood is given as

$$L(y_1, y_2, \dots, y_n, \theta) = \prod_{t=1}^n f(y_t, \theta)$$

Hence the likelihood of GPD is given by

$$L(\theta) = \prod_{t=1}^n \frac{1}{\beta_u} \left(1 - \frac{\varepsilon}{\beta_u} z_i \right)^{\frac{1}{\varepsilon} - 1}$$

Taking the log of both sides we have

$$\log L(\theta) = l(\theta) = \log \left(\left(\frac{1}{\beta_u} \right)^n \sum_{i=1}^n (1 - \beta_u z_i)^{\frac{1}{\varepsilon} - 1} \right)$$

$$l(\theta) = -n \log(\beta_u) + \left(\frac{1}{\varepsilon} - 1 \right) \sum_{i=1}^n \log \left(1 - \frac{\varepsilon}{\beta_u} z_i \right)$$

It is convenient to reparametrize (β_u, ε) of the GPD to (Φ, ε) , where $\Phi = \frac{\varepsilon}{\beta_u}$. Then, using estimates of (Φ, ε) , β_u can be estimated by $\hat{\beta}_u = \frac{\hat{\varepsilon}}{\hat{\Phi}}$. In terms of (Φ, ε) , the log-likelihood for the sample is

$$l(\Phi) = -n \log\left(\frac{\Phi}{\varepsilon}\right) + \left(\frac{1}{\varepsilon} - 1\right) \sum_{i=1}^n \log(1 - \Phi z_i)$$

It is easy to show that the estimating equations of the MLE for (Φ, ε) are equivalent to

$$1 - \varepsilon = \frac{n}{\sum_{i=1}^n (1 - \Phi z_i)^{-1}}$$

and

$$\varepsilon = -n^{-1} \sum_{i=1}^n \log(1 - \Phi z_i)$$

Using elimination method and eliminating ε from 3.11 and 3.11, it suffices to solve the equation for Φ

$$l(\Phi) = 1 - \frac{n}{\sum_{i=1}^n (1 - \Phi z_i)^{-1}} + n^{-1} \sum_{i=1}^n \log(1 - \Phi z_i) = 0, \Phi < \frac{1}{Y_{(n)}} \quad (3.23)$$

when $\hat{\Phi}_{MLE}$ is obtained, $(\hat{\beta}_\tau, \hat{\varepsilon})$ are estimated by

$$\hat{\varepsilon}_{MLE} = -n^{-1} \sum_{i=1}^n \log(1 - \hat{\Phi}_{MLE} z_i) \text{ and } \hat{\beta}_{MLE} = \frac{\hat{\varepsilon}_{MLE}}{\hat{\Phi}_{MLE}}$$

The numerical solution of Φ in 3.23 can be complex. First, the profile log-likelihood function $l(\Phi)$ may steadily decrease, as Φ decreases, from an infinity at $\Phi = \frac{1}{Y_{(n)}}$, so that no local maximum may be found. The situation of no maximum will occur with increasing probability as the true ε increases towards and beyond 1. Second, if there is a local maximum, it may be extremely close to $\Phi = \frac{1}{Y_{(n)}}$, for example, within 10^{-6} , for some data sets, and then the solution of 3.23 will easily be passed over or may give convergence problems, Hosking and Wallis (1987); Grimshaw (1993).

3.12 Threshold Selection

Threshold selection is still an area of ongoing research in the literature, which can be of critical importance. Coles (2001) states that the selection of the threshold process is always a trade-off between the bias and variance. If the threshold is

too low, the asymptotic arguments underlying the derivation of the GPD model are violated. By contrast, selecting a high threshold will lead to few data points to estimate the shape and scale parameter leading to a high variance. There are three graphical methods of estimating the threshold. These are:

1. Quantile-Quantile plot:

In this method, the quantile of the data is plotted against the quantile of a reference distribution. If the reference distribution is that of a medium tail (normal or exponential distribution) then data points will form straight line with a positive slope. If the right hand side of the graph is concave it indicates fat tail and if it is convex at the right hand side, it indicates short tail. If the sample is realization are from a distribution, which has the same form as the reference distribution but with difference scale and / or location parameters, then the QQ-plot, will still be linear and the intercept of QQ-plot will indicate the location while the slope indicates the scale parameter.

2. Mean Excess Plot (MEP):

The MEP is also known as mean residual life plot. It is one of the most commonly used graphical method. The theoretical reasons behind this methods reside in the fact that when the distribution of exceedances over a threshold u_1 is a GPD, then the distribution of exceedances over any threshold $u_1 > u_2$ is also a GPD with the same shape parameter ε and scale parameter $\beta_{u_2} = \beta_{u_1} - \varepsilon(u_1 - u_2)$. The Mean Excess plot is the representation of the empirical estimate of conditional expectation $E(Y - u|Y > u)$ as a function of the threshold. The optimal threshold u^0 , the distribution of the exceedances is GPD and the conditional mean excess is given by

$$E(Y - u|Y > u) = \frac{\beta_u}{1 + \varepsilon} = \frac{\beta_{u^0} - \varepsilon(u - u^0)}{1 + \varepsilon}, \quad \text{for } u > u^0 \quad (3.24)$$

The MEP will be roughly positive slope linear above a threshold u which indicates that the data follows GPD with a positive slope parameter ε . If data is medium tail then the MEF would be horizontal while if it is fat tail MEF will negatively sloped line. u is taken to be the beginning of the tail of the distribution under consideration. However, in practice, the use of an MEP is not always simple and detecting the linearity is subjective task. We explore the range of the linearity of graph using numerical techniques to select optimal threshold.

3. Hill Plot:

The Hill plot is done by ordering the statistics with respect their values, i.e $Y_{(1,n)}, Y_{(2,n)}, Y_{(3,n)}, \dots, Y_{(n,n)}$ and $Y_{(1,n)} \geq Y_{(2,n)} \geq \dots \geq Y_{(n,n)}$. Hill's estimator

of the tail index $\alpha = \frac{1}{\varepsilon}$ is given by

$$\hat{\alpha} = \left(\frac{1}{k} \sum_{j=1}^k \ln Y_{j,n} - \ln Y_{k,n} \right)^{-1} \quad (3.25)$$

where $Y_{k,n}$ is the upper order statistics and k is the number exceedance and n is the sample size. The plot is constructed by plotting the estimate of α or ε as a function of the k -upper statistics. A threshold is selected from the plot where the shape parameter α or ε is stable. Hill plot is constructed by plotting the estimate of ε as a function of k -upper order statistic or threshold. The threshold is selected from the plot where the shape parameter tail index is stable, Beirlant *et al* (1996)

3.13 Tail Estimation

Our interest is to build tail estimator that can be used to obtain the quantile. The method of non-parametric such as historical simulation may be used to estimate F_u as $\hat{F}_u = \frac{n-k}{n}$, where n is the total number of observations and k are the number of the observations above the threshold u . The MLE of the generalized Pareto distribution parameters give rise to the tail estimator formula as

$$\begin{aligned}
\hat{F}(y) &= \frac{k}{n} \left[1 - \left(1 - \hat{\varepsilon} \frac{(Y-u)}{\hat{\beta}_u} \right)^{\frac{-1}{\varepsilon}} \right] + \left(1 - \frac{k}{n} \right) \\
&= \frac{k}{n} - \frac{k}{n} \left(1 - \hat{\varepsilon} \frac{(Y-u)}{\hat{\beta}_u} \right)^{\frac{-1}{\varepsilon}} + 1 - \frac{k}{n} \\
&= 1 - \frac{k}{n} \left[\left(1 - \hat{\varepsilon} \frac{(Y-u)}{\hat{\beta}_u} \right)^{\frac{1}{\varepsilon}} \right]
\end{aligned} \tag{3.26}$$

3.14 Estimation of the Extreme Quantiles

Consider a random variable Y and a high probability level q , the quantile of random variable Y at probability level q is any real number e_q satisfying the following inequalities

$$P(X \leq e_q) \geq q \tag{3.27}$$

Now, defining the quantile e_q of distribution function F as inverse of distribution at a given probability level $q \in (0, 1)$ close to one

$$\hat{F}(e_q) = 1 - \frac{k}{n} \left[\left(1 - \hat{\varepsilon} \frac{(x-u)}{\hat{\beta}_u} \right)^{\frac{1}{\varepsilon}} \right] \tag{3.28}$$

The quantile estimate of an underlying distribution are obtained by simply inverting the above equation which gives

$$\hat{e}_q = u + \frac{\hat{\beta}_u}{\hat{\varepsilon}} \left[\left(\frac{n(1-q)}{k} \right)^{-\varepsilon} - 1 \right] \tag{3.29}$$

where $\hat{\varepsilon}$ and $\hat{\beta}_u$ are the estimates of ε and β_u shape and scale parameters respectively.

3.15 Value at Risk (VaR)

The Value-at-Risk answers the question how much we can lose, with given probability, over a certain time horizon. From a mathematical perspective VaR is simply a quantile of the profit and loss distribution of a given portfolio over a

particular holding period. Hence VaR quantifies the maximum loss for a portfolio under normal market condition over a given period with a certain confidence level. It is one of the major tools finance shareholders can use to assess level of riskiness of the market. It summarizes risk in a single number which makes it easier to understand the total exposure of the investment to market risk. It is given by

$$VaR_q = \hat{e}_q = u + \frac{\hat{\beta}_u}{\hat{\varepsilon}} \left[\left(\frac{n(1-q)}{k} \right)^{-\varepsilon} - 1 \right] \quad (3.30)$$

3.16 Conditional VaR

The conditional volatility provided by GARCH model and extreme quantile estimates are combined to obtain conditional value at risk. For extreme quantiles when q close to 1, the empirical quantiles are not efficient estimates of the theoretical quantiles. The conditional quantile is quantile of the predictive distribution for the return over the next h days. The conditional value at risk is given by $CVaR_q^{t+1} = \inf(x \in R, F_{x_{t+1}+x_{t+1}+\dots+x_{t+k}|\Pi_t}(x) \geq q)$.

Hence $CVaR_q^t = \mu_{t+1} + \sigma_{t+1}z_q$ which by assumption does not depend on t and the mean and volatility is estimated by volatility dynamic model. Where z_q is the upper q^{th} quantile of the marginal distribution of z_t . Conditional value at risk estimate is given by

$$CVaR_q^{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \left[u + \frac{\hat{\beta}_u}{\hat{\varepsilon}} \left(\left(\frac{n(1-q)}{k} \right)^{-\varepsilon} - 1 \right) \right] \quad (3.31)$$

Conditional VaR is defined as the q conditional quantile of the returns at $q \in (0.95, 0.995)$, Gouriéroux and Jasiak (2009). The conditional VaR estimate is also consistent since it is composed of consistent estimates

3.17 Expected Shortfall

Another measure of risk is expected shortfall (ES). It is used to estimate the potential size of the loss exceeding VaR at q probability. It addresses the question of how bad can things get. Mathematically it is given by $ES_q = E(X|X > VaR_q)$. It is estimated by the following formula

$$\hat{ES}_q = \frac{VaR_q}{1 - \hat{\epsilon}} + \frac{\beta + \hat{\epsilon}u}{1 - \hat{\epsilon}} \quad (3.32)$$

3.18 Conditional Expected Shortfall

The conditional VaR is not a sub-additive risk measure. Let ρ be a generic measure of risk that maps the riskiness of the profile to an amount of required reserves to cover losses that regularly occur and let D_1 and D_2 be portfolios of assets. For sub-additive property, the required reserves for the combination of two portfolios are less than the required reserves for each treated separately i.e $p(D_1 + D_2) \leq p(D_1) + p(D_2)$.

To overcome these shortcomings, Conditional ES which has better theoretical properties was introduced. It is defined as the conditional expectation of all the VaR violations. Conditional ES is defined as

$$CES_q^{t+1} = \mu_{t+1} + \sigma_{t+1}E[Z|Z > z_q] \quad (3.33)$$

where z_q is the upper q^{th} quantile of the marginal distribution of Z_t , which by assumption does not depend on t . The estimation of CES under extreme conditions requires estimation of volatility σ_t and using appropriate extreme value distribution to obtain quantiles. The estimator for the conditionals becomes

$$C\hat{ES}_q^{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \left(\frac{UVaR_q}{1 - \hat{\epsilon}} + \frac{\hat{\beta} + \hat{\epsilon}u}{1 - \hat{\epsilon}} \right) \quad (3.34)$$

where $\hat{\beta}$ and $\hat{\epsilon}$ are the scale and shape parameters respectively of the GPD distribution, and u is the threshold. It is a consistency estimate. The estimate of

the conditional ES is consistent since it comprises of consistent parameters.

3.19 Backtesting VaR and Expected Shortfall

In this section, the estimated CVaR and CES are backtested using different methods. In order to backtest a window of size w fix and stepping through the data day by day, using the past h observations to estimate the VaR and ES for the next day. let $(y_{t(1)}, \dots, y_{t(h)})$ be a window of raw data ordered by time which we use to estimate our models. In order to get the estimates of conditional VaR and conditional ES, GARCH(1,1) is first fitted to the data and finally the GPD is fitted to the standardized residuals. The standardized residuals are given by

$$(z_{t_1}, \dots, z_{t_w}) = \frac{x_{t_1} - \mu_{t_1}}{\sigma_{t_1}}, \dots, \frac{x_{t_w} - \mu_{t_w}}{\sigma_{t_w}} \quad (3.35)$$

The standardized residuals should have a mean zero and variance of one if the fitted model is plausible for the loss series.

To measure the performance of a specific model, the estimates should be compared to the actual outcomes to see how well the model fared over the days for which predictions of the ES were made. That is, the study wants to test whether the forecasts of the model are consistent with the assumptions underlying the model choice i.e. the distribution of the losses and/or residuals in the case of the study. Backtesting is generally described as finding a good way to make comparison between the reality and what our model gives. Unfortunately, the backtesting theory and methodology is not as developed for expected shortfall as it is for value at risk.

The backtesting of VaR is done using unconditional coverage, independence test and conditional coverage, while that of ES is implemented using bootstrap method used McNeil and Frey (2000) and what is known as the V test by Embrechts *et al.* (2005).

3.19.1 Backtesting CVaR

In this section the conditional coverage test by Christoffersen (2003) is used. At a given confidence level $q \in (0, 1)$, it is expected that Y_{t+1} will exceed estimated VaR_q^{t+1} only $100(1-q)\%$ of the time. These exceedances are called VaR violations. Particularly with a VaR model that adapts to recent losses and recent volatility (i.e. the GARCH(1,1)), it is expected the VaR violations to be independent of each other. It follows that a way to test the performance of a VaR model is to test if the model produces the expected number of VaR violations when testing the model on a set of data, and to test if the VaR violations are independent of each other. The hit sequence is formed using the indicator variables representing VaR violations as follows

$$I_{t+1} = \begin{cases} 1, & \text{if } Y_{t+1} > VaR_q^{t+1} \\ 0, & \text{if } Y_{t+1} \leq VaR_q^{t+1} \end{cases} \quad (3.36)$$

and a VaR violations means a hit and otherwise a miss. Suppose we have data set with T VaR predictions, then the hit sequence is given by $\{I_t\}_{t=1}^T$. The simplest possible null hypothesis is then that I_t are Bernoulli variables with success probability $\alpha = 1 - q$, so that $\{I_t\}_{t=1}^T$ is a sequence of i.i.d Bernoulli random variables. The probability mass function of a Bernoulli(p) variable is given by

$$f(1_t; p) = p^{I_t}(1 - p)^{1-I_t} \quad (3.37)$$

3.19.2 Unconditional Coverage Test

In this section, the unconditional test is used to check whether the amount of VaR violations produced by the model at a chosen coverage level α is as expected. This is done by using the conditional coverage test described by Christoffersen (2003). It done by comparing the theoretical sample fraction π of VaR violations to the promised fraction α under the null hypothesis for the unconditional coverage test is that $\pi = \alpha$. We do this comparison through the likelihood ratio test. Let T_1 and T_0 denote the number of hits and misses respectively in a sample of size

T, then the likelihood function under the null hypothesis is given by

$$L(\alpha) = \prod_{t=1}^T p^{I_t} (1-p)^{1-I_t} = p^{T_1} (1-p)^{T_0} \quad (3.38)$$

Next, π is estimated as $\hat{\pi} = \frac{T_1}{T}$, which is the maximum likelihood estimate of π .

The maximized likelihood for the sample is then given by

$$L(\hat{\pi}) = \left(\frac{T_1}{T}\right)^{T_1} \cdot \left(\frac{T_0}{T}\right)^{T_0} \quad (3.39)$$

The likelihood ratio statistic is then given by

$$LR_{uc} = -2 \frac{L(\alpha)}{L(\hat{\Phi})} \quad (3.40)$$

and is asymptotically (in T) distributed as a χ^2 random variable with one degree of freedom, hence the χ^2 distribution can be use to conduct the test.

According Christoffersen (2003), the number of observations T and/or the number of violations T_1 (especially for small α), may be too small hence the test might not be reliable. Hence Christoffersen recommends doing a Monte Carlo simulation to obtain reliable p-values for this test. The Monte Carlo simulation is done by generating 999 samples of i.i.d Bernoulli(α) variables and calculate the above test statistic for each sample, hence giving us a sequence $\{\hat{LR}_{uc}(i)\}_{i=1}^{999}$. The simulated P-value is the given by

$$p - value = \frac{1}{1000} \left[1 + \sum_{t=1}^{999} 1_{\{\hat{LR}_{uc}(i) > LR_{uc}\}} \right] \quad (3.41)$$

where $1_{\{\Delta\}}$ is an indicator function which is equal to one if the $\hat{LR}_{uc}(i) > LR_{uc}$ is true, and zero otherwise. If the simulated p-Value is too small then the null hypothesis is rejected indicating that the coverage rate of the model is correct.

3.19.3 Independence Test

In this section independent test Christoffersen (2003) is used to check if the VaR violations are independent of each other or if they come in clusters or not. If the VaR violations are not independent then the model does not adapt sufficiently

and quickly enough to large losses. This may possibly create a risk of bankruptcy within short period since losses are piling on. If the VaR violations are not independent, then we have the probability that violations happen tomorrow given it has happened today is greater than α . Christoffersen (2003) provided a method on how to test for independence of the VaR violations. Assume that the hit sequence $(I_t)_{t=1}^T$ is dependent, and that it can be described by a discrete-time Markov chain with transition probability matrix

$$\Theta_1 = \begin{bmatrix} \Phi_{00} & \Phi_{01} \\ \Phi_{10} & \Phi_{11} \end{bmatrix} = \begin{bmatrix} 1 - \Phi_{01} & \Phi_{01} \\ 1 - \Phi_{11} & \Phi_{11} \end{bmatrix} \quad (3.42)$$

where $\Phi_{i,j}(i, j \in \{0, 1\})$ is the probability that $I_{t+1} = j$ conditional on $I_t = i$. Here Φ_{11} is the probability that a VaR violations occurs tomorrow given that one occurred today. Under this method, only the outcome for today matters for the outcome tomorrow, with T observations. The maximum likelihood estimates for these probabilities are then given by

$$L(\Theta_1) = (1 - \Phi_{01})^{T_{00}} \Phi_{01}^{T_{01}} (1 - \Phi_{11})^{T_{10}} \Phi_{11}^{T_{11}} \quad (3.43)$$

where $T_{i,j}$ is the number of days for which a j followed an i in the hit sequence, with $i, j \in \{0, 1\}$. Then we maximum likelihood estimates for the probabilities as:

$$\hat{\Phi}_{01} = \frac{T_{01}}{T_{00} + T_{01}} \implies \hat{\Phi}_{00} = 1 - \hat{\Phi}_{01} \quad (3.44)$$

$$\hat{\Phi}_{11} = \frac{T_{11}}{T_{10} + T_{11}} \implies \hat{\Phi}_{10} = 1 - \hat{\Phi}_{11} \quad (3.45)$$

Hence the matrix of estimated transition probabilities is as follows

$$\hat{\Theta}_1 = \begin{bmatrix} \hat{\Phi}_{00} & \hat{\Phi}_{01} \\ \hat{\Phi}_{10} & \hat{\Phi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix} \quad (3.46)$$

Now, if the hit sequence is dependent, then $\Phi_{00} \neq \Phi_{11}$, otherwise $\Phi_{00} = \Phi_{11} = \Phi$, implying independent. Since the concerned is mostly about positive dependence that is $\Phi_{00} > \Phi_{11}$, Φ is estimated by $\hat{\Phi} = \frac{T_1}{T}$. Thus, under independence we get a transition probability matrix

$$\hat{\Theta} = \begin{bmatrix} 1 - \hat{\Phi} & \hat{\Phi} \\ 1 - \hat{\Phi} & \hat{\Phi} \end{bmatrix} \quad (3.47)$$

This has a likelihood function $L(\hat{\Phi})$ as that of the Unconditional coverage test. We then test the independence hypothesis $\Phi_{01} = \Phi_{11}$. The likelihood ratio test statistic is given by

$$LR_{cc} = -2 \frac{L(\alpha)}{L(\hat{\Theta}_1)} \quad (3.48)$$

This test statistic is asymptotically distributed as a χ^2 random variable with one degree of freedom. Christoffersen (2003) again recommended the use of Monte Carlo simulation to obtain an accurate p-value instead of using quantiles from the $\chi^2(1)$ distribution when testing the independence hypothesis. This is done in the same way as for the unconditional coverage test.

3.19.4 Conditional Coverage Test

Finally, in order to perform the conditional coverage test, unconditional and independent test are combined to jointly test for correct coverage and independence. Since LR_{uc} and LR_{ind} are each $\chi^2(1)$ -distributed (asymptotically), their sum should be $\chi^2(2)$ -distributed, and we have the test statistic as

$$LR_{cc} = LR_{uc} + LR_{ind} = -2 \frac{L(\alpha)}{L(\hat{\Phi})} \quad (3.49)$$

We again use Monte Carlo simulations to get a more accurate p-value.

3.19.5 Bootstrap Test for the Expected Shortfall

To backtest the ES estimates, the difference between the next day return Y_{t+1} and the estimate of the expected shortfall at time t , ES_q^{t+1} , conditional on Y_{t+1} exceeding the estimate of the q quantile of Y_{t+1} , i.e. Var_q^{t+1} is looked at. We introduce the notation $Y_t q := Var_q^{t+1}$, so that $y_q^t = Var_q^{t+1}$. As per the based model for losses and its assumptions, i.e that losses Y_t under all our models can be written in the form $Y_t = \mu_t + \sigma_t Z_t$, where the Z_t i.i.d with mean zero and unit variance and come from a location-scale family

$$\begin{aligned}
R_{t+1} &= \frac{Y_{t+1} - ES_q^{t+1}}{\sigma_{t+1}} \\
&= \frac{\mu_{t+1} + \sigma_{t+1}Z_{1+1} - (\mu_{t+1} + \sigma_{t+1}ES_q^{t+1}(Z))}{\sigma_{t+1}} \\
&= Z_{t+1} - E_t(Z|Z > z_q)
\end{aligned}$$

conditional on $Y_{t+1} > y_q^t$ or equivalently $Z_{t+1} > z_q$, being the q – *quantile* of Z . The R_t s are then i.i.d under the model and furthermore have an expected value of zero. Based on the data and the estimates of expected shortfall, residuals are constructed on day when $Y_{t+1} > \hat{y}_q^t$ i.e on days when VaR violation occur. Following McNeil and Frey (2000) these exceedance are call exceedance residuals, denoted by

$$r = \{r_{t+1}; \text{ for } t \text{ such that } y_{t+1} > \hat{y}_q^t\}, \text{ where} \quad (3.50)$$

$$r_{t+1} = \frac{y_{t+1} - \hat{ES}_q^{t+1}}{\hat{\sigma}_{t+1}} \quad (3.51)$$

and $|r| = m$, where m is the number of VaR violation from our model. Here the is that the estimates μ_{t+1}, σ_{t+1} , and the expected shortfall are correct and these residuals are i.i.d with mean zero. With alternative hypothesis as the residuals have a mean greater than zero, i.e. hence the expected shortfall is systematically underestimated, which as McNeil and Frey (2000) remarks is the more likely direction of failure. Underestimating ES is very dangerous to business since it leads to losses(as opposed to missing out on profit). The downside however, is that the test will tend to favor models that overestimate the expected shortfall, which is undesirable in the long-run. This test is done by a non-parametric bootstrap which is outlined by Efron and Tibshirani (1993). Here the null hypothesis that original residuals r , which are distributed according to some distribution function F , have mean $\mu_0 = 0$ tested. The test statistic is given by

$$T = t(r) = \frac{\hat{r} - \mu_0}{\sigma/\sqrt{(m)}} \quad (3.52)$$

where

$$\hat{r} = \frac{1}{m} \sum_{i=1}^m r_i \quad (3.53)$$

and

$$\hat{\sigma} = \frac{1}{m-1} \sum_{i=1}^m (r_i - \hat{r})^2 \quad (3.54)$$

Here the empirical distribution function is translated so that it has the desired mean μ_0 , by forming the shifted residuals

$$\hat{r}_i = r_i - \hat{r} + \mu_0 \quad i = 1, 2, \dots, m \quad (3.55)$$

From these, sampling $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m$ with replacement, and for each such bootstrap sample j (N_r of them in total) we compute the statistics

$$T^* = t(\hat{r}^*) = \frac{\hat{r}^* - \mu_0}{\sigma / \sqrt{(m)}} \quad (3.56)$$

Hence computing the p-value for our null hypothesis as

$$p - value = \frac{1 + \sum_{i=1}^{N_r} 1_{T_j^* > T}}{1 + N_r} \quad (3.57)$$

where $1_{\{\Delta\}}$ denotes the indicator function, which is 1 if $T_j^* > T$ is true and 0 otherwise. One is added to both numerator and denominator to avoid a p-value of 0. Models are chosen base on their p-value, high p-value speak in favor of a model, while low p-value speak against a model. we will take $N_r = 10000$.

3.19.6 V-test for the Expected Shortfall

Embrechts *et al.* (2005) introduced a couple of methods for evaluating the performance of different ES estimates based on the relative size of the test statistics. The first statistic V_1 simply takes the average of the difference between the actual return and the fore casted expected shortfall for days where the actual return exceeded the VaR estimate. This should lead to a value close to zero of V_1 if the model is good, since if the model is correct the expected value of this statistic is zero. For a chosen probability q , V_1 is thus given by

$$V_1 = \frac{\sum_{t=1}^T \left(y_{t+1} - \hat{E}S_q^{t+1} \right) 1_{y_{t+1} > y_q^t}}{1_{y_{t+1} > \hat{y}_q^t}} \quad (3.58)$$

where T is the total number of estimates of in the data set. According to Embrechts *et al.* (2005) the weakness of this measure is that it depends strongly on

the VaR estimates. With the unconditional expected Shortfall, the average size of a one in $\frac{1}{1-q}$ case is calculated. A measure which looks at these types of events is the measure V_2 defined by

$$V_2 = \frac{\sum_{t=1}^T \left(y_{t+1} - \hat{E}S_q^{t+1} \right) 1_{D_t > D_q}}{1_{D_t > D_q}} \quad (3.59)$$

where $D_t = \left(y_{t+1} - \hat{E}S_q^{t+1} \right)$ and D_q is the empirical q -quantile of $\{D_t, t = 1, 2, \dots, T\}$. D_t is expected to be negative in less than one in $\frac{1}{1-q}$ cases. A good estimator for ES would thus hopefully give an estimate close to zero. V_1 and V_2 can be combined into a third measure that strikes a balance between the theory-reliant V_1 measure and the more practically oriented V_2 measure. This third measure is defined as

$$V = \frac{|V_1| + |V_2|}{2} \quad (3.60)$$

and should again be closed to zero if it is good

Chapter 4

EMPIRICAL ANALYSIS AND DISCUSSIONS

4.1 Estimation of Volatility

4.1.1 Data Exploration

The daily Nairobi 20 share index were plotted to see the behavior of the data. The plot shows the daily fluctuations of the series. It shows that the daily NES 20 share index exhibit has a very high volatility since the graph has almost no smooth area. The plot reveals trends with high uncertainty in the NSE 20 share index.



Figure 4.1: The daily price series of NSE 20 Share index

Figure 4.1 shows a general trend of uncertainty in the NES 20 share index with extreme fluctuations especially half way of the diagram which corresponds to 2007-2008. These extreme fluctuations at that time period may be attributed to the post election violence or the world financial crisis during this period. The daily NES 20 share index has a significant difference between the minimum and maximum. It has a positive mean showing that the share prices are generally moving upwards. The negative skewness -0.3233 in Table 4.1 and excess kurtosis of -1.04891 clearly indicate the non-normality of the distribution. The large standard deviation shows a high variability in the data. Its standard deviation

of 1305.3 which is high indicating high variability in the data.

Table 4.1: Summary Statistics of Spot price

Statistics	mean	Max	Min	St. Dev	Kurtosis	Skewness
Value	3562	6161	1005	1305.742	-1.05039	-0.32269

4.1.2 Daily Returns

Most of the financial time series are decomposed into exponentially growing trend. It is important to transform our data using logarithm which gives us the log-returns. According Strong (1992) there are both theoretical and empirical reasons for preferring logarithmic returns. Theoretically, logarithmic returns are analytically more tractable when linking together sub-period returns to form returns over long intervals. Empirically, logarithmic returns are more likely to be normally distributed and so conform to the assumptions of the standard statistical techniques. Using the logarithmic returns in the study will help to test whether the daily returns were normally distributed or not. It also have good properties such as; it is very simple to aggregate the log-returns over time. In order to estimate the volatility in the daily returns we have used logarithm returns. The log-returns plots show that the data appear to be stationary in mean. This plot also reveal that the returns exhibit dependence structure where period of high returns are followed by high returns and periods of low returns followed by low returns. This is the evidence by volatility clustering in the data (short range dependence), which show that the data is not i.i.d. The clustering of the log-returns data indicates the presence of stochastic volatility. Figure 4.2 shows the existence of extreme losses.

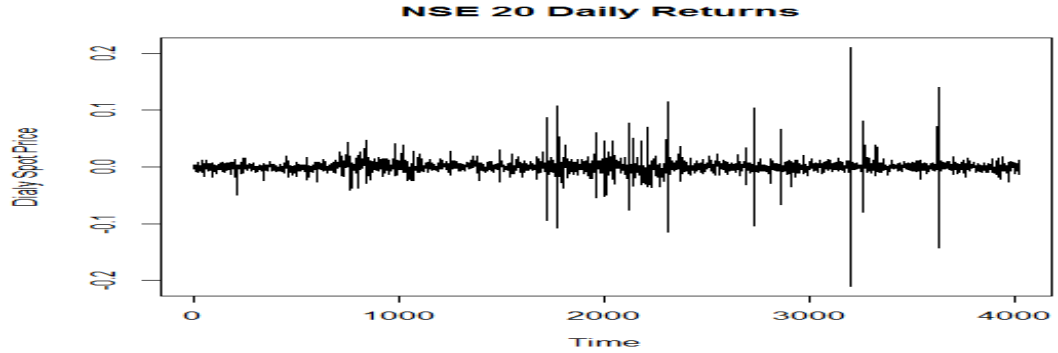


Figure 4.2: Graph of the daily log>Returns

The summary statistics of the log returns are given in table 4.2. The log-returns have mean of 0.000139 . The Kurtosis of the log-returns is 83.62801 which is far greater than 3 for the normal distribution. This indicates that the underlying distribution of the returns have a tail which is heavier than normal. The data exhibits negative skewness indicating frequent small gains and few extreme losses. Since the skewness which is 0.1159 is different from the zero for normal, the distribution of the returns is skewed towards the left.

Table 4.2: Summary Statistics of daily Returns

Statistics	mean	Max	Min	St. Dev	Ex. Kurtosis	Skewness
Value	$1.4e^{-4}$	0.2103	-0.2103	0.0115	83.61	0.1159

4.1.3 Test for Stationarity and Normality

Augmented Dickey Fuller test has been applied to test for the stationary of the data. The result from Table 4.3 revealed that null hypothesis which is the series are not stationary has been rejected since the p-value 2.2×10^{-16} is less than 5% level of confidence. The Jargue Bera test for normality rejects the null hypothesis that the distribution is normal, since the p-value is far less than 5%.

Table 4.3: Jargue Bera Test and ADF statistics

Jargue Bera Test		Augmented Dickey Fuller Test		
JB (p-value)	Sig level	ADF Statistic	ADF(P-value)	Sig level
$2.2e^{-16}$	5%	-14.552	0.01	5%

4.1.4 GARCH Models Selection

The Autocorrelation function and Partial Autocorrelation where applied to obtain the lags in the GARCH(p,q) model. ACF and PACF helps to know which of our past series values are most useful in predicting future values. The ACF helps to determine the length of our past conditional variance. It indicates the number of GARCH terms. The PACF determines the length of the past square innovation (q) where the lags which PACF cuts off is the indicated number of ARCH terms (p).

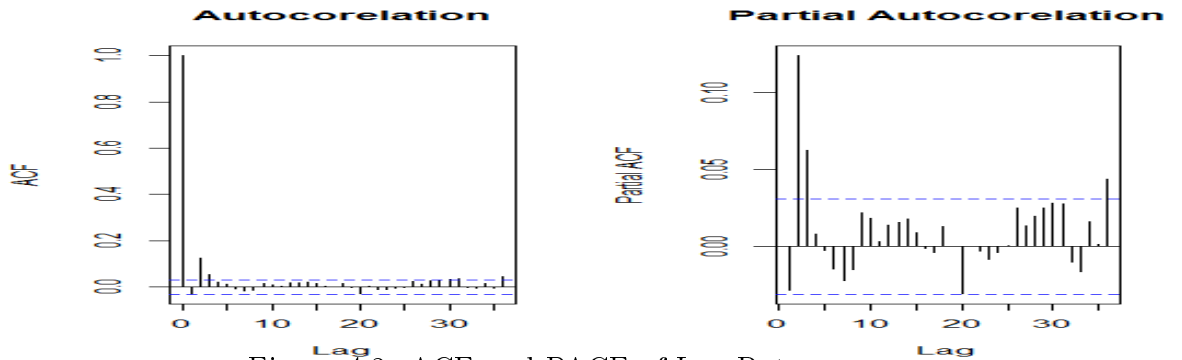


Figure 4.3: ACF and PACF of Log Returns

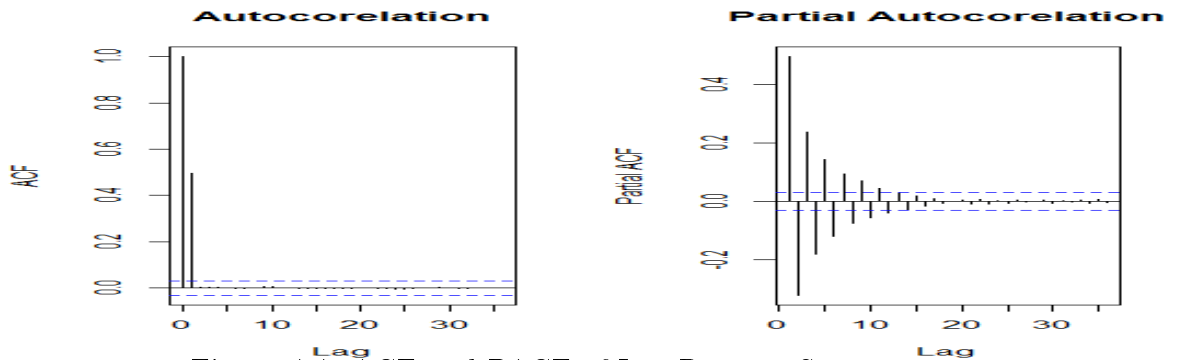


Figure 4.4: ACF and PACF of Log Returns Squares

According Figures 4.3 and 4.4, the most appropriate model is GARCH(1,1). The Akaike Information Criterion and Bayesian information criterion were used to select the best model for the data. The best model is the model with smallest AIC or BIC, Akiake (1973) and Schwarz (1978). Here the best model is Garch(1,1) which have all its parameters with p-values as 0.000 which is significant at 05% except the mean which has it p-value as 0.06 which is significant at 10% . All orders above Garch(1,1) have some of their parameters not significant at 10%.

Table 4.4: GARCH Model Selection (with student-t distribution innovation)

Parameter	GARCH(1,1)		GARCH(1,2)		GARCH(2,1)	
	Value	P-Value	Value	P-Value	Value	P-Value
δ_1	0.556518	0.0000	0.556272	0.0000	0.563300	0.0000
δ_2					0.015512	0.1301
β_1	0.298328	0.0000	0.299223	0.0000	0.269438	0.0000
β_2			0.020697	0.20134		
ω	0.000019	.06120	0.000445	0.00000	0.000323	0.40316
AIC	-7.1023		-7.1016		-7.1017	
BIC	-7.0944		-7.0922		-7.0923	

4.2 Conditional Volatility Estimation

Next, the parameters of the selected model are estimated using Quasi maximum likelihood procedure by fitting the model in our daily returns. Table 4.5 gives the estimate of the parameters of the GARCH(1,1)

Table 4.5: GARCH(1,1) (with student-t innovation)

δ_1	β_1	ω	$\delta_1 + \beta_1$	Observation
0.556518	0.298328	0.000019	0.854844	Low Persistent

Table 4.5 shows that the sum of ARCH and GARCH is less than one ,indicating that conditional variance is less persistent to shocks, Bollerslev 1996).The coefficients of both the ARCH and GARCH terms are all positive. The log returns show the existence of volatility clustering in the data. ACF and PACF confirmed existence of autocorrelation like the Dickey Fuller test did.

Table 4.6: JB and LM statistics for Residuals (with student-t innovation)

Jargue Bera test			Augmented Dickey Fuller Test		
JB-stat.	JB (p-value)	Sig. level	ADF Statistics	ADF (p-Value)	Sig. Level
1171240.184	$<2.2e^{-16}$	5%	-14.552	5%	5%

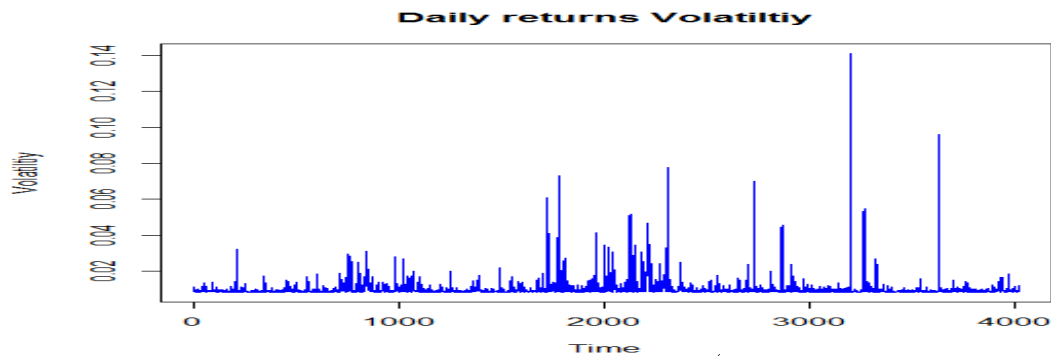


Figure 4.5: Daily Volatility NSE 20 Share index (with student-t distribution innovation)

Figure 4.5 gives the daily volatility of NSE 20 Share index. It shows that there are situations in with extreme high uncertainty in the Nairobi 20 Share index.

4.3 Estimation of Extreme Quantile

The aim of this section is to estimate the extreme quantiles in the daily returns using the extreme Value theory (EVT). The randomness in the model comes through the random variable e_t , which are called innovations of noise of the process and it is assumed to be independent and identically distributed with unknown distribution $F(e)$. Before estimating the extreme qauntile the residuals have to standardize first i.e $\hat{e}_t = \frac{\hat{y}_t}{\hat{\sigma}_t}$ where y_t is the returns series, and $\hat{\sigma}_t$ is the

estimated volatility from the returns. The ACF and PACF are plotted to show that the residuals are not auto correlated.

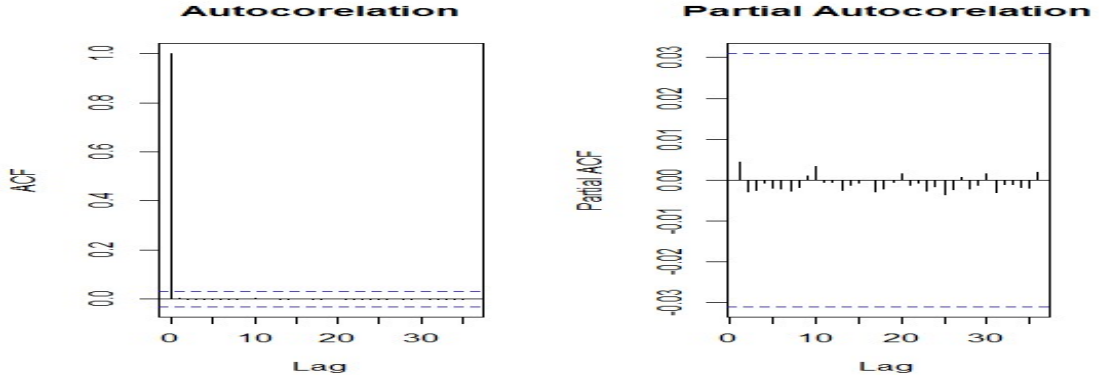


Figure 4.6: ACF and PACT of Square Residuals(GARCH with Student-t Distribution innovation)

Table 4.7: Jargue Bera and LM test for the Standardized Residuals(GARCH with Student-t distribution innovation)

Jargue Bera test		Arch Test (LM)	
JB-Stat.	p-value	χ^2 Statistic	P-Value
5732223.251	$<2.2e^{-16}$	1681.5	$2.2e^{-16}$

From Figure 4.6, it can be seen that the square residuals are approximately iid since all the lags are inside the confidence interval. This implies that GARCH(1,1) model used is an appropriate filter for the data. From Table 4.7 it can be seen that the standardized residuals are non-stationary and have a significant LM p-value.

4.3.1 Threshold selections

1. Quantile- Quantile Plot:

The QQ-plot is used for two reasons: first it confirms the Jargue Bera (JB) result on test for normality. This means that JB shows that residuals are not normal while the QQ-plot shows that the standardized residuals are heavy tail, which is the basis for the application of EVT. Secondly it can be

used to check if the data satisfies GPD. According to Picklands (1975) and Balkema (1974) if the empirical plots seem to follow a reasonable straight line with a positive slope above a certain threshold, then the data follows GPD. Hence according to Figure 4.7 the data follows a GPD with a scale parameter and shape parameter. Here the threshold chosen where the graph approximately linear in shape. According to the QQ-Plot of the residuals against the normal distribution, the standardized residuals are heavy tail, since the plot is convex on the left and concave on the right. Hence we cannot conclude that the data is conditional normal.

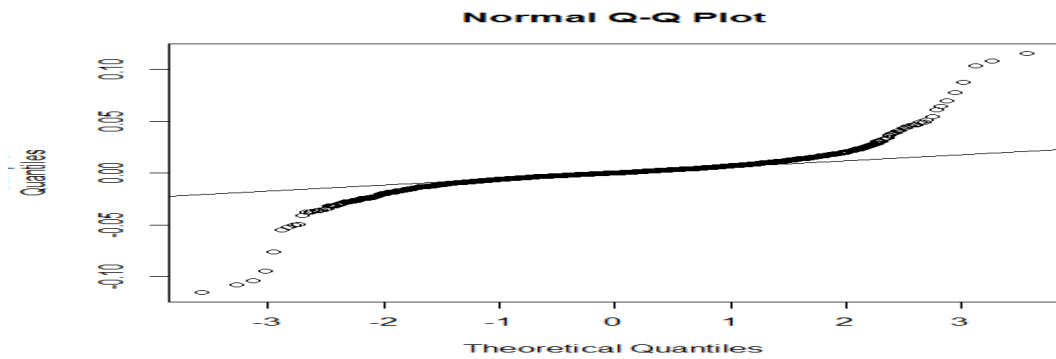


Figure 4.7: Q-Q plot of the data against the Normal plot

2. Mean Residuals life Plot:

Figure 4.8 is the Mean Residuals life plot of the negative returns. It is observe that the mean excess plot in Figure 4.8 shows an upward trend for the data, which indicates heavy tail behaviour. Since the graph seems to follow a straight line with positive slop above a certain threshold, this is evidence that the data follow a GPD. From Figure 4.8the thresholds can be chosen as 0.997 for the right tail based on the criterion of linearity in the MEP plots.

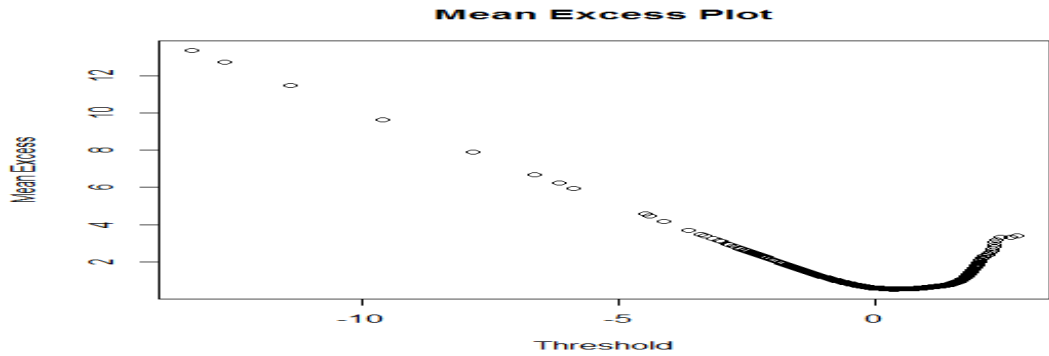


Figure 4.8: Mean Excess plot of data

3. **Hill Plot:** The plot of the shape parameter against the exceedances in the Figure 4.9 is helpful in threshold selection. The threshold is chosen where the line graph seems to be horizontal. If low threshold is selected which includes data from the center then the estimates become bias whereas for a very high threshold, the estimates become highly volatile.

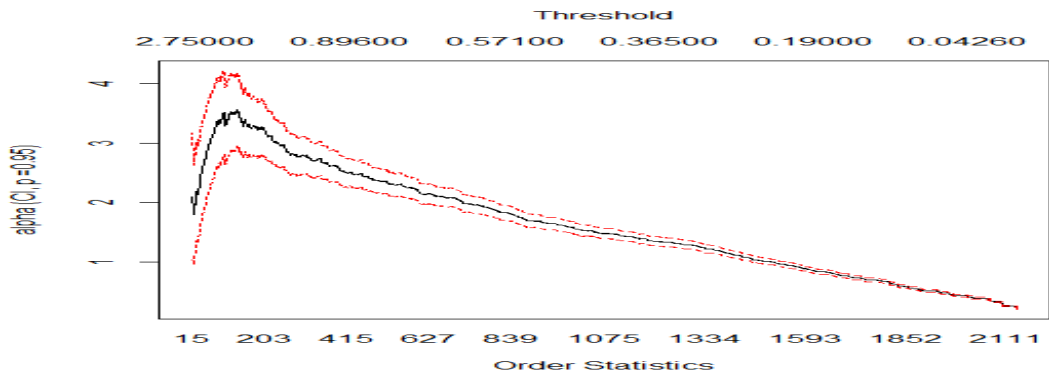


Figure 4.9: Hill Plot of data

From Figure 4.9 the threshold can be selected in the region from 0.85 to 1.2. This is a place on the graph where it is relative linear.

4.3.2 Estimation of GPD parameters

After the selection of threshold for the returns using any of the graphical threshold selection above, the number of exceedances over the threshold are used to determine the GPD parameter i.e the shape and scale parameters. The result of these estimates is presented in Table 4.8 below

Table 4.8: Estimates of GPD parameter (For GARCH with student-t distribution innovation)

$\hat{\varepsilon}$	$\hat{\beta}$	No. of exceedance	Threshold
0.3623955	0.4641953	458	0.997

In the Table 4.8 $\hat{\varepsilon}$ represents the shape parameter. It determines the type of distribution for our data. Since it is positive, this indicates that the data belongs to maximum domain of attraction of Frechet distribution which is a heavy tail distribution. $\hat{\beta}$ represents the scale parameter of the distribution.

4.3.3 The estimation of Extreme quantile

Using the GPD shape parameter and scale parameter estimated above, the quantile at extreme probability values can be calculated from the standardized residuals. Let \hat{e}_q denote the quantiles estimate of the innovation at probability q . Typically, the probability q is such that $0.99 \leq q < 1$. The quantile estimate is given by

$$\hat{e}_q = u + \frac{\hat{\beta}_u}{\hat{\varepsilon}} \left(\left(\frac{n(1-q)}{k} \right)^{-\varepsilon} - 1 \right) \quad (4.1)$$

where $\hat{\beta}_u$ and $\hat{\varepsilon}$ represent the estimates of the β_u and ε scale and shape parameters respectively. k represents the number of observation above the threshold u whereas n is the total number of observation. Table 4.9 gives the results with $q = 0.95, 0.99$ and 0.995 for normal innovations and student-t innovation. It can be seen that the estimates GARCH with student-t innovation are greater than those of GARCH with normal innovation which was expected.

Table 4.9: VaR and ES estimates (GARCH- with student-t distribution innovation)

	$\hat{e}_{0.95}$	$\hat{e}_{0.99}$	$\hat{e}_{0.995}$	$\hat{E}S_{0.95}$	$\hat{E}S_{0.99}$	$\hat{E}S_{0.995}$
$POT^{(t)}$	1.4672	2.8423	3.6876	3.1978	5.2109	6.4448

Table 4.9 give the values for VaR and ES. The VaR value gives the maximum that NSE 20 can lose under normal market conditions at a given confident level. Whereas the ES values gives the expected losses given that VaR violations have occurred. It shows that the model with student-t distribution innovations estimates VaR and CES better.

4.3.4 Estimates of Conditional VaR and Conditional Expected Shortfall

In Table 4.9 the estimates of the unconditional extreme quantiles for both GARCH-EVT with normal and student-t distribution innovation are shown. To obtain the conditional VaR the unconditional quantile e_q^t is combine with the conditional volatility estimates. Figures 4.10 and 4.11 give the plots of conditional VaR for both distributions

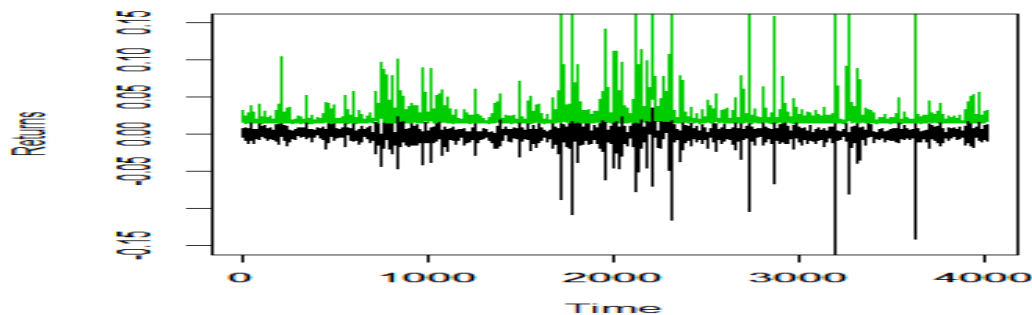


Figure 4.10: Conditional $VaR_{0.99}$ (GARCH model with normal distribution innovation)

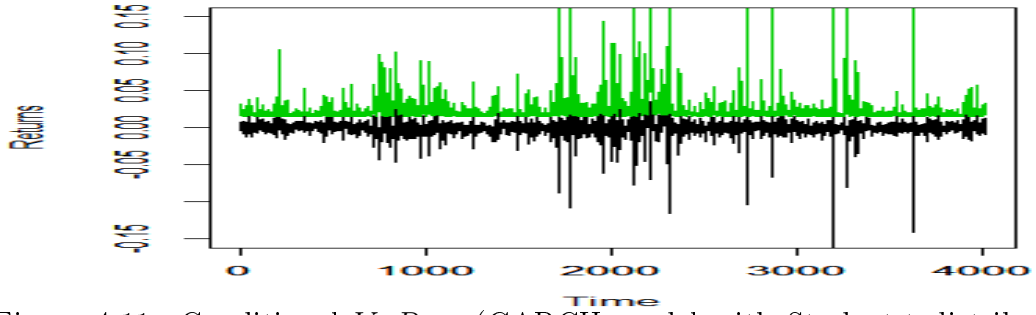


Figure 4.11: Conditional $VaR_{0.99}$ (GARCH model with Student-t distribution innovation)

Graphs 4.10 and 4.11 give the CVaR estimates at probability level $q = 0.99$ for both GARCH with normal and student-t innovations respectively. According to both the conditional VaR graphs smoothly trace the graph of lost distribution. This implies that the estimates using GARCH-EVT are all reliable but the one from GARCH-EVT with student-t distribution innovation is better since it appears that it has less places where the lost distribution crosses the graph of the CVaR. Expected Shortfall is a more desirable risk measure according to its more attractive theoretical properties. It overcomes the limitations of VaR. The estimation of CES under extreme conditions requires estimation of volatility σ_t and using appropriate extreme value distribution to obtain the quantile. Hence the CES for both GARCH-EVT with normal innovation and student-t innovation respectively are given by

$$CES_q^{t+1} = \mu_{t+1} + \hat{\sigma}_{t+1} \left(\frac{VaR_q}{1 - \hat{\epsilon}} + \frac{\hat{\beta}_u + \hat{\epsilon}u}{1 - \hat{\epsilon}} \right) \quad (4.2)$$

where VaR_q is the unconditional VaR which can be computed as $\hat{\epsilon}_q$. The CES for both GARCH-EVT with normal innovation and student-t distribution innovation respectively are given in Figures 4.12 and 4.13

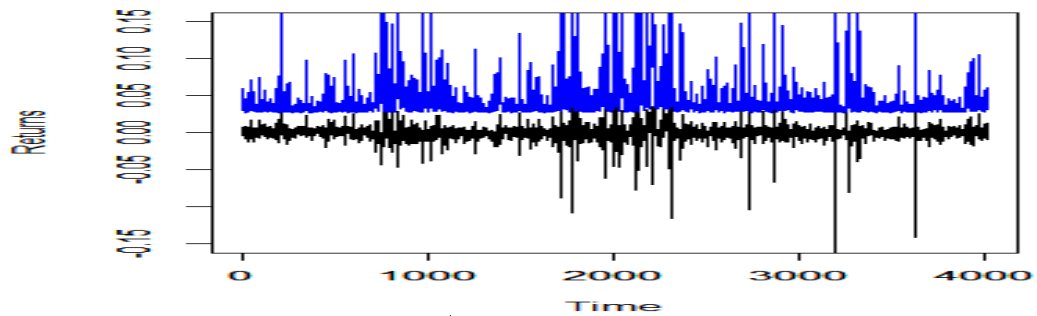


Figure 4.12: Conditional $ES_{0.99}$ (GARCH model with Student-t distribution innovation)

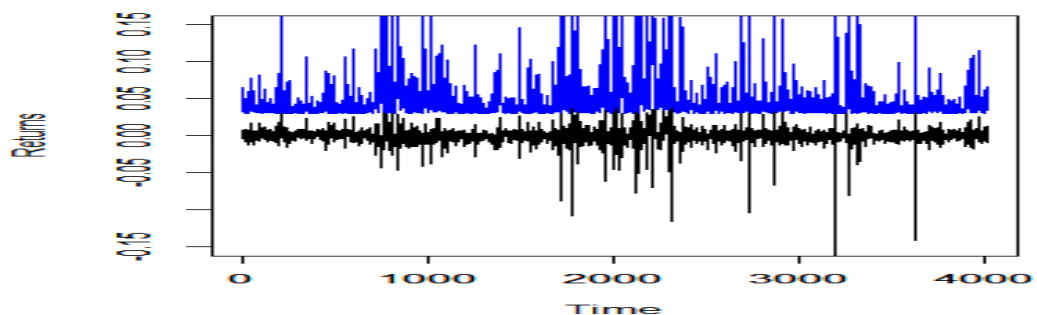


Figure 4.13: Conditional $ES_{0.99}$ (GARCH model with Student-t distribution innovation)

Figures 4.12 and 4.13 illustrate the average of all losses which are greater than or equal to VaR at each point for both GARCH-EVT with normal and student-t distribution innovation. Expected Shortfall is defined as the average of all losses which are greater or equal than VaR, i.e. the average loss in the worst $(1 - q)\%$ cases, where q is the confidence level. It gives the expected value of an investment in the worst $p\%$ of the cases. Again we have in both Figure 4.12 and Figure 4.13, the CES graph tracing the graph of the lost distribution but the one with student-t distribution innovation have fewer points of intersections of the two graphs. This implies the estimates under student-t innovation are better than that of normal innovation.

4.4 Backtesting VaR and ES

A window of a length of 1000 days is used in the backtesting with a degree of freedom of 4 for the t-distribution and bootstrap samples of 10,000 are used.

4.4.1 Test for VaR Estimates

Table 4.10 below gives the VaR-violations at different coverage levels. The VaR-violations are obtained by comparing the actual loss and forecasted VaR for each day. It shows that both models have lower Violations at lower confidence level than higher confidence level but the estimates with t-distribution distribution innovation perform better than the model with normal innovation all levels except at 99.5% where they produce the same number of violations.

Table 4.10: VaR Violations for Normal distribution innovation and student-t distribution innovation

	$POT^{(t)}$ (No.of Violations)	$POT^{(n)}$ (No.of Violations)
q=0.95	187	198
q=0.99	29	32
q=0.995	17	17

Unconditional Coverage

Here the unconditional coverage test is conducted described in Chapter 3. From Table 4.11 both the model with student-t distribution innovation and normal innovation have large P-value at a lower confidence levels. This implies that the models are producing the correct number of violation. At a higher confident level the model with normal innovation have a small p-value. This implies that the null hypothesis that the model with normal innovation have the correct coverage rate is rejected.

Independent Test

From Table 4.11 gives the results for independent test on both models. Here again at a lower confidence level it shows that violation produce by both models are independent since they all have large P-value. Whereas, at a higher confidence levels the model with normal with normal innovation is rejected.

Conditional Coverage

From Table 4.11 gives results for the conditional coverage test. Here again it shows that the model with student-t innovation perform better than the model with normal innovation.

Table 4.11: Results for Backtesting VaR (for both normal and student-t innovations)

	$POT^{(t)}$ (P-value)	$POT^{(n)}$ (P-value)
q=0.95		
Unconditional Coverage Test	0.349	0.515
Independent test	0.993	0.575
Conditional Coverage Test	0.641	0.543
q=0.99		
Unconditional Coverage Test	0.331	0.001
Independent test	0.883	0.015
Conditional Coverage Test	0.417	0.001
q=0.995		
Unconditional Coverage Test	0.936	0.001
Independent test	0.991	0.340
Conditional Coverage Test	0.564	0.001

4.4.2 Backtesting ES (for both normal and student-t innovations)

Bootstrap

Table 4.12 gives the results for testing ES using the bootstrapping method. Both the model with student-t distribution innovation and normal innovation performed satisfactorily at a lower confidence level. But at a higher confidence the model with student t-distribution innovation performed better.

Table 4.12: Result on backtesting ES (for both normal and student-t innovations)

	$POT^{(t)}$	$POT^{(n)}$
	(P-value)	(P-value)
q=0.95	0.9916008	0.993001
q=0.99	0.9923008	0.4158584
q=0.995	0.4039586	0.01269873

V-Test for Expected Shortfall(for both normal and student-t distributions innovations)

Table 4.13 gives the results for the V-test at different confident levels. The value of V-test is better the more it is closer to zero. From the V-test it is clear that the model with student-t distribution innovation performed better than the model with normal distribution innovation. In all confidence levels the model with student-t distribution have its values closer to zero, implying it gives a better estimate of the conditional expected Shortfall than the model with normal innovation.

Table 4.13: Result on V-test for backtesting ES (for both normal and student-t innovations)

	$POT^{(t)}$	$POT^{(n)}$
q=0.95		
V_1	-0.00548136	-0.008737709
V_2	-0.00470920	-0.007054860
V_3	0.00509528	0.007896285
q=0.99		
V_1	0.00521026	0.005536892
V_2	-0.001727765	-0.0020207837
V_3	0.011246335	0.0103807630
q=0.995		
V_1	0.01140205	0.06901005
V_2	-0.02458258	-0.0028334456
V_3	0.01799232	0.017617730

Chapter 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In this work conditional VaR and Conditional ES of Nairobi 20 shared index with normal and student-t innovations is estimated and the estimates backtested. These estimates are obtained by combining the GARCH model in the estimation of the volatility and extreme value theory. The Quasi-maximum likelihood procedure is used in estimating the GARCH parameters. These estimates are found to be consistent and asymptotically normal. The data exploration reveal that the data is not normally distributed but fat-tail. The distribution of the NSE 20 index is skewed to the left meaning investor can have frequent gain and few extreme losses.

The GPD was fitted to to the standardized residuals for both innovations and then the estimated distribution inverted to obtain extreme quantiles at 95% ,99% and 99.5% confident level. The MLE of the parameters are found to be consistent and asymptotically normal.

The VaR estimates are backtested using both the conditional and unconditional coverage. The model with normal innovation have high number of VaR violation at lower confident level than that of the student-t innovation. According to the unconditional coverage test both the model with normal and student-t innovation have high p-values at lower confidence level implying that they all produced the correct number of violations. But at a higher confident level the model with normal innovation is rejected. Producing the correct number of violations alone is not enough to justify whether a model is good or not. Hence it is vital to test if the violations produced are independent. Because if the violations are not independent, it means that the model do not adapt sufficiently and quickly

enough to large losses and may possibly create a risk of bankruptcy in a very short time as losses pile on. According to the p-value of the independent test the violations produce by the models are independent. Combining the unconditional and independent to give conditional coverage test. According to the conditional coverage test the model with student-t innovation out performed the model with normal innovation at higher confident level.

Two methods where used to backtest the conditional expected shortfall estimates i.e bootstrap method by McNeil and Fery (2000) and the V-test by Embrechts *et. al* (2005). According to the bootstrap result all the model perform better at a low confident level while the model with normal innovation fails at a higher confident level. All the V-test values revealed the model with student-t innovation performs better than the model with normal innovations since it have all its values closer to zero.

Being able to Estimate risk (VaR and ES) and backtest the estimates is very important since it allows policy makers and risk manager to be able to make good decisions about direction of portfolios. Backtesting the estimates is crucial in risk management because it helps to know whether the model used is given correct estimates.

5.2 Recommendation

1. Instead of using the symmetric GARCH the study recommends use of asymmetric GARCH models to see whether the method will predict CVaR and CES better.
2. Since it is a know fact the selection of a threshold is a trade off between bias estimator or high variance (Cole 2001). Hence the use of non-graphical

method in estimate the optimal threshold is also recommended.

3. Finally on backtesting Expected shortfall the study recommends the use of any of the models or both the models developed by Carlos *et. al* (2014) to see whether result arrived at will change.

References

- Acerbi, C., & Szekely, B. (2014). Backtesting expected shortfall. *MSCI Inc.*, 1-37.
- Akaike, H. (1973). Information theory and extension of the maximum likelihood theory.
- Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1997). Thinking coherently. *Risk*, 10(11), 68-71.
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.
- Bensalah, Y. (2000). Steps in applying extreme value theory to finance: A review. *Bank of Canada Working Paper*.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Brodin, C., & Kluppelberg, E. (2006). Extreme value theory in finance.
- Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(4), 478-484.
- Christoffersen, P. F. (2003). Elements of financial risk management. *Academic Press*.
- Davison, A. C. (1984). Modelling excess over threshold with an application, statistics extreme and application'. *NATO ASI Ser. C*, 131, 461-482.
- Du, Z., & Escanciano, J. C. (n.d.). Backtesting expected shortfall: Accounting for tail risk. *SSRN Electronic Journal*.
- Eagel, R. (2001). The use of arch/garch models in applied econometrics. *Journal of Economic Perspectives*, 15(4), 157-168.
- Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. *Chapman and Hall*.
- Embrechts, P., Kaufmann, R., & Patie, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications*, 32(1-2), 61-90.

- Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2), 30-41.
- Fisher, H., & Tippett, R. (1928). Limiting forms of the frequency distribution of the largest and smallest members of a sample. *Proceeding of the Cambridge philosophy Society*, 24, 180-190.
- Frad, H., & Ezzeddine, Z. (2014). Estimation of value at risk measures in the islamic stock market: Approach based on extreme value theory. *Journal of World Economic Research*, 3(2), 15-20.
- Gencay, R., Selcuk, F., & Ulugalyagci, A. (2003). High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics*, 33(2), 337-356.
- Gilli, M., & Ellezi, E. (n.d.). An application of extreme value theory for measuring financial risk. , 27.
- Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance*, 48(5), 1779-1801.
- Gnedenko, B. V. (1943). Sur la distribution limits du terme maximum d'une serie aleatoire. *Annal of Mathematics*, 44, 423-453.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of American Statistics Association*, 106(494).
- Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2).
- Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3).
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158-171.
- Kourouma, L., Dupre, D., Sanfilippo, G., & Taramasco, O. (n.d.). Extreme

- value at risk and expected shortfall during financial crisis. *SSRN Electronic Journal*.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models,. *Journal of Derivatives*, 3, 73-84.
- Lambert, N., Pennock, D. M., & Shoham, Y. (2008). Eliciting properties of probability distributions: the highlights. *ACM SIGecom Exchanges*, 7(3), 1-5.
- Lauridsen, S. (2001). Estimation of value at risk methods. *Goldman Sacks International*.
- Ljung, G., & Box, G. (1978). Measure of lack of fit in the series models. *Biometrika*.
- Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch(1 1) and covariance stationary garch(1 1) models. *Econometrica*, 64, 575-596.
- Ma, Y., Ye, F., & Bi, X. (2009). The empirical study of dynamic var on the index of shanghai security market based garch-evt.
- Maana, I., Kamau, A., & Kisinguh, K. (2015). Modelling extreme volatility in the daily exchange rates of the kenya shilling against the u.s. dollar. *Journal of Economics and International Finance*, 7(9), 192-203.
- Mancini, L., & Trojani, F. (n.d.). Robust value at risk prediction. *SSRN Electronic Journal*.
- Manganiél, S., & Eagle, R. (2001). Value at risk models in finance , european central bank working paper series, frankfurt. *Mathematical Finance*, 14, 75-102.
- Marimoutou, V., Raggad, B., & Trabelsi, A. (2009). Extreme value theory and value at risk: Application to oil market. *Energy Economics*, 31(4), 519-530.
- McNeil, A. (1997). The peaks over threshold method for estimating high quantiles of loss distribution.

- McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7(3-4), 271-300.
- Mises, R. V. (1955). . la distribution de la plus grande de n valeurs. selected papers. *Providence R. I.: Amer. Math. Soc, II*, 271-294.
- Murenzi, R., Thomas, K., & Mung'at, J. K. (2015). Modeling exchange market volatility risk in rwanda using garch-evt approach. *International Journal of Thesis Projects and Dissertations (IJTPD)*, 3(3), 67-80.
- Mwamba, J. W. M., Hammoudeh, H., & Gupta, R. (2014). Financial tail risks and the shapes of the extreme value distribution: A comparison between conventional and sharia-compliant stock indexes. *University of Pretoria , Department of Economics Working Paper Series*.
- Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics*, 45(1-2), 7-38.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2).
- Odening, M., & Hinrichs, J. (2000). Using extreme value theory to estimate value-at-risk.
- Osband, K. H. (1985). *Providing incentives for better cost forecasting* (Unpublished doctoral dissertation). University of California, Berkeley.
- Paul, S., & Karmakar, M. (2015). Intraday var using component garch-evt approach: An empirical study in select stock markets. *Proceedings of the Fourth International Conference on Global Business, Economics, Finance and Social Sciences (GB15Kolkata Conference)*.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Singh, A. K., Allen, D. E., & Powell, R. J. (2011). Value at risk estimation using extreme value theory. *Modelling and Simulation Society of Australia*

and New Zealand. Retrieved from <http://ro.ecu.edu.au/ecuworks2011/721>

- Smith, R. L. (1984). Threshold methods for sample extremes, in statistical extremes and application. *NATO ASI Ser., Ser. C, 131*, 621-638.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika, 72*, 67-90.
- Smith, W., Clifford, C., & Stykes, W. (1990). Managing financial risk. *Iowa State University Press*.
- Soltane, H. B., Karaa, A., & Bellalah, M. (2012). Conditional var using garch-evt approach: Forecasting volatility in tunisian financial market. *Journal of Computations and Modelling, 2(2)*, 95-115.
- Straumann, D., & Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics, 34(5)*, 2449-2495.
- Stuart, S. (2001). An introduction to statistical modeling of extreme values. *Springer-Verlag*.
- Su, E., & Knowles, T. W. (2006). Asian pacific stock market volatility modeling and value at risk analysis. *Emerging Markets Finance and Trade, 42(2)*, 18-62.
- Todorovic, P., & Zelenhasic, E. (1970). A stochastic model for flood analysis. *Water Resources Research, 6(6)*, 1641-1648.
- Uppal, J. Y., & Mangla, U. I. (2013). Extreme loss risk in financial turbulence-evidence from the global financial crisis. *Managerial Finance, 39(7)*, 653-666.
- Wang, W., Van Gelder, J., Vrijling, J., & Ma, J. (2005). Testing and modelling autoregressive conditional heteroskedasticity of stream flow processes. *Non-linear processes in Geophysics, 12*, 55-66.

APPENDIX

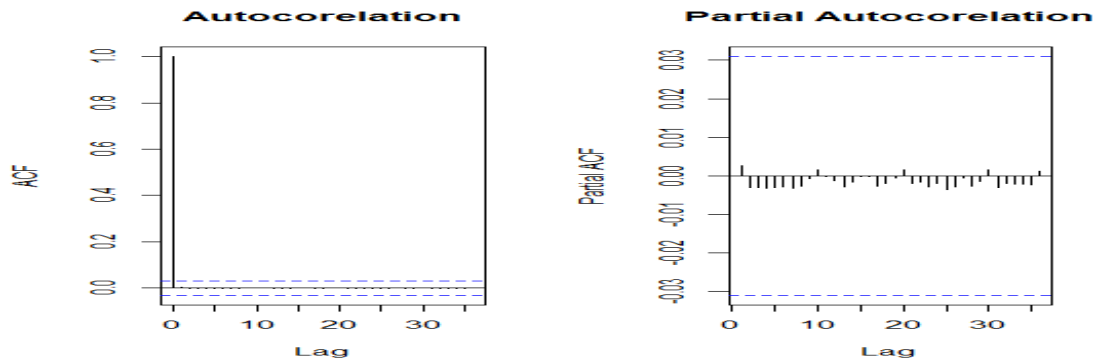
Appendix 1: GARCH(1,1) with Parameters with normal distribution innovations

Parameters	δ_1	β_1	ω
Values	0.062162	0.930566	0.000132
P-value	0.0000	0.0000	0.0724

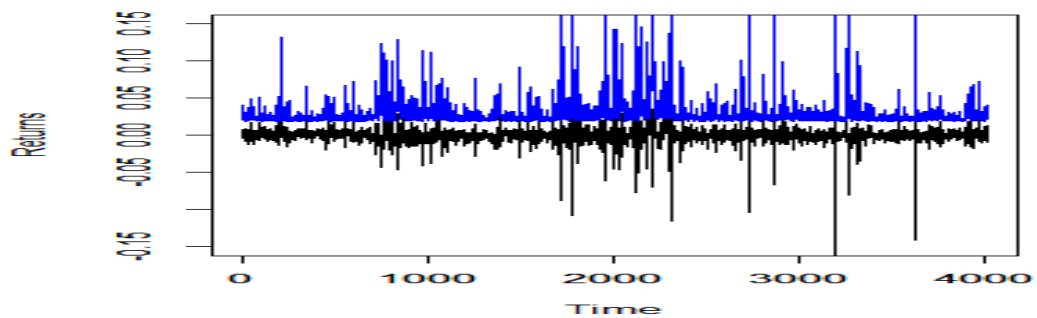
Appendix 2: GPD estimates for GARCH-EVT (with normal distribution innovations)

Parameter	u	k	β	ξ
Values	0.848	602	0.467156	0.3140057

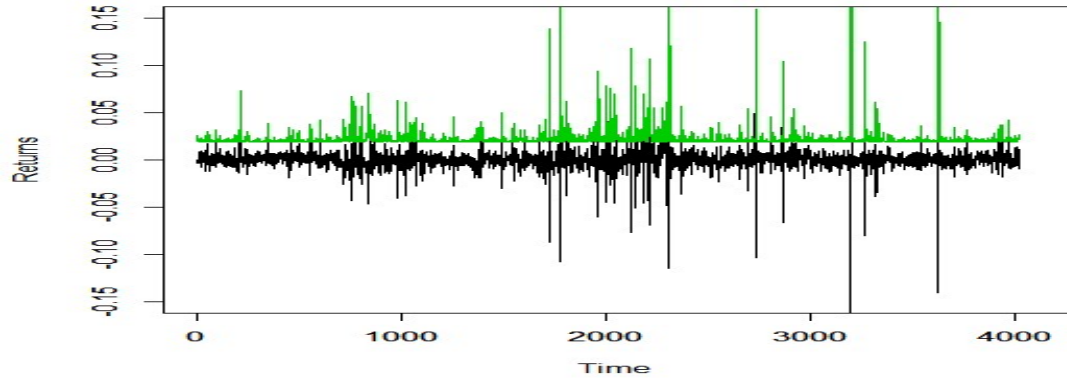
Appendix 3: ACF and PACF (GARCH model with Student normal distribution)



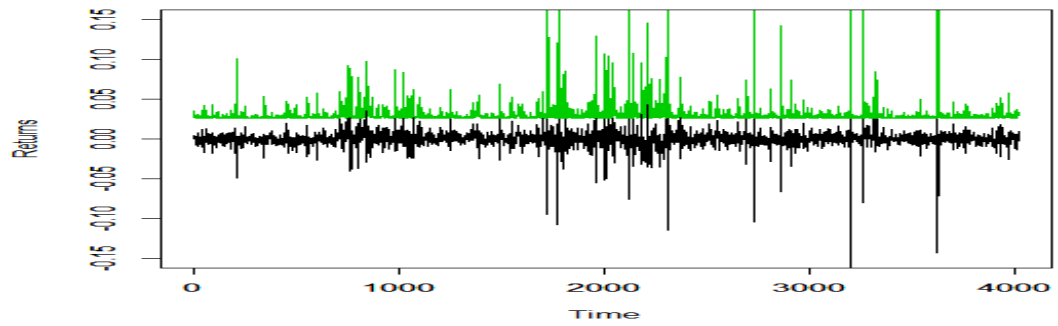
Appendix 4: Conditional $CVaR_{0.95}$ (GARCH model with Student normal dis-



Appendix 5: Conditional $ES_{0.95}$ (GARCH model with Student-t distribution)



Appendix 6: Conditional $CVaR_{0.995}$ (GARCH model with Student-normal dis-



Appendix 7: Conditional $ES_{0.995}$ (GARCH model with Student-t distribution)

