



**MODELLING SOUTH AFRICA'S MARKET RISK USING
THE APARCH MODEL AND HEAVY-TAILED
DISTRIBUTIONS**

By

ILUPEJU, YETUNDE ELIZABETH

Project submitted to the University of KwaZulu-Natal in fulfilment of the requirements for
the Master degree in Statistics

in the

University of KwaZulu-Natal

School of Mathematics, Statistics and Computer Science

Westville Campus, Durban, South Africa

December, 2016

Thesis advisors:

Mr. Retuis Chifurira

Mr. Knowledge Chinhamu

Prof. Michael Murray

“As the candidate’s Supervisor I agree/do not agree to the submission of this dissertation”.

Mr. Retuis Chifurira Signed: _____ Date: _____

Mr. Knowledge Chinhamu Signed: _____ Date: _____

Prof. Michael Murray Signed: _____ Date: _____

Declaration-Plagiarism

I, **Yetunde Elizabeth Ilupeju**, declare that:

1. The research reported in this dissertation, except where otherwise indicated, is my original research.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a. Their words have been re-written but the general information attributed to them has been referenced.

 - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

.....

Yetunde Elizabeth Ilupeju

Disclaimer

This document describes work undertaken as part of a master programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

ABSTRACT

Estimating Value-at-risk (VaR) of stock returns, especially from emerging economies has recently attracted attention of both academics and risk managers. This is mainly because stock returns are relatively more volatile than its historical trend. VaR and other risk management tools, such as expected shortfall (conditional VaR) are highly dependent on an appropriate set of underlying distributional assumptions being made. Thus, identifying a distribution that best captures all aspects of financial returns is of great interest to both academics and risk managers. As a result, this study compares the relative performance of the GARCH-type model combined with heavy-tailed distribution, namely Skew Student t distribution, Pearson Type IV distribution (PIVD), Generalized Pareto distribution (GPD), Generalized Extreme Value distribution (GEVD), and stable distribution in estimating Value-at-Risk of South African all share index (ALSI) returns. Model adequacy is checked through the backtesting procedure. The Kupiec likelihood ratio test is used for backtesting. The proposed models are able to capture volatility clustering (conditional heteroskedasticity), and the asymmetric effect (leverage effect) and heavy-tailedness in the returns. The advantage of the proposed models lies in their ability to capture volatility clustering and the leverage effect on the returns, though the GARCH framework and at the same time model their heavy tailed behaviour through the heavy-tailed distribution. The main findings indicate that APARCH model combined with this heavy-tailed distribution performed well in modelling South African market's risk at both the long and short position. It was also found that when compared in terms of their predictive ability, APARCH model combined with the PIVD, and APARCH model combined with GPD model gives a better VaR estimation for the short position while APARCH model combined with stable distribution give the better VaR estimation for long position. Thus, APARCH model combined with heavy-tailed distribution model provides a good alternative for modelling stock returns. The outcomes of this research are expected to be of salient value to financial analysts, portfolio managers, risk managers and financial market researchers, therefore giving a better understanding of the South African market.

Key Words:

Asymmetric power ARCH (APARCH), Value-at-Risk (VaR), Kupiec test, Pearson type IV distribution (PIVD), Generalized Pareto Distribution (GPD), Generalized Extreme Value Distribution (GEVD), and Stable distribution, All share index (ALSI).

ACKNOWLEDGEMENT

I am grateful unto the Lord God for the grace given me through Christ Jesus to complete my academic work especially this research. I also appreciate my supervisors Mr. R. Chifurira, Mr. K. Chinhamu, and Prof. M. Murray for successfully guiding me in the course of this research. In the short time I have spent at the University of Kwa-Zulu Natal, my knowledge has been refreshed. Everyone in the department has been of great help to me; the interaction has made me a better person. I can proudly say that I enjoyed every moment at the department.

I would like to appreciate everyone who contributed to this research work beginning with Prof North, a great mentor. Dr Lougue, thank you for your ideas, goodwill and advice. Prof Matthews, the mother of everyone and all the lecturers in the department who have enhanced my knowledge base.

I cannot but appreciate my dear husband who stood by me throughout the period of my academic activities. My children Moyosola, Ireoluwa, and Anuoluwa, thank you for your love and patience. My parents have been supporting, thank you all. To all my friends and families who have contributed to my success, thank you too.

I pray that God will continue to increase everyone, thank you for being there.

DEDICATION

I dedicate this work to my Lord and Saviour.

TABLE OF CONTENTS

Declaration-Plagiarism	iii
Disclaimer	iv
ABSTRACT.....	v
ACKNOWLEDGEMENT	vi
DEDICATION.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF APPENDICES.....	xiv
ABBREVIATION.....	xv
CHAPTER 1	1
INTRODUCTION	1
1.1 Introduction	1
1.2 Background	1
1.3 Literature Review	5
1.4 Stylized Facts of Asset Returns.....	10
1.5 Problem Statement	11
1.6 Aims and Objectives	11
1.7 Significance of the study	12
1.8 Research Layout.....	12
CHAPTER 2	13
VOLATILITY MODELS	13
2.1 Introduction	13
2.2 Basic Concept.....	13
2.2.1 Stationarity	13
2.1.2 Serial Correlation	14

2.2.3	Conditional Heteroskedasticity	14
2.3	Basic Structure of Financial Volatility Model	15
2.4	The Mean Model	16
2.5	The Volatility Models	17
2.5.1	Symmetric GARCH Models	18
2.5.2	Asymmetric GARCH Models	22
2.6	Maximum Likelihood Estimation (MLE)	26
2.6.1	Gaussian Quasi Maximum-Likelihood Estimation	27
2.6.2	Fat Tailed Maximum-Likelihood Estimation (MLE)	30
2.6.3	Skewed Maximum-Likelihood Estimation	31
2.7	Conclusion	32
CHAPTER 3	33
EXTREME VALUE THEORY (EVT)	33
3.1	Introduction	33
3.2	Extreme Value Modelling	33
3.2.1	Generalized Extreme Value Distribution (GEVD)	36
3.2.2	Generalized Pareto Distribution (GPD)	38
3.3	Parameter Estimation	41
3.3.1	Maximum Likelihood Estimation of GEVD	42
3.3.2	Maximum Likelihood Estimation of GPD	44
3.4	Problems with Extreme Value Modelling	46
3.4.1	Dependence of Extremes	46
3.4.2	Lacking extremal observation	47
3.4.3	Threshold selection and Common Approach	47
3.5	Conclusion	49
CHAPTER 4	50
STABLE DISTRIBUTION	50
4.1	Introduction	50
4.2	Definition and Properties of Stable Distributions	50
4.2.1	Characteristic Function Representation	51
4.2.2	Stable Probability Density Function	53
4.3	Parameter Estimation of Stable Distributions	55
4.4	Conclusion	56
CHAPTER 5	57

PEARSON TYPE IV DISTRIBUTION (PIVD)	57
5.1 Introduction	57
5.2 The Probability Density Function (pdf) of PIVD.....	57
5.3 The Cumulative Density Function (cdf) of Pearson Type IV Distribution.....	59
5.4 Parameter Estimation of Pearson Type IV Distribution.....	59
5.5 Conclusion.....	60
CHAPTER 6	61
METHODOLOGY	61
6.1 Introduction	61
6.2 Test for stationarity	61
6.2.1 Augmented Dickey-Fuller test.....	61
6.2.2 Phillips-Perron (PP) test.....	62
6.2.3 Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test.....	63
6.3 The VaR Model: Combining the GARCH type model and heavy-tailed distribution	64
6.3.1 Step-by-step method	66
6.4 Model selection criteria.....	66
6.4.1 The Akaike Information Criterion (AIC).....	67
6.4.2 The Bayesian Information Criterion (BIC).....	67
6.5 Model Diagnostics.....	68
6.5.1 Test for Serial Correlation and ARCH effect	68
6.5.2 Test for Leverage effect	70
6.5.3 Test for Independent and Identically Distribution (IID)	71
6.5.4 Test for Normality and Goodness-of-fit	74
6.6 Backtesting procedure	76
6.6.1 Kupiec Likelihood Ratio Test.....	76
6.7 Conclusion.....	77
CHAPTER 7	78
EMPIRICAL RESULTS.....	78
7.1 Introduction	78
7.2 Data Description.....	78
7.3 GARCH Parameter Estimation and Model Selection	83
7.4 Combining APARCH (1,1) with heavy tailed distribution.....	89
7.4.1 APARCH (1,1) model with skew Student t Distribution (sstd).....	89

7.4.2	APARCH (1,1)-GPD model	90
7.4.3	APARCH (1,1)- GEVD model	95
7.4.4	APARCH (1,1)-Stable Distribution model	97
7.4.5	APARCH (1,1)-PIVD model	98
7.5	Estimating Value at Risk (VaR).....	99
7.6	In-Sample Backtesting	100
7.7	Out of Sample Backtesting.....	101
7.8	Conclusion.....	102
CHAPTER 8		103
CONCLUSION.....		103
REFERENCES		105

LIST OF TABLES

Table 5: 1: Family of distribution for the Pearson System.....	58
Table 7. 1: Test for Stationarity for the ALSI returns.....	79
Table 7. 2 Descriptive statistics of the daily JSE ALSI returns.....	80
Table 7. 3: Test for normality of the ALSI returns	81
Table 7. 4: Test for serial correlation in the ALSI returns.....	82
Table 7. 5: Test for ARCH effect in the ALSI returns.....	82
Table 7. 6: ML Parameter estimates of GARCH (1,1) model and goodness-of- fit statistics .83	
Table 7. 7: Sign and size test	84
Table 7. 8: ML Parameter estimates of asymmetric GARCH models.....	85
Table 7. 9: Descriptive Statistics of standardized residuals of the APARCH (1,1) model.....	85
Table 7. 10: Test for normality of the standardized residuals.....	86
Table 7. 11: Test for serial correlation on the standardized residuals	87
Table 7. 12: Test for ARCH effect in the standardized residuals	88
Table 7. 13 : Test for independent and identically distributed of the standardized residuals..	88
Table 7. 14: Parameter estimates of APARCH (1,1) with sstd governing the innovation	89
Table 7. 15: ML Parameter estimates of GPD with standard errors in bracket.....	93
Table 7. 16: ML Parameter estimates of GEVD with standard error	95
Table 7. 17: ML Parameter estimates of stable distribution and goodness-of-fit statistics	97
Table 7. 18: ML Parameter estimates of PIVD and goodness-of-fit statistics.....	99
Table 7. 19: Value-at-Risk estimates at short and long position	99
Table 7. 20: In-sample backtesting: p -values of Kupiec likelihood ratio test	100
Table 7. 21: Out of sample backtesting: p -values of Kupiec likelihood ratio test.....	101

LIST OF FIGURES

Figure 7. 1: Time series plots.....	78
Figure 7. 2: Q-Q plot of JSE ALSI returns	80
Figure 7. 3: ACF plots of ALSI returns	81
Figure 7. 4: Empirical density plot and Q-Q plot of standardized residuals.....	86
Figure 7. 5: ACF plots of the Standardized residuals	87
Figure 7. 6: Empirical Density plot and Q-Q plot of APARCH with sstd.....	90
Figure 7. 7: Mean residual life plot of standardized residuals	91
Figure 7. 8: Parameter stability plot for positive standardized residuals.....	91
Figure 7. 9: Parameter stability plot for negative standardized residuals	92
Figure 7. 10 : Pareto quantile plot of standardized residuals	92
Figure 7. 11: GPD Diagnostic plot of positive standardized residuals	94
Figure 7. 12: GPD Diagnostic plot of negative standardized residuals	94
Figure 7. 13: GEVD Diagnostic plot of positive standardized residuals	96
Figure 7. 14: GEVD Diagnostic plot of positive standardized residuals	97
Figure 7. 15: Variance stabilized P-P plot of the standardized residuals.....	98

LIST OF APPENDICES

Appendix A: CDF of PIVD in hyper geometric function.....	120
Appendix B: Model estimate results.....	124
Appendix C: R Codes	125

ABBREVIATION

AD test	Anderson-Darling test
ADF test	Augmented Dickey-Fuller test
AIC	Akaike Information Criterion
ALSI	All Share Price Index
APARCH	Asymmetric Power ARCH
AR model	Autoregressive model
ARCH	Autoregressive Conditional heteroskedasticity
BDS test	Brock, Dechert and Scheinkman test
BIC	Bayesian Information Criterion
BM	Block maxima
cdf	Cumulative distribution function
DW	Dublin Watson test
EGARCH	Exponential GARCH Model
ETV	Extreme value theory
FTSE	Financial time stock exchange
GARCH	Generalized autoregressive conditional heteroskedasticity
GEVD	Generalized extreme value distribution
GPD	Generalized Pareto distribution
IGARCH	Integrated GARCH Model
IID	Independent and identically distribution
JB test	Jarque-Berea test
JSE	Johannesburg Stock Exchange
KPSS test	Kwiatkowski, Phillips, Schmidt and Shin test
MA model	Moving average model
MLE	Maximum likelihood estimation
pdf	Probability density function
PIVD	Pearson type IV distribution
POT	Peaks over threshold
pp plot	Probability plot
PP test	Phillips-Perron test
Q-Q plot	Quantile-Quantile plot

sstd	skew student t distribution
TGARCH	Threshold GARCH Model
VaR	Value-at-risk

CHAPTER 1

INTRODUCTION

1.1 Introduction

In this chapter, the background of this study is discussed, the literature review, the stylized facts of returns, problem statement, aim and objectives and significance of the study are also discussed.

1.2 Background

South Africa today is one of the most diverse and promising emerging markets globally. It is the sixth most outstanding in the emerging economies category, with vast opportunities within her border. It is a gateway to the rest of the African continent (a market of more than one billion people) therefore; it is a key investment location. South Africa is strategically located at the tip of the continent, with vibrant manufacturing and financial sectors. It is the economic powerhouse of Africa and forms part of BRICS group of countries which includes Brazil, Russia, India and China. South African stock market – Johannesburg Stock Exchange (JSE) – is Africa’s largest stock exchange with more than 400 listed firms and offering a wide range of products. The stock market’s estimate is about double the turnout of the country. A white paper by the South African Reserve Bank revealed that the South African Stock market is significantly a robust one and able to make the list of the first twenty largest stock markets in the world consistently Hassan (2013). As of the year 2013, it boasts of an excess of \$US0. 9 trillion market capitalization value. This market value is unavoidably significant among world stock indexes, including contributing about 8% MSCI Emerging Markets index, the fifth most robust country weight, making it respond significantly to the global economic meltdown surrounding emerging markets. The value of this market economically enhances macro-economic sensitivity to price fluctuations with a possible eventual collapse (Hassan, 2013).

This study employs the FTSE/JSE All Share Price Index (ALSI) to reflect the South African stock market. It has about 164 listed companies and it is about 100% of the South African market capitalization value. ALSI as an equity index portrays the operational activities of a typical ordinary share in the South African market. The ALSI positioned against world criteria and standards which are majorly a pointer to the market situation. This ALSI also evaluates the operationalization of the entire market (Makhwiting, 2014). The major volume of all securities listed on the JSE is an integral function of the market index because the share prices flow of the listed companies is what makes the market.

However, this market is highly volatile and unpredictable, making it a very risky market. This may be as a result of an authentic/stylized fact that well describes stock returns in both emerging and developed stock markets (Stavroyiannis et al., 2012), known to exhibit stochastic processes with volatility clustering and heavy-tailed (Mandelbrot, 1963; Fama, 1965). There are many types of empirical models which have been used to describe the stylized facts in stock returns. These include, ARCH (Engle, 1982), GARCH (Bollerslev, 1986), IGARCH (Engle and Bollerslev, 1986), EGARCH (Nelson, 1991), TARARCH (Glosten et al., 1993a), APARCH (Ding et al., 1993), FIGARCH (Baillie et al., 1996), FIGARCHC (Chung, 1999), FIEGARCH (Bollerslev and Mikkelsen, 1996), FIAPARCH (Tse, 1998), FIAPARCHC (Chung, 1999), and HYGARCH (Davidson, 2004). Additionally, there are also some probability density function (pdf) such as normal distribution which is symmetric and does not exhibit heavy-tailed phenomenon (Engle, 1982); Student-t distribution (Bollerslev, 1987) with symmetric but heavy-tailed behaviour. Furthermore, the list also includes the generalized error distribution (GED) (Subbotin, 1923), with heavy-tailed behaviour, but accommodates flexibility more than Student-t distribution. Developed models which are a function of symmetric density distribution will perform less in the error term because the pdf of asset returns is non-symmetric, (Giot and Laurent, 2003a). Lambert and Laurent (2000) also applied the skewed Student-t distribution introduced by Fernandez and Steel (1998) which is non-symmetric and exhibit heavy-tailed phenomenon.

Risk is an elementary recipe for profit making related activities in a market environment when properly managed (Stambaugh, 1996). With the continuous increase witnessed in trading, both emerging and developed markets have commissioned financial and economic experts, searching a well-structured approach in measuring techniques associated with risk used in

calculating expected loss a financial body may experience. The emergence of global economic and financial issues from the 1980s to 2000s precipitated the need to develop a sustainable model that would be able to predict risk factors in the investment world. It would be recollected that in 1987, there was a global stock market fall. Mexico stock crisis followed in 1995: between 1997 and 1998 was Russia's financial turmoil, the dotcom bubble as well as the most felt 'the Lehman Brothers'. Variability in the financial world is on the increase, exposing financial institutions to incur losses due to unconsidered, unpredicted and unforeseen market fluctuations. As a way of providing solutions to some of these challenges, the Basel I and II agreements which were more of Committee on Banking Supervision were introduced in 1996 and executed in 2007. In this scheme, the financial institutions were motivated to lift assets of high risk value from their balance sheets. Investment trading regulation which captures changes in price was also obtained after Basel I allowing insured assets as government securities which attracts zero risk. With this, there was explosive credit default swaps (CDS) market because banks took advantage of the policy. Basel II on the other hand allows banks to use a developed institutionalised risk management model which is extremely backtested and stress tested.

The tool Value at risk (VaR) is employed by financial institutions as a control over their transactions whilst it allows regulatory authorities to standardise boundaries for the future (Bhattacharyya et al., 2008). Ever since the Basel Committee adopted VaR, it has become a valuable and useful measure for risk in financial institutions (Chen and Liang, 2008). Value at risk (VaR) can be expressed as a single value of the amount at which the risk situation of an institution could diminish as a result of market flow or fluctuations in a given period (Bhattacharyya et al., 2008). It is described as the most regretful loss with a bounded target within a given level of confidence (Chen and Liang, 2008; Jorion, 2007). VaR measures the value of the maximum loss certain investments can experience in a specified time of transaction at a given level of confidence (Jorion, 2007). In statistical terms, it is quantile of the returns distribution at a given confidence interval or probability level. VaR was elaborately applied to conventional financial markets (Duffie and Pan, 1997; Dowd, 1998; Holton, 2003; Jorion, 2007). VaR at the long position is associated with the risk that comes from the price drop. It is the left side of the distribution of the returns corresponding to negative returns while VaR at the short position is associated with the risk that comes from the price increase. It is the right side of the distribution of the returns corresponding to positive returns. Many of the models concentrated on VaR computation taking into consideration the left tail of the pdf. In recent

times, more models centred on formulating VaR function which considers short and long positions. Measurement of VaR is employed using variance-covariance method, historical simulation method, dynamic risk models and filtered historical simulation method (Bhattacharyya et al., 2008). Each of these methods are further explained below.

Variance-covariance method: This assumes that returns are normally distributed. However, the assumption of normal return distribution results in undervalue tails, which in turn undermine excess kurtosis and skewness in the returns (Soltane et al., 2012).

Historical simulation method: It assumes that the volatility of the returns is constant over time, therefore estimating the future volatility base on the past volatility (Bhattacharyya et al., 2008). To assume that there is a perfect volatility is indefensible due to previous market volatility clustering records. Research discoveries have shown that in financial markets seasons of undulation between high volatility and relatively stable one is periodically experienced (Soltane et al., 2012). A motionless VaR may be unnecessary in a time of stability, but this will not be able to give a correct estimate during extreme volatility.

Dynamic risk models: These models consider both the past and present volatility. Many risk models have been developed which include the simple moving averages, exponentially weighted moving average, Risk Metrics and GARCH (Bhattacharyya et al., 2008).

Filtered historical simulation method: It is a simulation developed by Barone-Adesi et al. (1998) based on generalised historical values. This method integrates GARCH models and historical simulation method so as to overcome the deficiencies of conventional historical simulation. Filtered historical simulation is simple in capturing conditional volatility, volatility clustering, and asymmetric effect. This has strong predictive power for unstable market situations. McNeil and Frey (2000) also combined GARCH model with extreme value theory, whilst Bhattacharyya et al. (2008) proposed a combination of the Pearson's Type IV distribution and the GARCH (1, 1) framework.

In order to investigate the sufficiency and reliability of VaR estimates, the Basle Committee specifically suggests that backtesting should be implemented (Baharul-Ulum et al., 2012).

Backtesting is a comparison between losses as assumed by the VaR model and the real value within a time frame of examination.

1.3 Literature Review

Over the years, in depth studies have been conducted in the area of developing and analysing VaR models. These have resulted in generating many methodologies with their associated pros and cons. More complex methodologies are being developed so as to enhance the predictability and correctness of VaR models. Such complex developments include different aspects of Autoregressive Conditional Heteroskedastic (ARCH) models, Extreme Value Theory (EVT), stable distributions, and Pearson IV distribution (PIVD). These approaches have been employed to investigate different categories of assets in various parts of the world. Outcomes prove that there is no single best model that effectively describes all assets. Therefore, it is better to test various models on a specific asset and from the outcomes pick the one that best handles the particular market financial characteristic. This is best done by employing latest financial data.

Engle (1982), proposed a volatility process with time varying conditional variance known as Autoregressive Conditional heteroskedasticity (ARCH) model. It is the first model that stipulates a systematic framework for volatility modelling. Previous studies by Engle (1982) was further modified by Bollerslev (1986) who generalised the ARCH model to obtain a new model known as a GARCH model in which time-varying volatility depends on previous volatility and previous innovations. GARCH model has been widely used to study volatility since its inception. The volatility in Government-Sponsored Enterprise (GSE) indices was modelled with specifications like EGARCH, TGARCH, GARCH and RW (Magnus and Fosu, 2006). Rafique and Kashif (2011) examined Karachi Stock Exchange 100 Index (*KSE-100 Index*) considering excess kurtosis, heavy tails and volatility clustering using the ARCH, GARCH and EGARCH processes. The GARCH model has been proved to be able to effectively capture the tenacity in volatility and EGARCH successfully overcame the leverage effect specification in *KSE-100 index*. Granger and Andersen (1978), revealed the theoretical features of a few and up-to-date GARCH specifications that possess the component of leverage effect. Parameters were compared to assure the form/state of positivity, stationarity, and finite fourth order moment. The results disclosed that EGARCH specification is the most flexible

and a detailed description of TGARCH is accompanied with boundaries and deficiencies when it has finite kurtosis. The GARCH model and its subsequent variants were used to model and elucidate the instability in the financial market risk from daily reflections from Israel (TASE-100 index) and Egypt (CMA General Index). The Egyptian CMA index is said to be the most volatile series due to the economy and price fluctuation during the time phase under consideration (Floros, 2008). Many studies have been carried out using the GARCH model in the area of VaR studies. Investigations involving volatility using the APARCH model are still few. One such research was performed by Giot and Laurent (2003a) who carried out a study on VaR to assess both long and short trading positions. He applied Skew Student t -APARCH, RiskMetrics, normal-APARCH and Student t -APARCH models for French CAC40, US NASDAQ, Swiss SMI, Japanese Nikkei, and German DAX stock indexes and found that for VaR forecast, either long or short trading positions, the use of an APARCH model with a skewed Student t distribution is best. Using three VaR models; Risk matrices, Student t -APARCH and normal-APARCH, Huang and Lin (2004) tested their forecasting performance based on Taiwan Stock index futures. From the study, the normal-APARCH VaR model appeared to be best at the lower confidence point, Student t -APARCH appeared to display highest accuracy at the high confidence point when compared to Risk matrices as well as a normal-APARCH model. Angelidis and Degiannakis (2008) predicted one-day ahead VaR for Athens Stock market. The symmetric and asymmetric ARCH models were assessed to determine their predicting performance. The GARCH, the symmetric ARCH VaR model was tested with the three innovations; normal, Student t and skewed Student t distributions and the same procedure was conducted for the APARCH, the asymmetric ARCH VaR model. It was concluded that, the best performing model for this index was the skew Student t -APARCH VaR model. Degiannakis et al. (2012) evaluated VaR of specific markets before and after- year 2008 global financial crises. They applied three alternative VaR models; Risk metrics (EMWA), normal-GARCH and skew Student t -APARCH. The tests were conducted on developed and emerging markets. The Greece and Turkey markets represent the emerging while USA, UK and Germany represent developed markets. The normal-GARCH VaR model had satisfactory performance before the 2008 crisis but afterwards only the skew student t provided satisfactory forecasts.

A new aspect of review on VaR modelling centres on Extreme Value Theory (EVT) which is often employed in risk modelling of extreme occurrence. Models that give a good description

of distribution patterns of rare events can be presented using the theory of extreme value. One outstanding characteristic of an extreme value is aimed at quantifying the stochastic conduct of procedure with both extremely large and small stages. Extreme Value Theory (EVT) can be described as a statistical and theoretical distribution which models sample maximum and sample minimum extremes behaviour. The characteristics of this statistical arrangement are sometimes measured using the precise distribution function or the limiting distribution function, the asymptotic distribution function when the size of observation increase toward infinity (Vicente, 2012). The account of the EVT application to describe extreme event distribution is traced to early work by Fisher and Tippett (1928) followed by Gnedenko (1943). These researchers laid a solid foundation in the theories and classic boundary laws guiding distributions of extremes. EVT has been employed in various fields of study and proved to be a potent tool in extreme event distributions. Torrielli et al. (2013), used EVT in meteorology to analyse extreme wind prediction. In hydrology, Fernandes et al. (2010) employed EVT in floods and draught prediction. Rocco (2011) with Jonathan and Ewans (2013) likewise used it in the financial crisis and wave height predictions respectively. Kuester et al. (2005), studied VaR where they employed out of sample performance of recognised models compared with other developed alternative models to forecast VaR. They applied those models on the NASDAX composite index. The outcomes revealed that a hybrid approach which mixes heavy-tailed GARCH with EVT based method outperform other VaR methods which undervalue risk. Singh et al. (2001), estimated VaR by employing EVT. He modelled VaR using dynamic EVT with GARCH (1,1) model, GARCH (1,1), and Risk matrices, for ASX – all ordinaries (Australian) index and S&P 500 (USA) index. GARCH (1,1) and Risk matrices could not match the performance of dynamic EVT with GARCH (1,1) model. Sigauke et al. (2014) used GDP to model a conditional heteroskedasticity in the JSE ALSI index. The distribution ALSI returns as well as an approximation of extreme tail quantiles were modelled with a comparison between ARMA-GARCH-GDP models and ARMA GARCH models. ARMA-GARCH-GDP models generated more precise predictions of extreme returns than the other. Ozun et al. (2010) estimated VaR for Istanbul Stock Exchange by comparing eight filtered EVT models, GARCH (1,1) with normal, student t and skew student t innovation and FIGARCH model. He found that the filtered EVT models outperform the other models.

Mean-variance portfolio theory and pricing of financial derivatives and some other financial applications depend on the type of distribution of financial returns involved. Due to outcomes

of Mandelbrot (1963) and Fama (1965), normality of returns was rejected because heavy tail is a prominent feature when talking about financial returns. As a result of this development, many other authors developed several distributions which consider excess kurtosis (Press, 1967; Praetz, 1972; Blattberg and Gonedes, 1974; Peiró, 1994). This has led to investors preferring positive first and third moments while there is dislike for the second and the fourth moment, making the skewness significant for financial returns modelling (Kon, 1984; Hansen, 1994; Young and Graff, 1995; Peiró, 1999; Rachev and Mittnik, 2000; Aparicio and Estrada, 2001). Therefore, due to the empirical evidence, stable distribution was introduced as a substitute model (Mandelbrot, 1963; Fama, 1965). The theory of stable distribution was studied and introduced by Paul Levy, a French mathematician. He examined a stable distribution now referred to as the Levy distribution (Yang, 2012). While some members in the stable family lack distribution function expressible elementarily, developing the relevant theories is very difficult leading to more difficult computational applications. All the same, (Mandelbrot and Taylor, 1967; Fama, 1965; Samuelson, 1967) equally concluded that theoretical and empirical results affirms non-normal stable distributions in specific financial models which include financial asset returns, risk management and portfolio management. There are other heavy-tailed distributions for financial models, for example, Student t , hyperbolic, normal inverse Gaussian, or truncated stable. At least modelling financial variables using stable distributions is good because they are supported by the generalized Central Limit Theorem (CLT) (Borak et al., 2005). This theory states “that stable laws are the only possible limit distributions for properly normalized and centred sums of independent, identically distributed random variables”. Subsequently, stable distributions have potentials to accommodate the heavy-tails and asymmetry which often give a very good match to empirical data. They are also useful models for data sets in the class of extreme events including market crashes or usual catastrophes making it a very useful tool for financial or insurance analyst (Borak et al., 2005).

The Pearson type-IV distribution (PIVD) which is heavy-tailed with skewness was developed as a stationary distribution of Pearson systems (Pearson, 1895; Sato, 2014). In order to design a family of probability distribution that will capture the different range of skewness and kurtosis, Karl Pearson developed a family of distributions which has a small number of parameters to capture any given pair of skewness and kurtosis (Pearson, 1895). The scope of PIVD in the skewness–kurtosis axis is quite large. Therefore, in practice this has given the PIVD the capacity to fit financial data which most times kurtosis is in excess while skewness

is moderate. Premaratne and Bera (2001) innovated this distribution in GARCH models for capturing heavy-tail and asymmetry in financial data. A few years later, Yan (2005) employs PIVD so as to fit time-varying parameters by following autoregressive conditional density (ARCD) models. In the case of the innovated GARCH models with PIVD, (Premaratne and Bera, 2001; Yan, 2005) employed parameters of both constants and time variables in modeling for skewness and kurtosis. In the research outcome of Premaratne and Bera (2001), the dynamic parameter model does not rank better to a static parameter model. While with Yan (2005) it was discovered that time-varying shape parameters model behaved better compared to constant parameters models. In their research, Bhattacharyya et al. (2008) developed and applied PIVD in the estimation of conditional VaR using about fourteen international stock market indices. Non-normality of returns were modelled to give a reason for the dynamic volatility using a combination of GARCH model and PIVD. As time goes on, many researches that applied the Pearson method of distributions came out with reports that PIVD enhanced the log-likelihood, and accurate VaR results at high confidence levels can be obtained, which indicated that it is better in performance. Stavroyiannis et al. (2012) and Stavroyiannis (2013) examined a TGARCH model of Glosten et al. (1993b), considering the notable daily Standard and Poor's (S&P500) index. Stavroyiannis (2013) -produced results proving that at a high confidence level, the PIVD outshines the skewed t-Student. Stavroyiannis (2016), tested the efficiency of the APARCH model for a residual after standardized modelling. It was on record that the APARCH model alongside the standardized Pearson type IV distribution generated a great accuracy.

As a result of the political instability in many African countries which has a huge effect on their market structure, stock market research has been limited. Brooks et al. (1997) worked on examining the effect of political change in the South African stock market. They discovered that ARCH/GARCH model is adequate for South African market and more complex GARCH model is needed for the post 1990 period. Makhwiting et al. (2012) employed the ARMA (0,1)-GARCH (1,1) model in modelling the volatility of this market and most recently, Huang et al. (2014) assessed the performance of different heavy-tailed distributions. However, none considered a model that can capture volatility clustering, asymmetric effect and heavy-tailed behaviour. Thus, this article uses the APARCH (1,1) model to capture volatility clustering and asymmetric effect and uses the heavy-tailed distribution to capture the heavy-tailedness.

We are not aware of any literature relating to an application of APARCH (1,1)-heavy tailed model to the South Africa market. Since the performance of VaR model depends on the quality of the distributional assumptions. Therefore, the main contribution of this study is to compare the predictive power of APARCH model combined with different heavy-tailed distributions such as: Pearson Type-IV distribution, stable distribution, GPD and GEVD and also compare it with an APARCH model with skewed Student t innovation that has been proven in the literature to be a good model.

1.4 Stylized Facts of Asset Returns

Many asset returns share some common statistical properties which include

- Absence of autocorrelations: Apart from the cases of intraday activities, the significance of autocorrelation of asset returns is almost negligible.
- Heavy-tails: the tails of the unconditional distribution of returns tends to be asymptotically equivalent to a Pareto law which means exhibiting power-law behaviour and the lower values of the tail index express heavy-tailed. However, making the tail shape determination strenuous.
- Gain/loss asymmetry: Stock prices tend to exhibit large downward fall, but do not produce an equal upward rise.
- Volatility clustering: It is observed from studies that events with high volatility cluster over time. This is linked to the fact that the change in volatility is usually positively autocorrelated within a few days.
- Conditional heavy tails: Upon regularising returns as a result of volatility clustering, heavy tails is still experienced in the residuals time series.
- Leverage effect: most times, changes in volatility of an asset exhibit negative correlation with the changes in returns of the asset.

It is important to develop a model that will account for most of these stylized facts in any asset returns, however it is rare to get a model capturing all the stylized properties of asset returns.

1.5 Problem Statement

An adequate Value-at-Risk (VaR) model for predicting stock price enhances investment. In order to attract foreign business investors in emerging markets, it is vital to have an overview of the market buoyancy so as to predict its risk rate and frailty nature (Bucevska and Bucevska, 2012). It has been found in literature that an adequate and robust VaR model must be able to capture current volatility clustering. Brooks and Persaud (2003) stated the significance of asymmetry in the VaR framework, and recommended it be integrated into volatility specification models. However, Ortiz and Arjona (2001) studied various stock markets using GARCH models and observed that there is no single GARCH model capable of efficiently describing the volatility of stock returns in the listed markets. Ortiz and Arjona (2001), suggested that using different GARCH models for each market that is 'the best models seem adequate' is most appropriate. In addition, Mittnik and Paolella (2000) recommended extended structures useful in enhancing a VaR forecast in terms of the distribution and the volatility operation. It has also been extensively proven in the literature that the performance of VaR model depends on the quality of the distributional assumptions. In an emerging market, a model that assumes normal distribution is a weak model (Živković and Aktan, 2009). It is not a good thing to have an unreliable risk estimate. It is a known fact that there is no singular model that is most efficient in estimating VaR for every market condition. Thus, the purpose of this study is to identify an appropriate VaR model which captures volatility via an appropriate GARCH framework combined with a heavy-tailed distribution for South African market.

1.6 Aim and Objectives

The main aim of embarking on the current research was to obtain the best possible robust VaR model for the South African market, by assessing the applicability of APARCH model combined with heavy-tailed distribution on the South African market similar to the proposed filtered historical simulation by Barone-Adesi et al. (1998).

This was achieved by:

- Fitting an APARCH model to capture volatility clustering in the ALSI returns.
- Extracting the standardized residuals of the APARCH model.

- Fitting the heavy-tailed distributions, namely Generalized Pareto Distribution (GPD), Generalized Extreme Value Distribution (GEVD), Pearson type IV distribution (PIVD) and stable distribution to the standardized residuals.
- Estimating VaR for the ALSI returns.
- Backtesting the VaR models to check for model adequacy and selecting the robust model.

1.7 Significance of the study

This study is important in ascertaining the applicability of APARCH-GPD model, APARCH-GEVD, APARCH-PIVD and APARCH-stable model to South African market's risk. It also adds candidate models for estimating VaR for the South African market. The final result of this study will help investors, risk managers, portfolio managers and academics who are interested in the risk associated with the South African market.

1.8 Research Layout

Chapter 2 of this research work reviews the theoretical framework of financial volatility model. The theoretical foundation of extreme value theory, stable distribution, and Pearson Type IV distribution are reviewed in chapter 3, chapter 4 and chapter 5 respectively. In chapter 6, the methods of application employed in this thesis is discussed. In chapter 7 we applied the APARCH model and the heavy-tailed distributions to the ALSI returns. Chapter 8 contains the conclusion part as a result of an engaging combination of APARCH model and heavy-tailed distribution as a research tool. It also outlines the possible extensions and improvements to the used models.

CHAPTER 2

VOLATILITY MODELS

2.1 Introduction

The aim of this chapter is to review the theoretical framework of a financial volatility model. Section 2.2 some basic concepts in financial time series is discussed. Section 2.3 - the basic structure of financial volatility model is discussed. In Section 2.4, the mean model is considered while Section 2.5 treated the volatility model by reviewing some GARCH model and their properties. Section 2.6 reviewed parameter estimation of GARCH models based on maximum likelihood estimation (MLE).

2.2 Basic Concept

In this section, some basic concepts in financial time series are defined in details below.

2.2.1 Stationarity

Stationarity is a critical part of time series and most of the analysis completed on financial time series involve stationarity. It is a suitable assumption that allows the description of statistical properties in a time series (Nason, 2008). The statistical properties of a stationary time series remain constant through time period of interest that is the periodic variations or seasonality does not occur. In other words, a time series is said to be stationary if the mean and variance remains constant over time and the covariance (correlation) between the series X_t and X_{t+k} depend on the time difference, (lag) only. For instance, if the joint distribution of the observation $X_t, X_{t+1}, \dots, X_{t+n}$ and $X_{t+k}, X_{t+k+1}, \dots, X_{t+k+n}$ is the same then, the time series is said to be strictly stationary. That is:

$$E(X_t) = \mu_t = \mu$$

$$\text{Var}(X_t) = \sigma_t^2 = \sigma^2$$

and

$$\gamma_k = \text{cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

where γ_k , $k = 0, 1, 2, \dots$ is the auto-covariance function at lag k , and the auto-covariance at lag $k = 0$, i.e. γ_0 is the variance of the time series.

2.1.2 Serial Correlation

There are numerous circumstances, especially in finance, where sequential observations of a stochastic time series will exhibit correlation. Correlation occurs when the behaviour of sequential observations influence one another in a dependent way. One noteworthy case happens in mean-reverting pairs trading. Mean-inversion appears as the correlation between successive observations in time series (Halls-Moore, 2015b).

It is important to be able to recognize the structure of these correlations, as they will permit us to extraordinarily enhance our forecasts. Moreover, it will enhance the authenticity of any simulated time series in light of the model. This is greatly valuable for enhancing the adequacy of risk management techniques. If the sequential observations of a time series possess correlation, then the series exhibit serial correlation (or autocorrelation). Mathematically, the serial correlation or autocorrelation of lag k , ρ_k of a second order stationary time series is given

by: $\rho_k = \frac{\gamma_k}{\gamma_0}$

2.2.3 Conditional Heteroskedasticity

Most assets price possesses conditional heteroskedastic. The fundamental need for examining conditional heteroskedasticity is volatility of asset returns. In the event that we have a series of observations, we say that the series is heteroskedastic if there are sure subsets, of observations inside a set of the series that have a variance different from the other observations. For example, in a non-stationary time series that shows seasonality or trending effects, whereby variance increases as the seasonality or the trend increase. Be that as it may, in finance time series there are numerous reasons why an increment in variance is correlated to a further increment in variance (Halls-Moore, 2015a).

Accordingly, when a time series shows autoregressive conditional heteroskedasticity, it can be said that it has the ARCH effect or exhibit volatility clustering.

2.3 Basic Structure of Financial Volatility Model

A common idea of volatility modelling with GARCH is the serially uncorrelated nature of log return or manifesting a negligibly small degree of serial correlation (Tsay, 2013). As a way of putting volatility model in the right perspective, it will enrich knowledge by identifying conditional mean and variance of the log return in this review.

Let r_t represent- the daily log return of the financial time series at time t , given by;

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (2.1)$$

where P_t is the stock price at time t .

Then, Tsay (2013) defined the conditional mean of the r_t given F_{t-1} as;

$$\mu_t = E(r_t|F_{t-1})$$

and conditional variance of the r_t given F_{t-1} as;

$$\sigma_t^2 = \text{Var}(r_t|F_{t-1}) = E([r_t - \mu_t]^2|F_{t-1}),$$

where, F_{t-1} represents the past information at pre-set time, $t - 1$, which comprises all linear functions of the past returns.

If $r_t \sim \text{ARMA}(p, q)$, then

$$r_t = \mu_t + Z_t \quad (2.2)$$

where

$$\mu_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{j=1}^q \theta_j Z_{t-j} . \quad (2.3)$$

This is referred to as the mean equation for r_t and Z_t is the return innovation/shock at time t given as:

$$Z_t = \sigma_t \varepsilon_t \quad (2.4)$$

where ε_t is sequence of identical and independent random variable with mean zero and variance one. While;

$$\sigma_t^2 = \text{Var}(r_t|F_{t-1}) = \text{Var}(Z_t|F_{t-1}). \quad (2.5)$$

This is referred to as the volatility equation for r_t . The manner in which the conditional variance σ_t^2 evolves over time distinguishes one volatility model from one another.

2.4 The Mean Model

The model for μ_t in Equation (2.3) is referred to as the mean model for r_t . The removal of possible linear dependence in the data is achieved by specifying the mean model; this is the same as to remove sample mean from the data. This is practised when the sample mean is substantively not zero. Another reason is that the residuals of the ARMA model can be relevantly useful in identifying the presence of ARCH effects. The mean models of choice usually give residual with white noise and at the same time having the ARCH effect.

ARMA Models: The ARMA models came into existence through Box et al. (2015). They are generally employed in the analysis involving time series as a result of its ability to estimate many stationary processes. In finance return series, ARMA models are not usually used but the idea is of high relevance on modelling volatility (Tsay, 2013). In the real sense, ARMA combines and summarises the knowledge of AR and MA models into a simple form which keeps the number of parameters involved relatively small, resulting in closeness in parameter description.

Autoregressive (AR) Model is a model that exclusively estimates future values based on the previous values of the time series. An AR model for $p \geq 1$ can be defined using

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + Z_t$$

where Z_t is a sequence of independent and identically distributed random variable with mean zero and variance σ^2 known as white noise.

Taking a vivid look at the model illustrated above, it is not a task to demonstrate that X_t is a linear regression function through past values. Since its implementation is easy, this method remains the most employed model of time series.

Moving Average (MA) Model: is a model where estimates of future values are exclusively estimated based on past shocks. An MA process with a for $q \geq 1$ is likewise defined using

$$r_t = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$$

An MA model suggests that a time series is a moving average of a white noise process while r_t and r_{t-h} remain uncorrelated. AR(p) may be identified as an infinite-order of a MA process and vice versa.

Autoregressive Moving Average (ARMA) Model: is a model that combines AR and MA model ARMA model is defined as:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$$

ARMA models are commonly used in the time series analysis because of their flexibility in estimating many stationary processes. But, they do not have advantages in nonlinear phenomena (Fan and Yao, 2003).

2.5 The Volatility Models

Many new models with distinctive features have been proposed since the invention of Bollerslev's GARCH model. The existing models can be categorized into symmetric and asymmetric models. In the former model, the conditional variance does not depend on the sign of the underlying assets r_t , but only on the magnitude. This property is always in line with empirical results where leverage effect is frequently present. In other words, volatility increases for negative return shocks more than positive return shocks of the same magnitude, i.e. bad

news produces volatility more than good news reduces the volatility. These features are more or less captured in asymmetric models.

2.5.1 Symmetric GARCH Models

2.5.1.1 ARCH Model

The meaning of Autoregressive Conditional heteroskedasticity (ARCH process) can be summarized as follow (Karlsson, 2002):

- Autoregressive property on a fundamental level implies that old occasions leave waves behind a certain time after the real time of the action. The process relies on its past.
- Conditional heteroskedasticity implies the variance (conditional on the available data) shifts and relies on old estimations of the process. Therefore, one can say that the process has a transient memory and that the process' behaviour is affected by this memory.

The two underpinning ideas of the ARCH model are: the shock Z_t of an asset return is serially uncorrelated but dependent and the dependency of Z_t can be explained with a simple quadratic function of its lagged values (Tsay, 2013). In ARCH model, the volatility of the process at time t is given as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2 \quad (2.6)$$

under the condition that $\alpha_0 > 0$ and $\alpha_i \geq 0$ for any value of i .

The setbacks of ARCH model are identified by Karlsson (2002):

- It requires many parameters to correctly describe the volatility process
- It cannot model asymmetric effects of both positive and negative shocks because it only uses the squared shocks as variable to model conditional variance.

- Lastly, it imposes restrictive intervals for the parameters only if it is finite fourth moments and does over envisage volatility since it reacts slowly to large isolated shocks.

2.5.1.2 GARCH Model

However, the empirical study indicates that high ARCH order has to be picked in order to capture the dynamic nature of the conditional variance. High ARCH order means that many parameters have to be estimated and the calculations become more involving.

After four years of introducing Engel's ARCH process, Bollerslev (1986) purported a natural solution to the problem of high ARCH orders using generalised ARCH (GARCH) model and the use of infinite ARCH specification. This specification allows to significantly reduce estimated parameters in number.

The ARCH model was expanded to the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) by Bollerslev (1986) which possesses the same main properties with ARCH but only differs in the sense that it requires less parameters to correctly model the volatility process.

The volatility equation for the GARCH (p,q) model is given as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (2.7)$$

under the condition that $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ and $\alpha + \beta < 1$

The parameters $\alpha_i, i = 1, 2, \dots, p$ measures volatility reaction to developments in the market and $\beta_j, j = 1, 2, \dots, q$ capture the degree of shocks persistence brought about by extreme values of conditional variance. It shows the conditional variance as a linear function of previous data permitting the returns' conditional heteroskedasticity (Curto et al., 2009).

The GARCH (1,1) model is adequate to catch all the volatility clustering in the data (Brooks, 2008). In most observational applications (French et al., 1987; Pagan and Schwert, 1990;

Franses and Van Dijk, 1996; Gokcan, 2000) the fundamental GARCH (1,1) model fits the varying conditional variance of most financial time series sensibly well. The primary formulation of the GARCH (1,1) model depicts the ARCH component and the second one shows dynamic average. Specifically, the GARCH (1,1) model is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2.8)$$

The properties of GARCH models can easily be observed from basic model Equation (2.8). The GARCH (1,1) shows that large Z_{t-1}^2 or σ_{t-1}^2 will lead to a large σ_t^2 . GARCH models can model volatility clustering but they cannot model the asymmetric effect. A GARCH model is akin to ARCH model in the modelling of volatility if not for the addition of lagged conditional variances, σ_{t-1}^2 coupled with lagged squared returns, Z_{t-1}^2 . This helps to decrease the quantity of parameters that need to be estimated. Conditional variance when considering GARCH (1, 1) model can be modified as below;

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

It demonstrates that the GARCH (1,1) model is similar to an ARCH(∞) model with a certain structure for the parameters' value of the lagged returns Z_{t-l}^2 .

Besides, as the ARCH(1) could be written as an AR(1) model of the squared returns, the GARCH(1,1) model can similarly be written as an ARMA(1,1) model on the squared returns.

Suppose $\eta_t = Z_t^2 - \sigma_t^2$, then $\sigma_t^2 = Z_t^2 - \eta_t$.

By substituting $\sigma_{t-1}^2 = Z_{t-1}^2 - \eta_{t-1}$ into GARCH (1,1) in Equation 2.8, then;

$$Z_t^2 - \eta_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 [Z_{t-1}^2 - \eta_{t-1}]$$

$$Z_t^2 = \alpha_0 + (\alpha_1 + \beta_1) Z_{t-1}^2 - \beta_1 \eta_{t-1} + \eta_t$$

This shows an ARMA (1,1) for the squared returns of Z_t^2 . This connection to ARMA model suggests that the hypothesis behind the GARCH model may be firmly identified with that of ARMA model, which is entirely simple and broadly known (Karlsson, 2002). According to

Tsay (2013), by using the unconditional mean of ARMA model, the unconditional variance of Z_t can be written as

$$E(Z_t^2) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \quad (2.9)$$

The unconditional kurtosis of Z_t is given as

$$\frac{E(Z_t^4)}{[E(Z_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3 \quad (2.10)$$

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

Since the unconditional kurtosis is greater than 3, then GARCH (1,1) has a tail heavier than that of normal distribution.

2.5.1.3 The Integrated GARCH Model (IGARCH)

The name Integrated GARCH (IGARCH) was authored by Bollerslev et al. (1994). Here, "integrated" means that there may be a unit root issue which could prompt the existence of a non-stationary form of the time series X_t , -i.e. it exhibits an infinite variance.

Therefore, IGARCH models are unit-root GARCH models. Compared to ARMA models, a key element of IGARCH models is that the effect of the past square shocks on Z_t^2 is constant (Tsay, 2013). The volatility process of the IGARCH (1,1) can be written as:

$$\sigma_t^2 = \alpha_0 + (1 - \beta_1)Z_{t-1}^2 + \beta_1\sigma_{t-1}^2 \quad (2.11)$$

The significant contrast between GARCH and IGARCH is that the unconditional variance of Z_t and that of r_t are not characterized under the IGARCH (1,1) model. This appears to be difficult to justify for log return series. From the theoretical purpose of view, the IGARCH process may be brought about by occasional level shifts in volatility (Tsay, 2013).

2.5.2 Asymmetric GARCH Models

In various literature studies, it has been shown that the sign of the shock is consequential (Bekaert and Guojun, 2000; Bucevska, 2013). The conclusion of the comprehensive literature studies show that negative returns have a greater volatility when compared to positive returns of equivalent size. In simplified words, bad news increases volatility than good news (Angabini and Wasiuzzaman, 2011).

For asymmetry to be captured in return volatility, a new class of models was developed and called asymmetric ARCH models. These are the Exponential GARCH (EGARCH), Threshold GARCH (TGARCH) models, and Asymmetric Power ARCH (APARCH).

2.5.2.1 The Exponential GARCH Model (EGARCH)

Nelson (1991), suggested the first extension to GARCH called Exponential GARCH (EGARCH) to model the asymmetric effects of both positive and negative asset returns. The volatility process of the EGARCH (p, q) can be written as:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^p \alpha_i g(\varepsilon_{t-1}) + \sum_{j=1}^q \beta_j \ln \sigma_{t-1}^2 \quad (2.12)$$

where

$$g(\varepsilon_t) = \gamma_1 \varepsilon_t + \gamma_2 [|\varepsilon_t| - E(|\varepsilon_t|)]$$

So that $\gamma_1 \varepsilon_t$ is the sign effect and $\gamma_2 [|\varepsilon_t| - E(|\varepsilon_t|)]$ is the magnitude effect, γ_1 and γ_2 are real constants. The conditional variance of the EGARCH model is in logarithmic form that guarantees its non-negativity without the need to force extra non-negativity limitations.

The mean of $[g(\varepsilon_t)]$ is zero since the mean of ε_t and $[|\varepsilon_t| - E(|\varepsilon_t|)]$ are zero. As a result $g(\varepsilon_t)$ permit the conditional variance to response to asymmetric effects which can clearly be seen by rewriting $g(\varepsilon_t)$ in the following form:

$$g(\varepsilon_t) = \begin{cases} (\gamma_1 + \gamma_2)\varepsilon_t - \gamma_2 E(|\varepsilon_t|), & \varepsilon_t \geq 0 \\ (\gamma_1 - \gamma_2)\varepsilon_t - \gamma_2 E(|\varepsilon_t|), & \varepsilon_t < 0 \end{cases}$$

If $\varepsilon_t \geq 0$ the positive shocks have an impact of $(\gamma_1 + \gamma_2)$ on the conditional variance and if $\varepsilon_t < 0$ the negative shocks have an impact of $(\gamma_1 - \gamma_2)$ on the condition variance. To ascertain a non-negative conditional variance, no restrictions are required on the parameters which are contrary to GARCH models. As a result, it is able to model asymmetrical effect, volatility persistence and mean reversion. The pertinent importance of EGARCH over GARCH is that it permits positive and negative shocks to have different impact on the volatility (Karlsson, 2002).

2.5.2.2 The Threshold GARCH Model (TGARCH)

Glosten et al. (1993a) proposed another means of modelling the asymmetric effects of positive and negative asset returns called TGARCH or GJR-GARCH. The volatility equation is shown below:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \omega_i G_{t-i}) Z_{t-i}^2 + \sum_{j=1}^n \beta_j \sigma_{t-j}^2 \quad (2.13)$$

where G_{t-i} is an indicator variable for negative Z_{t-i} , given as

$$G_{t-i} = \begin{cases} 1, & Z_{t-i} < 0 \\ 0, & Z_{t-i} \geq 0 \end{cases}$$

Under the condition that $\alpha_0 > 0$, $\alpha_i \geq 0$, $\omega_i \geq 0$, and $\beta_j \geq 0$ in order to ensure nonnegative conditional variance. From this model, the impact of Z_{t-i}^2 on σ_t^2 is dependent on the sign of Z_{t-i} which permits the model to accommodate asymmetric effect. As a result, the positive shock have an impact of $\alpha_i Z_{t-i}^2$ on the conditional variance while the negative shock has an impact of $(\alpha_i + \omega_i) Z_{t-i}^2$ on the conditional variance.

The GJR-GARCH model are fundamentally the same to the EGARCH model which both have the capacity to capture the impact of both positive and negative shocks. As a result, the TGARCH and the EGARCH might both be considered for the same data hence, it is important to find a criterion for choosing between the two models (Karlsson, 2002).

2.5.2.3 Asymmetric Power ARCH (APARCH) MODEL

Ding et al. (1993), introduced the APRACH model as an extension of the GARCH model. The APARCH generalized both the ARCH and GARCH models. The structure of the volatility equation is given as (Tsay, 2013)

$$\sigma_t^\delta = \omega + \sum_{i=1}^p \alpha_i (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta \quad (2.14)$$

Under the condition that $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $0 \leq \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \leq 1$, where α_i and β_j are respectively the ARCH and GARCH coefficients, γ_i is the leverage coefficient such that when γ_i is positive it implies that the negative shocks has stronger impact on price volatility than the positive shocks, and δ is the positive real number which functions as the symmetric power transformation of σ_t . Considering the case, where $\delta = 1$ for $p = q = 1$, then the volatility equation becomes:

$$\sigma_t = \omega + \alpha_1 (|Z_{t-1}| + \gamma_1 Z_{t-1}) + \beta_1 \sigma_{t-1}$$

In this case, the model uses volatility directly and it is more robust to outliers. Recall, Equation (2.5) shows that the conditional variance of r_t is given as: $\sigma_t^2 = \text{Var}(r_t | F_{t-1}) = \text{Var}(Z_t | F_{t-1})$. Thus, the conditional variance of r_t for the APARCH model is

$$\sigma_t^\delta = \text{Var}(r_t | F_{t-1})$$

Therefore, if $t \rightarrow \infty$, then the conditional variance of Z_t would be

$$\sigma_t^\delta = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i (1 - \gamma_i) - \sum_{j=1}^q \beta_j} \quad (2.15)$$

The idea behind the APARCH model is in the introduction of the parameter δ . The power factor δ enhances the flexibility character of the proposed GARCH-type model, allowing a previous choice of the arbitrary power to be avoided. The APARCH model is a nested model which include:

- ARCH Model of Engle when $\delta = 2$, $\beta = 0$, and $\gamma = 0$
- GARCH Model of Bollerslev when $\delta = 2$, and $\gamma=0$
- TGARCH Model of Glosten, Jagannathan, and Runkle when $\delta=2$
- TS-GARCH (Taylor and Schwert GARCH) Model of Taylor and Schwert when $\delta=1$ and $\gamma=0$
- TARCH (Threshold ARCH) Model of Zakoian when $\delta=1$
- NARCH (Nonlinear ARCH) Model of Higgens and Bera when $\beta=0$, and $\gamma=1$
- EGARCH Model of Nelson when $\delta \rightarrow 0$

The Stationarity of the APARCH (1,1) model.

Given that $\{Z_{t-i} > 0\} = \{\varepsilon_{t-i} > 0\}$, then according to Francq and Zakoian (2010),

$$\sigma_t^\delta = \omega + \sum_{i=1}^{\max[p,q]} a_i(\varepsilon_{t-i})\sigma_{t-i}^\delta \quad (2.16)$$

where

$$\begin{aligned} a_i(k) &= \alpha_i(|k| - \gamma k)^\delta + \beta_i \\ &= \alpha_i(1 - \gamma_i)^\delta |k|^\delta \mathbb{I}_{[k>0]} + \alpha_i(1 + \gamma_i)^\delta |k|^\delta \mathbb{I}_{[k<0]} + \beta_i \end{aligned}$$

for $i = 1, \dots, \max[p, q]$

Thus, the strict stationarity condition of the process is

$$E[\log[\alpha_1(1 - \gamma_1)^\delta |\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t>0]} + \alpha_1(1 + \gamma_1)^\delta |\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t<0]} + \beta_1]] < 0 \quad (2.17)$$

For the APARCH (1,0) model

$$\begin{aligned} &\log[\alpha_1(1 - \gamma_1)^\delta |\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t>0]} + \alpha_1(1 + \gamma_1)^\delta |\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t<0]}] \\ &= \log(1 - \gamma_1)^\delta \mathbb{I}_{[\varepsilon_t>0]} + \log(1 + \gamma_1)^\delta \mathbb{I}_{[\varepsilon_t<0]} + \log \alpha_1 |\varepsilon_t|^\delta. \end{aligned}$$

It shows that if the distribution of ε_t is symmetric, then the strict stationarity condition becomes

$$|1 - \gamma_1|^{\delta/2} |1 + \gamma_1|^{\delta/2} \alpha_1 < e^{-E(\log|\varepsilon_t|^\delta)}.$$

In the case of $|\gamma_1| = 1$, the model is strictly stationary for any value of α_1 . The solution for strict stationarity condition given in Equation 2.17 is

$$Z_t = \sigma_t \varepsilon_t, \quad \sigma_t^\delta = \omega + \sum_{m=1}^{\infty} k_1 \varepsilon_t \dots k_1 \omega \varepsilon_{t-m+1} \quad (2.18)$$

Assuming $E|\varepsilon_t|^\delta < \infty$, the condition for the existence of $E(Z_t^\delta)$ and $E(\sigma_t^\delta)$ is

$$E(k_1 \varepsilon_t) = \alpha_1 \left\{ (1 - \gamma_1)^\delta E|\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t > 0]} + (1 + \gamma_1)^\delta E|\varepsilon_t|^\delta \mathbb{I}_{[\varepsilon_t < 0]} \right\} + \beta_1 < 1 \quad (2.19)$$

which reduces to

$$\frac{1}{2} E|\varepsilon_t|^\delta \alpha_1 \left[(1 - \gamma_1)^\delta + (1 + \gamma_1)^\delta \right] + \beta_1 < 1 \quad (2.20)$$

when the distribution of ε_t is symmetric, with

$$E|\varepsilon_t|^\delta = \sqrt{\frac{2^\delta}{\pi}} \Gamma\left(\frac{1 + \delta}{2}\right)$$

given that ε_t follows a normal distribution.

2.6 Maximum Likelihood Estimation (MLE).

Maximum-likelihood estimation (MLE) method is the most common method for the estimation of GARCH model (Karlsson, 2002). Let $L(\eta|Z_1, Z_2, \dots, Z_T)$ to be the likelihood function defined as a function of the parameters with the data, where $\eta = (\gamma, \delta, \theta)$ are the set of parameters needed to be estimated in the case of APARCH (p, q) model, given that γ and θ are defined as $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$ and $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$. Also, given that the data Z_1, Z_2, \dots, Z_T are not independent, then the joint density function is the product of the conditional density functions are given as:

$$f(Z_1, Z_2, \dots, Z_T | \eta) = f(Z_T | F_{T-1}) f(Z_{T-1} | F_{T-2}) \dots f(Z_1) \quad (2.21)$$

So that the likelihood function can be written as:

$$L(\eta|F_{t-1}) = \prod_{t=1}^T f(Z_t|F_{t-1}) \quad (2.22)$$

Then, the log-likelihood function is:

$$\ell(\eta|F_{t-1}) = \log[L(\theta|F_{t-1})] = \log \left[\prod_{t=1}^T f(Z_t|F_{t-1}) \right] \quad (2.23)$$

where F_t is information at time t , and f is the density function of ε_t .

Hence, the distribution of the error term will determine the likelihood function that will be used for estimation. Therefore, to specify any GARCH model, it is important to assume a specific distribution for the error term.

In Equation 2.4, the innovation Z_t is given as;

$$Z_t = \sigma_t \varepsilon_t.$$

where ε_t is assumed to follow one of the following distribution namely; standard normal, standardized Student t , and skewed Student t distributions. Given the types of distribution assumed for the error term, the log-likelihood function can be divided into three, namely: Gaussian Quasi Maximum-Likelihood Estimation, Fat-Tailed Maximum-Likelihood Estimation, and the Skewed Maximum-Likelihood Estimation.

2.6.1 Gaussian Quasi Maximum-Likelihood Estimation

Suppose that the error term ε_t is assumed to be normally distributed, then the innovation Z_t also follows normal (Gaussian) distribution with zero mean and variance σ_t^2 . Therefore, the probability density function Z_t is defined as:

$$f(Z_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{Z_t^2}{2\sigma_t^2}\right) \quad (2.24)$$

So that the likelihood function from Equation 2.22 becomes:

$$L(\eta|F_{T-1}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{Z_t^2}{2\sigma_t^2}\right) \quad (2.25)$$

Consequently, the log-likelihood equation is given as:

$$\ell(\eta|F_{T-1}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^T \ln(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^T \frac{Z_t^2}{\sigma_t^2} \quad (2.26)$$

Since $\ln(2\pi)$ does not involve any parameter the log-likelihood can be simplified and re-written as:

$$\ell(\eta|F_{T-1}) = -\sum_{t=1}^T \left(\frac{\ln(\sigma_t^2)}{2} + \frac{Z_t^2}{2\sigma_t^2} \right) \quad (2.27)$$

Therefore, we have:

$$\frac{\partial \ell(\eta|F_{T-1})}{\partial \eta} = \frac{1}{2} \sum_{t=1}^T \left(-\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \eta} - \frac{1}{\sigma_t^2} \frac{\partial Z_t^2}{\partial \eta} + \frac{Z_t^2}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \eta} \right)$$

$$\frac{\partial \ell(\eta|F_{T-1})}{\partial \eta} = \frac{1}{2} \sum_{t=1}^T \left(\frac{Z_t^2 - \sigma_t^2}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \eta} - \frac{Z_t}{\sigma_t^2} \frac{\partial Z_t}{\partial \eta} \right)$$

In order to differentiate the log-likelihood function with respect to η , it requires the computation of $\frac{\partial \sigma_t^\delta}{\partial \eta}$ in the case of APARCH model which is given as

$$\frac{\partial \sigma_t^\delta}{\partial \eta} = \frac{\partial (\omega + \sum_{i=1}^p \alpha_i (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta)}{\partial \eta} \quad (2.28)$$

Rewriting σ_t^2 as $(\sigma_t^\delta)^{2/\delta}$ leads to

$$\frac{\partial \sigma_t^2}{\partial (\theta', \gamma')} = \frac{2\sigma_t^2}{\partial \sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial (\theta', \gamma')}$$

and

$$\frac{\partial \sigma_t^2}{\partial \delta} = \frac{2\sigma_t^2}{\partial \sigma_t^\delta} \left[\frac{\partial \sigma_t^\delta}{\partial \delta} - \frac{\sigma_t^\delta \ln(\sigma_t^\delta)}{\delta} \right]$$

Thus, to obtain a tractable solution then σ_t^δ must be differentiated with respect to each parameter. According to Laurent (2003) it is better to set unobserved components to the sample average at the start of a recursion of the Equation 2.12 . That is

$$|Z_{t-i}| + \gamma_i Z_{t-i} = \frac{1}{T} \sum_{s=1}^T (|Z_s| + \gamma_i Z_s)^\delta \quad \text{for } t \leq i$$

and

$$\sigma_t^\delta = \left(\frac{1}{T} \sum_{s=1}^T Z_s^2 \right)^{\frac{\delta}{2}} \quad \text{for } t \leq 0$$

- Laurent (2003) obtained the $\frac{\partial \sigma_t^\delta}{\partial \gamma}$ as

$$\frac{\partial \sigma_t^\delta}{\partial \gamma} = \frac{\partial \omega}{\partial \gamma} + \sum_{i=1}^p \frac{\partial \alpha_i (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta}{\partial \gamma} + \sum_{j=1}^q \frac{\beta_j \sigma_{t-j}^\delta}{\partial \gamma} \quad (2.29)$$

where

$$(|Z_{t-i}| + \gamma_i Z_{t-i})^\delta = \begin{cases} (|Z_{t-i}| + \gamma_i Z_{t-i})^{\delta-1} Z_{t-i} & \text{if } t > j \\ \frac{\delta}{T} \sum_{s=1}^T (|Z_s| + \gamma_i Z_s)^{\delta-1} Z_s & \text{if } t \leq j \end{cases}$$

and $\frac{\partial \sigma_t^\delta}{\partial \gamma} = 0$, for $t \leq 0$. As t changes the derivative changes.

- Laurent (2003) obtained $\frac{\partial \sigma_t^\delta}{\partial \delta}$ as

$$\frac{\partial \sigma_t^\delta}{\partial \delta} = \frac{\partial \omega}{\partial \delta} + \sum_{i=1}^p \frac{\partial \alpha_i (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta}{\partial \delta} + \sum_{j=1}^q \frac{\beta_j \sigma_{t-j}^\delta}{\partial \delta} \quad (2.30)$$

then

$$\begin{aligned}
\frac{\partial \sigma_t^\delta}{\partial \delta} &= \delta \sum_{i=1}^p [\alpha_i (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta \ln(|Z_{t-i}| + \gamma_i Z_{t-i})]^{F_{t-i}} \\
&\quad \times \left[\frac{1}{T} \sum_{s=1}^T [(|Z_{t-i}| + \gamma_i Z_{t-i})^\delta \ln(|Z_s| + \gamma_i Z_s)]^{1-F_{t-i}} \right] \\
&\quad + \sum_{j=1}^q \beta_j [\sigma_{t-j}^\delta \ln(\sigma_{t-j})]^{F_{t-j}} \left[-\frac{1}{T} \left(\frac{1}{T} \sum_{s=1}^T Z_s^2 \right)^{\frac{\delta}{2}} \ln \left(\frac{1}{T} \sum_{s=1}^T Z_s^2 \right) \right]^{1-F_{t-j}} \quad (2.31)
\end{aligned}$$

where $F_t = \begin{cases} 1, & t > 0 \\ 0, & t \leq 0 \end{cases}$

- Finally, $\frac{\partial \sigma_t^\delta}{\partial \theta}$ is given as

$$\frac{\partial \sigma_t^\delta}{\partial \theta} = \frac{\partial \sigma_t^\delta}{\partial \theta} = \frac{\partial \omega}{\partial \theta} + \sum_{i=1}^p \frac{\alpha_i \partial (|Z_{t-i}| + \gamma_i Z_{t-i})^\delta}{\partial \theta} + \sum_{j=1}^q \frac{\beta_j \sigma_{t-j}^\delta}{\partial \theta} \quad (2.32)$$

and $\frac{\partial \sigma_t^\delta}{\partial \gamma} = 0$ for $t \leq 0$.

2.6.2 Heavy-tailed Maximum-Likelihood Estimation (MLE)

The best way to manage a non-Gaussian error term is by assuming a non-Gaussian distribution for the error term that can explain the behaviour of the data better. Hence, the parameters will be estimated by using the non-Gaussian distribution in the likelihood function. This section will consider a non-Gaussian distribution known as Student t Distribution.

2.6.2.1 MLE for Student t Distribution

The Student t distribution was proposed by Praetz (1972) and Blattberg and Gonedes (1974) for modeling financial returns. If the error terms are assumed to follow a student t distribution then the innovation Z_t also follows symmetric student t distribution with w degree of freedom and has a mean of zero and variance of $\frac{w}{w-2}$ for $w > 2$. The probability density function is given as:

$$f(Z_t|w) = \frac{\Gamma\left(\frac{w+1}{2}\right)}{\Gamma\left(\frac{w}{2}\right)\sqrt{(w-2)\pi}\sigma_t} \frac{1}{\left(1 + \frac{Z_t^2}{(w-2)\sigma_t^2}\right)}, \quad \text{for } w > 2 \quad (2.33)$$

Then, the log-likelihood function can be written as:

$$\ell(\theta|F_{T-1}) = -\frac{1}{2} \sum_{t=1}^T \ln(\sigma_t^2) - \frac{w+1}{2} \sum_{t=1}^T \ln\left(1 + \frac{Z_t^2}{(w-2)\sigma_t^2}\right) \quad (2.34)$$

Therefore

$$\begin{aligned} \frac{\partial \ell(\eta|F_{T-1})}{\partial \eta} &= -\frac{1}{2} \sum_{t=1}^T \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \eta} - \frac{w+1}{2} \sum_{t=1}^T \frac{\partial \ln\left(1 + \frac{Z_t^2}{(w-2)\sigma_t^2}\right)}{\partial \eta} \\ \frac{\partial \ell(\eta|F_{T-1})}{\partial \eta} &= -\frac{1}{2} \sum_{t=1}^T \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \eta} - \frac{w+1}{2(w+1)} \frac{1}{1 + \frac{Z_t^2}{(w-2)\sigma_t^2}} \sum_{t=1}^T \frac{\partial Z_t^2}{\partial \eta} \end{aligned}$$

where $\frac{\partial Z_t^2}{\partial \eta} = \sigma_t^2 \frac{\partial \varepsilon_t^2}{\partial \eta} + \varepsilon_t^2 \frac{\partial \sigma_t^2}{\partial \eta} = 2\varepsilon_t \sigma_t^2 \frac{\partial \varepsilon_t}{\partial \eta} - 2\varepsilon_t \sigma_t^3 \frac{\partial \sigma_t}{\partial \eta}$

From the above $\frac{\partial \sigma_t^2}{\partial (\theta', \gamma')} = \frac{2\sigma_t^2}{\partial \sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial (\theta', \gamma')}$ and $\frac{\partial \sigma_t^2}{\partial \delta} = \frac{2\sigma_t^2}{\partial \sigma_t^\delta} \left[\frac{\partial \sigma_t^\delta}{\partial \delta} - \frac{\sigma_t^\delta \ln(\sigma_t^\delta)}{\delta} \right]$

The derivative of σ_t^δ with the respect to γ , δ , and θ will follow the same steps as normal distribution above.

2.6.3 Skewed Maximum-Likelihood Estimation

Since Kon (1984) presented the Skewed Maximum-Likelihood Estimation, it has been set as standard practice in considering asymmetry of financial returns. Broadening the density function as described in Equation 2.18 by λ , which is the skewness parameter gives skewed student t distribution and account for skewness and excess kurtosis. Hansen (1994), also described the density of the skewed student t distribution as positive skewness if λ is positive or negative skewness for negative λ . The probability density function is given as:

$$f(Z_t|\lambda, w) = \begin{cases} \frac{2}{\lambda + \frac{1}{\lambda}} sg(\lambda[sZ_t + m]|w), & Z_t < -\frac{m}{s} \\ \frac{2}{\lambda + \frac{1}{\lambda}} sg\left(\frac{sZ_t + m}{\lambda} |w\right), & Z_t \geq -\frac{m}{s} \end{cases} \quad (2.35)$$

where $g(\cdot)$ is the pdf of the standardized student t distribution, m and s is the mean and variance of non-standardized skew student respectively given as

$$m = \frac{\Gamma\left(\frac{w-1}{2}\right)\sqrt{w-2}}{\sqrt{\pi}\Gamma\left(\frac{w}{2}\right)}\left(\lambda - \frac{1}{\lambda}\right)$$

and

$$s^2 = \left(\lambda^2 + \frac{1}{\lambda^2} - 1\right) - m^2$$

According to Hansen (1994), if $\lambda = 0$, then skewed student t distribution becomes a student t distribution.

2.7 Conclusion

In this chapter, basic concepts of financial time series, basic structure of financial volatility model, the mean model, the volatility model and MLE of GARCH model were reviewed.

CHAPTER 3

EXTREME VALUE THEORY (EVT)

3.1 Introduction

The aim of this chapter is to review the theoretical foundation of parametric EVT and the modelling technique for extreme events. Section 3.2 considers various modelling techniques used to describe an extreme event. The mathematical concepts and results that constitute the theoretical foundation of the extreme value modelling approach are presented. Sections 3.3 consider the parameter estimation of the extreme value models based on maximum likelihood estimation. Sections 3.4 consider the problems with extreme value modelling.

3.2 Extreme Value Modelling

The peaks over threshold (POT) and the block maxima (BM) approaches are the two aspects to modelling events in extreme distribution (Gilli and Kellezi, 2006). Suppose X_i , where $i = 1, \dots, n$ are random variables which represents data of daily returns. The characteristic of each approach is identified in detail- below:

Block maxima approach (BM): This type of approach demands selection of maximum values in every consecutive period. The selected observations are named block maxima while matching periodic extreme events.

The block maxima approach which is used to generate generalized extreme value distribution is not efficient and therefore considered as data waster especially if complete data is available (Hu, 2013). For example in an instance where 20 daily observations are considered with a block size n , that is the number of observations for each block, given 4 blocks then every block will consist of 5 days. Figure 2.1a below described the observations X_3, X_7, X_{15} and X_{19} as the block maxima or the extreme events in block $m = 1, 2, 3$ up to 4 respectively; m symbolizes a specific block (Gilli and Kellezi, 2006).

Peaks over threshold approach (POT): centres on events exceeding a big threshold value ‘ u ’. It is necessary to choose a threshold value ‘ u ’ which is large enough and equally has many events above it.

The POT which is used to generate generalized Pareto distribution is better because it makes use of every extreme data (Hu, 2013). Now, if another 20 daily observations are considered as revealed in Figure 2.1b, assume 80th percentile of “ u ” is chosen with 4 observations above ‘ u ’. It is observed from the figure that X_3 , X_7 , X_{18} , and X_{19} are the extreme event since they exceed the threshold of ‘ u ’. For each of these four observations, the exceedances above ‘ u ’ can be computed as: $X_3 - u$, $X_7 - u$, $X_{18} - u$ and $X_{19} - u$, which are random observations characterized by a distribution of excesses (Gilli and Kellezi, 2006).

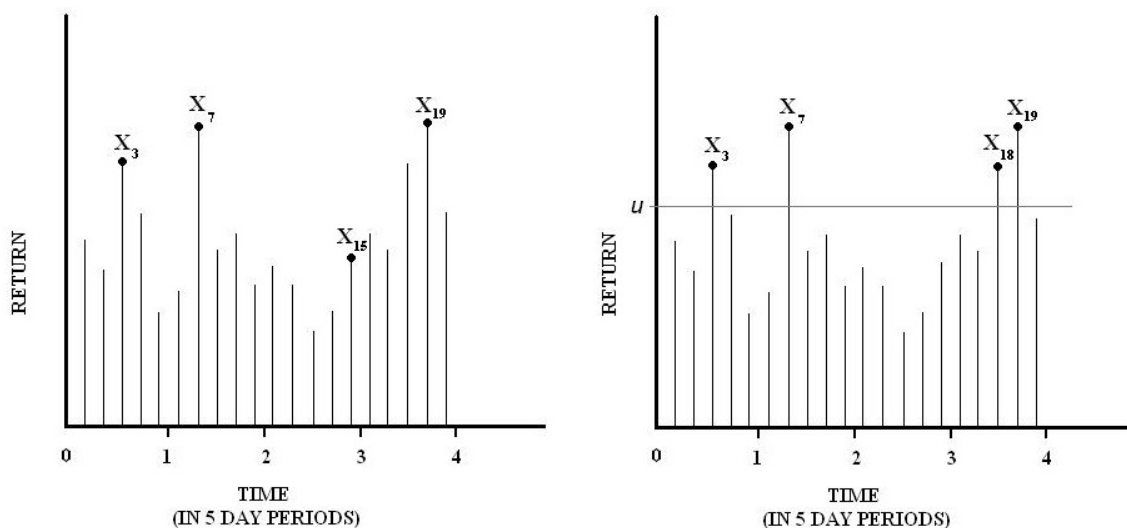


FIGURE 3. 1: (A) BLOCK MAXIMA AND (B) PEAKS OVER THRESHOLD (GILLI AND KELLEZI, 2006)

These approaches were compared by using simulation data of over 200 years by Caries (2009) and results from the research show the estimate accuracy outcomes from each of the two approaches have similar figures. For data obtained within a time span of fewer than 200 years, POT approach has better performance compared to the BM approach. Jaruskova and Hanek (2006), made a conclusion after working on data of daily discharges from two rivers for forty-five years and daily precipitations forty years that both approaches correlate based on their results. It was also noted that in a situation where the number of observation is moderately

large, the BM approach shows good results whereas the POT approach would give a poor outcome.

Coles (2001) illustrates that if X_1, \dots, X_n is a sequence of independent stochastic variables with F representing the common population distribution and M_n is the process maxima (or minima) over block size n , that is, $M_n = \max\{X_1, \dots, X_n\}$, then the distribution M_n in the case of a maxima is given as

$$\begin{aligned} P\{M_n \leq x\} &= P\{X_1 \leq x, \dots, X_n \leq x\} \\ &= P\{X_1 \leq x\} \times \dots \times P\{X_n \leq x\} \\ &= \prod_{i=1}^n P(X \leq x) \\ &= [P(X \leq x)]^n \\ &= [F(x)]^n \end{aligned}$$

Two problems can possibly be developed from this approach. The first one, $F(x)^n$ approaches zero as n approaches ∞ since $F(x) < 1$ (Pocernich, 2002). The second problem is that the distribution of F may be unknown. The values can be renormalized in a way that $M_n^* = \frac{M_n - d_n}{c_n}$ as a solution to the first problem for $c_n > 0$. c_n and d_n are also sequences of constants.

Theorem 3.1: (Extremal types theorem) - If there exist a sequence of constants $c_n > 0$ and d_n such that,

$$\lim_{n \rightarrow \infty} P \left[\frac{M_n - d_n}{c_n} \leq x \right] \rightarrow H(x)$$

where H is a non-degenerate distribution function, then H belongs to one of following extreme value distributions (Coles, 2001).

$$\text{Gumble: } H(x) = \exp \left\{ -\exp \left[-\left(\frac{x - \mu}{\sigma} \right) \right] \right\}, \quad -\infty < x < \infty;$$

$$\text{Fréchet: } H(x) = \begin{cases} 0, & x \leq u; \\ \exp \left\{ -\left(\frac{x - \mu}{\sigma} \right)^{-\alpha} \right\}, & x > u; \end{cases}$$

$$\text{Weibull: } H(x) = \begin{cases} \exp\left\{-\left[\left(\frac{x-\mu}{\sigma}\right)^\alpha\right]\right\}, & x < u, \\ 1, & x \geq u, \end{cases}$$

for parameters for $c > 0$, d and $\alpha > 0$.

Leadbetter et al. (1983), provided a proof that validates this theorem as it was concluded that for every population whose distribution is unknown, applying a definite type of limiting distribution is not appropriate because uncertainties attributed to family distribution is disregarded. Therefore, it is better to apply a wider extreme value distribution which covers the three types.

3.2.1 Generalized Extreme Value Distribution (GEVD)

An approach to unify the three aspects of extreme value distribution was proposed by von Mises (1954) and Jenkinson (1955), leading to what is termed ‘generalized extreme value distribution’ GEVD mostly used for modelling block maxima distribution. Coles (2001), gives GEV family as;

$$H(x) = \exp\left\{-\left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\} \quad (3.1)$$

defined on $\left\{x: 1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$, where $\sigma > 0, \mu, \varepsilon \in \mathbb{R}$ and $\varepsilon = \frac{1}{\gamma}$ such that $\gamma = \varepsilon^{-1}$ is the rate of tail decay. The meaning of three parameters are given below:

- μ is defined as the location parameter, correspondent to the mean.
- σ is the scale parameter corresponding to the standard deviation.
- ε is defined as the shape parameter/tail index, which defines the thickness at the tails.

There are three cases of Equation (3.1) which include:

1. Gumbel distribution which has two main parameters μ and σ as $\varepsilon \rightarrow 0$. It comprises distributions such as;

- Exponential distribution: if $F(x) = 1 - e^{-x}$, then

$$H(x) = [1 - e^{-x}n^{-1}]^n \rightarrow \exp(-e^{-x})$$

which is the Gumbel distribution with $\mu = 0$ and $\sigma = 1$.

- Normal, Logistic and Gamma distributions

These mentioned cases are identified as light-tailed distribution, therefore, $\varepsilon \rightarrow 0$ is related to light-tailed Gumbel order of distribution. It is also called the double exponential distribution. The tail decreases progressively in an exponential distribution and is unbounded, which means every moment exists.

2. A Weibull distribution with $\varepsilon < 0$ comprises of uniform and beta distributions which are identified as short-tailed distribution. The Weibull type of distribution is characterised by a bounded tail with a finite endpoint, even though not all of its moments are finite.
3. A heavy tail distribution of Frechet type with $\varepsilon > 0$ includes Pareto, Log-gamma, Student t and Cauchy. Frechet type of distribution is described by a polynomial tail decay, having moments only up to γ . Both the Weibull and Frechet distribution have three parameters.

A few properties of the Generalised Extreme Value (Smith, 2003) include:

- if $\varepsilon < 1$, then the mean exist, thus the expectation is;

$$E(x) = \mu + \frac{\sigma}{\varepsilon} \{\Gamma(1 - \varepsilon) - 1\} \quad (3.2)$$

- if $\varepsilon < \frac{1}{2}$, then the variance exist which is given by;

$$\text{Var}(x) = E \left\{ X - (E(X))^2 \right\} = \frac{\sigma^2}{\varepsilon^2} \{ \Gamma(1 - 2\varepsilon) - \Gamma^2((1 - \varepsilon)) \} \quad (3.3)$$

- if $\varepsilon = 0$, the then expectation of the mean and the variance are;

$$E(x) = \mu + \sigma\gamma \quad (3.4)$$

and

$$\text{Var}(x) = \frac{\sigma^2 \pi^2}{6} \quad (3.5)$$

where $\gamma \approx 0.5772$ (Euler's constant).

- The q th quantile of the GEVD is

$$x_q = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\varepsilon}} \{[-n \ln(1-q)]^{-\hat{\varepsilon}} - 1\}, & \varepsilon \neq 0 \\ \hat{\mu} - \hat{\sigma} \ln[-n \ln(1-q)], & \varepsilon \rightarrow 0 \end{cases} \quad (3.6)$$

3.2.2 Generalized Pareto Distribution (GPD)

The GPD is used to model the peaks over threshold which was proposed by Pickands (1975) and a comprehensive treatment of the model was given later by Davison and Smith (1990). Given that X_1, \dots, X_n is a sequence of IID random variables with F common distribution, consider the X_i 's over the given threshold ' u ' as extreme events. Given that X is the arbitrary term in the sequence X_i and let $y = X - u$, be the value of an exceedance over the threshold ' u '. Then, by using the law of conditional probability, the cumulative distribution function of y is;

$$\begin{aligned} F_u(y) &= P(X - u \leq y | X > u) \\ &= \frac{P(X \leq u + y | X > u)}{P(X > u)} \\ &= \frac{F(y + u) - F(u)}{1 - F(u)}, \quad y > 0 \end{aligned} \quad (3.7)$$

Coles (2001), shows that as $u \rightarrow \infty$, $F_u(y)$ is approximately a generalized Pareto family. By Theorem 3.1, for large n we have;

$$F^n(x) \approx \exp \left\{ - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}} \right\}, \quad \mu, \sigma > 0$$

Thus, taking the natural log, we have:

$$n \log F(x) \approx - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}} \quad (3.8)$$

Now, by Taylor expansion;

$$\begin{aligned} F(x) &= 1 - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}} + \dots \\ &= 1 - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}} \end{aligned}$$

for large x

By rearrangement, we get;

$$-(1 - F(x)) = \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}}$$

Thus, we have;

$$\ln F(x) \approx -(1 - F(x)) \quad (3.9)$$

Substituting Equation (3.9) into (3.8), we have;

$$n(1 - F(x)) \approx \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}}$$

$$(1 - F(x)) \approx \frac{1}{n} \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}}$$

Thus;

$$F(u) \approx 1 - \frac{1}{n} \left[1 + \varepsilon \left(\frac{u - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}}$$

$$F(y + u) - F(u) \approx -\frac{1}{n} \left[1 + \varepsilon \left(\frac{y + u - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}} + \frac{1}{n} \left[1 + \varepsilon \left(\frac{u - \mu}{\sigma} \right) \right]^{-\frac{1}{\varepsilon}}$$

Therefore;

$$F_u(y) = \left\{ 1 - \left(1 + \frac{\varepsilon y}{\sigma^*} \right)^{-1/\varepsilon} \right\}$$

where $\sigma^* = \sigma + \varepsilon(u - \mu)$.

Theorem 3.2:

Supposed X_1, \dots, X_n is the sequence of independent and identically distributed random variables with common distribution function, F (Coles, 2001), and let

$$M_n = \max\{X_1, \dots, X_n\}$$

Given that the common cumulative distribution function, F of X_1, \dots, X_n satisfies Theorem 3.1, so that for large n ,

$$P\{M_n \leq x\} \approx H(x)$$

where $H(x)$ is given in Equation 3.1 above.

Then for large “ u ”, if $\varepsilon \neq 0$, then

$$\lim_{u \rightarrow \infty} P(X - u \leq y | X > u) \approx K(y) = \left\{ 1 - \left(1 + \frac{\varepsilon y}{\sigma^*} \right)^{-1/\varepsilon} \right\}, \quad (3.10)$$

defined on $\left\{ y: y > 0 \text{ and } \left(1 + \frac{\varepsilon y}{\sigma^*} \right) > 0 \right\}$, where

$$\sigma^* = \sigma + \varepsilon(u - \mu). \quad (3.11)$$

and for large “ u ”, if $\varepsilon = 0$, then

$$\lim_{u \rightarrow \infty} P(X - u \leq y | X > u) \approx K(y) = \left\{ 1 - \exp\left(\frac{y}{\sigma^*}\right)^{-1/\varepsilon} \right\}, \quad y > 0 \quad (3.12)$$

The Equations 3.10 and 3.12 defined above are the generalized Pareto distribution. The deduced outcome shows that the GPD shape parameter is similar to that of the GEVD making the dependence of the GPD scale σ^* and threshold ‘ u ’ extremely distinct.

The properties of GPD include:

- The mean and the variance of the GPD are given by (Pocernich, 2002) as follows;

$$\blacksquare E(Y) = \frac{\sigma^*}{1-\varepsilon} \quad (3.13)$$

$$\blacksquare \text{Var}(Y) = \frac{(\sigma^*)^2}{(1-\varepsilon)^2(1-2\varepsilon)} \quad (3.14)$$

$$\blacksquare E(Y - u|y > w) = \frac{\sigma^* + \varepsilon w}{1-\varepsilon} \quad (3.15)$$

where $w > 0$

- The q th quantile of the GEVD is

$$x_q = \begin{cases} u + \frac{\widehat{\sigma}^*}{\widehat{\varepsilon}} \left[\left(q \frac{N}{n} \right)^{-\widehat{\varepsilon}} - 1 \right], & \varepsilon \neq 0 \\ u - \widehat{\sigma}^* \ln q \frac{N}{n}, & \varepsilon \rightarrow 0 \end{cases} \quad (3.16)$$

3.3 Parameter Estimation

Different methods are used in the parameter estimation determination. Musah (2010) listed some of these as Method of Moments Estimation (MME), equivalently L-Moments (LM) or Probability Weighted Moments (PWM), Bayesian methods and Maximum Likelihood Estimation (MLE). Zhao (2010) described the ML method as the most widely used of all the approaches in spite of even likelihood for only $\varepsilon > \frac{1}{2}$. In dealing with finance data, L- moment has proven to be a preferred approach because it obtains reliable estimates for events with heavy tails. Alternatively, Bayesian inference can help in implicit sparsity of extreme events through explicit processing of known expert information.

MLE is not suitable when handling smaller sample sizes, specifically, for $n < 50$ (Musah, 2010). This is because it is unstable and as a result prone to giving uncertain shape parameter estimates. The following researchers, (Hosking and Wallis, 1987; Coles and Dixon, 1999; Martins and Stedinger, 2000; Martins and Stedinger, 2001; Madsen et al., 1997) all disputed this instead and argued that the MME quantile estimators give lower root mean square error when the true shape parameter values zero.

The method of moments estimation (MME) and Probability weighted moments (PWM) are inapplicable when $\varepsilon \leq -0.5$ because moments ≤ 2 are non-existent, therefore values for using PWM or MME also do not exist (Castillo and Hadi, 1997). Apart from the stated effect, both

can equally result in non-consistent estimates involving the observed data. Castillo and Hadi (1997), used simulations to compare results of using these stated methods.

The integration of covariate data into parameter estimates is achieved easily with MLE. In addition, MLE allows a fairly easy way of obtaining error bounds for parameter estimates when compared with other methods. Using Bayesian estimation for extreme-value analysis has been carried out by Coles (2001); Stephenson and Tawn (2004) and Cooley et al. (2007). It is noted that the Gumbel distribution does not hold when data is to be fitted to a GEV due to the fact that Gumbel distribution is scaled down to a unit point in an uninterrupted parameter space. A usual approach used is to a preliminary hypothesis in determining which of the extreme indicator is most appropriate after which data is fitted into it. It is evident that this method will not give explanation for uncertainty encountered in the choice of tail type on the resultant inference, which are usually large. Stephenson and Tawn (2004) therefore proposed that Bayesian method in estimation of parameters which accords the Gumbel distribution to be accomplished with positive probability. The outcomes can be extremely dependent on the selection of prior distributions.

3. 3.1 Maximum Likelihood Estimation of GEVD

Suppose X_1, \dots, X_m is a sequence of a block maxima for 'm' blocks, given that X is a random variable defined as:

$$X \equiv M_k = \max\{X_1, \dots, X_k\}$$

Assume GEVD is fitted to X_1, \dots, X_m , then the probability density function of Equation 3.1 is (by differentiation of $H(x)$ with respect to x), if $\varepsilon \neq 0$;

$$H'(x) = \frac{1}{\sigma} \left\{ \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-1-1/\varepsilon} \right\} \exp \left\{ - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\varepsilon} \right\}$$

If $\varepsilon \rightarrow 0$;

$$H'(x) = \frac{1}{\sigma} \exp \left[-\exp \left(- \left(\frac{x - \mu}{\sigma} \right) \right) - \left(\frac{x - \mu}{\sigma} \right) \right]$$

Thus, the log-likelihood as a function of X_1, \dots, X_m is given as;

$$\ell(\mu, \sigma, \varepsilon) = \ln L(\mu, \sigma, \varepsilon) = \ln \prod_{j=1}^m H'(x_j)$$

Therefore, for $\varepsilon \neq 0$;

$$\ell(\mu, \sigma, \varepsilon) = -m \ln \sigma - \left(1 + \frac{1}{\varepsilon}\right) \sum_{j=1}^m \ln \left[1 + \varepsilon \left(\frac{x_j - \mu}{\sigma}\right)\right] - \sum_{j=1}^m \left[1 + \varepsilon \left(\frac{x_j - \mu}{\sigma}\right)\right]^{-1/\varepsilon} \quad (3.17)$$

and, for $\varepsilon \rightarrow 0$;

$$\ell(\mu, \sigma, \varepsilon) = -m \ln \sigma - \sum_{j=1}^m \exp\left(-\frac{x_j - \mu}{\sigma}\right) - \sum_{j=1}^m \left(\frac{x_j - \mu}{\sigma}\right) \quad (3.18)$$

The estimates of $(\mu, \sigma, \varepsilon)$ are obtained by optimising Equation 3.17 or its equivalent Equation 3.18. By differentiating these equations, the score expressions are obtained with no definite solutions. Even though ML has attractive advantages, the ML method is not suitable for solving some EVT problems. The asymptotic tendencies of ML estimators make it effective under normal ML theory specifics (Cox and Hinkley, 1974). Since GEVD applications depends on unknown parameters, the normal conditions are not satisfied, even though adequate numerical processes is available in obtaining the ML estimates, but they still lack the normal asymptotic properties of ML estimation. The normal asymptotic condition of ML estimators used in GEVD is a function of the unknown EV index value (Smith, 1985). This was further established that the ML estimator exist for $\varepsilon > -1$, however classical asymptotic property of consistency and asymptotic normality hold only for $\varepsilon > -0.5$, therefore the case $-1 < \varepsilon \leq -0.5$ remain unsolved. For example, if we let $\varepsilon > -0.5$, we have;

$$\sqrt{m}[(\hat{\mu}, \hat{\sigma}, \hat{\varepsilon}) - (\mu, \sigma, \varepsilon)] \xrightarrow{d} z \approx N(0, I^{-1}), \quad m \rightarrow \infty$$

where I is the Fisher Information matrix.

In actual sense, a sample having an extremely short tail $\varepsilon < -0.5$ is uncommon. In the financial application aspect, most models show a positive tail index of $\varepsilon > 0$, making ML a relevant tool for GEVD calculation (Smith, 1985).

When considering $\varepsilon \leq -1$, ML approach is inapplicable because the log-likelihood function does not have local maximum. It was observed that the density has a J-shape and the corresponding log-likelihood function approach $+\infty$ along some path in the log-likelihood space. It was observed that this shortcoming poses a negligible application importance. This anchor on the fact that, distributions with $\varepsilon \leq -1$ with a very light upper tail are hardly experience in typical EVT analysis. Zhou (2009) and Zhou (2010) proffered solutions to problem discovered by Smith (1985) by establishing that the ML estimator meets the two listed asymptotic properties for $\varepsilon \leq -1$. Another problem associated with using ML estimation is identified as the convergence of iterative process of maximisation. This can be linked to the fact that computational process do not converge, making it difficult to find a suitable estimator. Upon its numerous challenges, estimating with ML handles processes with missing data, temporary dependence, and non-stationarity with minor alteration which will be difficult using other estimators, in fact almost impossible.

3.3.2 Maximum Likelihood Estimation of GPD

Suppose y_1, \dots, y_m is excess over a sufficiently high threshold “ u ” from the original random variable x_1, \dots, x_n , $m < n$. That is, $y_i = x_i - u$. Assume GPD is fitted to the excess values y_i , then the probability density function obtained by differentiation of $K(y)$ with respect to y from Equation 3.10 and 3.12 are;

For $\varepsilon \neq 0$; we have

$$k'(y) = \frac{1}{\sigma^*} \left(1 + \frac{\varepsilon y}{\sigma^*}\right)^{-1-1/\varepsilon}$$

For $\varepsilon \rightarrow 0$; we have

$$k'(y) = \frac{1}{\sigma^*} \exp\left(-\frac{y}{\sigma^*}\right)$$

Thus, the log-likelihood function of y_1, \dots, y_m is given in Vicente (2012) as;

$$\ell(\sigma^*, \varepsilon) = \ln L(\sigma^*, \varepsilon) = \ln \prod_{i=1}^m K'(y_i)$$

For $\varepsilon \neq 0$; we have

$$\ell(\sigma^*, \varepsilon) = -m \ln \sigma^* - \left(1 + \frac{1}{\varepsilon}\right) \sum_{i=1}^m \ln \left(1 + \frac{\varepsilon y_i}{\sigma^*}\right) \quad (3.19)$$

where $1 + \frac{\varepsilon y_i}{\sigma^*} > 0$.

For $\varepsilon \rightarrow 0$; we have

$$\ell(\sigma^*, \varepsilon) = -m \ln \sigma^* - \frac{1}{\sigma^*} \sum_{i=1}^m y_i \quad (3.20)$$

By defining $\tau = \frac{\varepsilon}{\sigma^*}$, the log-likelihood function could be represented as:

$$\ell(\tau, \varepsilon) = -m \ln \varepsilon + m \ln \tau - \left(1 + \frac{1}{\varepsilon}\right) \sum_{i=1}^m \ln(1 + \tau y_i)$$

where $1 + \tau y_i > 0$.

Then, the estimates of $(\hat{\varepsilon}, \hat{\tau})$ is given as:

$$\frac{1}{\hat{\tau}} - \left(1 + \frac{1}{\hat{\varepsilon}}\right) \frac{1}{m} \sum_{i=1}^m \frac{y_i}{1 + y_i \hat{\tau}} = 0 \quad (3.21)$$

where $\hat{\varepsilon} = \frac{1}{m} \sum_{i=1}^m \ln(1 + y_i \hat{\tau})$.

The re-parameterisation of the log-likelihood function recommended by Davison (1984) is essential in getting $\hat{\varepsilon}$ explicitly as a function of $\hat{\tau}$, which is computed mathematically using Equations 3.21. with the replacement of $\hat{\varepsilon} = \frac{1}{m} \sum_{i=1}^m \ln(1 + y_i \hat{\tau})$. For $\varepsilon \rightarrow 0$, there is a definite example of the exponential distribution, yielding $\sigma^* = \bar{Y}$.

Referring to Zhou (2009) and Zhou (2010) in Section 3.3.1, it is demonstrated that the asymptotic normality and consistency of the estimates of maximum likelihood within the EVT for $\varepsilon > -1$ is specified to yield:

$$\sqrt{m}[(\hat{\sigma}^*, \hat{\varepsilon}) - (\sigma^*, \varepsilon)] \xrightarrow{d} z \approx N(0, V), \quad m \rightarrow \infty$$

$$\text{where } V = \begin{bmatrix} (1 + \varepsilon)^2 & -\sigma^*(1 + \varepsilon) \\ -\sigma^*(1 + \varepsilon) & -2(\sigma^*)^2(1 + \varepsilon) \end{bmatrix}$$

3.4 Problems with Extreme Value Modelling.

Extreme value models are characterised with challenges ranging from dependency of extremes, threshold selection and to extremal observations deficiency. An overview of the listed challenges is provided in the section below (Zhao, 2010).

3.4.1 Dependence of Extremes

The theory behind generalised extreme value distribution establishes a limiting distribution of a series of random variables which are IID and whose upper or lower limits justification are asymptotical. Extreme events usually take place in clusters as a result of dependency in data. Beirlant et al. (2004), stated that under specific conditions, the distribution of the extreme levels lies along the same GEVD family as a result of minimum range within events. However, these conditions are not usually satisfied, also the dependent extreme sequences has less information than an IID sequences and the statistical inference should be adjusted in situation of dependence. It is logical to attempt minimising dependence while selecting extremes for a given sample in order to reduce the effect of dependence.

Financial returns mostly demonstrate clusters of events also called Auto Regressive Conditional Heteroskedastic (ARCH) process and the general form also known as GARCH process. Latest developments in the finance review shows that a two level model is used to apply extreme models to analyse dependency in events (McNeil and Frey, 2000; Chan et al., 2007; Zhao et al., 2010a)

The first stage uses GARCH model to stimulate clusters volatility by addressing dependence in the returns while the second stage uses extreme value model to develop independent residual concept. A major setback is as a result of uncertainty estimation.

3.4.2 Lacking extremal observation

A distinct difficulty attached to using extreme value models is linked to the fact that extremal data is sparsely distributed. This can easily lead to model classification as well as parameter estimation difficulties, especially when considering a system with complex model. Bayesian inference is an approach that can make use of the advantage prior information in a complex extreme model.

3.4.3 Threshold selection and Common Approach

Prior to introducing GPD into certain data, it is necessary to choose an appropriate threshold. Coles (2001), recommended that selecting threshold process accommodates a compromise between the variance and bias.

For a low threshold, the asymptotic arguments that form the foundation GPD model derivation are violated. By contrast, extremely high threshold generates less exceedances capable of estimating the shape (pattern) and scale parameter which leads to a very high variance. For this reason, it is required during threshold selection to consider if the limiting model offers a sufficiently good approximation against the variance of the parameter estimate. Three diagnostics were listed by Coles (2001) in determining threshold choice. These are;

1. Mean residual life plot,
2. Parameter stability plot,
3. Model fits diagnostics plots.

3.4.3.1 Mean Residual Life Plot

This approach considered the mean of the GPD (Coles, 2001). If the GPD of Y has parameters σ and ε then

$$E(Y) = \frac{\sigma}{1 - \varepsilon}$$

For excess $X - u$ approximated by a *GPD*, for a suitable w , by Equation 3.12 the mean excess is:

$$E(X - w | X > w) = \frac{\sigma_w}{1 - \varepsilon}$$

At every higher threshold $u > w$, by Equation 3.11, mean excess can be defined as:

$$\begin{aligned} E(X - u|X > u) &= \frac{\sigma_u}{1 - \varepsilon} \\ &= \frac{\sigma_w + \varepsilon(u - w)}{1 - \varepsilon}, \quad \text{for } \varepsilon < 1 \end{aligned} \quad (3.22)$$

This implies that, mean excesses: $E(X - u|X > u)$ gives a linear function of u , when a desirable high threshold u has been attained.

The sample mean residual life plot points, which was drawn using

$$\left[\left(u, \frac{1}{N_u} \sum_{j=1}^{N_u} x_{(j)} - u \right) : u < x_{(max)} \right]$$

where;

$x_{(j)}$ is the observation over ‘ u ’, N_u is the number of observation over ‘ u ’ and $x_{(max)}$ is the largest of the observations $x_{(j)}$. If threshold ‘ u ’ is adequately high, then all excesses $u > w$ in the mean residual life plot changes directionally with u . This attribute potentially makes ways to deciding the threshold value. As soon as the sampling variability is included, the model threshold would be decided relative to the mean excess for the entire higher thresholds being linear. Therefore, in the conclusion of Coles (2001), the interpretation given to such plot becomes more complex.

3.4.3.2 Parameter Stability Plot

If GPD is true for excesses over threshold ‘ w ’ with ε and σ_w , then for higher threshold $u > w$, these excesses also adopt a GPD with ε which has a scale parameter given as:

$$\sigma_u = \sigma_w + \varepsilon(u - w)$$

By re-parameterising the scale parameter σ_u

$$\sigma^* = \sigma_u - \varepsilon u$$

Then, σ^* no more depends on ‘ u ’, because w is positioned at a threshold of reasonably high value (Coles, 2001). Parameter stability plot conforms GPD over specific values of the thresholds in contrast to scale and shape parameter. The model threshold is selected at the spot where the shape and the scale parameter remain fixed even upon considering sampling variability. As soon as the most suitable threshold is obtained, the exceedance follows a GPD.

3.4.3.3 Model Fit Diagnostic Plot

Probability plot, return level plot, quantile plot, and empirical versus fitted density comparison plot are part of standard statistical model diagnostic plots used in checking models fit as well as threshold choice suitability. Most of the checks are post-calculation diagnostic plots and are therefore based on a chosen threshold. This property makes it more a suitable alternative quantifying assess performance (Coles, 2001).

3.5 Conclusion

In this chapter, reviews of POT and BM models, GEVD, GPD, and the parameter estimation of both GEVD and GPD were presented. Also reviewed was the problem with extreme value modelling.

CHAPTER 4

STABLE DISTRIBUTION

4.1 Introduction

The aim of this chapter is to review the theoretical background of stable distributions. Section 4.2 describes stable distributions and their properties. In Sections 4.3, the parameter estimation of stable distributions is outlined.

4.2 Definition and Properties of Stable Distributions

Stable distributions fall in the class of probability laws with attractive theoretical and practical properties for economic models. This is the reason why their application in financial modelling is based on the fact that they generalize the normal (Gaussian) distribution having heavy tails and skewness, which are regular features in financial data (Nolan, 2003).

Stable distributions are distributions that keep their complete shape under addition. For instance; if Y, Y_1, Y_2, \dots, Y_m are independent and identical distributed stable random variables, then for all m .

$$Y_1 + Y_2 + \dots + Y_m \doteq a_m Y + b_m \quad (4.1)$$

where $a_m > 0$ and b_m are constants.

Equation 4.1 implies that the left hand side of the equation have the same distribution with right hand. This law is identified simply as strictly stable if $b_m = 0$ for every m . In other words, Y is a stable distributed random variable if for positive real numbers d_1 and d_2 there exist real numbers b and $a > 0$, such that

$$d_1 Y_1 + d_2 Y_2 \doteq a Y + b \quad (4.2)$$

Stable distributions can be classified into normal and non-normal stable distributions. A Normal stable distribution has a finite variance such as the normal distribution while a non-normal stable distribution has infinite variance which includes the Cauchy distribution and Levy distribution (Nolan, 2003).

4.2.1 Characteristic Function Representation

Broadly speaking, a stable distribution has neither a probability density function nor cumulative distribution function that can be expressed in a closed form but can be easily be described by its characteristic function of four parameters (Yang, 2012). These four parameters are the index of stability or the tail index, tail exponent or characteristic exponent (α), and skewness (β), scale (γ), and location (μ) parameters (Nolan, 2003; Borak et al., 2005). As a result of multiple parametric conditions for stable distribution, this led to different representations mix-up. Series of past developments in solving problems associated with analysing special forms of stable distributions led to generation of various formulas (Yang, 2012).

Here two different parameterizations will be described, which are $\mathbf{S}(\alpha, \beta, \gamma, \mu_0; 0)$ which refer to the 0- parameterization and $\mathbf{S}(\alpha, \beta, \gamma, \mu_1; 1)$ refer to as 1- parameterization. The parameters α, β , and γ are of the same meaning in the parameterizations, location parameter μ , is the only one different (Nolan, 2003).

Definition 1 Nolan $\mathbf{S}(\alpha, \beta, \gamma, \mu_0; 0)$: A random variable Y is described as from a stable distribution with parameters: α, β, γ , and μ_0 if it has characteristic function given as:

$$E(e^{izY}) = \begin{cases} \exp\{-\gamma^\alpha |z|^\alpha [1 + i\beta(\tan \frac{\pi\alpha}{2})(\text{sign } z)(|\gamma z|^{1-\alpha} - 1)] + i\mu z\}, & \alpha \neq 1 \\ \exp\{-\gamma |z| [1 + i\beta(\frac{2}{\pi})(\text{sign } z) \ln|\gamma z|] + i\mu z\}, & \alpha = 1 \end{cases} \quad (4.3)$$

Definition 2 Nolan $\mathbf{S}(\alpha, \beta, \gamma, \mu_1; 1)$: A random variable Y is described as from a stable distribution with parameters: α, β, γ , and μ_1 if it has characteristic function given as:

$$E(e^{izY}) = \begin{cases} \exp\{-\gamma^\alpha |z|^\alpha [1 - i\beta(\tan \frac{\pi\alpha}{2})(\text{sign } z)] + i\mu z\}, & \alpha \neq 1 \\ \exp\{-\gamma |z| [1 + i\beta(\frac{2}{\pi})(\text{sign } z) \ln|z|] + i\mu z\}, & \alpha = 1 \end{cases} \quad (4.4)$$

The location parameters are connected using

$$\mu_0 = \begin{cases} \mu_1 + \beta\gamma(\tan \frac{\pi\alpha}{2}), & \alpha \neq 1 \\ \mu_1 + \beta(\frac{2}{\pi})\gamma \ln \gamma, & \alpha = 1 \end{cases}$$

$$\mu_1 = \begin{cases} \mu_0 + \beta\gamma(\tan \frac{\pi\alpha}{2}), & \alpha \neq 1 \\ \mu_0 + \beta(\frac{2}{\pi})\gamma \ln \gamma, & \alpha = 1 \end{cases}$$

The four parameters in the characteristic function are described as follows:

- **Index of stability (α):** It defines the rate at which the tails of the distribution taper away. This exists in the range $0 < \alpha \leq 2$. The constant a_m indicated in equation 4.1 must be of the form $n^{1/\alpha}$. For $\alpha = 2$, it becomes the Gaussian distribution but β loses its influence. For $\alpha < 2$, the variance becomes infinite while the tails tend to be asymptotically equivalent to a Pareto law which means exhibiting power-law behaviour and the lower values of α express heavy tails (Nolan, 2003). When $\alpha > 2$, the mean of the distribution exists which is equal to μ , while for $\alpha < 1$, it means the stable distribution has no mean either.
- **Skewness (β):** This is expected to fall within $-1 \leq \beta \leq 1$. So for $\beta = 0$, it is said that the distribution is symmetric. In case $\beta > 0$, it is skewed to the right hand direction while for left skewness, $\beta < 0$. Therefore β and α are the parameters responsible for determining the shape of the distribution (Nolan, 2003).
- **Scale (γ):** This is responsible for width determination and can always be any positive real number.
- **Location (μ):** is identified as the shift of the mode of the density. It is expected to fall within $-\infty \leq \mu \leq \infty$. It shifts the distribution right if $\mu > 0$, and left if $\mu < 0$. A distribution is said to be standard stable when $\gamma = 1$ and $\mu = 0$ (Nolan, 2003).

It is of importance to mention that for $\beta = 0$, parameterization is similar. For $\alpha \neq 1$ and $\beta \neq 0$, there is a shift in parameterization by $\beta\gamma \tan \frac{\pi\alpha}{2}$ which increases towards infinity, as α tend towards 1. For instance, as α tend towards 1, the mode of $\mathbf{S}(\alpha, \beta, \gamma, \mu_1; 1)$ density approaches ∞ if $\text{sign}(\alpha - 1)\beta > 0$ or $-\infty$ if $\text{sign}(\alpha - 1)\beta < 0$ (Nolan, 2003).

When α is close to 1, working out probability density functions and cumulative distribution function in the numeric class range is usually difficult and the estimated parameters is undependable. With $\alpha = 1$, the 0-parameterization falls into a simple unit standard, unlike the 1-parameterization which is not. Considering the application to practical events, it is preferred to use $\mathbf{S}(\alpha, \beta, \gamma, \mu_0; 0)$ parameterization, because it uses continuously the four parameters. The reason for using $\mathbf{S}(\alpha, \beta, \gamma, \mu_1; 1)$ appears to be historic algebraic simplicity.

4.2.2 Stable Probability Density Function.

Stable random variables have probability density functions which are continuous and unimodal but do not have a closed form except for Normal, Cauchy and Levy distributions (Belov, 2005).

Normal or Gaussian Distributions: is a stable distribution with parameters $\alpha = 2$, $\beta = 0$, $\gamma = \frac{\sigma}{\sqrt{2}}$ and $\mu = 0$ for both 0-parameterization and 1-parameterization. Therefore, it is symmetric with finite variance. For normal distribution all moments exist. The probability density is given as:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right], \quad -\infty < y < \infty$$

Cauchy Distributions: is a stable distribution with parameters $\alpha = 1$, $\beta = 0$, γ and μ for both 0-parameterization and 1-parameterization. This implies that it is symmetric with infinite variance. For Cauchy distribution, not all moment exists. The probability density is given as:

$$f(y) = \frac{\gamma}{\pi[\gamma^2 + (y - \delta)^2]}, \quad -\infty < y < \infty$$

Levy Distributions: is a stable distribution with parameters $\alpha = 1/2$, $\beta = 1$, γ and $\mu = \gamma + \mu$ for 0-parameterization and $\alpha = 1/2$, $\beta = 1$, γ and μ for 1-parameterization. This implies that it is non-symmetric with infinite variance. Also, for the Levy distribution, not all moments exist. The probability density is given as:

$$f(y) = \frac{\sqrt{\gamma}}{\sqrt{2\pi}(y - \delta)^{\frac{3}{2}}} \exp\left[-\frac{\gamma}{2(y - \delta)}\right], \quad \delta < y < \infty$$

In general terms, linear amalgamations of individual stable laws having same α are described to be stable (Belov, 2005). For instance: Y_j is a stable random variable for any independent stable distribution j with parameters $\alpha, \beta_j, \gamma_j, \mu_j$ for any k -parameterization, where $j = 1, 2, \dots, m$, then;

$$d_1 Y_1 + d_2 Y_2 + \dots + d_m Y_m \sim \mathcal{S}(\alpha, \beta, \gamma, \mu; k) \quad (4.5)$$

where

$$\beta = \frac{[\sum_{j=1}^m \beta_j (\text{sign} d_j) |d_j \gamma_j|^\alpha]}{\sum_{j=1}^m |d_j \gamma_j|^\alpha}$$

$$\gamma^\alpha = \sum_{j=1}^m |d_j \gamma_j|^\alpha$$

and

$$\mu = \begin{cases} \sum_{j=1}^m \mu_j + \beta \gamma (\tan \frac{\pi \alpha}{2}), & k = 0, \alpha \neq 1 \\ \sum_{j=1}^m \mu_j + \beta (\frac{2}{\pi}) \gamma \ln \gamma, & k = 0, \alpha = 1 \\ \sum_{j=1}^m \mu_j, & k = 1 \end{cases}$$

One important part is that all α 's are the same and adding two stable variables which differs in α 's will not make the sum stable (Nolan, 2003). This result in generalised Equation (3.1) which means that different skewness, scales, and locations are allowed in the terms.

The stable density is supported in the domain $(-\infty, \infty)$ or a half-line. The half-line happens only if $0 < \alpha < 1$ or $\beta = \pm 1$. Moreover, according to Nolan (2003), the support of the stable

density for stable random variable Y given parameters $\alpha, \beta, \gamma, \mu$ with any K-parameterization is given as:

$$f(y|\alpha, \beta, \gamma, \mu; k) = \begin{cases} [\mu - \gamma(\tan \frac{\pi\alpha}{2}), \infty), & \alpha < 1, \beta = 1, k = 0 \\ (-\infty, \mu + \gamma(\tan \frac{\pi\alpha}{2})], & \alpha < 1, \beta = -1, k = 0 \\ [\mu, \infty), & \alpha < 1, \beta = 1, k = 1 \\ (-\infty, \mu], & \alpha < 1, \beta = -1, k = 1 \\ (-\infty, \infty), & \text{elsewhere} \end{cases}$$

Aside from the normal distribution, every stable distribution has heavy tail with an asymptotic Pareto. Samorodnitsky and Taquq (1994), affirm the asymptotic tail behaviour of stable distribution is Pareto when $\alpha < 2$ is given as:

$$\lim_{y \rightarrow \infty} P(Y > y) = y^{-\alpha} a_{\alpha} (1 - \beta) \gamma^{\alpha}$$

where $a_{\alpha} = \frac{\Gamma(\alpha) \sin(\frac{\pi\alpha}{2})}{\pi}$;

At the point when $\beta = -1$, the right tail decays speedier than any power. The left tail behaviour is comparable on the grounds $f(y|\alpha, -\beta, \gamma, \mu) = f(-y|\alpha, \beta, \gamma, -\mu)$, (Nolan, 1999).

4.3 Parameter Estimation of Stable Distributions

The fundamental problem of estimation in stable distributions is to estimate the four parameters α, β, γ , and μ . Many approaches have been used in estimating this basic problem; McCulloch (1986), proposed a quantile method, Ma and Nikias (1995), developed a fractional moment method, while sample characteristic function (SCF) method was introduced by Kogon and Williams (1998), which was a product of the foundation built by DuMouchel (1973b) on maximum likelihood (ML) estimation. Ojeda (2001), extensively compared these approaches where he concluded that ML estimates showed the most accurate measure or estimate. The second best is the SCF, followed by the quantile method and the moment method. This ML approach makes it easily for one to give large sample confidence intervals for the parameters, thus making it the more preferred method.

If X_1, X_2, \dots, X_n are independent and identically distributed stable samples, Nolan et al. (2001) defined the log likelihood function as:

$$\ell(\alpha, \beta, \gamma, \mu_0) = \sum_{i=1}^n \ln(X_i | \alpha, \beta, \gamma, \mu_0)$$

The lack of closed form formulas for general stable densities is a major challenge in evaluating this equation.

The suggested ML approaches reviewed are different in the selection of their approximating algorithm. At the same time, they have a close common characteristic which is, having the ML estimator to be asymptotically normal under specific conditions.

It should be mentioned here that, new innovative ML estimation techniques employ two methods which are; the direct integration method (Nolan et al., 2001) and the fast Fourier transform (FFT) method for approximating the stable pdf (Mittnik et al., 1999). The two approaches can be evaluated based on the efficiency terms while the types of approximation algorithms differentiate both.

Nolan (1997), suggested a stable program which establishes authentic computations of stable densities ranging between $\alpha > 0.1$ and any β , γ , and δ_0 .

4.4 Conclusion

Review of the properties of stable distribution, characteristic function representation and parameter estimation of stable distribution was achieved.

CHAPTER 5

PEARSON TYPE IV DISTRIBUTION (PIVD)

5.1 Introduction

The aim of this chapter is to review the theoretical framework of Pearson Type IV distribution. The mathematical concepts of probability density function (pdf) and the cumulative density function (cdf) of PIVD are reviewed in section 5.2 and section 5.3 respectively. Section 5.4 considers the parameter estimation based on maximum likelihood estimation.

5.2 The Probability Density Function (pdf) of PIVD

The Pearson system Pearson (1916) which is a generalization of the differential equation

$$\frac{f'(x)}{f(x)} = \frac{\alpha - x}{a_0 + a_1x + a_2x^2} \quad (5.1)$$

whose solution was given as (Stavroyiannis, 2013),

$$f(x) = (a_0 + a_1x + a_2x^2)^{-1/2a_2} \exp \left[\tan^{-1} \left(\frac{a_1 + 2a_2x}{4a_0a_2 - a_1^2} \right) \frac{a_1 + 2a_2x}{a_2\sqrt{4a_0a_2 - a_1^2}} \right]. \quad (5.2)$$

Depending on the coefficients of the a_i and $4a_0a_2 - a_1^2$ in Equation 5.2. The Pearson system provides any of the known distribution provided in Table 5.1.

Table 5:1: Family of distribution for the Pearson System

Distribution	Type
Normal distribution	0
Beta	I
Continuous uniform distribution	II
Chi-squared, Gamma, and Exponential distributions	III
Cauchy (or Lorentz, or Breit-Wigner)	IV
Inverse Gamma, and the inverse chi-squared distributions	V
F-distribution	VI
t-Student location scale distribution	VII
Monotonically decreasing power distribution	VIII

If $a_0 + a_1x + a_2x^2$ in Equation 5.1 is negative, then $4a_0a_2 - a_1^2$ in Equation 5.2 are real, thus rearranging Equation 6.2 resulting in Pearson type IV distribution given as (Nagahara, 1999); (Nagahara, 2007).

$$f(x) = k \left(1 + \left(\frac{x - \lambda}{a} \right)^2 \right)^{-m} \exp \left[-v \tan^{-1} \left(\frac{x - \lambda}{a} \right) \right] \quad (5.3)$$

where $a > 0$ is the scale parameter, λ is the location parameter, $m > 1/2$ controls the kurtosis, so that the normalization coefficient exist, v is the asymmetry of the distribution. The distribution is negatively skewed for $v > 0$ and positively skewed for $v < 0$ while for $v = 0$ reduces to the Student's t-distribution (Pearson Type VII) with $v = 2m - 1$, k is the normalization constant which is chosen in order to ensure that this function is a probability density function, is given as (Pearson, 1895; Nagahara, 2007). The expression for k is:

$$\begin{aligned}
k &= \frac{2^{2m-2} |\Gamma(m - iv/2)|^2}{\pi a \Gamma(2m - 1)} \\
&= \frac{\Gamma(m)}{\sqrt{\pi} a \Gamma(m - 0.5)} \left| \frac{\Gamma(m - iv/2)}{\Gamma(m)} \right|^2
\end{aligned} \tag{5.4}$$

As a result, the mean and variance of Pearson type IV is given as:

$$\mu = \lambda - \frac{av}{2m - 2}, \quad m > 1 \tag{5.5}$$

$$\sigma^2 = \frac{a^2}{2m - 3} \left[1 + \frac{v^2}{4(m - 1)^2} \right], \quad m > 3/2 \tag{5.6}$$

5.3 The Cumulative Density Function (cdf) of Pearson Type IV Distribution.

The cumulative distribution defined as:

$$F(x) = \int_{-\infty}^x f(t) dt$$

Therefore the cdf of Pearson type IV distribution is given as:

$$F(x) = \int_{-\infty}^x k \left(1 + \left(\frac{t - \lambda}{a} \right)^2 \right)^{-m} \exp \left[-v \tan^{-1} \left(\frac{t - \lambda}{a} \right) \right] dt \tag{5.7}$$

According to Heinrich (2004) the cdf of Pearson type IV that is F(x) in Equation (5.7) can be expressed in terms of the hyper-geometric function; See Appendix A

5.4 Parameter Estimation of Pearson Type IV Distribution

As at the time Karl Pearson's classes of distribution was developed, the necessity for maximum likelihood methodology was not identified. He based his approach on method of moments, later

proven to be inadequate. The method of moment is usually applied in setting good foundation in fitting a maximum likelihood. In that case, setting the preliminary estimates in this manner is less important. It is discovered that in few cases where the moment method estimates are inadequate for starting the maximum likelihood approach, then using a mere typical parameter will be more proficient.

Let X_1, \dots, X_n be independent and identically distributed values obtained in a Pearson Type IV distribution. Then, the likelihood function is given as

$$\begin{aligned}
 L(m, v, a, \lambda) &= \prod_{j=1}^n f(m, v, a, \lambda; x_j) \\
 &= \prod_{j=1}^n k \left(1 + \left(\frac{x_j - \lambda}{a} \right)^2 \right)^{-m} \exp \left[-v \tan^{-1} \left(\frac{x_j - \lambda}{a} \right) \right]
 \end{aligned} \tag{5.23}$$

Then, the log-likelihood is given by

$$\begin{aligned}
 \ell(m, v, a, \lambda) &= \ln L(m, v, a, \lambda) \\
 &= \sum_{j=1}^n \ln f(m, v, a, \lambda; x_j) \\
 &= n \ln k - v \sum_{j=1}^n \tan^{-1} \left(\frac{x_j - \lambda}{a} \right) - m \sum_{j=1}^n \ln \left(1 + \left(\frac{x_j - \lambda}{a} \right)^2 \right)
 \end{aligned} \tag{5.24}$$

where p is the number of observed data points x_j . The parameter estimates of $m, v, a,$ and λ is obtained by minimizing the negative log likelihood of Equation 5.24 which can be done numerically.

5.5 Conclusion

This chapter successfully reviewed the pdf of PIVD, cdf of PIVD and the parameter estimation of PIVD.

CHAPTER 6

METHODOLOGY

6.1 Introduction

The aim of this chapter is to give an overview of the applied methods used in the thesis. Section 6.2 provides a brief description of stationarity test. The approach for the Value-at-risk (VaR) model employed in this analysis is provided in Section 6.3 Section 6.4 describes the criteria used for selecting the best model. Section 6.5 describes the method employed in model diagnostics. Section 6.6 discusses the backtesting method employed in this study.

6.2 Test for stationarity

It is necessary to ensure that the data are a stationary form before analysis, so that the statistical properties of the data are constant over time. In this study, the unit root test is used to check if the data is stationary or not. The unit root test employed are:

- Augmented Dickey-Fuller (ADF) test
- Phillips-Perron (PP) test
- Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test

6.2.1 Augmented Dickey-Fuller test

The ADF is used to accommodate ARMA (p, q) model with unknown orders. It is based on the regression equation

$$X_t = \phi X_{t-1} + \sum_{j=1}^p \alpha_j \Delta X_{t-1} + Z_t \quad (6.1)$$

where the error Z_t is assumed to be homoscedastic.

The AR(p) in equation 6.1 is used to get rid of the serial correlation in the errors, by setting the value of the lag p of the difference term ΔX_{t-1} large enough to allow the approximation of the

ARMA(p, q) process such that the errors Z_t are serially uncorrelated. Based on the regression equation estimate the test statistics is given as

$$ADF_t = t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})} \quad (6.2)$$

While the normalized bias statistics is

$$ADF_n = \frac{T(\hat{\phi} - 1)}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p} \quad (6.3)$$

Under the null hypothesis that the data has a unit root (that is $\phi = 1$) against an alternative hypothesis that the data is stationary (that is $\phi > 1$).

6.2.2 Phillips-Perron (PP) test

The PP test is based on the regression equation

$$X_t = \phi X_{t-1} + Z_t \quad (6.4)$$

where Z_t is assumed to be stationary and heteroscedastic.

The PP test is used to correct for any serial correlation and heteroskedasticity in the errors Z_t in the regression Equation 6.4 by applying the modified test statistics given as

$$P_t = \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \right)^{1/2} t_{\phi=1} - \frac{1}{2} \left(\frac{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}{\hat{\sigma}_2^2} \right) \left(\frac{T - SE(\hat{\phi} - 1)}{\hat{\sigma}_1^2} \right) \quad (6.5)$$

and

$$P_{\hat{\phi}} = T_{\hat{\phi}} - \frac{1}{2} \frac{T^2 SE(\hat{\phi} - 1)}{\hat{\sigma}_1^2} (\hat{\sigma}_2^2 - \hat{\sigma}_1^2) \quad (6.6)$$

where variance parameters $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are given as

$$\sigma_1^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(Z_t^2)$$

$$\sigma_z^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^T E \left(\frac{S_T^2}{T} \right)$$

and $S_T = \sum_{t=1}^T Z_t$

Under the null hypothesis that $\phi = 1$, the modified test statistics P_t and $P_{\hat{\phi}}$ have the same asymptotic distributions as the ADF t-statistic and normalized bias statistics.

6.2.3 Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test

The KPSS is based on the model

$$X_t = \varepsilon_t + Z_t$$

where Z_t is assumed to be stationary and heteroskedastic and ε_t is a random walk given as

$$\varepsilon_t = \varepsilon_{t-1} + v_t$$

and with v_t is independent and identically distributed with mean zero and variance σ_v^2 . For $\sigma_v^2=0$, then ε_t is a constant for all value of t and X_t is stationary. Thus, using the regression equation

$$X_t = \hat{\mu} + \hat{Z}_t$$

The test statistics is

$$KPSS = \frac{\sum_{t=1}^T \hat{S}_t^2 / T^2}{\hat{\lambda}^2} \tag{6.7}$$

The test statistic is the Lagrange multiplier (LM) for testing the null hypothesis of $\sigma_v^2 = 0$ against the alternative hypothesis of $\sigma_v^2 > 0$

6.3 The VaR Model: Combining the GARCH type model and heavy-tailed distribution

The VaR model used in this thesis is formed by combining the GARCH-type model with the heavy-tailed distribution. The heavy tailed distributions considered are: Generalized Pareto Distribution (GPD), Generalized Extreme Value Distribution (GEVD), stable distribution, and Pearson Type IV distribution (PIVD). This method is similar to that of McNeil and Frey (2000) and Bhattacharyya et al. (2008). The VaR for the long and short position is considered. As indicated in the introduction, VaR at the long position is associated with the left side of the distribution of the returns corresponding to negative returns. It is the left quantile of the distribution. In finance, traders at the long position incur a loss when prices drop. These are traders buying a particular equity. Correspondingly, the VaR at the short position is associated with the right side of the distribution of the returns corresponding to positive returns. It is the right quantile of the distribution. In finance, traders at the short position incur a loss when prices increase. These are traders selling a particular equity.

Mathematically, VaR is defined as the q th quantile of the distribution F . Thus, VaR at the long position is given as

$$\text{VaR}_q = F^{-1}(q)$$

where F^{-1} is the inverse of F called quantile function and $0 < q < 1$.

Likewise, VaR at the short position is given as

$$\text{VaR}_{1-q} = F^{-1}(1 - q)$$

Let P_t be the stock price on day t . A 1-day VaR at the long position on day t is the solution to

$$P(P_t - P_{t-1} \leq \text{VaR}_q^t) = q$$

where t is used to indicate a varying time.

Similarly, a 1-day VaR at the short position on day t is defined as

$$P(P_t - P_{t-1} \geq VaR_q^t) = q$$

Traders at the long position incur a loss when $\Delta P_t = P_t - P_{t-1} < 0$, while traders at the short position incur a loss when $\Delta P_t = P_t - P_{t-1} > 0$. Considering the value at risk for returns series, 1-day log returns on day t is defined as

$$r_t = \log(P_t) - \log(P_{t-1})$$

Given that the past information at pre-set time $t - 1$ is F_{t-1} . Thus, the VaR of a returns series is given as

$$P(r_t \leq VaR_q^t | F_{t-1}) = q$$

The process of the stock returns is modelled as follows

$$r_t = \mu_t + Z_t$$

With conditional mean of the r_t given F_{t-1} as; $\mu_t = E(r_t | F_{t-1})$ and conditional variance of the r_t given F_{t-1} as; $\sigma_t^2 = var(r_t | F_{t-1})$. And if the standardized residuals of the returns model is fitted to a heavy-tailed distribution; then, the 1- day ahead VaR at day t for a given probability level is given as:

$$VaR_q^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} F^{-1}(q) \quad (6.8)$$

for the long position. The VaR the short position is given as:

$$VaR_q^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} F^{-1}(1 - q) \quad (6.9)$$

where μ_{t+1} is the conditional mean forecast at time $t + 1$, σ_{t+1} is the conditional variance forecast at time $t + 1$ and, $F^{-1}(\cdot)$ is the VaR for the given heavy tailed distribution at a given probability level.

6.3.1 Step-by-step method

The step-by-step method for this research procedure is described in Figure 6.1 as follows:

- The first step is to choose the best GARCH-type model that adequately capture the properties of the all share returns. The selection of best GARCH-type model is based on model selection criteria.
- Then, the best GARCH-type model is fitted to the all share returns by Gaussian quasi maximum-likelihood estimation. That is, the log-likelihood is maximized by assuming a normal distribution innovation.
- The standardized residuals of this best GARCH-type model is extracted and fitted to the four heavy-tailed distributions.
- Finally, the forecast VaR is computed based on Equation 6.8 and 6.9 and compared by backtesting.

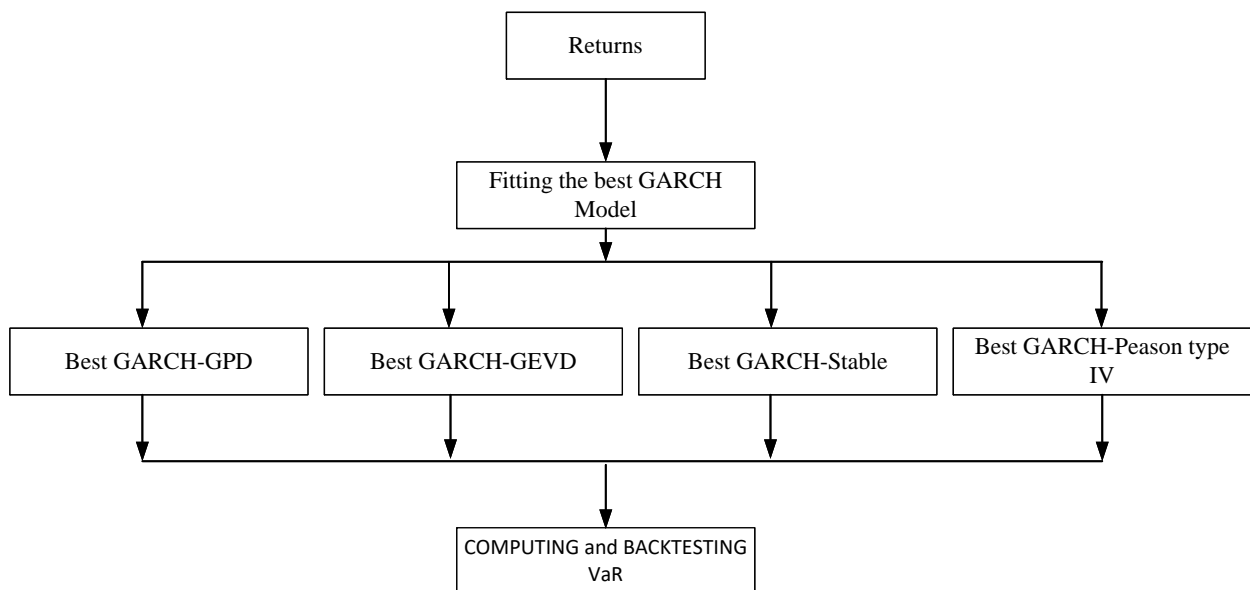


Figure 6. 1: Step-by-step method.

6.4 Model selection criteria

Model selection criteria are used to select the best model from candidate GARCH-type to describe the JSE all share returns. It is a useful tool for assessing if a fitted model provides an

optimal balance between parsimony and goodness-of-fit. It helps to find the best GARCH-type model that is either too simple or too complex to accommodate the JSE all share returns. The model selection criteria employed in this thesis are: Akaike information criterion (AIC) and Bayesian information criteria or Schwarz-Bayesian criteria (BIC).

6.4.1 The Akaike Information Criterion (AIC)

AIC measures how well the evaluated model fits with the data in respect to candidate models. Given a GARCH-type models of different structures, each model is fitted to the JSE all share returns using Gaussian quasi maximum likelihood. The AIC is computed for each model (Akaike, 1974).

$$AIC = -\frac{2}{k}\log(\text{likelihood}) + \frac{2}{k}(p + q) \quad (6.10)$$

where k is the sample size and $p + q$ is the number of parameters in the model. The model with the smallest number of parameters and with the largest likelihood has the minimum AIC. The model with the smallest AIC is regarded as the best model for the data (Tsay, 2013).

6.4.2 The Bayesian Information Criterion (BIC)

BIC is concerned with the Bayes factor. The BIC of a model is given as:

$$BIC = -2\log(\text{likelihood}) + [(p + q) + (p + q)\log k] \quad (6.11)$$

where k is the sample size and $p + q$ is the number of parameters in the model. BIC allows comparison of multiple models by penalising complex models (model with many parameters) relative to simpler models. The model that has the largest posterior probability has the minimum BIC and is regarded as the best model for the data.

6.5 Model Diagnostics

The model diagnostic is important for checking for any possible model inadequacy. It is divided into sections which include: test for serial correlation and ARCH effect, test for leverage effect, and the test for goodness-of-fit.

6.5.1 Test for Serial Correlation and ARCH effect

The standardized residuals of the GARCH-type model are assumed to be independent and identically distributed therefore, if the model is adequate the standardized residuals are expected not to exhibit autocorrelation (serial correlation) and conditional heteroskedasticity (ARCH effect). In this study, the autocorrelation plot and the partial autocorrelation plot are used to check for the presence of autocorrelation and conditional heteroskedasticity (ARCH effect). Then, formal tests are also applied which include: Ljung-Box test and ARCH test.

6.5.1.1 Autocorrelation plot and Partial autocorrelation plot

These are the graphical techniques used to examine if the standardized residuals exhibit serial correlation and conditional heteroskedasticity. The plot of autocorrelation function (ACF) of the standardized residuals against the lags and the plot of the partial autocorrelation function (PACF) of the standardized residuals against the lags are used to assess the presence of serial correlation in the standardized residuals. While, the plot of ACF of the squared standardized residuals against the lags and the plot of PACF of the squared standardized residuals against the lags are used to check for conditional heteroskedasticity. The plots also include a middle horizontal reference line at zero and the confidence bands at 95%. If the autocorrelation or partial autocorrelation at several lags fall outside the 95% confident bands, then they are said to exhibit serial correlation or conditional heteroskedasticity.

6.5.1.2 Ljung-Box test

This test is used for both autocorrelation and conditional heteroskedasticity. It is based on the sample autocorrelation of the standardized residuals. When the Ljung-Box test is used to test whether the first L lags of the ACF of the standardized residuals are zero, the test statistic is

$$Q(L) = N(N + 2) \sum_{k=1}^L \frac{\hat{\rho}_k^2(z)}{N - k} \quad (6.12)$$

where N is the sample size, L is the total number of autocorrelation, $\hat{\rho}_k^2$ is the squared sample autocorrelation of standardized residual (z_t) at lag k .

The test statistic $Q(L)$ follows a chi-squared with L degree of freedom under the null hypothesis of no serial correlation. The null hypothesis is rejected if $Q(L) > \chi_{1-\alpha}^2$ where $\chi_{1-\alpha}^2$ is the $1-\alpha$ quantile of the chi-squared distribution with L degrees of freedom or if p -value is less than α , significant level. It is also used to test whether the first L lags of the ACF of the squared standardized residuals are zero, the test statistic is

$$Q = N(N + 2) \sum_{k=1}^L \frac{\hat{\rho}_k^2(z^2)}{N - k} \quad (6.13)$$

It also follows a chi-squared with L degree of freedom, but under the null hypothesis of no conditional heteroskedasticity (that is, no ARCH effect). The null hypothesis of no ARCH effect is rejected if $Q(L) > \chi_{1-\alpha}^2$ or if p -value is less than α , significant level.

6.5.1.3 Lagrange multiplier ARCH (ARCH-LM) test

It is also used to examine the presence of ARCH effect in the standardized residuals based on the linear regression.

$$z_t^2 = \alpha_0 + \alpha_1 z_{t-1}^2 + \dots + \alpha_m z_{t-m}^2 + e_t \quad (6.14)$$

where $t = m + 1, \dots, T$

The test statistic is given as:

$$LM = TR^2 \quad (6.15)$$

where T is the sample size, R = is the sample multiple correlation coefficient obtained from the regression Equation 6.14 using estimated residuals.

The test statistic follows a chi-squared with m degrees of freedom under the null hypothesis of no ARCH effect. It should also be noted that the test for serial correlation and ARCH effect can also be applied to the returns before estimation in order to examine if the time series exhibit autocorrelation and conditional heteroskedasticity. The purpose of this is to decide if there is a need for GARCH-type model.

6.5.2 Test for Leverage effect

In this study, Engle and Ng test known as sign and size test is employed to test for leverage effect. The test is used to determine whether an asymmetric GARCH-type model is needed for the ALSI returns or whether the symmetric GARCH model is adequate for the given returns. The test is applied to the standardized residuals of the GARCH model fit to the returns. Engle- Ng test is a joint test of sign and size bias which is based on regression.

$$\hat{\mu}_t^2 = \beta_0 + \beta_1 S_{t-1}^- + \beta_2 S_{t-1}^- \mu_{t-1} + \beta_3 S_{t-1}^+ \mu_{t-1} + e_t \quad (6.16)$$

where e_t is the independent and identically distributed error term, S_{t-1}^- is an indicator dummy variable given as:

$$S_{t-1}^- = \begin{cases} 1, & \mu_{t-1} < 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{and } S_{t-1}^+ = 1 - S_{t-1}^-.$$

The coefficient $\beta_i, i = 1,2,3$ follow student-t distribution. If β_1 is significant, it indicates the presence of sign bias implying that positive and negative shocks impact differently upon future volatility. If β_2 or β_3 is significant, it indicates the presence of size bias, meaning the size of the negative or positive affect the response of volatility from being symmetric. The joint test statistic is

$$TR^2$$

where, R is based on the regression Equation 6.15 and T is the sample size.

The test follows a chi-squared distribution with 3-degree of freedom under the null hypothesis of no asymmetric effect.

6.5.3 Test for Independent and Identically Distribution (IID)

If the GARCH-type model has successfully captured linear dependence in the returns, then the standardized residuals are expected to be IID or random. Therefore, this study employs the following test:

- Bartels' rank test
- Cox and Stuart test
- Brock, Dechert and Scheinkman test

6.5.3.1 Bartels' rank test

The Bartels' rank test is based on the rank of standardized residuals in ascending order. The ranks are sequential number of Z_i : Rank (Z_i). All the possible set of rank arrangement of standardized residuals is given as $N!$. Under the null hypothesis of randomness each rank arrangement is equally likely to occur. The test statistic is given as

$$NM = \sum_{i=1}^{N-1} (\text{rank}(Z_i) - \text{rank}(Z_{i+1}))^2 \quad (6.17)$$

For large sample size, test statistics is

$$RVM = \frac{\sum_{i=1}^{N-1} (\text{rank}(Z_i) - \text{rank}(Z_{i+1}))^2}{n(n^2 - 1)/12} \quad (6.18)$$

6.5.4.2 Cox and Stuart test

The Cox-Stuart test is based on trend. Given a set of standardized residuals Z_1, \dots, Z_t which are assumed to be arranged in the order of occurrence. They are grouped into pairs of the following form

$$(Z_1, Z_{1+k}), (Z_2, Z_{2+k}), \dots, (Z_{t-k}, Z_t)$$

So that

$$k = f(x) = \begin{cases} \frac{t}{2}, & \text{if } t \text{ is even} \\ \frac{t+1}{2}, & \text{if } t \text{ is odd} \end{cases}$$

A sign test is then computed by defining

$$\text{sign}(Z_i, Z_{i+k}) = \begin{cases} +, & \text{if } Z_i < Z_{i+k} \\ 0, & \text{if } Z_i = Z_{i+k} \\ -, & \text{if } Z_i > Z_{i+k} \end{cases} \quad (6.19)$$

To test if $P(Z_i < Z_{i+k}) = P(Z_i > Z_{i+k})$, the tied pairs $Z_i = Z_{i+k}$ are omitted. Let

T = total number of +'s

N = total number of +'s and -'s

Under the null hypothesis of no trend, $T \sim \text{Bin}(N, 1/2)$. Hence, the null hypothesis is rejected at a given level of significance if $T \leq m$ or $T \geq N - m$, where m is chosen such that

$$\sum_{j=0}^m \binom{N}{j} (1/2)^n \approx \frac{\alpha}{2}$$

While for large N ,

$$\left(\frac{T}{N}\right) \sim N\left[\frac{1}{2}, \frac{N}{4}\right]$$

Hence, the null hypothesis is rejected if

$$T \leq \frac{1}{2} - Z_{\alpha/2} \sqrt{\frac{1}{4N}} \quad \text{or} \quad T \geq \frac{1}{2} + Z_{\alpha/2} \sqrt{\frac{1}{4N}}$$

6.5.3.4 Brock, Dechert and Scheinkman (BDS) Test

The BDS test is used to examine if the standardized residuals are IID. The standardized residuals are embedded into m - dimensional vectors by taking each m successive point in the series such that

$$Z_t^m = (Z_t, Z_{t+1}, \dots, Z_{t+m-1})$$

The correlation integral with m dimensions and distance ε is given as

$$C_{m,\varepsilon} = \lim_{T \rightarrow \infty} \binom{T-m}{2}^{-1} \sum_{\forall s < t} I_\varepsilon(Z_t^m, Z_s^m)$$

where

$$I_\varepsilon(Z_t^m, Z_s^m) = \begin{cases} 1, & \|Z_t^m - Z_s^m\| \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

The correlation integral is used to measure the proportion of pairs of any m -vectors (Z_t^m and Z_s^m) with the distance ε . If the standardized residuals are IID, the Z_t^m will show pattern in the m -dimensional space, that is

$$C_{m,\varepsilon} \approx [C_{1,\varepsilon}]^m$$

The BDS test statistic is

$$BDS_{m,\varepsilon} = \frac{\sqrt{N}(C_{m,\varepsilon} - [C_{1,\varepsilon}]^m)}{\sqrt{V_{m,\varepsilon}}} \quad (6.20)$$

where $V_{m,\varepsilon}$ is a consistent estimator of the asymptotic standard deviation, $\sigma_{m,\varepsilon}$ of $\sqrt{N}(C_{m,\varepsilon} - [C_{1,\varepsilon}]^m)$.

Under the null hypothesis that the standardized residuals are independent and identically distributed, $BDS_{m,\varepsilon}$ follows a normal distribution with mean 0 and variance 1. Hence, the null

hypothesis is rejected, if the test statistic is greater than or less than the critical values (e.g. if $\alpha=0.05$, the critical value = ± 1.96)

6.5.4 Test for Normality and Goodness-of-fit

Test for goodness of fit is necessary to examine how well the standardized residuals fit a given distribution. Jarque-Bera (JB) Test, Shapiro-Wilk Test, Anderson-Darling test (AD test), probability plot and quantile-quantile plot were used for the goodness-of-fit test.

6.5.4.1 Jarque-Bera (JB) Test

The JB test is used to examine if the standardized residuals have skewness and kurtosis corresponding to that of normal distribution. It tests whether the skewness and excess kurtosis of the standardized residuals are zero. The test statistic is

$$JB = n \left[\frac{S^2}{6} + \frac{K^2}{24} \right] \quad (6.21)$$

where S is the skewness parameter given as

$$S = \frac{1}{n} \frac{\sum_{i=1}^n (Z_i - \bar{Z})^3}{(\hat{\sigma}^2)^{3/2}}$$

and K is the excess kurtosis parameter given as

$$K = \frac{1}{n} \frac{\sum_{i=1}^n (Z_i - \bar{Z})^4}{(\hat{\sigma}^2)^2} - 3$$

Under the null hypothesis of standardized residuals are normally distributed, JB follows a chi-square distribution with two degree freedom. Hence, the null hypothesis is rejected if $JB > \chi_{\alpha,2}^2$ or p -value is less than the given significant level.

6.5.4.2 Shapiro-Wilk Test

The Shapiro-Wilk test is used to test if the standardized residuals follows a normal distribution. The test statistics is given as

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})} \quad (6.22)$$

where $x_{(i)}$ are the i th ordered statistics, \bar{x} is the sample mean, a_i are constants obtained from the means, variances and covariance of the order statistics of a sample of size n from a normal distribution.

6.5.3.2 Anderson-Darling test (AD test)

The AD test is also to test if the standardized residuals came from a population with specified distribution. This test is based on the difference between an observed CDF and the expected CDF. The test statistic is

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N (2i - 1) \{ \ln F(Y_i) + \ln(1 - F(Y_{N+1-i})) \} \quad (6.23)$$

The test statistic, A , is defined under the null hypothesis of standardized residuals follow the specified distribution. It is a one-sided test and the null hypothesis is rejected, if A is greater than the critical value (given by the table of AD test) or p-value is less than α , significant level.

6.5.3.4 Probability Plot (PP plot)

The PP plot is a graphical method for examining if standardized residuals follow a given distribution. The empirical distribution of the data is plotted against the specified theoretical distribution in such a way that the points should form approximately a straight line. If there is a departure from the straight line, then it indicates the departure of the data from the specified distribution, implying that that the data does not follow the specified distribution.

6.5.3.5 Quantile-Quantile plot (Q-Q plot)

It is a graphical method for assessing if the empirical distribution and the specified theoretical distribution come from population with common distribution. It is a plot of the quantiles of the empirical distribution against the quantiles of theoretical distribution. A qq-line is also plotted. A qq-line is a 45-degree reference line that represents a perfect match between the empirical and theoretical distribution. The departure from the qq-line indicates that the empirical distribution and the theoretical distribution come from population of different distribution.

6.6 Backtesting procedure

In order to analyse the predictive ability of the model, the data is divided into two periods: the in-sample periods and out-of-sample period. The in-sample period, which is from 20-May-2005 to 31-Dec-2013, is used for the model estimation and for forecasting risk. The out-of-sample period from 2-Jan-2014 to 31-May-2016 is used for testing Value-at-risk (VaR) forecast. As a result, the estimation window has about 2155 observations, the testing window has 602 observations, and thus the total observations are 2757. The in-sample and the out-of-sample backtesting is conducted. The in-sample backtesting is used to check for the adequacy of the VaR estimate obtained from the heavy-tailed distributions. This is done by backtesting on the standardized residuals of the in-sample period which consists of 2155 observations. The out-of-sample backtesting: is used to test for adequacy and predictive ability of the VaR models. The VaR model is used to compute an out-of-sample forecast for the 602 observations sequentially. The period $[t + 1; t + h]$ is used for the VaR forecast where $h = 1 \text{ day}$ is the time horizon of the VaR forecasts. The one-day-ahead VaR (both for long and short positions) is then compared with the 602 observed returns by employing the statistical test known as Kupiec likelihood ratio test.

6.6.1 Kupiec Likelihood Ratio Test

The backtesting method employed in this study is the Kupiec test. Kupiec test is used to examine the frequency of losses along the tail. It is based on the fact that a sufficient model should have a boundary exceedance of VaR estimates around the matching tail probability level

(Baharul-Ulum et al., 2012). The approach allows computing the probability of observed returns x in sample size N around the VaR estimate level β , by using the binomial distribution.

$$f(x) = \binom{N}{x} \beta^x (1 - \beta)^{N-x}$$

The test evaluates the operation of a VaR model while also assuming independence. Kupiec (1995), then recommends an examination that makes use of likelihood ratio so as to create a concession between type 1 and 2 errors. The null hypothesis for Kupiec test is that the expected proportion of exceedances is equal to β with a test statistic (Chinhamu et al., 2015);

$$LR_{uc} = 2 \ln \left[\frac{\left(1 - \frac{x}{N}\right)^{N-x} \left(\frac{x}{N}\right)^x}{(1 - \beta)^{N-x} \beta^x} \right] \approx \chi^2(1) \quad (6.24)$$

Therefore, we reject the null hypothesis if the expected proportion of exceedances x is less than LR_{uc} , (Dowd, 2005).

6.7 Conclusion

This chapter discussed the methods and various statistical tests employed in carrying out the analysis for this study.

CHAPTER 7

EMPIRICAL RESULTS

7.1 Introduction

In this chapter, data source and description are reported. GARCH type models combined with heavy-tailed distribution are fitted and VaR estimates were backtested using the Kupiec likelihood test for ALSI returns. Lastly robust VaR model is selected. Section 7.2 describes the characteristics of ALSI returns. In section 7.3, how APARCH (1,1) model was chosen as the best possible GARCH-type model was justified. The combination of APARCH (1,1) with heavy-tailed distribution are discussed in section 7.4. Section 7.5 considers the VaR estimate. Finally, section 7.6 and 7.7 discuss the backtesting result for the in-sample and out-sample respectively.

7.2 Data Description

The data is made up of the daily closing price of the all share index (ALSI) from 20-May-2005 to 31-May-2016 obtained from INET. Figure 7.1(a) shows the time series plots of both the daily closing price of JSE ALSI and ALSI log returns.

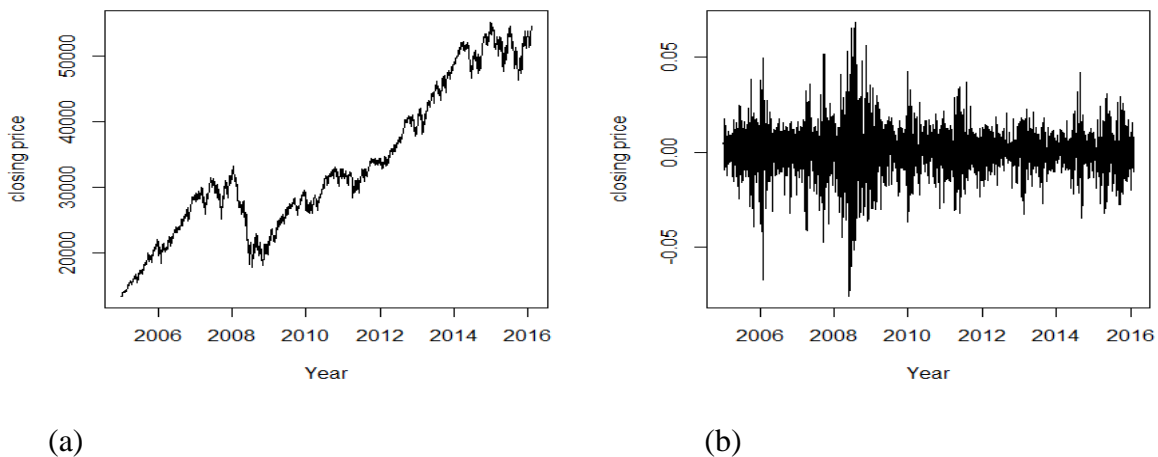


Figure 7. 1: Time series plots

**(a) Time series plot of the daily ALSI index, (b) Time series plot of the daily ALSI log returns.*

From Figure 7.1, it is observed that the daily ALSI index data does not seem to be stationary both in mean and in variance. The data seems to exhibit a stochastic trend, and it suggests the presence of heteroskedasticity. Therefore, we obtain the daily log returns (r_t). The log returns is given by

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

where,

r_t , is the natural logarithmic return of daily price of ALSI at time t

P_t is the daily closing price of ALSI at time t

P_{t-1} is the daily closing price of ALSI at time $t - 1$

The time series plots of daily ALSI log returns in Figure 7.1 (b), shows that log returns appear to be stationary. However, the variance appears not to be constant over time indicating volatility clustering. In order to confirm the stationarity of the JSE ALSI returns, the unit root tests are employed, namely: Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) Unit root test, and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test. The unit root test statistics with their corresponding p -values are presented in Table 7.1.

Table 7. 1: Test for Stationarity for the ALSI returns

Test	Statistics	P -value
ADF	-13.6101	0.0100
PP	-2457.7280	0.0100
KPSS	0.1418	0.1000

From Table 7.1 the ADF and the PP test statistics have a p -value which are less 0.05, therefore the null hypothesis of non-stationarity of ALSI returns is rejected at the 5 % level of significance. While the KPSS test statistics has a p -value which is greater than 0.05, therefore the null hypothesis of stationarity of the ALSI returns is not rejected at 5% level of significance. In conclusion, the results of ADF, PP and KPSS tests confirm that the ALSI returns are stationary. The descriptive statistics of the ALSI returns are reported in Table 7.2.

Table 7. 2 Descriptive statistics of the daily JSE ALSI returns

N	Mean	Minimum	Maximum	Medium	Skewness	Excess Kurtosis
2757	0.0005	-0.0758	0.0683	0.0010	-0.1922	3.5236

Table 7.2 indicate that the JSE daily ALSI returns consists of 2757 observations ranging from -0.0758 to 0.0683 with an average of -0.0005 and a medium 0.001. The average is quite large, implying that the overall returns were slightly on the increase. A positive skewness is also known as, right-skewed indicates a distribution with asymmetric side which tends in the direction of more positive numbers. On the other hand, negative skewness, also left-skewed expresses a distribution that tends asymmetrically in the left direction. In this case, as reported in the Table 7.2, the skewness is negative and indicating that losses (left tail) of JSE ALSI returns is larger than the profits (right tail). Kurtosis is a measure of its levelness in comparison to the frequency distribution peak, which is actually the values of excess kurtosis with reference to Amir (1993) who described a positive kurtosis as one that demonstrates a moderately peaked distribution, also called leptokurtic, while a negative kurtosis is viewed as a comparatively flat distribution, also known as Platykurtic. From Table 7.2, it is observed that the excess kurtosis value of JSE ALSI returns is positive indicating leptokurtic behaviour of the returns. The kurtosis value is 3.5236 which is greater than 3. It suggests that the empirical distributions of the daily JSE ALSI returns have a much heavier tail than that of normal distribution. This suggests that the JSE ALSI returns follow a heavy tailed distribution. To check for the non-normality of the returns, the Q-Q plot and numerical normality test, namely: Jarque-Bera (JB) test and Shapiro-Wilk tests are employed. Figure 7.2 shows the Q-Q plot of the daily JSE ALSI returns.

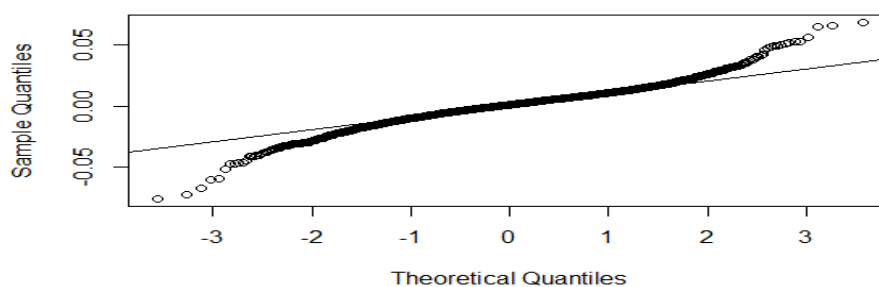


Figure 7. 2: Q-Q plot of JSE ALSI returns

From Figure 7.2, the Q-Q plot indicates that ALSI returns seem to diverge from the normal distribution at both tails of the distribution. This is confirmed by the Jarque-Bera (JB) and Shapiro-Wilk test statistics reported in Table 7.3.

Table 7. 3: Test for normality of the ALSI returns

Test	Statistics	p -value
Shapiro-Wilk	0.9591	< 0.0001
Jarque-Bera	1447.1069	< 0.0001

From the Table 7.3, both test statistics have a p -value less than 0.05, thus we reject the null hypothesis of normality at 5% significance level. This confirms that ALSI returns are not normality distributed. Thus, there is strong evidence of modelling the JSE ALSI returns incorporating a heavy tailed distribution. This study also tests for serial correlation in ALSI returns. The Autocorrelation (ACF) plot, Ljung-Box statistics $Q(L)$ of the returns and the Dublin Watson (DW) test are used to investigate serial correlation of ALSI returns. The ACF plot of JSE ALSI returns is shown in Figure 7.3(a).

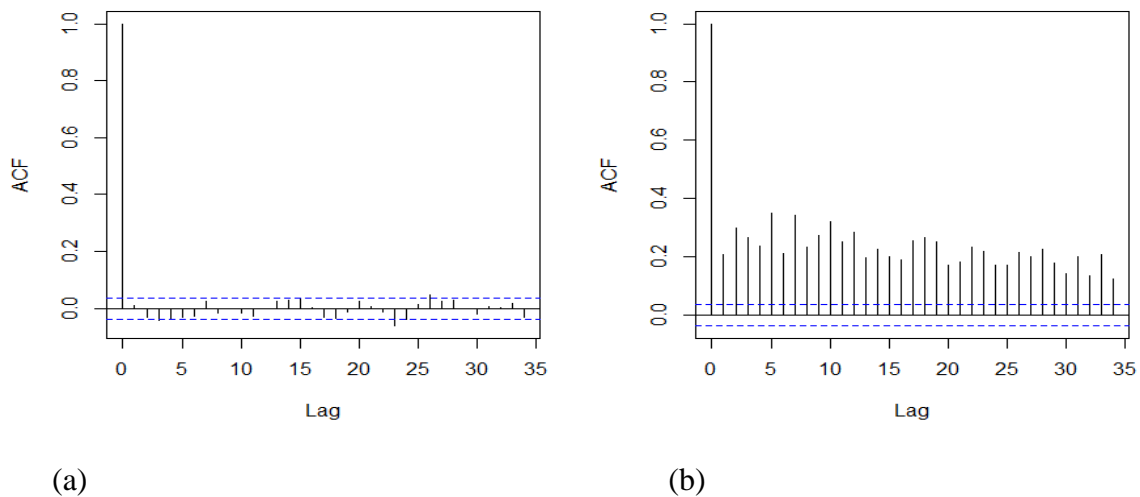


Figure 7. 3: ACF plots of ALSI returns
 *(a) ACF plot of ALSI returns (b) ACF plot of squared ALSI returns

From Figure 7.3(a) ACF plot of JSE ALSI returns does not have any significant spike at any lag, this suggests that the JSE ALSI returns are not serially correlated. Table 7.4 shows the p -value of the Ljung-Box and the DW statistics.

Table 7. 4: Test for serial correlation in the ALSI returns

Test	Statistics	<i>p</i>-value
Ljung-Box	0.65047	0.4199
DW test	1.9792	0.2920

From Table 7.4, Ljung-Box $Q(L)$ and DW statistics of JSE ALSI returns have a p -value greater than 0.05, therefore the null hypothesis of no serial correlation is not rejected at 5% level of significance. This confirms that the JSE ALSI returns are not serially correlated. In order to further examine for higher order serial correlation that is the heteroskedasticity in the ALSI returns, this study employed the ACF plot of the squared returns, Ljung-Box statistics, $Q^2(L)$ of the squared returns and the ARCH Lagrange Multiplier (ARCH-LM) test. From Figure 7.3(b), the sample ACF plot of squared ALSI returns shows significant spikes at several lags suggesting high order serial dependence in JSE ALSI returns. This suggests that the ALSI returns seem to be heteroskedastic. Table 7.5 shows the p -value of the Ljung-Box and the ARCH-LM statistics.

Table 7. 5: Test for ARCH effect in the ALSI returns

Test	Statistics	<i>p</i>-value
Ljung-Box, $Q^2(20)$	3642.3000	< 0.0001
ARCH-LM	734.5172	< 0.0001

From Table 7.5, Ljung-Box, $Q^2(L)$ and the ARCH-LM statistics have p -values less than 0.05, therefore the null hypothesis of no heteroskedasticity in the returns is rejected at 5% level of significance. This confirms ALSI returns exhibit conditional heteroskedasticity or ARCH effect.

In summary, results from Table 7.1- 7.5 show that the ALSI returns are stationary, asymmetric, have an Arch effect and not serially correlated, thus a GARCH type model can be fitted to capture volatility clustering and the asymmetric effect on the returns. Results also have shown

that the returns are non –normal, this suggests that a heavy-tailed distribution can be fit to be the data to capture normality property of the ALSI returns.

7.3 GARCH Parameter Estimation and Model Selection

This section involves the selection of the best GARCH-type model that can adequately capture the volatility clustering and the leverage effect in JSE ALSI returns. In this study, the data series used for the model estimation is from 20-May-2005 to 31-Dec-2013, that is 2155 observations so that the others will be used for out-of-sample forecasting. Firstly, GARCH (1,1) model was fitted to JSE ALSI returns. According to Yang et al. (2016), it has proved to be the most common model for predicting volatility. Table 7.6 shows the maximum likelihood (ML) parameter estimates of GARCH (1,1) model with normal distribution governing the innovations.

Table 7. 6: ML Parameter estimates of GARCH (1,1) model and goodness-of- fit statistics

	$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	Ljung-Box, Q²	ARCH LM
statistics	0.0009	$2e^2$	0.1006	0.8861	15.3300	14.949
<i>p</i> -value	0.0000	0.0976	0.0000	0.0000	0.7572	0.7793

From Table 7.6, the parameters are significant at 5% level of significance expect $\hat{\alpha}_0$. It is also observed that the GARCH (1,1) model has successively captured the volatility clustering with Ljung-Box and ARCH-LM *p*-values greater than 0.05, therefore the null hypothesis of no heteroscedasticity in the returns is accepted at 5% level of significance. To further check if GARCH (1,1) has successively captured the asymmetric effect in JSE ALSI returns. The sign and size test is employed. Table7.7 displays the result of the asymmetric test.

Table 7. 7: Sign and size test

Test	t-value	p-value
Sign Bias	2.2256	0.0026
Negative Sign Bias	0.8455	0.3979
Positive Sign Bias	2.2542	0.0242
Joint Effect	23.6960	0.0003

From Table 7.7, the sign bias has p -value less than 0.05, thus the null hypothesis of no sign bias is rejected at 5% level of significance. This indicates the presence of sign bias implying that positive and negative shocks have different impacts upon future volatility. The p -value of the negative sign bias is greater than 0.05, thus the null hypothesis of no negative bias is not rejected. This implies no presence of negative size bias. The p -value of the positive sign bias is less than 0.05, thus the null hypothesis of no positive bias is rejected. This suggests the presence of positive size bias, implying that the size of the positive shocks affects the response of volatility from being symmetric. The joint effect test has a p -value less than 0.05, thus the null hypothesis of no asymmetric effect is rejected at 5% level of significance. It implies that there is a combined effect of the sign and size on future volatility. This shows that the sign bias, the positive sign and the joint effect provide some evidence of bias. This suggests that the GARCH (1,1) model has not successively captured the asymmetric effect in ALSI returns. Therefore, GARCH (1,1) model may not be able to represent the ALSI returns adequately. Thus, an asymmetric GARCH model is required for ALSI returns. Secondly, the following asymmetric GARCH models: EGARCH (1,1), TGARCH (1,1) and APARCH (1,1) models are fitted to the JSE ALSI returns using the MLE method. The Table 7.8 shows the results of maximum likelihood estimates of the asymmetric GARCH models with normal distribution innovation. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) model selection criteria are also reported in Table 7.8.

Table 7. 8: ML Parameter estimates of asymmetric GARCH models

	EGARCH (1,1)	TGARCH (1,1)	APARCH (1,1)
$\hat{\mu}$	0.0005 (0.0247)**	0.0005 (0.0195)**	0.0004 (0.0048)***
$\hat{\alpha}_0$	-0.1616 (0.0000)***	0.0000 (0.0897)*	0.0002 (0.0000)***
$\hat{\alpha}_1$	-0.1023 (0.0000)***	0.0105 (0.3905)	0.0713 (0.0000)***
$\hat{\beta}_1$	0.9819 (0.0000)***	0.9057 (0.0000)***	0.9251 (0.0000)***
$\hat{\gamma}_1$	0.1369 (0.0000)***	0.1319 (0.0000)***	0.7932 (0.0000)***
δ	-	-	1.0000
AIC	-6.1498	-6.1459	-6.1533
BIC	-6.1367	-6.1327	-6.1402

NOTE: *, **, *** indicates (p-value) that significant at 10%, 5%, and 1% level of significant .respectively.

From Table 7.8, it is observed that the ML parameters estimates for the three asymmetric GARCH models fitted to the ALSI returns are significant at least at 10% level of significance. The APARCH (1,1) model has the least AIC and BIC values, this is selected as the best GARCH type model. Finally, the standardized residual of the best GARCH type model is used for checking for model adequacy. Table 7.9 shows the descriptive statistics of standardized residuals.

Table 7. 9: Descriptive Statistics of standardized residuals of the APARCH (1,1) model

Mean	Minimum	Maximum	Median	Skewness	Excess Kurtosis
0.0063	-4.4260	3.4712	0.0684	-0.3748	0.4203

From Table 7.9, it is observed that the excess kurtosis value of the standardized residuals of APARCH (1,1) with normal distribution innovation is greater than zero. This indicates that there is still relatively more value in the tail, therefore the standardized residuals seem to have a tail heavier than that of normal distribution. To check for the non-normality of the standardized residuals, the Q-Q plot, Jarque-Bera (JB) test and Shapiro-Wilk tests are

employed. Figure 7.4 shows the empirical density plot and a Q-Q plot of the standardized residuals.

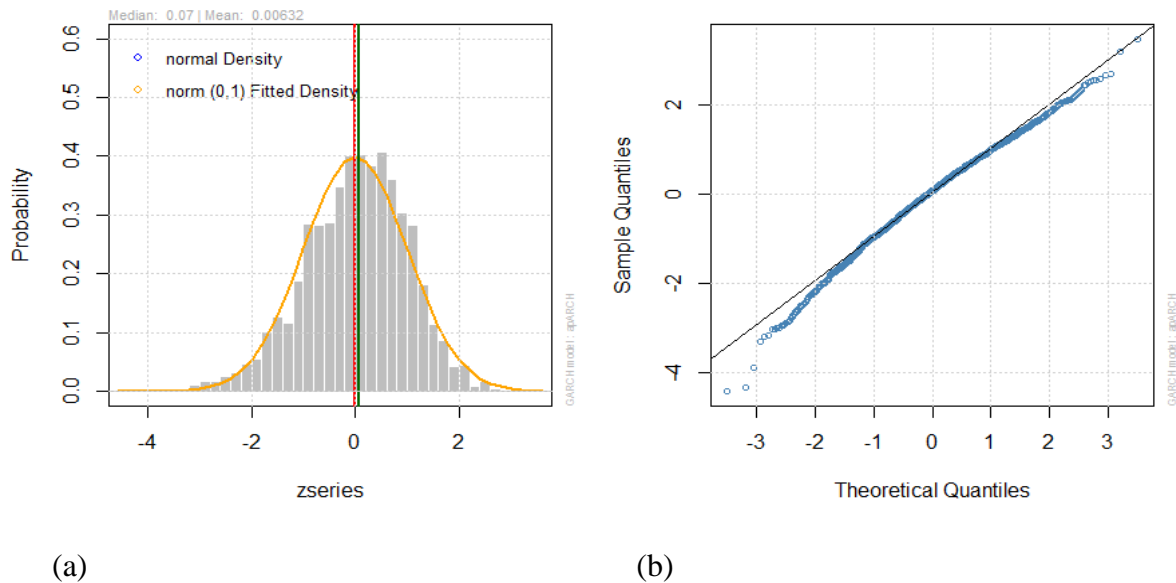


Figure 7. 4: Empirical density plot and Q-Q plot of standardized residuals

**(a) Empirical density plot of standardized residuals (b) Q-Q plot of standardized residuals*

From Figure 7.4, the empirical density and Q-Q plot suggest that standardized residuals seem not to be the normal distribution. This is confirmed by the Jarque-Bera (JB) and Shapiro-Wilk test statistics reported in Table 7.10.

Table 7. 10: Test for normality of the standardized residuals

Test	Statistics	p -value
Shapiro-Wilk	0.99182	< 0.0001
Jarque-Bera	66.6210	< 0.0001

From Table 7.10, it is observed that both tests have p -values less than 0.05, thus the null hypothesis of normality is rejected at 5% level of significance. This confirms that the standardized residuals of APARCH (1,1) have a much heavier tail than that of the normal distribution. To test for serial correlation in the standardized residuals, the autocorrelation (ACF) plot, Ljung-Box statistics $Q(L)$ of the standardized residuals and the DW tests are employed. The ACF plot of standardized residuals is shown in Figure 7.5(a).

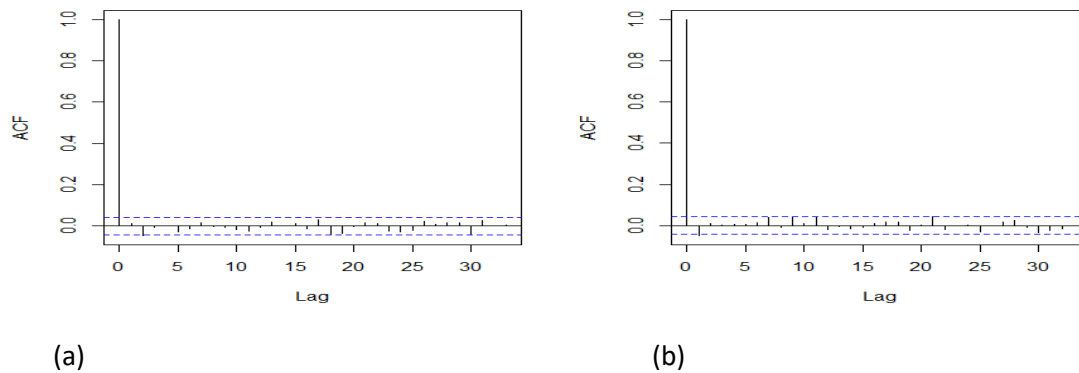


Figure 7. 5: ACF plots of the Standardized residuals

**(a) ACF plot of ALSI standardized residuals (b) ACF plot of squared standardized residuals*

From Figure 7.5(a), it is observed that the sample ACF of the standardized residuals APARCH (1,1) model shows no significant spike, suggesting that the standardized residuals of APARCH (1,1) model are not serially correlated. The Ljung-Box statistics $Q(L)$ of the standardized residuals and the DW test is used to test the null hypothesis of no serial correlation in the residuals. Table 7.11 shows the p -values of the Ljung-Box and DW statistics.

Table 7. 11: Test for serial correlation on the standardized residuals

Test	Statistics	p -value
Ljung-Box, $Q(20)$	19.8670	0.4663
DW test	1.9763	0.2913

From Table 7.11, Ljung-Box $Q(L)$ and DW statistics of the standardized residuals have p -value greater than 0.05, therefore the null hypothesis of no serial correlation is not rejected at 5% level of significance. This confirms that the standardized residuals are not serially correlated. In order to examine the presence of ARCH effect in the standardized residuals, the ACF plot of the squared standardized residuals, the Ljung-Box statistics, $Q^2(L)$ of the squared standardized residuals and the ARCH-LM test are examined for heteroscedasticity in the standardized residuals. It is observed from Figure 7.5(b) that the ACF plot of squared standardized residuals shows no significant spike. This suggests that the standardized residuals do not exhibit ARCH effect. Table 7.12 shows the results of the ARCH tests of the standardized residuals.

Table 7. 12: Test for ARCH effect in the standardized residuals

Test	Statistics	<i>p</i>-value
Ljung-Box, $Q^2(20)$	21.8100	0.3509
ARCH-LM	21.3940	0.3743

From Table 7.12, both the tests have p -values greater than 0.05, therefore the null hypothesis of no arch effect in the squared standardized residuals of APRACH (1,1) model, is not rejected at 5% levels of significance. This confirms that the APRACH (1,1) has adequately captured conditional heteroskedasticity/ARCH effect on JSE ALSI returns. The standardized residuals are also examined for randomness. The Bartels' rank, Cox Stuart, and BDS tests are employed for testing. Table 7.13 shows statistics of independent and identically distributed (IID) tests with their corresponding p -values.

Table 7. 13 : Test for independent and identically distributed of the standardized residuals

Test	Statistics	<i>p</i>-value
Bartels' rank	-0.1109	0.9117
Cox Stuart	537.0000	0.9514
BDS	-1.6930	0.0904

From Table 7.13, it is observed that the Bartels' rank, Cox Stuart and the BDS tests have a p -value greater than 0.05, the null hypothesis of standardized residuals are independent and identically distributed (random) is not rejected at 5% level of significance. This implies that the standardized residuals of APARCH (1,1) model are white noise.

From results reported in Table 7.9, 7.10, 7.11, 7.12 and 7.13, the standardized residuals are not serially correlated, do not have any arch effect and are IID, however they seem to exhibit heavy-tailness. This suggests that the APARCH (1,1) model, although a fairly good model failed to adequately capture non-normality property of the ALSI returns exhibited in the standardized residuals.

7.4 Combining APARCH (1,1) with heavy tailed distribution

The heavy-tailed distributions are fitted to the standardized residuals of APARCH (1,1) model. This is like combining the APARCH model with the heavy-tailed distributions. The four heavy-tail distributions fitted are: Pearson Type IV distribution (PIVD), stable distribution, GPD and GEVD. The APARCH (1,1) with skew Student t (sstd) innovation is also considered. All the heavy-tailed distributions are fitted to the standardized residuals using the MLE method.

7.4.1 APARCH (1,1) model with skew Student t Distribution (sstd)

The JSE ALSI returns, fit to APARCH model with heavy-tail innovations. The AIC and BIC are used to compare these models. See Appendix B for the result of the model selection criteria. The APARCH (1,1) model with skew Student's t distribution minimized both the AIC and the BIC. Table 7.14 shows the parameter estimates of APARCH (1,1) with sstd innovation.

Table 7. 14: Parameter estimates of APARCH (1,1) with sstd governing the innovation

$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	δ	Skew	shape	AD Test
0.0004	0.0002	0.0652	0.9331	0.9197	1.0000	0.8041	23.46	0.0010
(0.004)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.9832)

NOTE: statistics (p -value)

From Table 7.14, all parameters are significant at 5% levels of significance. The skewness parameter is less than 1 indicating that the distribution is negatively skewed. The shape parameter is high, implying that the distribution is heavy-tailed. The empirical density and the Q-Q plot shown in Figure 7.6 seem to follow a skew Student t distribution. This is confirmed by the Anderson-Darling (AD) tests statistics whose p -value is greater than 0.05. This implies that the model is adequate.

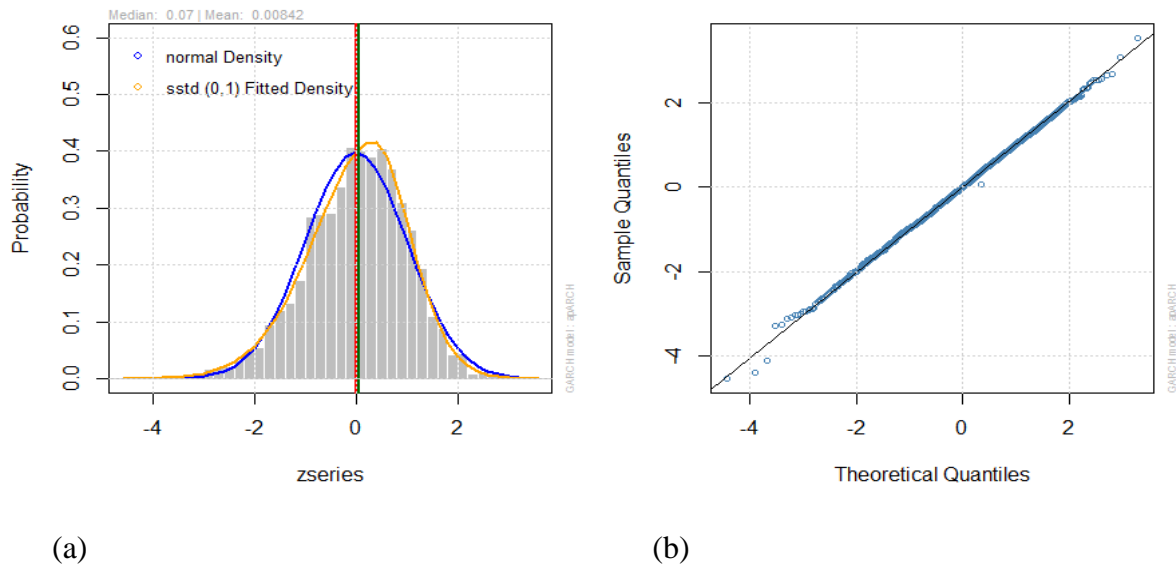


Figure 7. 6: Empirical Density plot and Q-Q plot of APARCH with sstd
 *(a) Empirical density plot of standardized residuals (b) Q-Q plot of standardized residuals

7.4.2 APARCH (1,1)-GPD model

Since the standardized residuals of APARCH (1,1) are now independent and identically distributed, it is therefore considered as suitable for the application of the extreme value analysis. The GPD is fitted to the standardized residuals for both the upper (gains) and lower tails (losses). The mean residual life plot, parameter stability plot will be used for the selection of proper threshold. The Pareto quantile plot is also used to confirm the threshold selected. To obtain the threshold and fit the losses, the standardized residuals was multiplied by -1, that is $Z_t^* = -Z_t$. This converts the minimum values to become maximum values. The mean residual life plot and the Pareto quantile plot select the highest possible threshold on the upper tail of the distribution. Figure 7.7 shows the mean residual life plot of the standardized residuals of both the positive standardized residuals (upper tail) and the negative standardized (lower tail).

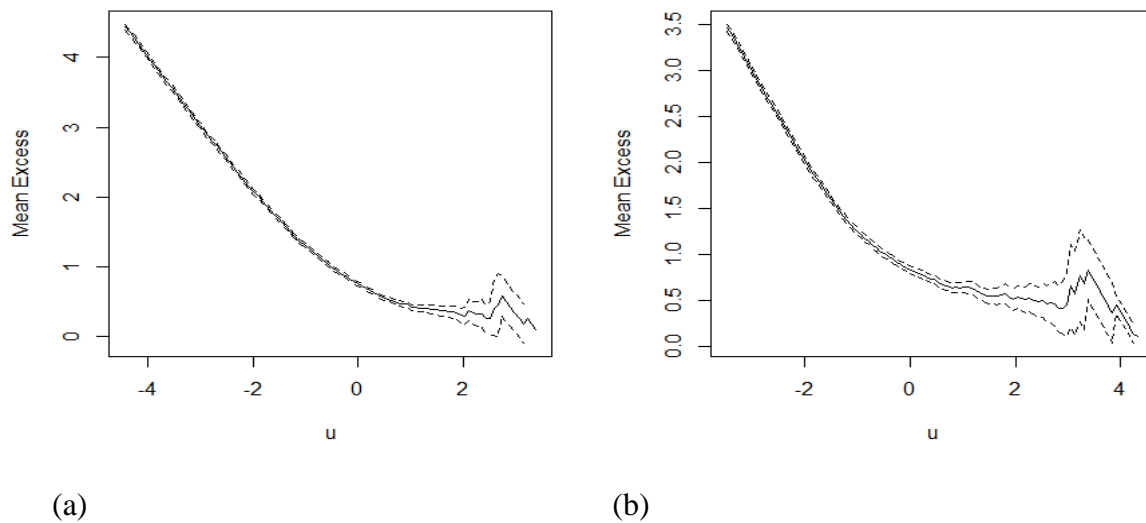


Figure 7. 7: Mean residual life plot of standardized residuals
 *(a) Mean residual life plot of positive standardized residuals (b) Mean residual life plot of negative standardized residuals

In Figure 7.7, it can be seen that the 95% confidence bands (dotted lines) around the mean excesses have been superimposed for both the gains and losses. The suitable threshold must lie where there is a positive gradient change in the mean excess. Thus, the selected threshold u should lie around 2 for both plots. Unfortunately, and this is often encountered empirically, a definite choice for the threshold value can hardly be deduced from this kind of plot. Therefore, to further explore what the proper threshold should be, the parameter stability plot is used. Figure 7.8 and Figure 7.9 show the parameter stability plots for the positive and negative standardized residuals respectively.

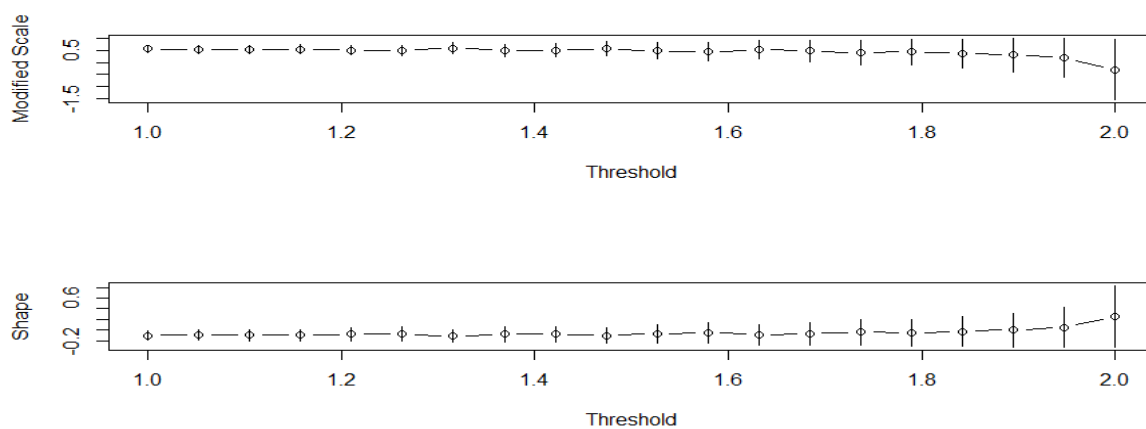


Figure 7. 8: Parameter stability plot for positive standardized residuals.

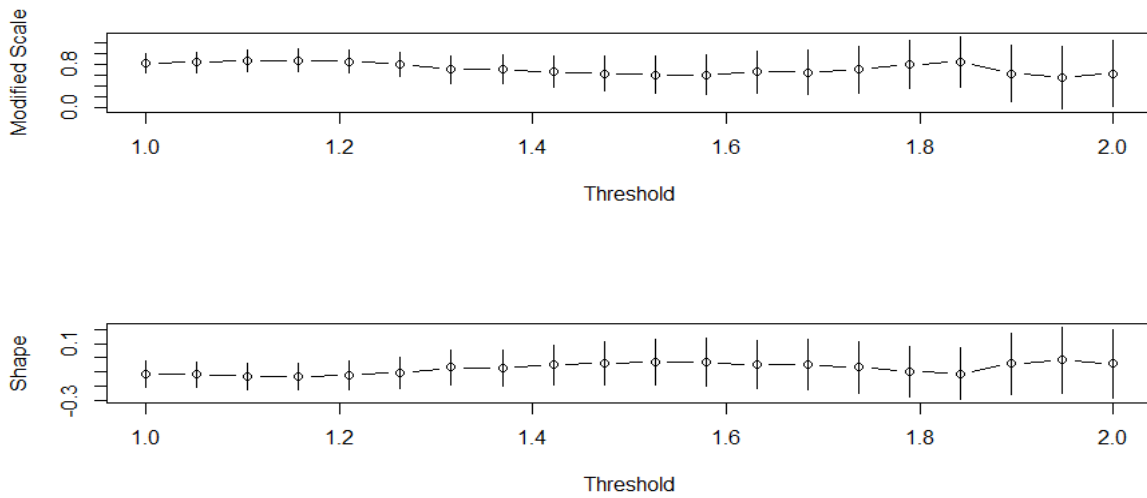


Figure 7. 9: Parameter stability plot for negative standardized residuals

From Figure 7.8 and Figure 7.9, the parameter stability plots show the plot of the optimized shape and scale parameters of possible threshold values from 1.0 to 2.0, it can be seen that the estimated parameters are more or less stable when $u \geq 1.2$ for the Z_t while for the Z_t^* when $u \geq 1.5$. To confirm the threshold the Pareto quantile plot is employed. Figure 7.10, shows the Pareto quantile plots for both the positive and negative standardized residuals.

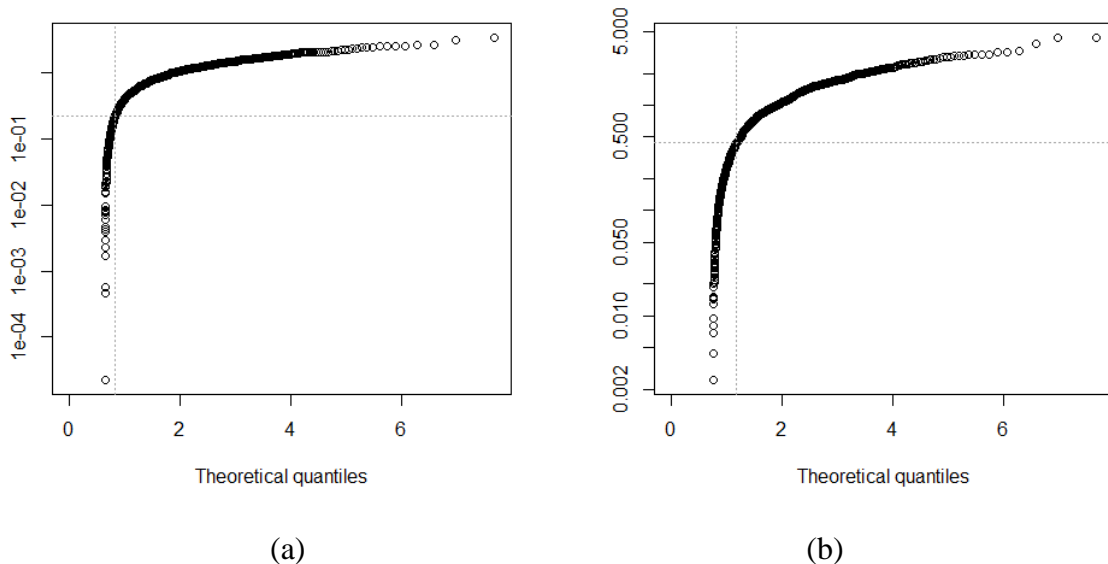


Figure 7. 10 : Pareto quantile plot of standardized residuals

**(a) Pareto quantile plot for positive standardized residuals (b) Pareto quantile plot for negative standardized residuals*

From Figure 7.10(a), the threshold is $u = \exp(0.2261406) = 1.25$ for the Z_t and Figure 7.10(b), show that $u = \exp(0.4400882) = 1.56$ for Z_t^* . The number of observations above the selected thresholds is 197 and 146 for positive (upper) and negative (lower) standardized residuals respectively. The GPD is fitted to both the positive and negative standardized residuals using MLE. Table 7.15, shows the ML parameter estimates with corresponding standard errors in brackets.

Table 7. 15: ML Parameter estimates of GPD with standard errors in bracket

	Threshold,(u)	No of exceedances, (Y)	$\hat{\epsilon}$ Se($\hat{\epsilon}$)	$\hat{\sigma}$ Se($\hat{\sigma}$)
Z_t	1.25	197	-0.0806 (0.0647)	0.4337 (0.0417)
Z_t^*	1.56	146	-0.0434 (0.0820)	0.571 (0.0665)

It is observed from Table 7.15, that the shape parameter is negative, which also suggests that standardized residuals of APARCH (1,1) for JSE ALSI returns follow a GPD for both the upper and left tail. However, the 95% confidence interval of the shape parameters is (0.046212,-0.207412) for the upper tail and is (0.041664,-0.128464) for the lower tail. This indicates that there is no evidence to reject the null hypothesis - the upper and lower tail follows Gumbel distribution or Weibull distribution. Given the impact of shape parameter on the upper and lower tail, there is much uncertainty regarding its characteristics. Figures 7.11 and 7.12 show the GPD model fit diagnostic plot of the positive and negative standardized residuals.

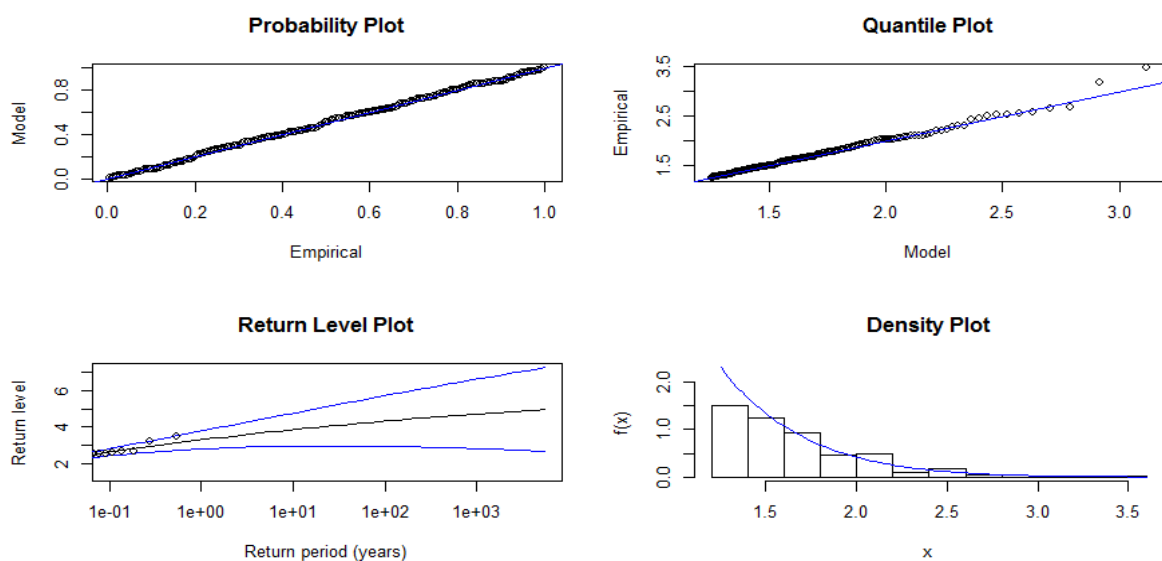


Figure 7. 11: GPD Diagnostic plot of positive standardized residuals

* PP plot (on the upper left panel), (b) Q-Q plot (on the upper right panel), (c) Return level plot (on the lower left panel), (d) Density plot (on the lower right panel)

From Figure 7.11, it is observed that the positive standardized residuals seem to follow the GPD. The Q-Q and the PP plots do not show any serious divergences from the straight lines. The empirical density plot and the return level estimates of the GPD all suggest that the positive standardized residuals follow a GPD at the upper tail.

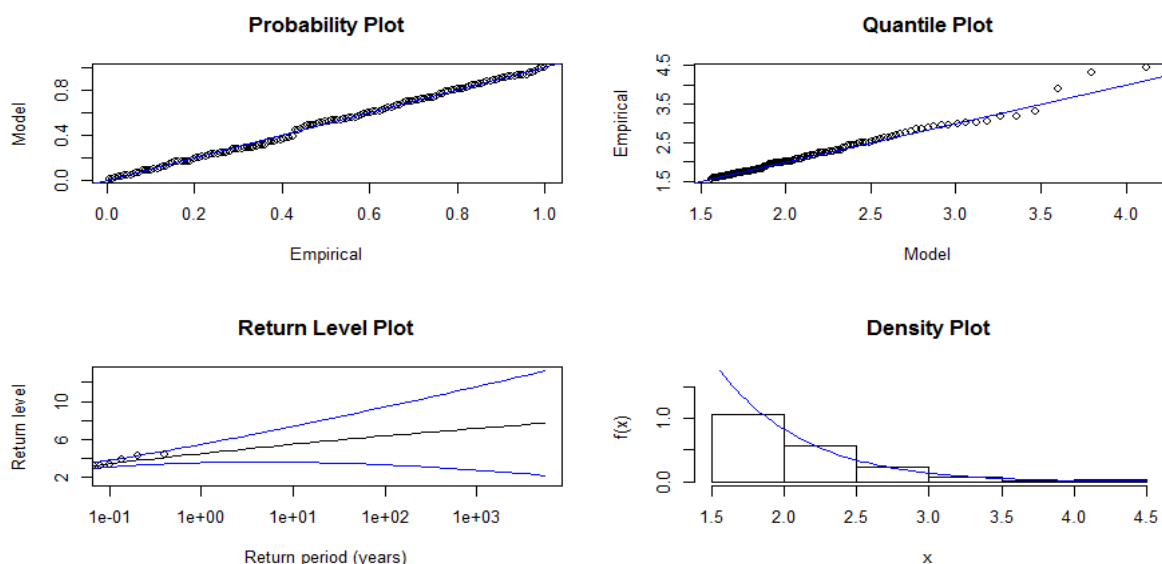


Figure 7. 12: GPD Diagnostic plot of negative standardized residuals

* PP plot (on the upper left panel), (b) Q-Q plot (on the upper right panel), (c) Return level plot (on the lower left panel), (d) Density plot (on the lower right panel)

From Figure 7.12, the Q-Q and the PP plot do not show any serious divergences from the straight lines. The empirical density plot and the return level estimates of the GPD all suggest that the negative standardized residuals follow a GPD at the lower tail. This implies that the negative standardized residuals follow the GPD.

7.4.3 APARCH (1,1)- GEVD model

The GEVD is fitted to the positive and negative standardized residuals using MLE. The block size of 5 is used to perform a block maxima method due to fact that less accuracy is attached to estimates with larger block sizes in accordance with asymptotic property as noted by Coles (2001). Table 7.16 shows the ML parameter estimates of the GEVD with corresponding standard errors in brackets.

Table 7. 16: ML Parameter estimates of GEVD with standard error

	Maxima,(m)	$\hat{\varepsilon}$, Se($\hat{\varepsilon}$)	$\hat{\mu}$ Se($\hat{\mu}$)	$\hat{\sigma}$ Se($\hat{\sigma}$)	AD test (p-value)
Z_t	431	-0.1797 (0.0185)	0.9102 (0.0287)	0.5547 (0.0191)	0.3744 (0.9681)
Z_t^*	431	-0.0852 (0.0296)	0.9027 (0.0245)	0.6535 (0.0348)	0.2551 (0.9676)

From Table 7.16, the shape parameter is negative suggesting that the standardized residuals follow Weibull distribution. This is supported by the 95% confidence interval of the shape parameters which is (-0.14344,-0.21596) for the upper tail and (-0.027184,-0.143216) for the lower tail. Figures 7.13 and Figure 7.14 show the GEVD model fit diagnostic plot of the positive and negative standardized residuals.

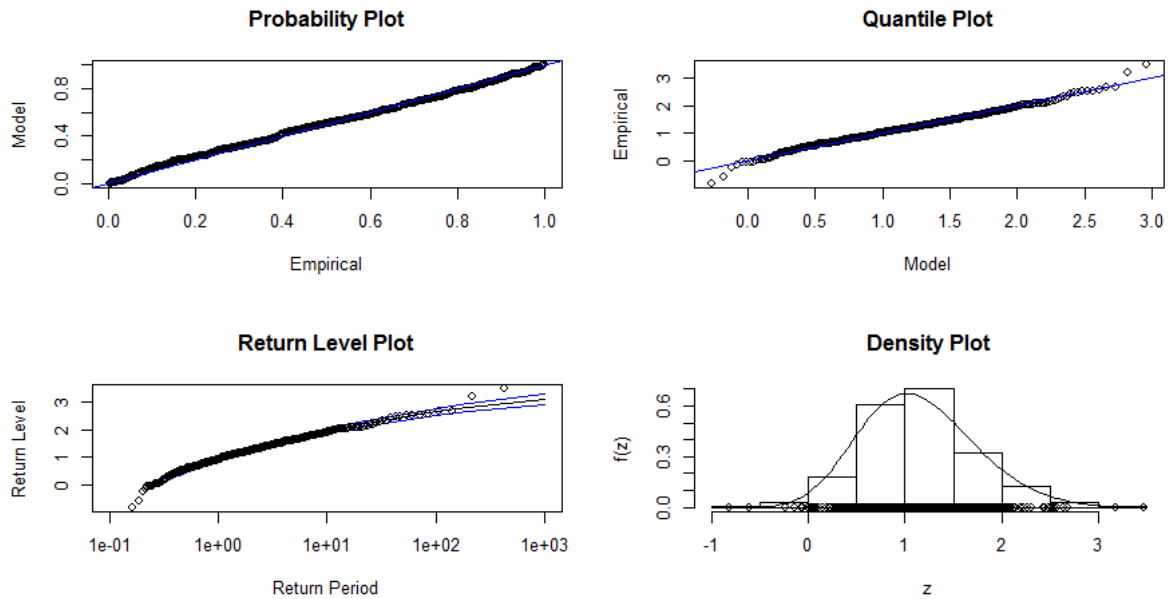


Figure 7. 13: GEVD Diagnostic plot of positive standardized residuals

* PP plot (on the upper left panel), (b) Q-Q plot (on the upper right panel), (c) Return level plot (on the lower left panel), (d) Density plot (on the lower right panel)

From Figure 7.13, the Q-Q and the PP plot do not show any serious divergences from the straight lines. The empirical density plot and the return level estimates of the GPD all suggest that the positive standardized residuals seem to follow a GEVD at the upper tail. This is confirmed by the AD statistics whose p-value is greater than 0.05. Thus, the GEVD distribution is a good fit for the upper tail.

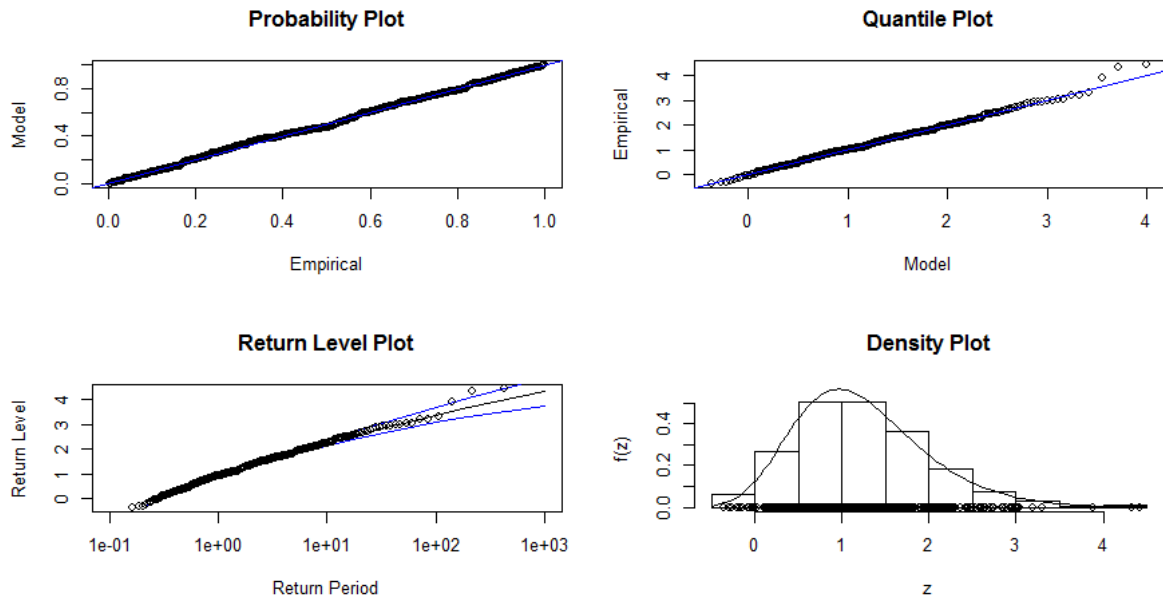


Figure 7. 14: GEVD Diagnostic plot of positive standardized residuals
 * PP plot (on the upper left panel), (b) Q-Q plot (on the upper right panel), (c) Return level plot (on the lower left panel), (d) Density plot (on the lower right panel)

From Figure 7.14, it is observed that the Q-Q and the PP plot do not show any serious divergences from the straight lines. The empirical density plot and the return level estimates of the GPD all suggest that the negative standardized residuals seem to follow a GEVD at the lower tail. This is confirmed by the AD statistics which p-value is greater than 0.05. Thus, the GEVD distribution is a good fit for the lower tail.

7.4.4 APARCH (1,1)-Stable Distribution model

The stable distribution is also fitted to the extracted standardized residuals of the APARCH (1,1) model. The model is referred to as APARCH (1,1)-stable distribution model. Table 7.17 shows the ML parameter estimates of a stable distribution fitted to the standardized residuals of APARCH (1,1) model.

Table 7. 17: ML Parameter estimates of stable distribution and goodness-of-fit statistics

$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	δ	AD test (p-value)
1.9163	-1.0000	0.6784	0.0778	1.0652 (0.3248)

From Table 7.17, the value of the index of stability ($\hat{\alpha}$) is 1.92 which is less than 2. This suggested that the tail of the standardized residuals follows a Pareto law indicating the distribution is heavy-tailed and also has infinite variance. The stable skewedness ($\hat{\beta}$) is -1, meaning that it is skewed to the left. From Figure 7.15, it is observed that the variance stabilized P-P plot of the standardized residuals does not show any divergence from the straight line. This suggested that the standardized residuals seem to follow a stable distribution. The AD statistics has a p -value greater than 0.05, thus the null hypothesis of the standardized residuals follow a stable distribution is not rejected at 5% level of significance. This confirmed that the stable distribution is a good fit for the standardized residuals.

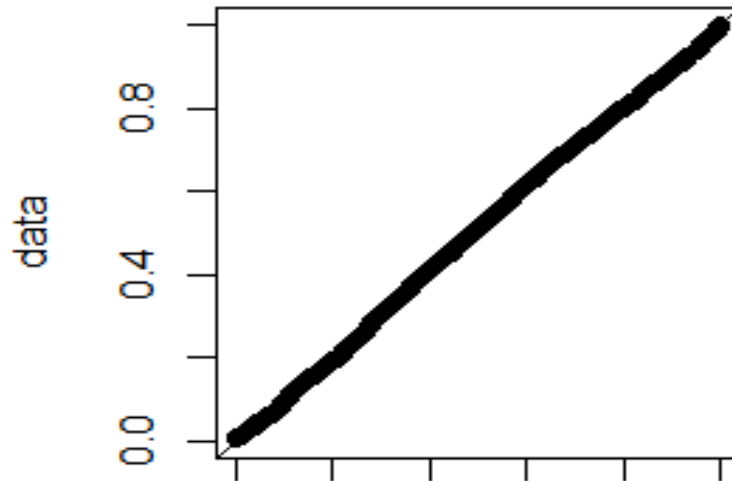


Figure 7. 15: Variance stabilized P-P plot of the standardized residuals

7.4.5 APARCH (1,1)-PIVD model

Finally, PIVD, is fitted to the standardized residuals from extracted from APARCH (1,1) Table 7.18 shows the ML parameter estimates of PIVD.

Table 7. 18: ML Parameter estimates of PIVD and goodness-of-fit statistics

\hat{m}	$\hat{\nu}$	$\hat{\lambda}$	$\hat{\alpha}$	AD-test (<i>p</i> -value)
12.6666	11.7361	-2.1298	4.2221	0.2868 (0.9477)

From Table 7.18, it can be seen that kurtosis parameter is 12.6666 which is greater than 2.5. Therefore, it satisfies the conditions for a PIVD. The AD statistics has a *p*-value greater than 0.05, the null hypothesis of standardized residuals follow a PIVD is not rejected. Thus, the PIVD is a good fit for the standardized residuals.

7.5 Estimating Value at Risk (VaR)

The VaR is estimated for both the long and the short position. The VaR for the short position is associated with the right quantiles of the distribution at a given probability level. The VaR for the long position is associated with is the left quantiles of the distribution at a given probability level. Table 7.19 shows of the VaR estimates for both the short and long position respectively.

Table 7. 19: Value-at-Risk estimates at short and long position

Model	Long position				short position			
	0.10	0.05	0.025	0.01	0.90	0.95	0.975	0.99
APARCH(1,1)-sstd	-1.3006	-1.7367	-2.1326	-2.6170	1.2068	1.5126	1.7834	2.1090
APARCH(1,1)-PIVD	-1.2861	-1.7261	-2.1348	-2.6466	1.2253	1.5368	1.8045	2.1150
APARCH(1,1)-stable	-1.2160	-1.6378	-2.0569	-2.6967	1.2811	1.6123	1.8975	2.2270
APARCH(1,1)GEVD	-1.3103	-1.7423	-2.1411	-2.6279	1.2460	1.5798	1.8679	2.1936
APARCH(1,1)-GPD	-1.3357	-1.7323	-2.117	-2.6084	1.2109	1.5054	1.7839	2.1290

Note: Values in **bold blue** and **red** are the highest and smallest VaR estimates at given levels respectively

From Table 7.19, it is observed that the stable distribution has the highest VaR estimate for both short and long positions except at 1% level of the long position where the GPD has the highest VaR estimate. The sstd has the smallest VaR estimate at the short position except at 95% level where the GPD has the smallest VaR estimate. For the long position, the smallest VaR differs at difference level, the GPD has the smallest VaR at 10%, sstd distribution has the smallest VaR at 5%, GEVD has the smallest VaR at 2.5% and the stable distribution has the smallest VaR at 1%

7.6 In-Sample Backtesting

In order to check model adequacy in estimating the VaR estimate, the VaR estimates are backtested using the Kupiec likelihood ratio test. Table 7.20 shows the p -value of the Kupiec likelihood ratio test at different levels for the in-sample data.

Table 7. 20: In-sample backtesting: p -values of Kupiec likelihood ratio test

Model	Long position				Short position			
	0.10	0.05	0.025	0.01	0.90	0.95	0.975	0.99
APARCH(1,1)-sstd	0.9714	0.4383	0.9862	0.9227	0.9714	0.7850	0.9036	0.7341
APARCH(1,1)-PIVD	0.5921	0.8623	0.7947	0.9048	0.3278	0.4384	0.6890	0.5731
APARCH(1,1)-stable	0.1280	0.0786	0.4064	0.7341	0.0137	0.0197	0.1200	0.1336
APARCH(1,1)GEVD	0.8578	0.7849	0.7947	0.9227	0.1783	0.1107	0.1596	0.3064
APARCH(1,1)-GPD	0.5883	0.7849	0.9862	0.7562	0.8011	0.7849	0.9862	0.4289

Note: Values in **Bold blue** are the highest p -value at a given level.

From Table 7.20, the p -values for all the fitted models are greater than 0.05, thus the null hypothesis of model adequacy is not rejected at all levels under investigation. The best model is selected at different levels using the p -value of the Kupiec likelihood test statistics. The model with the highest p -value at a given level is selected as the best (robust) model. The sstd distribution has the highest p -value at all probability levels except at 97.5% where GPD has outperformed it. For the long position, sstd distribution has the highest p -value at 10% level,

the PIVD has the highest at 5% level, the GPD and sstd distribution highest p -value at 2.5% level and finally the GEVD and sstd distribution has the highest p -value at 1% level.

7.7 Out of Sample Backtesting

In this section, adequacy and predictive ability of the VaR model is examined by backtesting VaR models. Table 7.21 shows the p -value of the Kupiec test for both long and short position.

Table 7. 21: Out of sample backtesting: p -values of Kupiec likelihood ratio test

Model	Long position				Short position			
	0.10	0.05	0.025	0.01	0.90	0.95	0.975	0.99
APARCH(1,1)-sstd	0.2005	0.2377	0.0882	0.3782	0.0059	0.0018	0.0436	0.1707
APARCH(1,1)-PIVD	0.7637	0.3265	0.1607	0.3782	0.0022	0.0018	0.0882	0.3782
APARCH(1,1)-stable	0.7056	0.4329	0.2672	0.3782	0.0000	0.0003	0.0191	0.0558
APARCH(1,1)GEVD	0.3921	0.2377	0.1607	0.3782	0.0002	0.0008	0.0191	0.0558
APARCH(1,1)-GPD	0.1546	0.3265	0.1607	0.3782	0.0059	0.0078	0.0882	0.3782

Note: Values in **bold red** are p -values that is not significant, Values in **Bold blue** are the highest p -values at a given level.

From Table 7.21, it is observed that for the short position, all the models rejected the null hypothesis of model adequacy at lower VaR levels while at higher levels the models fail to reject the null hypothesis of model adequacy. This implies that at the short position, all the VaR models are adequate at high levels. For long positions, all the VaR models did not reject the null hypothesis of model adequacy at all levels. This implies that at the long position, the VaR models are adequate at all levels. At the short position, the APARCH (1,1)-PIVD model and the APARCH (1,1)-GPD model have the highest p -values at 97.5% and 99% levels respectively. This implies that they outperform the other models at 97.5% and 99% levels. It is noted that the APARCH (1,1)-PIVD model and APARCH (1,1)-GPD model seems to produce similar results at 97.5% and 99% levels. At the long position, all the models produced similar results at 1% and the APARCH (1,1)-PIVD models and APARCH (1,1)-GPD model produce results that are quite close except at the 10% level. The APARCH (1,1)-stable distribution

model has the highest p -values at 5% and 2.5% levels while the APARCH (1,1)-PIVD model has the highest p -value at 10% levels. This implies that at the long position, the APARCH (1,1)-stable distribution model outperforms the other models at all levels except at 10% where APARCH (1,1)-PIVD model outperforms the other models.

7.8 Conclusion

In this chapter, the main finding of the fitted APARCH (1,1)-GPD model, APARCH (1,1)-GEVD model, APARCH (1,1)-PIVD models, and APARCH (1,1)-stable distribution model was presented. APARCH (1,1) model with sstd governing the innovation is also fitted to the ALSI returns. VaR is estimated for the distributions and backtesting is performed to assess the adequacy of the VaR estimates and the VaR models.

CHAPTER 8

CONCLUSION

In order to find an adequate Value-at-risk (VaR) model for South Africa's market risk, this study examines the combination of APARCH (1,1) model and heavy-tailed distribution: generalized Pareto distribution (GPD), generalized extreme value distribution (GEVD), stable distribution and Pearson Type IV distribution (PIVD) on JSE all share index (ALSI) returns. The APARCH (1,1) model has the ability to capture both volatility clustering and leverage effect while the heavy-tail feature is captured by the heavy-tailed distributions.

The asymmetric GARCH model -APARCH (1,1) was found to be the best possible model to capture both volatility clustering and the leverage effect on ALSI returns based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). It has been confirmed by the Anderson-Darling test (AD test) and Kolmogorov-Smirnov test (K-S test) that these heavy tailed distributions adequately fitted the standardized residuals of APARCH (1,1) model. Besides, the APARCH (1,1) model with skew Student t (sstd) innovation is also found to be an adequate model.

The adequacy of these heavy-tailed distributions is examined through in-sample backtesting by employing Kupiec test. It was found that both in the short position and the long position, all the heavy-tailed distributions have adequately produced an adequate VaR estimation for the innovation of APARCH (1,1). At the short position, the skew Student t distribution is found to be the most adequate distribution for VaR estimate of the innovation of APARCH (1,1) except at 97.5% VaR level where the GPD outperforms it. While for the long position, it was found that best distribution for the VaR estimation for the innovation of APARCH (1,1) differ at different VaR levels. The sstd distribution outperforms others at 10% VaR level, the PIVD has outperformed others at 5% VaR level, the GPD and sstd distribution outperforms at 2.5% VaR level and finally the GEVD and sstd distribution outperforms at 1% VaR level.

The relative performance of APARCH (1,1)-GPD model, APARCH (1,1)-GEVD, APARCH (1,1)-stable, APARCH (1,1)-Pearson Type IV model, and APARCH (1,1) with skew student t

innovation are compared in terms of their predictive ability of the South Africa market risk. It was found from the out of sample backtesting that all the models performed well at higher VaR levels in the short position and APARCH (1,1)-GPD model and APARCH (1,1)-Pearson Type IV model outperformed the other models. While at the long position, all the models performed well at all VaR levels. At the long position, the APARCH (1,1)-stable model outperformed the other models at 5% and 2.5% VaR levels while the APARCH (1,1)-Pearson Type IV models has outperformed the other models at 10% probability levels.

For further research, we can examine these models on other emerging markets.

REFERENCES

- ABRAMOVITZ, M. & A., S. I. 1964. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, US Department of Commerce.
- AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Transaction on Automation Control*, 19, 716-723.
- AMIR, D. A. 1993. *Complete Business Statistics*, Richard D. Irwin.
- ANGABINI, A. & WASIUZZAMAN, S. 2011. GARCH Models and the Financial Crisis-A Study of the Malaysian Stock Market. *The International Journal of Applied Economics and Finance*, 5, 226-236.
- ANGELIDIS, T. & DEGIANNAKIS, S. 2008. Forecasting One-day-ahead VaR and Intra-Day Realized Volatility in the Athens Stock Exchange Market. *Managerial Finance*, 34, 489-497.
- APARICIO, F. M. & ESTRADA, J. 2001. Empirical distributions of stock returns: European securities markets, 1990–95. *European Journal of Finance*, 7, 1–21.
- BAHARUL-ULUM, Z. K. A., BIN AHMAD, I. & BINTI SALAMUDIN, N. 2012. Assessing the Accuracy of Risk Models in the Malaysian market. *Asian Journal of Business and Management Sciences*, 1, 48-59.
- BAILLIE, R. T., BOLLERSLEV, T. & MIKKELSEN, H. O. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.
- BARONE-ADESI, G., BOURGOIN, F. & GIANNOPOULOS, K. 1998. Don't look back. *Risk*, 11, 100-104.

- BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. & TEUGELS, J. 2004. *Statistics of Extremes: Theory and Applications*, Wiley Series in Probability and Statistics.
- BEKAERT, G. & GUOJUN, W. 2000. Asymmetric Volatility and Risk in Equity Markets. *The Review of Financial Studies*, 13, 1-42.
- BELOV, I. A. 2005. On the computation of the probability density function of α -stable distributions. *Mathematical Modelling and Analysis 2005*.
- BHATTACHARYYA, M., CHAUDHARY, A. & YADAV, G. 2008. Conditional VaR estimation using Pearson's type IV distribution. *European Journal of Operational Research*, 191, 386–397.
- BLATTBERG, R. C. & GONEDES, N. J. 1974. A comparison of the stable and student distributions as statistical models for stock prices. *J. Bus*, 47, 244–280.
- BOLLERSLEV, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- BOLLERSLEV, T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69, 542–547.
- BOLLERSLEV, T. & MIKKELSEN, H. O. 1996. Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, 73, 151–184.
- BOLLERSLEV, T., R.F., E. & NELSON, D. B. 1994. ARCH Models. *Handbook of Econometrics*, 4, 2959-3038.
- BORAK, S., HÄRDLE, W. & WERON, R. 2005. *Stable Distributions*. Berlin: Humboldt-Universität zu Berlin.
- BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. & LJUNG, G. M. 2015. *Time Series Analysis: Forecasting and Control*, Prentice Hall.

- BROOKS, C. 2008. *Introductory Econometrics for Finance*, Cambridge University Press.
- BROOKS, C. & PERSAND, G. 2003. The effect of asymmetries on stock index return Value-at-Risk estimates. *Journal of Risk Finance*, 4, 29-42.
- BROOKS, R. D., DAVIDSON, S. & FAFF, R. W. 1997. An Examination of the Effects of Major Political Change on Stock Market Volatility: The South African Experience. *Journal of International Financial Markets Institutions and Money*, 7, 255 - 275.
- BUCEVSKA, V. 2013. An Empirical Evaluation of GARCH Models in Value-at-Risk Estimation: Evidence from the Macedonian Stock Exchange. *Business Systems Research*, 4, 49-64.
- BUCEVSKA, V. & BUCEVSKA, J. 2012. Estimating and forecasting stock market volatility under crisis: empirical evidence from Macedonian stock exchange. *International Scientific Conference*. University of Belgrade.
- CARIES, S. 2009. A Comparative Simulation Study of the Annual Maxima and the Peaks-over-threshold Methods. Delft University of Technology.
- CASTILLO, E. & HADI, A. S. 1997. Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, 92, 1609–1620.
- CHAN, N. H., DENG, S., PENG, L. & XIA, Z. 2007. Interval estimation of Value-at-Risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, 137, 556–576.
- CHEN, L. & LIANG, X. 2008. *Empirical results from VaR prediction using Pearson's Type IV distribution*. MSc, Simon Fraser University.
- CHINHAMU, K., HUANG, C.-K., HUANG, C.-S. & CHIKOBVU, D. 2015. Extreme Risk, Value-At-Risk And Expected Shortfall In The Gold Market. *International Business & Economics Research Journal*, 14, 107-122.

- CHUNG, C. F. 1999. Estimating the fractionally integrated GARCHmodel. Taiwan: National Taiwan University.
- COLES, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag: London.
- COLES, S. & DIXON, M. 1999. Likelihood-Based Inference for Extreme Value Models. *Extremes* 2, 5-23.
- COOLEY, D., NYCHKA, D. & NAVEAU, P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of American Statistics Association*, 102, 824–840.
- COX, D. R. & HINKLEY, D. V. 1974. *Theoretical Statistics*, London, Chapman and Hall.
- CURTO, J. D., PINTO, J. C. & TAVARES, G. N. 2009. Modelling Stock Markets' Volatility Using GARCH Models with Normal, Student's t and Stable Paretian Distributions. *Statistical Papers*, 50, 311-321.
- DAVIDSON, J. 2004. Moment and memory properties of linear conditional heteroscedasticitymodels, and a new model. *Journal of Business and Economic Statistics*, 22, 16-29.
- DAVISON, A. C. 1984. Modelling exceedances over high thresholds: In J. Tiago de Oliveira (ed.). *Statistical Extremes and Applications (Dordrecht: Reidel)*, 461–482.
- DAVISON, A. C. & SMITH, R. L. 1990. Models for exceedence over high thresholds. *Journal of the Royal Statistical Society*, 52, 393–442.
- DEGIANNAKIS, S., FLOROS, C. & LIVADA, A. 2012. Evaluating Value-at-Risk Models before and after the Financial Crisis of 2008: International Evidence. *Managerial Finance*, 38, 436-452.
- DING, Z., GRANGER, C. W. J. & ENGLE, R. F. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.

- DOWD, K. 1998. *Beyond Value at Risk: The New Science of Risk Management*, Chichester, Wiley.
- DOWD, K. 2005. *Model Risk, Measuring Market Risk*, West Sussex, England, John Wiley & Sons Ltd.
- DUFFIE, D. & PAN, J. 1997. An overview of value at risk. *The Journal of Derivatives*, 5, 7-49.
- DUMOUCHEL, W. H. 1973b. Stable distributions in statistical inference. 1:Symmetric stable distributions compared to other symmetric long-tailed distributions. *JASA* 68, 469-477.
- ENGLE, R. F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.
- ENGLE, R. F. & BOLLERSLEV, T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews*, 5, 1–50.
- FAMA, E. F. 1965. The behavior of stock-market prices. *Journal of Business*, 38, 34–105.
- FAN, J. & YAO, Q. 2003. *Nonlinear time series: nonparametric and parametric methods*, Springer.
- FERNANDES, W., NAGHETTINI, M. & LOSCHI, R. 2010. A Bayesian Approach for estimating Extreme Flood Probabilities with Upper-bounded Distribution Functions. *Stochastic Environmental Research and Risk Assessment*, 24, 1127–1143.
- FERNANDEZ, C. & STEEL, M. 1998. On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, 93, 359–371.
- FISHER, R. A. & TIPPETT, L. H. C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

- FLOROS, C. 2008. Modeling Volatility using GARCH Models: Evidence from Egypt and Israel. *Middle Eastern Finance and Economics*, 31-41.
- FRANCQ, C. & ZAKOIAN, J.-M. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*, United Kingdom, Wiley and Sons.
- FRANSES, P. H. & VAN DIJK, D. 1996. Forecasting Stock Market Volatility Using (Non-Linear) GARCH Models. *Journal of Forecasting*, 15, 229-235.
- FRENCH, K., SCHWERT, G. & STAMBAUGH, R. 1987. Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19, 3-29.
- GILLI, M. & KELLEZI, E. 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27, 1-23.
- GIOT, P. & LAURENT, S. 2003a. Value-at-Risk for Long and Short Trading Positions. *Journal of Applied Econometrics*, 18, 641-664.
- GLOSTEN, L. R., JAGANNATHAN, R. & RUNKLE, D. E. 1993a. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48, 1779–1801.
- GLOSTEN, L. R., JAGANNATHAN, R. & RUNKLE, D. E. 1993b. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779–1801.
- GNEDENKO, B. V. 1943. Sur la distribution limite du terme maximum d'une seriealeatoire. *Annals of Mathematics* 44, 423–453.
- GOKCAN, S. 2000. Forecasting Volatility of Emerging Stock Markets: Linear Versus Non-Linear GARCH Models. *Journal of Forecasting*, 19, 499-504.
- GRANGER, C. W. J. & ANDERSEN, A. P. 1978. *Introduction to Bilinear Time Series Models*, Göttingen, Vandenhoeck and Ruprecht.

- HALLS-MOORE, M. 2015a. Generalised Autoregressive Conditional Heteroskedasticity GARCH(p, q) Models for Time Series Analysis. *QuantStart.com* [Online]. Available from: <https://www.quantstart.com/articles/Generalised-Autoregressive-Conditional-Heteroskedasticity-GARCH-p-q-Models-for-Time-Series-Analysis> 2016].
- HALLS-MOORE, M. 2015b. Serial Correlation in Time Series Analysis. *QuantStart.com* [Online]. Available from: <https://www.quantstart.com/articles/Serial-Correlation-in-Time-Series-Analysis> 2016].
- HANSEN, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review*, 35, 705–730.
- HASSAN, S. 2013. South African Capital Markets: An Overview. *In: BANK, S. A. R. (ed.)*. Pretoria.
- HEINRICH, J. 2004. A guide to the Pearson type IV distribution.
- HOLTON, G. A. 2003. *Value-at-risk: Theory and practice*, San Diego, Academic Press.
- HOSKING, J. R. M. & WALLIS, J. R. 1987. Parameter and quantile estimation for the generalised Pareto distribution. *Technometrics*, 29, 339-349.
- HU, Y. 2013. *Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance*. MSc, University of Canterbury.
- HUANG, C.-S., HUANG, C.-K. & CHINHAMU, K. 2014. Assessing The Relative Performance Of Heavy-Tailed Distributions: Empirical Evidence From The Johannesburg Stock Exchange. *The Journal of Applied Business Research*, 30, 1263-1286.
- HUANG, Y. C. & LIN, B.-J. 2004. Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations. *Review of Quantitative Finance and Accounting*, 22, 79-95.

- JARUSKOVA, D. & HANEK, M. 2006. Peaks over threshold method in comparison with block maxima method for estimating high return levels of several Northern Moravia precipitation and discharges series. *Journal of Hydrology and Hydromechanics* 54, 309–319.
- JENKINSON, A. F. 1955. The frequency distribution of the annual maximum (or minimum) value of meteorological events. *Journal of the Royal Meteorological Society*, 81, 158-172.
- JONATHAN, P. & EWANS, K. 2013. Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, 62, 91–109.
- JORION, P. 2007. *Value at Risk - The New Benchmark for Managing Financial Risk*, USA, McGraw-Hill.
- KARLSSON, L. 2002. *GARCH – Modelling Theoretical Survey, Model Implementation and Robustness Analysis*. Master degree, Royal Institute of Technology (KTH).
- KOGON, S. M. & WILLIAMS, D. B. 1998. Characteristic function based estimation of stable parameters. In: ADLER, R., FELDMAN, R. AND TAQQU, M. (EDS.) (ed.) *A Practical Guide to Heavy Tailed Data*. Boston, MA: Birkhäuser.
- KON, S. 1984. Models of stock returns—a comparison. *Journal of Finance*, 39, 147–165.
- KUESTER, K., MITTIK, S. & PAOLELLA, M. 2005. Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics*, 4, 53-89.
- KUPIEC, P. H. 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3, 73–84.
- LAMBERT, P. & LAURENT, S. 2000. Modelling skewness dynamics in series of financial data. In: STATISTIQUE, L.-L.-N. I. D. (ed.) *Discussion Paper*.

- LAURENT, S. 2003. Analytical derivatives of the APARCH model. *Computational Economics*, 24, 51-57.
- LEADBETTER, M. R., LINDGREN, G. & ROOTZEN, H. 1983. *Extremes and Related Properties of Random Sequences and Processes*, New York, Springer-Verlag.
- MA, X. & NIKIAS, C. L. 1995. On blind channel identification for impulsive signal environments. *Proceedings of ICASSP'95*. Detroit, MI: IEEE
- MADSEN, H., RASMUSSEN, P. F. & ROSBJERG, D. 1997. Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events, 1, At-site modelling. *Water Resource Research*, 33, 747–758.
- MAGNUS, F. J. & FOSU, O. 2006. Modeling and Forecasting Volatility of on Returns on the Ghana Stock Exchange Using GARCH Models. *Journal of Applied Science*, 3, 2042-2048.
- MAKHWITING, M. R. 2014. *Modelling volatility and financial market risks of shares on the Johannesburg stock exchange*. Master of Science, University of Limpopo.
- MAKHWITING, M. R., LESAOANA, M. & SIGAUKE, C. 2012. Modelling volatility and financial market risk of shares on the Johannesburg stock exchange. *African Journal of Business Management*, 6, 8065-8070.
- MANDELBROT, B. 1963. The variation of certain speculative prices. *Journal of Business*, 36, 394–419.
- MANDELBROT, B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*, 1057-1062.
- MARTINS, E. S. & STEDINGER, J. R. 2000. Generalized maximum-likelihood generalized extreme value quantile estimators for hydrologic data. *Water Resource Research*, 36, 737–744.

- MARTINS, E. S. & STEDINGER, J. R. 2001. Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resour. Res.*, 37, 2551–2557.
- MCCULLOCH, J. H. 1986. Simple consistent estimators of stable distribution parameters. *Communication in Statistics- Simulation and computations*, 15, 1109-1136.
- 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, 271–300.
- MITTNIK, S. & PAOLELLA, M. S. 2000. Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting*, 19, 313-333.
- MITTNIK, S., RACHEV, S. T., DOGANOGLU, T. & CHENYAO, D. 1999. Maximum likelihood estimation of stable Paretian models. *Mathematical and Computer Modelling*, 29, 275-293.
- MUSAH, A. I. 2010. *Application of Extreme Value Theory for estimating daily Brent Crude Oil Prices*. Master of Science, Kwame Nkrumah University of Science and Technology.
- NAGAHARA, Y. 1999. The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. *Statistics and Probability Letters*, 43, 251-264.
- NAGAHARA, Y. 2007. A Method of Calculating the Downside Risk by Multivariate Nonnormal Distributions. *Asian Pacific Financial Markets*, 15, 175-184.
- NASON, G. 2008. *Wavelet methods in statistics with R*, Verlag New York, Springer.
- NELSON, D. B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59, 347–370.
- NOLAN, J. P. 1997. Numerical calculation of stable densities and distribution functions. *Communications in Statistics-Stochastic Models*, 13, 759-774.

- NOLAN, J. P. 1999. An algorithm for evaluating stable densities in Zolotarev's (M) parametrization. *Mathematical and Computer Modelling*, 29, 229-233.
- NOLAN, J. P. 2003. Modeling financial data with stable distributions. In: RACHEV, S. T. (ed.) *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*. Netherlands: Elsevier Science.
- NOLAN, J. P., PANORSKA, A. K. & MCCULLOCH, J. H. 2001. Estimation of stable spectral measures. *Mathematical and Computer Modelling*, 34, 1113-1122.
- OJEDA, D. 2001. *Comparison of stable estimators*. Ph.D, American University.
- ORTIZ, E. & ARJONA, E. 2001. Heteroscedastic Behaviour of the Latin American Emerging Stock Markets. *International Review of Financial Analysis*, 10, 287-305.
- OZUN, A., CIFTER, A. & YILMAZER, S. 2010. Filtered extreme-value theory for value-at-risk estimation: evidence from Turkey. *Journal of Risk Finance*, 11, 164-179.
- PAGAN, A. R. & SCHWERT, G. W. 1990. Alternative Models for Conditional Stock Volatility. *Journal of Econometrics*, 45, 267-290.
- PEARSON, K. 1895. Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London A* 186, 343–414.
- PEARSON, K. 1916. Mathematical contributions to the theory of evolution.-XIX. Second Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society of London A*, 216, 429-457.
- PEIRÓ, A. 1994. The distribution of stock returns: international evidence. *Applied Financial Economics*, 4, 431–439.
- PEIRÓ, A. 1999. Skewness in financial returns. *Journal of Banking and Finance*, 23, 847–862.

- PICKANDS, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- POCERNICH, M. J. 2002. *Application of Extreme Value Theory and Threshold Models in Hydrauligical Events*. MSc, University of Colorado in Denver.
- PRAETZ, P. D. 1972. The distribution of share price changes. *Journal of Business*, 45, 49–55.
- PREMARATNE, G. & BERA, A. 2001. Modeling Asymmetry and Excess Kurtosis in Stock Return Data. *Working paper*. USA: University of Illinois.
- PRESS, J. 1967. A compound events model for security prices. *Journal of Business*, 40, 317–335.
- RACHEV, S. T. & MITTNIK, S. 2000. *Stable Paretian models in finance*, New York, Wiley.
- RAFIQUE, A. & KASHIF, R. 2011. Comparing persistency of different frequencies of stock returns volatility in an emerging market: A case study of Pakistan. *African Journal of Business Management*, 59-67.
- ROCCO, M. 2011. Extreme Value Theory for Finance: A Survey. *Economic Research and International Relations Area, Bank of Italy*. Italy: Questioni di Economia e Finanza, (Occasional papers).
- SAMORODNITSKY, G. & TAQQU, M. S. 1994. *Stable Non-Gaussian Random Processes*, New York, Chapman and Hall.
- SAMUELSON, P. A. 1967. Efficient portfolio selection for Pareto-Levy investments. *Journal of Financial and Quantitative Analysis*, 2, 107-122.
- SATO, A. H. 2014. *Applied Data-Centric Social Sciences*, Japan Springer

- SIGAUKE, C., MAKHWITING, R. M. & LESAONA, M. 2014. Modelling conditional heteroscedasticity in JSE stock returns using the Generalized Pareto Distribution. *African Review of Economics and Finance*, 6, 41-55.
- SINGH, A. K., ALLEN, D. E. & POWELL, R. J. Value at Risk Estimation Using Extreme Value Theory. 19th International Congress on Modelling and Simulation, 12–16 December, 2011 2001 Perth, Australia. 1-7.
- SMITH, R. L. 1985. Maximum likelihood estimation in a class of non regular cases. *Biometrika*, 72, 67-90.
- SMITH, R. L. 2003. Statistics of Extremes, with Applications in Environment, Insurance and Finance. In: FINKENSTADT, B., ROOTZEN, H. & HALL., C. A. (eds.) *Extreme Values in Finance, Telecommunications and the Environment*. London: CRC Press.
- SOLTANE, H. B., KARAA, A. & BELLALAH, M. 2012. Conditional VaR using GARCH-EVT approach: Forecasting Volatility in Tunisian Financial Market. *Journal of Computations & Modelling*, 2, 95-115.
- STAMBAUGH, F. 1996. Risk and value at risk. *European Management Journal*, 14, 612–621.
- STAVROYIANNIS, S. 2013. On a GJR-GARCH Model with the Standardized Pearson Type IV Distribution. Available at SSRN: <http://ssrn.com/abstract=2345972> or <http://dx.doi.org/10.2139/ssrn.2345972>, 1-17.
- STAVROYIANNIS, S. 2016. Value-at-Risk and backtesting with the APARCH model and the standardized Pearson type IV distribution. 1-21.
- STAVROYIANNIS, S., MAKRIS, I., NIKOLAIDIS, V. & ZARANGAS, L. 2012. Econometric modeling and value-at-risk using the Pearson type-IV distribution. *International Review of Financial Analysis*, 22, 10–17.
- STEPHENSON, A. G. & TAWN, J. A. 2004. Bayesian inference for extremes: accounting for the three extremal types. *Extremes*, 7, 291–307.

- SUBBOTIN, M. T. 1923. On the law of frequency of error. *Matematicheskii Sbornik*, 31, 296–301.
- TORRIELLI, A., REPETTO, M. P. & SOLARI, G. 2013. Extreme wind speeds from long-term synthetic records. *Journal of Wind Engineering and Industrial Aerodynamics*, 115, 22–38.
- TSAY, R. S. 2013. *An introduction to Analysis of Financial Data with R*, New Jersey, US, John Wiley and Sons.
- TSE, Y. K. 1998. The conditional heteroskedasticity of the yen–dollar exchange rate. *Journal of Applied Econometrics*, 193, 49–55.
- VICENTE, S. 2012. *Extreme value theory: An application to sports*. MSc, University of Lisbon.
- VON MISES, R. 1954. La distribution de la plus grande de n valeurs. *American Mathematical Society: Providence RI II*, 271, 294.
- YAN, J. 2005. Asymmetry, Fat-tail, and Autoregressive Conditional Density in Financial Return Data with Systems of Frequency Curves. *Technical Report 355*. USA: Department of Statistics and Actuarial Science, University of Iowa.
- YANG, G. 2012. *The energy goodness-of-fit test for univariate stable distributions*. PhD, Bowling Green State University.
- YANG, X., CHATPATANASIRI, R. & SATTAYATHAM, P. 2016. Dynamic Risk Measurement of Financial Time Series with Heavy-Tailed: A New Hybrid Approach. *Thai Journal of Mathematics*, 14, 269.
- YOUNG, M. & GRAFF, R. 1995. Real estate is not normal: a fresh look at real estate return distributions. *Journal of Real Estate Finance and Economics*, 10, 225–259.

- ZHAO, X. 2010. *Extreme Value Modelling with Application in Finance and Neonatal Research*. PhD, University of Canterbury.
- ZHAO, X., SCARROTT, C. J., OXLEY, L. & REALE, M. 2010a. Extreme value modelling for forecasting market crisis impacts. *Applied Financial Economics* 20, 63–72.
- ZHOU, C. 2009. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100, 794-815.
- ZHOU, C. 2010. The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 101, 971-983.
- ŽIVKOVIĆ, S. & AKTAN, B. 2009. Global financial crisis and VaR performance in emerging markets: A case of EU candidate state - Turkey and Croatia. *Zbornik radova Ekonomskog fakulteta u Rijeci (Proceedings of Water Faculty of Economics)*, 27, 145-170.

Appendix A: CDF of PIVD in hyper geometric function

Let $u = \frac{t-\lambda}{a}$ in Equation 5.7, so that:

$$F(x) = ka \int_{-\infty}^{\frac{x-\lambda}{a}} (1+u^2)^{-m} e^{-v \tan^{-1} u} du$$

Replacing u with $\tan \theta$, so that:

$$F(x) = ka \int_{-\frac{\pi}{2}}^{\tan^{-1}\left(\frac{x-\lambda}{a}\right)} \cos^{2m-2} \theta e^{-v\theta} d\theta \quad (5.8)$$

Now, let $\phi = \theta + \frac{\pi}{2}$, then Equation 5.8 becomes:

$$F(x) = ka \int_0^{\tan^{-1}\left(\frac{x-\lambda}{a} + \frac{\pi}{2}\right)} \sin^{2m-2} \phi e^{v\frac{\pi}{2} - v\phi} d\phi \quad (5.9)$$

Given that:

$$I = \int_0^y \sin^r \phi e^{-v\phi} d\phi \quad (5.10)$$

and let $w = 1 - e^{-2i\phi} = -2ie^{i\phi} \sin \phi$, so that:

$$e^{-v\phi} = (1-w)^{\frac{iv}{2}}$$

and

$$\sin^r \phi = \left(\frac{i}{2}\right)^r w^r e^{-ri\phi} = (-2i)^{-r} w^r (1-w)^{-\frac{r}{2}}$$

also

$$d\phi = (-2i)^{-1} (1-w)^{-1} dw$$

From Equation 5.10, ϕ ranges from 0 to y , w ranges from 0 to z and so that:

$$z = 1 - e^{-2iy} = -2i e^{iy} \sin y \quad (5.11)$$

Therefore, Equation 5.10 becomes:

$$I = (-2i)^{-(r+1)} \int_0^z w^r (1-w)^{\frac{(iv-r-2)}{2}} dw \quad (5.12)$$

Equation 5.12 is an incomplete Beta function, which is related to the hyper-geometric function by:

$$\begin{aligned} I &= \frac{z^\alpha}{\alpha} F(\alpha, 1 - \beta; \alpha + 1; z) \\ &= \frac{z^\alpha (1-z)^\beta}{\alpha} F(1, \alpha + \beta; \alpha + 1; z) \end{aligned} \quad (5.13)$$

Now, substitute $\alpha = r + 1$ and $\beta = \frac{iv-r}{2}$ into Equation 5.13, so that:

$$I = \frac{z^{r+1} (1-z)^{\frac{iv-r}{2}}}{r+1} F\left(1, \frac{iv+r+2}{2}; r+2; z\right) \quad (5.14)$$

Since from Equation 5.11:

$$z^{r+1} = (-2i e^{iy} \sin y)^{r+1} = (-2i)^{r+1} e^{iy} e^{riy} \sin^{r+1} y$$

and

$$(1-z)^{\frac{iv-r}{2}} = [1 - (1 - e^{2iy})]^{\frac{iv-r}{2}} = e^{-vy} e^{-riy}$$

Hence, Equation 5.14 becomes:

$$I = \int_0^y \sin^r \phi e^{-v\phi} d\phi = \frac{e^{-vy} \sin^{r+1} y}{r+1} e^{iy} F\left(1, \frac{iv+r+2}{2}; r+2; -2i e^{iy} \sin y\right) \quad (5.15)$$

Applying Equation 5.15 to Equation 5.9, we have:

$$F(x) = ka \frac{e^{v\frac{\pi}{2}-vy} \sin^{2m-1} y}{2m-1} e^{iy} F\left(1, m + \frac{iv}{2}; 2m; -2i e^{iy} \sin y\right) \quad (5.16)$$

$$\text{where, } y = \tan^{-1}\left(\frac{x-\lambda}{a} + \frac{\pi}{2}\right)$$

so that:

$$e^{-vy} = e^{-v \tan^{-1}\left(\frac{x-\lambda}{a} + \frac{\pi}{2}\right)} \quad (5.17)$$

and,

$$e^{iy} = e^{i \tan^{-1}\left(\frac{x-\lambda}{a} + \frac{\pi}{2}\right)} = -\frac{(x-\lambda) + ia}{\sqrt{a^2 + (x-\lambda)^2}} \quad (5.18)$$

also,

$$\sin y = \sin\left[\tan^{-1}\left(\frac{x-\lambda}{a} + \frac{\pi}{2}\right)\right] = \frac{a}{\sqrt{a^2 + (x-\lambda)^2}} \quad (5.19)$$

then,

$$-2i e^{iy} \sin y = \frac{2}{1 - \frac{x-\lambda}{a} i} \quad (5.20)$$

Hence, applying Equation 5.17, 5.18, 5.19 and 5.20 to Equation 5.16 yields the cdf, $F(x)$ of Pearson type IV in form hyper-geometric function which is given as:

$$F(x) = \frac{ka}{2m-1} \left(1 + \left(\frac{x-\lambda}{a}\right)^2\right)^{-m} \exp\left[-v \tan^{-1}\left(\frac{x-\lambda}{a}\right)\right] \\ \times \left(i - \frac{x-\lambda}{a}\right) F\left(1, m + \frac{iv}{2}; 2m; \frac{2}{1 - \frac{x-\lambda}{a}i}\right)$$

which can also be written as:

$$F(x) = \frac{ka}{2m-1} f(x) \left(i - \frac{x-\lambda}{a}\right) F\left(1, m + \frac{iv}{2}; 2m; \frac{2}{1 - \frac{x-\lambda}{a}i}\right) \quad (5.21)$$

which converges absolutely when $x < \lambda - a\sqrt{3}$ but for the case of $x > \lambda + a\sqrt{3}$, () suggest the use of identity given as:

$$F(m, v, a, \lambda; x) = 1 - F(m, -v, a, -\lambda; x)$$

and when $|x - \lambda| < a\sqrt{3}$ a linear transformations described in Abramovitz and A. (1964) can be used, as a result Equation 5.21 becomes:

$$F(x) = f(x) \frac{ia}{2m - iv - 2} \left[1 + \left(\frac{x-\lambda}{a}\right)^2\right] F\left(1, 2 - 2m; 2 - m + \frac{iv}{2}; \frac{1 - \frac{x-\lambda}{a}i}{2}\right) \\ + \frac{1}{1 + \exp[-(v + 2im)\pi]} \quad (5.22)$$

Appendix B: Model estimate results

APACH (1,1) model with heavy- tailed distributions

The JSE ALSI returns are fitted to APARCH (1,1) with heavy-tail innovations: student t distribution (std), skewed student t distribution (sskew), generalized error distribution (ged), skew generalized error distribution (sged), inverse Gaussian distribution (nig), generalized hyperbolic distribution (ghyp), and Johnson's S_u distribution (jsu). The AIC and BIC are used to compare these models. Table B.2 shows the result of the model selection criteria.

Table B.1: Model selection of APACH (1,1) model with heavy- tailed distribution

Information criteria	std	sstd	ged	sged	nig	ghyp	jsu
AIC	-6.1571	-6.1762	-6.1546	-6.1749	-6.1753	-6.1747	-6.1753
BIC	-6.1413	-6.1578	-6.1388	-6.1565	-6.1569	-6.1536	-6.1568

Appendix C: R Codes

```
#####Test for stationarity#####
data<-(ALSI.WF)[,c("Close")]
ALSI=diff(log(data))
par(mfrow=c(1,2))
cod=ts(data,frequency=249,start=c(2005,5))
plot(cod, type= 'l', xlab= 'Year', ylab= "closing price")
cod2=ts(ALSI,frequency=249,start=c(2005,5))
plot(cod2, type= 'l', xlab= 'Year', ylab= "closing price")
par(mfrow=c(1,1))
adf.test(ALSI)
pp.test(ALSI)
kpss.test(ALSI)
#####Descriptive Statistics#####
basicStats(ALSI)
jarqueberaTest(ALSI)
qqnorm(ALSI);qqline(ALSI)
#####Checking for serial correlation and volatility clustering of returns#####
par(mfrow=c(1,2))
acf(ALSI)
Box.test(ALSI, lag =20, type = "Ljung-Box", fitdf = 0)
dwttest(ALSI ~ ALSI)
acf(ALSI^2)
Box.test(ALSI^2, lag = 20, type = "Ljung-Box", fitdf = 0)
ArchTest(ALSI, lags=20, demean = FALSE)
#####Test for Asymmetric effect#####
spech = ugarchspec(variance.model = list(garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE))
garchr= ugarchfit(spec=spech,data= ALSI[1:2155])
signbias(garchr)
```

```

#####Fitting asymmetric GARCH models#####
spech = ugarchspec(variance.model = list(model="eGARCH",garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE))
Egarchr= ugarchfit(spec=spech,data= ALSI[1:2155])
Tspech = ugarchspec(variance.model = list(model="gjrGARCH",garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE))
Tgarchr= ugarchfit(spec=Tspech,data= ALSI[1:2155])
speca = ugarchspec(variance.model = list(model="apARCH",garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE),fixed.pars=list(delta=1))
Agarchr2= ugarchfit(spec=speca,data= ALSI[1:2155])
#####Model diagnosis#####
resi= residuals(Agarchr2, standardize= TRUE)
basicStats(resi)
resi=as.vector(resi)
Resi=resi*-1
basicStats(Resi)
#####test for autocorrelation and volatility clustering#####
dwtest(resi ~ resi)
acf(resi)
Box.test(resi, lag = 20, type = "Ljung-Box", fitdf = 0)
acf(resi^2)
Box.test(resi^2, lag =20, type = "Ljung-Box", fitdf = 0)
ArchTest(resi, lags=20, demean = FALSE)
#####Test for IID#####
bartels.rank.test(resi)
bds.test(resi)
cox.stuart.test(resi)
#####test for normality#####
shapiro.test(resi)
jarque.bera.test(resi)
##### fitting APARCH with sstd#####

```

```

spec = ugarchspec(variance.model = list(model="apARCH",garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE),distribution.model =
"sstd",fixed.pars=list(delta=1))

ssAgarchr= ugarchfit(spec=spec,data= ALSI[1:2155])

#####Goodnees-of-fit for sstd#####

ssresi= residuals(ssAgarchr, standardize= TRUE)

basicStats(ssresi)

ssresi=as.vector(ssresi)

ad.test(ssresi, psstd,0.008367,0.997462,23.459483,0.804057)

#####Threshold selection#####

mrl.plot(resi)

mrl.plot(Resi)

gpd.fitrage(resi,1.00,2.00,nint=20)

gpd.fitrage(Resi,1.00,2.00,nint=20)

paretoQPlot(resi)

paretoQPlot(Resi)

#####fitting GPD and model diagnosis#####3

out <- gpd(resi, 1.25)

SieGpd <- gpd.fit(out$data,1.25)

gpd.diag(SieGpd)

Rout <- gpd(Resi, 1.56)

SieGpd2 <- gpd.fit(Rout$data,1.56)

gpd.diag(SieGpd2)

#fitting GEVD and model diagnosis#

out.gev <- gev(resi, 5)

SieGEV <- gev.fit(out.gev$data)

gev.diag(SieGEV)

ad.test(out.gev$data, pgev,xi =-0.08523683, mu =0.90270379, sigma =0.65348487)

Rout.gev <- gev(Resi, 5)

SieGEV2 <- gev.fit(Rout.gev$data)

gev.diag(SieGEV2)

ad.test(Rout.gev$data, pgev,xi =-0.08523683, mu =0.90270379, sigma =0.65348487)

```

```

#####Fitting stable distribution and model diagnosis#####
stable.fit(resi,method=1,param=0)
theta=c(1.9159194 ,-1.0000000,0.6775130,0.1205831)
stable.ks.gof(resi,theta,method=0,param=0)
ad.test(resi,      pstable,alpha=1.9159194,      beta=-1.0000000,      gamma=0.6775130,
delta=0.1205831)
ppstable(resi, theta1, var.stabilized = FALSE, param = 0)
#####FittingPearsonTypeIVandmodeldiagnosis#####
pearsonFitML(resi)
ad.test(Aresi, ppearsonIV,m=12.66659,nu=11.73608,location= 2.129829,scale=4.222132)
#####In-sample Backtesting#####
#####In-samples_skewshort#####
IsstdVaR1 = qdist("sstd", 0.90,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR2 = qdist("sstd", 0.95,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR3 = qdist("sstd", 0.975,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR4 = qdist("sstd", 0.99,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
VT1=VaRTest(0.90,resi,rep(IsstdVaR1,2155))
VT2=VaRTest(0.95, resi,rep(IsstdVaR2,2155))
VT3=VaRTest(0.975, resi,rep(IsstdVaR3,2155))
VT4=VaRTest(0.99, resi,rep(IsstdVaR4,2155))
#####In-samples_skewlong#####
IsstdVaR1 = qdist("sstd", 0.1,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR2 = qdist("sstd", 0.05,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR3 = qdist("sstd", 0.025,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR4 = qdist("sstd", 0.01,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)
IsstdVaR5 = qdist("sstd", 0.005,mu =0.008418, sigma = 0.997427,skew = 0.803948,
shape=23.415740)

```



```

VT1=VaRTest(0.1,ssresi,rep(IsstdVaR1,2155))
VT2=VaRTest(0.05,ssresi,rep(IsstdVaR2,2155))
VT3=VaRTest(0.025,ssresi,rep(IsstdVaR3,2155))
VT4=VaRTest(0.01,ssresi,rep(IsstdVaR4,2155))
#####In-samples_pearsonshort#####
IpearsonVaR1=qpearsonIV(0.90,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR2=qpearsonIV(0.95,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR3=qpearsonIV(0.975,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR4=qpearsonIV(0.99,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
VT1=VaRTest(0.90,resi,rep(IpearsonVaR1,2155))
VT2=VaRTest(0.95,resi,rep(IpearsonVaR2,2155))
VT3=VaRTest(0.975,resi,rep(IpearsonVaR3,2155))
VT4=VaRTest(0.99,resi,rep(IpearsonVaR4,2155))
#####In-samples_pearsonlong#####
IpearsonVaR1=qpearsonIV(0.1,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR2=qpearsonIV(0.05,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR3=qpearsonIV(0.025,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR4=qpearsonIV(0.01,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
IpearsonVaR5=qpearsonIV(0.005,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
VT1=VaRTest(0.1,resi,rep(IpearsonVaR1,2155))
VT2=VaRTest(0.05,resi,rep(IpearsonVaR2,2155))
VT3=VaRTest(0.025,resi,rep(IpearsonVaR3,2155))
VT4=VaRTest(0.01,resi,rep(IpearsonVaR4,2155))
#####Insampleshort_stable#####
IstableVaR1 = qstable(0.90,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)
IstableVaR2 = qstable(0.95,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

```

IstableVaR3 = qstable(0.975,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

IstableVaR3

IstableVaR4 = qstable(0.99,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

VT1=VaRTest(0.90,resi,rep(IstableVaR1,2155))

VT2=VaRTest(0.95,resi,rep(IstableVaR2,2155))

VT3=VaRTest(0.975,resi,rep(IstableVaR3,2155))

VT4=VaRTest(0.99,resi,rep(IstableVaR4,2155))

#####in-samplelong_stable#####

IstableVaR1 = qstable(0.1,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

IstableVaR2 = qstable(0.05,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

IstableVaR3 = qstable(0.025,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

IstableVaR4 = qstable(0.01,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

VT1=VaRTest(0.1, resi,rep(IstableVaR1,2155))

VT2=VaRTest(0.05, resi,rep(IstableVaR2,2155))

VT3=VaRTest(0.025,resi,rep(IstableVaR3,2155))

VT4=VaRTest(0.01,resi,rep(IstableVaR4,2155))

#####Insampleshort_GEVD#####

IgevVaR1 =qgev(0.90, xi = -0.179686,mu =0.9101711,sigma =0.5547224)

IgevVaR2 =qgev(0.95, xi = -0.179686,mu =0.9101711,sigma =0.5547224)

IgevVaR3 =qgev(0.975, xi = -0.179686,mu =0.9101711,sigma =0.5547224)

IgevVaR4 =qgev(0.99, xi = -0.1796867, mu =0.9101711, sigma = 0.5547224)

VT1=VaRTest(0.90,resi,rep(IgevVaR1,2155))

VT2=VaRTest(0.95,resi,rep(IgevVaR2,2155))

VT3=VaRTest(0.975,resi,rep(IgevVaR3,2155))

VT4=VaRTest(0.99,resi,rep(IgevVaR4,2155))

#####longposition_GEVD#####

IgevVaR1 =qgev(0.1, xi = -0.08524132,mu =0.90272571,sigma =0.65347411)

IgevVaR2 =qgev(0.05, xi = -0.08524132,mu =0.90272571,sigma =0.65347411)

IgevVaR3 =qgev(0.025, xi = -0.08524132,mu =0.90272571,sigma =0.65347411)

IgevVaR4 =qgev(0.01, xi = -0.08524132,mu =0.90272571,sigma =0.65347411)

VT1=VaRTest(0.1,Resi,rep(IgevVaR1,2155))

```

VT2=VaRTest(0.05,Resi,rep(IgevVaR2,2155))
VT3=VaRTest(0.025,Resi,rep(IgevVaR3,2155))
VT4=VaRTest(0.01,Resi,rep(IgevVaR4,2155))
#####Insampleshortposition_GPD#####
riskmeasures(out, c(0.900,0.950,0.975,0.990))
VT1=VaRTest(0.90,resi,rep(1.210929,2155))
VT2=VaRTest(0.95,resi,rep(1.505440,2155))
VT3=VaRTest(0.975,resi,rep(1.783939,2155))
VT4=VaRTest(0.99,resi,rep(2.128996,2155))
#####Insamplelongposition_GPD#####
riskmeasures(Rout, c(0.900,0.950,0.975,0.990 ))
VT1=VaRTest(0.1, Resi,rep(-0.003206271,2155))
VT2=VaRTest(0.05,Resi,rep(-0.037792434,2155))
VT3=VaRTest(0.025,Resi,rep(-0.054437553,2155))
VT4=VaRTest(0.01,Resi,rep(-0.064229732,2155))
#####OUT-SAMPLEBACKTESTING#####
#####shortpostion_skew#####
spec2=spec
setfixed(spec2)<-as.list(coef(ssAgarchr))
filt = ugarchfilter(spec2, ALSI[2156:2757])
actual = ALSI[2156:2757]
sstdVaR1 = fitted(filt) + sigma(filt)*qdist("sstd", 0.90,mu =0.008418, sigma = 0.997427,skew
= coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR2 = fitted(filt) + sigma(filt)*qdist("sstd", 0.95,mu =0.008418, sigma = 0.997427,skew
= coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR3 = fitted(filt) + sigma(filt)*qdist("sstd", 0.975,mu =0.008418, sigma =
0.997427,skew = coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR4 = fitted(filt) + sigma(filt)*qdist("sstd", 0.99,mu =0.008418, sigma = 0.997427,skew
= coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
VT1=VaRTest(0.90, as.numeric(actual), as.numeric(sstdVaR1))
VT2=VaRTest(0.95, as.numeric(actual), as.numeric(sstdVaR2))
VT3=VaRTest(0.975, as.numeric(actual), as.numeric(sstdVaR3))
VT4=VaRTest(0.99, as.numeric(actual), as.numeric(sstdVaR4))

```

```
#####longposition_skew#####
sstdVaR1 = fitted(filt) + sigma(filt)*qdist("sstd", 0.1, mu = 0, sigma = 1,skew =
coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR2 = fitted(filt) + sigma(filt)*qdist("sstd", 0.05, mu = 0, sigma = 1,skew =
coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR3 = fitted(filt) + sigma(filt)*qdist("sstd", 0.025, mu = 0, sigma = 1,skew =
coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
sstdVaR4 = fitted(filt) + sigma(filt)*qdist("sstd", 0.01, mu = 0, sigma = 1,skew =
coef(ssAgarchr)["skew"], shape=coef(ssAgarchr)["shape"])
VT1=VaRTest(0.1, as.numeric(actual), as.numeric(sstdVaR1))
VT2=VaRTest(0.05, as.numeric(actual), as.numeric(sstdVaR2))
VT3=VaRTest(0.025, as.numeric(actual), as.numeric(sstdVaR3))
VT4=VaRTest(0.01, as.numeric(actual), as.numeric(sstdVaR4))
#####shortposition_pearsonIV#####
spec2=spec2
setfixed(spec2)<-as.list(coef(Agarchr2))
filt = ugarchfilter(spec2, ALSI[2156:2757])
actual = ALSI[2156:2757]
pearsonVaR1 = fitted(filt) +
sigma(filt)*qpearsonIV(0.90,m=12.66659,nu=11.73608,location= 2.129829,scale=4.222132)
pearsonVaR2 = fitted(filt) +
sigma(filt)*qpearsonIV(0.95,m=12.66659,nu=11.73608,location= 2.129829,scale=4.222132)
pearsonVaR3 = fitted(filt) +
sigma(filt)*qpearsonIV(0.975,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
pearsonVaR4 = fitted(filt) +
sigma(filt)*qpearsonIV(0.99,m=12.66659,nu=11.73608,location= 2.129829,scale=4.222132)
VT1=VaRTest(0.90, as.numeric(actual), as.numeric(pearsonVaR1))
VT2=VaRTest(0.95, as.numeric(actual), as.numeric(pearsonVaR2))
VT3=VaRTest(0.975, as.numeric(actual), as.numeric(pearsonVaR3))
VT4=VaRTest(0.99, as.numeric(actual), as.numeric(pearsonVaR4))
#####longposition_pearsonIV#####
pearsonVaR1 = fitted(filt) + sigma(filt)*qpearsonIV(0.1,m=12.66659,nu=11.73608,location=
2.129829,scale=4.222132)
```

pearsonVaR2= fitted(filt) + sigma(filt)*qpearsonIV(0.05,m=12.66659,nu=11.73608,location=2.129829,scale=4.222132)

pearsonVaR3=fitted(filt)+sigma(filt)*qpearsonIV(0.025,m=12.66659,nu=11.73608,location=2.129829,scale=4.222132)

pearsonVaR4= fitted(filt) + sigma(filt)*qpearsonIV(0.01,m=12.66659,nu=11.73608,location=2.129829,scale=4.222132)

VT1=VaRTest(0.1, as.numeric(actual), as.numeric(pearsonVaR1))

VT2=VaRTest(0.05, as.numeric(actual), as.numeric(pearsonVaR2))

VT3=VaRTest(0.025, as.numeric(actual), as.numeric(pearsonVaR3))

VT4=VaRTest(0.01, as.numeric(actual), as.numeric(pearsonVaR4))

#####shortposition_stable#####

stableVaR1=fitted(filt)+sigma(filt)*qstable(0.90,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR2=fitted(filt)+sigma(filt)*qstable(0.95,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR3=fitted(filt)+sigma(filt)*qstable(0.975,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR4=fitted(filt)+sigma(filt)*qstable(0.99,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

VT1=VaRTest(0.90, as.numeric(actual), as.numeric(stableVaR1))

VT2=VaRTest(0.95, as.numeric(actual), as.numeric(stableVaR2))

VT3=VaRTest(0.975, as.numeric(actual), as.numeric(stableVaR3))

VT4=VaRTest(0.99, as.numeric(actual), as.numeric(stableVaR4))

#####longposition_stable#####

stableVaR1=fitted(filt)+ sigma(filt)*qstable(0.1,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR2=fitted(filt)+sigma(filt)*qstable(0.05,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR3=fitted(filt)+sigma(filt)*qstable(0.025,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

stableVaR4=fitted(filt)+sigma(filt)*qstable(0.01,1.9159194,-1.0000000 ,0.6775130,0.1205831 ,0)

VT1=VaRTest(0.1, as.numeric(actual), as.numeric(stableVaR1))

VT2=VaRTest(0.05, as.numeric(actual), as.numeric(stableVaR2))

VT3=VaRTest(0.025, as.numeric(actual), as.numeric(stableVaR3))

```

VT4=VaRTest(0.01, as.numeric(actual), as.numeric(stableVaR4))
#####shortposition_GEVD#####
gevVaR1 =fitted(filt)+sigma(filt)*qgev(0.90, xi = -0.179686,mu =0.9101711,sigma
=0.5547224)
gevVaR2 =fitted(filt)+sigma(filt)*qgev(0.95, xi = -0.179686,mu =0.9101711,sigma
=0.5547224)
gevVaR3 = fitted(filt)+sigma(filt)*qgev(0.975, xi = -0.179686,mu =0.9101711,sigma
=0.5547224)
gevVaR4 = fitted(filt)+sigma(filt)*qgev(0.99, xi = -0.1796867, mu =0.9101711, sigma =
0.5547224)
VT1=VaRTest(0.90, as.numeric(actual), as.numeric(gevVaR1))
VT2=VaRTest(0.95, as.numeric(actual), as.numeric(gevVaR2))
VT3=VaRTest(0.975, as.numeric(actual), as.numeric(gevVaR3))
VT4=VaRTest(0.99, as.numeric(actual), as.numeric(gevVaR4))
#####longposition_GEVD#####
RgevVaR1 = fitted(filt) + sigma(filt)*qgev(0.1,xi =-0.08523683, mu =0.90270379, sigma
=0.65348487)
RgevVaR2 = fitted(filt) + sigma(filt)*qgev(0.05,xi =-0.08523683, mu =0.90270379, sigma
=0.65348487)
RgevVaR3 = fitted(filt) + sigma(filt)*qgev(0.025,xi =-0.08523683, mu =0.90270379, sigma
=0.65348487)
RgevVaR4 = fitted(filt) + sigma(filt)*qgev(0.01,xi =-0.08523683, mu =0.90270379, sigma
=0.65348487)
VT1=VaRTest(0.1, as.numeric(actual), as.numeric(RgevVaR1))
VT2=VaRTest(0.05, as.numeric(actual), as.numeric(RgevVaR2))
VT3=VaRTest(0.025, as.numeric(actual), as.numeric(RgevVaR3))
VT4=VaRTest(0.01, as.numeric(actual), as.numeric(RgevVaR4))
#####shortposition_GPD#####
gpdVaR1 = fitted(filt) + sigma(filt)*1.210929
gpdVaR2 = fitted(filt) + sigma(filt)*1.505440
gpdVaR3 = fitted(filt) + sigma(filt)* 1.783939
gpdVaR4 = fitted(filt) + sigma(filt)*2.128996
VT1=VaRTest(0.90, as.numeric(actual), as.numeric(gpdVaR1))
VT2=VaRTest(0.95, as.numeric(actual), as.numeric(gpdVaR2))

```

```
VT3=VaRTest(0.975, as.numeric(actual), as.numeric(gpdVaR3))
VT4=VaRTest(0.99, as.numeric(actual), as.numeric(gpdVaR4))
#####longposition_GPD#####
RgpdVaR1 = fitted(filt) + sigma(filt)*-0.003206271
RgpdVaR2 = fitted(filt) + sigma(filt)*-0.037792434
RgpdVaR3 = fitted(filt) + sigma(filt)*-0.054437553
RgpdVaR4 = fitted(filt) + sigma(filt)*-0.064229732
VT1=VaRTest(0.05, as.numeric(actual), as.numeric(RgpdVaR1))
VT2=VaRTest(0.01, as.numeric(actual), as.numeric(RgpdVaR2))
VT3=VaRTest(0.005, as.numeric(actual), as.numeric(RgpdVaR3))
VT4=VaRTest(0.001, as.numeric(actual), as.numeric(RgpdVaR4))
```