Application of Volatility and Extreme Value Theory in Exchange Rate Risk Estimation

Ntawihebasenga Jean De Dieu

MF 300-0006/12

A Thesis submitted in partial fulfillment for the degree of Master of Science in Mathematics – Financial Option in the Pan African University Institute for Basic Sciences Technology and Innovation

2014

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university
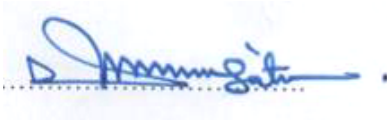
Signature …………………… Date ……………………………….

Ntawihebasenga Jean De Dieu

This thesis has been submitted for examination with our approval as the University supervisors

Signature …………………………. Date ………………………….

Prof. Peter N. Mwita

Sugnature ………………………. Date 3 | 11 | 2014.

Dr. Joseph K. Mung'atu

JKUAT, Kenya

DEDICATION

This work is dedicated to my Brother and my sister for giving me easy moment during my studies; it is dedicated to the family of Mr. Nsengimana J. Serge for supporting me in many ways.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FREQUENTLY USED NOTATIONS

AIC   Akaike Information Criterion

ARCH   Autoregressive Conditional Heteroscedasticity

BIC   Bayesian Information Criterion

BMM Block Maxima Model

CES_φ^t Conditional Expected Shortfall at φ Probability level

$CVaR_\varphi^t$ Conditional Value-at-Risk at φ Probability level

ES Expected Shortfall

EVT Extreme Value Theory

GARCH   Generalized ARCH

GEV Generalized Extreme Value

GPD Generalized Pareto Distribution

M_n Limit law for the block maxima

POT Peak over Threshold

UES_φ Unconditional at φ Probability level

VaR Value-at-Risk

$UVaR_\varphi$ Unconditional Value-at-Risk at φ Probability level

r_t	Daily Exchange Rate return at time t
f	Probability density function
F_τ	Conditional excess distribution function
θ	Vector parameter
υ_τ	Exceedance random variables over threshold τ
ψ	Scale parameter
ℓ	Location parameter
$\mathfrak{D}$	Maximum Domain of Attraction
Θ	Parameter vector space
μ	Conditional Expectation of the exchange rate returns
δ	Tail index of the distribution
ζ	Shape parameter
$\mathbb{R}$	Set of real numbers
CQ_φ^t	Conditional Quantile at time t with probability level φ
$\mathcal{H}$	Non-generate distribution function
e_φ	Quantile at probability level φ

ABSTRACT

The importance of conditional Value-at-Risk and conditional Expected Shortfall to estimate extreme risk in financial time series data cannot be exaggerated. This study applies these tools to estimate extreme risk in exchange rate returns. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is applied to estimate the current volatility in daily exchange rate returns over the period of 10 years and Extreme Value Theory (EVT) approach to estimate quantiles of innovations. First, Autoregressive (AR) model is fitted with GARCH errors to the daily exchange rate returns using Quasi-Maximum likelihood Estimate (Q-MLE) to get the current volatility. Second, Generalized Pareto Distribution (GPD) approach is fitted to the excess returns assuming the residuals are independent and identically distributed. The asymptotic properties of the estimators are given. Finally, the estimated volatility and estimated quantiles are combined to obtain Conditional Value at Risk and Conditional estimates. Results are applied to real data to estimate extreme risk in Rwanda exchange rate process.

CHAPTER ONE

1.0 INTRODUCTION AND BACKGROUND INFORMATION

1.1 Background Information

Exchange rates are a challenging concept due to the fact that one has to deal with foreign exchange rates whenever he/she travels to foreign country. They play a crucial role in a country's level of trade. Markets of Exchange rates are world decentralized market places that determine the relative values of different currencies.

Risk is a random variable transforming unforeseen future states of the world into values representing profits and losses. It is a common phenomenon in all areas of finance. Risk in foreign exchange can be defined as exposure to uncertainty and it cannot be dismissed in exchange markets since both importers and exporters of goods and services are affected by exchange rates fluctuations. This risk refers to a financial risk posed by an exposure to unanticipated changes in the exchange rate between two currencies. It may also be defined as the variability of a portfolio's value caused by uncertain fluctuations in the exchange rates.

Exchange rate risk is related to the effect of unexpected exchange rate changes on the value of a firm (Madura, 1989). The value of any currency fluctuates as its demand and supply fluctuates, this means that if demand decrease or supply increase this can cause depreciation of the currency's value. On other hand if supply decreases and demand increases this can cause appreciation of the value of currency. There are three main types of exchange rate risk: *Transaction risk (*cash flow risk and deals with the

effect of exchange rate), *Translation risk* (balance sheet exchange rate risk) and *Economic risk (*reflects the risk to the firm's present value of future operating cash flows from exchange rate movements). To deal with exchange rate risk, a firm needs to determine the specific type of currency risk exposure, factors influencing exchange rate and also find out a suitable technique for risk estimation.

In this study; we are concerned with the estimation of extreme risk due to the exchange rates fluctuations. The modern era of risk measurement and estimation for exchange rate positions started in 1973. The risk managers and investors became concerned about the impact of exchange rate fluctuations on portfolios. Thereafter, exchange rates are among the most watched analyzed and governmentally manipulated economic measures. Exchange rate fluctuations have become an essential subject in macroeconomic analysis and have received a great deal of interest from academics, financial economists and policymakers, particularly after the collapse of the Bretton Woods agreement of fixed exchange rates among major industrialized countries. This prompted the research for more appropriate methodologies to deal with rare events that have big effects.

As the financial world focus on risk management, various models have been developed to estimate risk in financial data. Most common used risk measures are value at risk and Expected Shortfall. Value at risk summarizes the worst loss over a target horizon that will not be exceeded with a given level of confidence. In other words, Value at Risk answers the question about, how much one can lose over a certain time horizon with a given probability. It also summarizes in a single number

the overall market risk faced by a financial institution, (Jorian, 2007). Expected Shortfall (coherent risk measure) is an expected value of the loss, given that a VaR violation occurred. In other word, Expected Shortfall estimates the potential size of the loss exceeding the VaR at φ probability level (Delbaen's, 2002).

Last decades many techniques of estimating risk in exchange rates have been developed, but those that seem to be more successful and popular are: Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models and extreme value theory (EVT) approach. Empirical evidence found that financial return series such as exchange rate returns exhibit certain stylized facts such as volatility clustering, heavy-tailedness, heteroskedasticity and non-linearity (Gravril & Altar, 2009). The GARCH family models were introduced to deal with these problems of pattern facts in financial data. GARCH models manage changing volatility assuming normality of the data. However, the assumption of conditional normality does not seem to hold for real data since the VaR based on such models has difficulties in capturing the extreme events (Bollerslev, 1986).

VaR estimations are only related to the tails of a probability distribution while extreme value theory (EVT) focuses directly on the tails. The Extreme value theory (EVT) provides a solid framework to formally study the behavior of extreme observations and using it in estimating VaR could give better forecasts of risk. However, applying EVT to the exchange rate return series is inappropriate as they are not independently and identically distributed and the current volatility background is not taken into account (Soltane *et al*, 2012).

In addition, Gravril & Altar (2009) applied exchange rate returns of selected countries versus Euro to test the fit of EVT as complementary risk management tool for stressed market conditions and analyzed various VaR models. They showed that in extreme market conditions, extreme measures are required and no single model can perform proper for both the centre and the tails of an exchange rate distribution. For this reason, this study, applies GARCH model to estimate conditional volatility and Extreme Value theory (EVT) particularly, the peak over threshold (POT) model where tails are estimated by fitting Generalized Pareto Distribution (GPD) to estimate conditional Value-at-Risk and Conditional Expected Shortfall.

1.2   Statement of the Problem

In the recent years, the exchange rate movements and fluctuations have become a serious challenge to the growing economy and have received a great deal of interest from academicians, financial economists, and policy makers (Papaioannou, 2006). There has been considerable amount of research on estimation of exchange rate risk using different approaches in estimation of extreme Value-at-Risk; however, no single model has performed well for both the centre and the tails of exchange rate distribution (i.e capturing volatility clustering and extreme events) , for example see (Gravril & Altar, 2009). In order to capture volatility clustering and extreme events, this study has used conditional Value-at-Risk and conditional Expected Shortfall estimates obtained by combining conditional volatility with extreme quantiles estimates.

1.3   Objectives of the Study

1.3.1   Main objective

The overall objective is to estimate extreme risk in exchange rate process with application to Rwandese exchange rates

1.3.2   Specific objectives

1.   To estimate exchange rate volatility using GARCH model

2.   To use Generalized Pareto Distribution to estimate extreme quantile assuming the residuals are independent and identically distributed data.

3.   To study asymptotic properties of the estimated parameters

4.   To combine the estimated volatility with estimated extreme quantiles to obtain conditional Value at Risk and conditional Expected Shortfall estimates for the risk.

1.4   Justification

The estimation of extreme risk in exchange rate will assist risk managers, portfolio managers, traders, investors and market makers in different manners. For instance, it can help the risk practitioners to monitor the exposure of market risk, therefore, protecting their institution from collapsing. The conditional Value at risk and conditional Expected Shortfall estimates help risk managers to understand the position of their institutions thus making them actively involved in policies on risk management. They can also be used to ensure that financial institutions can still be in business after a catastrophic event. In general, results of this work will contribute a lot to understanding of how changes in exchange rate affect the prices of goods and services in international trade.

1.5   Scope of the Study

The work used daily exchange rates of Rwanda francs (Frw) against Kenya Shillings (Ksh), US Dollars (USD), Euro and sterling GBP (GBP) respectively for the time period between 1st January 2002 and 31st December 2012. Data were provided by The National Bank of Rwanda (BNR).

1.6   Outline of the Thesis

This work is organized as follow, Chapter two focuses on literature review, model specification and definitions of some basics concepts applied in this work. Chapter three involves estimation of exchange rate volatility using Generalized Autoregressive Conditional Heteroskedasticity model with assumption that innovations follow a normal distribution. The chapter also studies asymptotic properties of the estimators. Chapter four deals with estimation of tails of distribution using Generalized Pareto Distribution with assumption that the residuals are independent and identically distributed. This chapter also provides asymptotic properties of estimated parameters and it obtains the conditional Value at Risk and conditional Expected Shortfall estimates. Chapter five focuses on empirical analysis of Rwanda exchange rate returns and discussion of the results. The last chapter concludes and gives recommendations for further study, based on this work.

1.7   Summary

This chapter has introduced Generalized Autoregressive Conditional Heteroscedasticity and Extreme Value Theory and their application in risk estimation. The statement of the problem, research questions, objectives of the thesis, justification of the study and the scope of the study were also addressed in this chapter.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Introduction

The main purpose of this section is to have an overall view on estimation of exchange rate volatility and a more detailed presentation of basic Extreme Value Theory framework. This enables to gain an insight of our research on risk estimation. The model and other basics concepts used in this research have been defined.

2.2 Review on Exchange Rate Volatility

Volatility refers to the spread of all likely outcomes of an uncertain variable. Statistically, volatility is often measured as the sample standard deviation and can be defined mathematically as follows

$$\hat{\sigma} = \left\{ \frac{1}{T-1} \sum_{t=1}^{T} (r_t - \mu)^2 \right\}^{1/2}, \text{t=1, 2,...,T} \tag{2.1}$$

Where r_t represents daily returns to the exchange rate at time t and μ represents the average return over the T days period. Sometimes the variance σ^2 is also considered as volatility. The volatility and expected returns estimates are based on an Autoregressive model with GARCH innovations.

Volatility is related to but not exactly the same as risk. Risk is associated with undesirable outcome, but volatility can be defined as a measurement of the change in price over a given period of time. Exchange rate volatility is a measure of the

fluctuations in an exchange rate. When volatility in exchange rate increases it leads to uncertainty in pricing and this hurts importers who spend more for the same quantity while exporters benefit from this depreciation of local currency. The volatility in prices has implications on the profits and survival of business enterprises (Smith *et al*., 1990).

2.2.1 Stylized Facts about Volatility

Financial time series such as exchange rates, stocks returns and other financial time series are known to exhibit certain stylized patterns which are crucial for correct model specification, estimation and forecasting. The most common stylized facts are fat tails, volatility clustering & persistence and leverage effects.

a) Fat tails

The fourth central moment that measures the tail behavior of a continuous random variable X is called kurtosis denoted by $K(x)$. Mathematically $K(x)$ can be defined as

$$K(x) = E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right] \tag{2.2}$$

Where μ_x represents the first central moment called mean and σ_x^4 represents square of second central moment called variance. The underlying distribution is also referred to as positive excess kurtosis (leptokurtosis) if the quantity of kurtosis is greater than three $i.e. K(x) - 3 > 0$. This indicates that the underlying distribution has Fat tails behavior since for normal distribution $K(x) - 3 = 0$. On the other hand, the distribution with negative excess kurtosis $i.e. K(x) - 3 < 0$ has short tails and such distribution is called platykurtic (Ruey, 2005).

9

Researchers in the past years found that most of the times, the distributions of financial assets returns specifically, exchange rate returns are not normal, (Hull & White, 1998). Some researchers argue that the distribution should have fat tails, ((Longin, 1996), (Neftci, 2000)) and others argue that it should not be symmetric (Glosten *et al*., 1993).

b)　Volatility clustering and persistence

In financial time series volatility clustering means that small and large changes in the series tend to occur in clusters. This means that the large changes tend to be followed by large changes and small changes tend to be followed by small changes. $i.e.$ When volatility is high it is likely to remain high and when it is low it is likely to remain low (Engle, 2004). Manganelli & Engle (2001) showed that exchange rates and interest rate returns are not normally distributed, suffer from volatility clustering and are not independent.

c)　Leverage Effects

In financial market, leverage effects is a stylized fact in which depreciation (a downward movement) is always followed by higher volatility (volatility is high after negative shocks) while appreciation (upward movement) is followed by low volatility (volatility is low after positive shocks). There is evidence that volatility is higher after negative shocks than after positive shocks of the same magnitude (Nelson, 1991).

Maana *et al*. (2010) estimated exchange rate volatility of Kenya Markets using GARCH (1, 1) model. They showed that the importers and exporters of goods and services are both affected by exchange rate fluctuations. Andersen & Sorensen

(1996), Ghysels *et al*. (1996) and Sandmann & Koopman (1998) estimated the volatility as non-constant and non-symmetric with left fat tail. They argued that the true volatility cannot be estimated because there is no relationship between prior, current, and future volatilities for financial assets. If so, approaches utilizing volatility in estimating VaR should be invalid.

The foreign exchange rates can be subject to considerable daily fluctuations (up to 5 percent within one day); this can cause serious losses on open overnight positions. The risk can be quantified by focusing on the tails of the distribution and using estimations one can compute limits that a risk manager can set to open positions to avoid unexpected huge losses (Blum & Dacorogna, 2002). Some researchers have found that the distributions of financial assets returns are not constant over time. Such findings are related to another field of research in finance: the prediction of volatility of financial assets. There have been a lot of debate about the attributes of volatility; whether volatility is time-varying or constant, whether it should be weighed through time or not, or what time interval from the past is relevant for current volatility (Nelson, 1991).

2.3   Review on Extreme Value Theory

Extreme Value Theory is a well developed theory in the field of probability that studies the distribution of extreme realizations of a given distribution function, or of a stochastic process, satisfying suitable assumptions. The foundations of the theory were set by Fisher and Tippett (1928) and Gnedenko (1943), who proved that the distribution of the extreme values of an independent and identically distributed (iid)

sample, can converge to one out of only three possible distributions (Acerbi *et al.*, 2001).

Many researchers have oriented their work towards more efficient tail-oriented models of risk, namely Extreme Value Theory (EVT) approach. The superiority of EVT has been extensively demonstrated by many researchers, in fields like insurance or financial risk management. The EVT approach was applied to assess fat tails of different time series, like hydrologic, insurance and financial data, supported by a very detailed and complex mathematical framework (Embrechts *et al.*,1997). Similar work is found in Resnick (2007), who studied extreme events in data networks, finance and insurance. McNeil (1997a, 1997b, 1998 and 1999) also studied the performance of the methods in insurance and finance. His studies focused on the POT method, i.e. fitting a Generalized Pareto Distribution to excesses over a high threshold. He also applied Block Maxima to financial time series.

Gravril & Altar (2009) applied exchange rate returns of four currencies against the Euro to analyze the relative performance of several VaR models and Extreme Value Theory. They revealed that in extreme market conditions, extreme measures are needed and their studies came up with the evidence that no single measure can perform proper for both the centre and the tails of an exchange rate distribution. Ammann and Reich (2000) examined Value-at-Risk for Nonlinear Financial Instruments; they found that the VaR estimates by the variance covariance approach sometimes do not differ greatly from other simulations even for some optioned portfolios. However, they concluded that for heavily optioned portfolio, Variance-covariance approaches with linear model are not appropriate to be applied.

Angelidis & Degiannakis (2004) Modeled risk for long and short trading positions show that different risk management techniques produce different VaR forecasts and therefore, these risk estimates might be imprecise. It is now widely believed that VaR is not the best risk measure. The use of Extreme Value Theory (EVT) proposed by McNeil and Frey (2000) estimate residuals distributions under the assumption that the tail of the condition distribution of GARCH innovations is well approximated by the heavy tailed distribution. The innovations estimates distribution are based on the Maximum Likelihood fitting of GARCH models to estimate the conditional volatility and extreme value theory for estimating the tail of the innovations distribution of the GARCH model.

Hendricks (1996) analyzed the performance of twelve different VaR models using historical data on exchange rate returns. His work agreed to the works of Hols & De Vries (1991), Huisman *et al* (1998), Wagner & Marsh (2003) and others who showed that financial data are fat-tailed and EVT methodology is superior in estimating tail risks but does not capture *volatility clustering*.

A considerable amount of research has also been dedicated to more specific issues of Extreme Value Theory, e.g. tail index and graphical tools of the framework, like mean excess function plot, Hill plot, QQ plots etc. Tail index estimation is yet a very widely debated problem of EVT. Starting with the work of Hill (1975) and Pickands (1975), many studies have tried to establish a measure of the tail thickness of fat-tailed distributions. Dekkers *et al* (1989) improve the Hill estimator and prove

consistency as well as asymptotic normality. Artzner *et al*. (1998) studied Coherent measures of risk and revealed that the VaR of a portfolio may be greater than the sum of individual VaRs and therefore, managing risk by using it may fail to automatically stimulate diversification. Furthermore, VaR does not indicate the size of the potential loss, given that this loss exceeds the VaR. Due to inconsistent patterns of distributions, some argue that VaR does not give an appropriate risk measurement, and its estimation is subject to large estimation errors.

To remedy these shortcomings of VaR, Delbaen (2002) introduced the Expected Shortfall which equals the expected value of the loss, given that a Value at Risk violation occurred.

2.4   Definitions of Basics Concepts

2.4.1   Exchange rate returns

To study any asset returns, and specifically, daily exchange rate, practitioner usually derives daily log-returns. As in most of empirical finance literature, the variable to be modeled is daily exchange rate return which is the first difference of the natural logarithm of the exchange rate. Mathematically exchange rate returns at time t can be modeled as follows:

$$r_t = log\left(\frac{EX_{t-1}}{EX_t}\right), t = 1,2,\dots,T \tag{2.3}$$

r_t represents the daily percentage return to the exchange rate at time $t \in T$ where T is the total number of observations. EX_t and EX_{t-1} denote the exchange rate at the current day and that of previous day respectively. The positive returns will therefore

denote losses at time t. A good parametric example is an AR (1)-GARCH (1, 1) given as

$$r_t = \mu_t + \sigma_t e_t, t = 1,2, \dots, T \qquad (2.4)$$

Where $\mu_t = \sigma r_{t-1}$

$$\sigma_t^2 = \omega + \alpha (r_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2, \qquad t \in \mathbb{Z} \text{ and } \omega, \alpha, \beta > 0, \quad \alpha + \beta < 1 \text{ and}$$

$|\sigma| < 1$

2.4.2   Model Specification

The term model means any assumptions about the structure of the population (Smith , 1994). Mathematically exchange rate returns can be modeled as Autoregressive model with conditionally heteroskedastic financial time series as follows:

$$r_t = \mu(y_t, z_t) + \sigma(y_t, z_t)e_t, t = 1,2, \dots, T \qquad (2.5)$$

$$y_t = (r_{t-1}, r_{t-2}, \dots, r_{t-\pi})$$

Where y_t represents endogenous variables in the model, z_t represents Explanatory variables consisting of information other than the past of the returns, μ represents Conditional expected return which may be arbitrary, σ represents Conditional volatility of daily exchange rate returns, π represents the order of Autoregressive and e_t represents standardized return i.e. independent and identically distributed random variable with $E(e_t) = 0$ and $E(e_t^2) = 1$.

2.5   Research Gaps

In literature many methods of estimating extreme risk in exchange rate process have been developed. However, none has performed properly both for the central and the tails of the exchange rate distribution. The most popular models applied by many researchers and academicians are GARCH models and Extreme Value Theory approach. The GARCH models introduced by Bollerslev (1986) were applied to estimate Value at Risk; the models capture volatility clustering and persistence. However, these models often fail to fully capture the fat tails observed in exchange rate return series. Wagner and Marsh (2003) showed that even if EVT methodology is superior in estimating tail risks it fails to capture volatility clustering. To deal with these two major shortcomings of these models, this study combines both GARCH model with EVT approach to estimate extreme risk in Exchange rate process.

2.6   Summary

This section reviews what other researchers have done on the estimation of exchange rate volatility and the use of the Extreme Value Theory approach in estimating Value at Risk. It has also explained how some assumptions can affect the estimation of the model and therefore, the direction of our study has taken. Some basics concepts applied in this study have also been defined at the end of this section.

CHAPTER THREE

3.0 CONDITIONAL VOLATILITY ESTIMATION IN GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY

3.1 Introduction

The conditional volatility in exchange rate returns is considered as the origin of exchange rates risk and has certain significances on the volume of international trade. Exchange rate series exhibits non-normal characters as we described them in the second chapter. In literature, many different models in econometrics have been proposed to deal with those stylized facts including models from the ARCH/GARCH family. Engle (1982) and Bollerslev (1986) suggested ARCH and GARCH models respectively to resolve the problem of volatility clustering in financial data. ARCH/GARCH models play a crucial role in estimation of conditional volatility. These models manage changing volatility with the assumption of conditional normality.

3.2 ARCH model and its properties

A stochastic process is called Autoregressive Conditional Heteroscedasticity if its time varying conditional variance is heteroscedastic with auto-regression;

$$\begin{cases} r_t = \sigma_t e_t, \quad e_t \sim N(0,1) \\ \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2 \end{cases} \tag{3.1}$$

Where e_t is $i.i.d$ process with $E(e_t) = 0$ and $var(e_t) = 1$, and $\omega > 0$ and $0 \leq \alpha_i < 1$ to ensure that conditional variance is strictly positive for all t and that the variance stationary as it can be seen below:

Consider the ARCH model given by

$$r_t = \sigma_t e_t$$

$$= \sqrt{\omega + \alpha r_{t-1}^2} . e_t$$

$$E(r_t^2) = E[(\omega + \alpha r_{t-1}^2)e_t^2]$$

$$= E(\omega + \alpha r_{t-1}^2)E(e_t^2|r_{t-1})$$

$$= \omega + \alpha E(r_t^2)$$

$$E(r_t^2) = \frac{\omega}{1-\alpha} \qquad\qquad\qquad (3.2)$$

It is clear that a value of α need to be less than 1 to make the equation (3.2) stable and hence finite variance. The conditional disturbance variance is the variance of r_t, conditional on the given information at time $t - 1$.

$$i.e. \sigma_t^2 = var(r_t/y_t) , y_t = r_{t-1}, \dots, r_{t-q}$$

$$= E(r_t^2/y_t)$$

$$= E_{t-1}(r_t^2) ,$$

Where E_{t-1} represents the expectation conditional on all information up to the end of period $t - 1$, it is now easy to see the recent disturbance influence the variance of

current disturbance. The ARCH terms can be interpreted as news about volatility from past periods. The estimation and testing are natural extensions of ARCH (1) model. To show that $r_t \sim N(0, \sigma_t^2)$, consider the ARCH (1) model where

$$E(r_t^2) = E\left[e_t\sqrt{\sigma_t^2}\right]$$

$$= E(e_t)E\left[\sqrt{\sigma_t^2}\right] = 0, \text{ since } e_t \text{ is } i.i.d \text{ with mean zero and variance } 1.$$

In other words, the disturbance r_t is conditionally heteroscedastic with respect to r_{t-1}. The short-run volatility of the exchange rate process is a function of the immediate past value of the error term. The ARCH can describe volatility clustering since the conditional variance of r_t is an increasing function of r_t^2. This means that large shock cluster together and exchange rate return goes through a period of large volatility and a period of small volatility.

3.2.1   The Properties of ARCH Model

The Autoregressive Conditional Heteroscedasticity models of r_t have the following properties:

a) $E(r_t/y_t) = 0$

b) $Var(r_t/y_t) = \sigma_t^2, \ y_t = r_{t-1}, \dots, r_{t-p}$

Where $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2 \, , \, 0 \le \sum_{i=1}^{p} \alpha_i \, < 1, \ \omega > 0$, the conditional variance σ_t^2 is a nontrivial positive valued parametric function of $r_{t-1}, r_{t-2} \ \dots, r_{t-p}$

c) $e_t = {r_t}/{\sigma_t}$ are $i.i.d$ and also independent of $r_{t-1}, r_{t-2} \dots, r_{t-p}$, the sequence

(r_t) may be observed directly, or it may be an error or innovation sequence of an

econometric model. The ARCH process of order p, abbreviated as ARCH (p) can be

used to describe volatility clustering in exchange rate process.

3.3   GARCH Model and its Properties

The ARCH models capture the mentioned stylized facts behavior of real return data.

The ARCH (p) is more flexible than ARCH (1). In order to have a good fit to real life

exchange rate data one needs a large number of parameters, however, large lags

reduce data required for estimation. Bollerslev (1986) proposed the GARCH model

by adding the concept that the volatility for tomorrow depends not only on the past

realizations but it depends too on the errors of the volatility predicted. The advantage

of the GARCH model over the ARCH model is that it can capture the series

correlation in squared residuals using a smaller number of parameters. For this reason

GARCH models have found extremely wide use since they integrate the two main

characteristics about financial returns series, volatility clustering and unconditional

non-normality.

The general framework of GARCH (p, q) model is represented by allowing the

current conditional variance to depend on the first q past Conditional variance as well

as the p past squared innovations. In general case the volatility presented in (3.1) can

be defined by

$$
\begin{cases}
r_t = \epsilon_t = \sigma_t e_t, \epsilon_t \sim (0, \sigma_t^2) \\
\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \, \sigma_{t-j}^2
\end{cases}
\tag{3.4}
$$

Equation (3.4) is a GARCH (p, q) process where Autoregression in its squared residuals has an order of $p \geq 0$ and $q \geq 0$ is the number of lagged of variance terms and is the number of lagged r^2 terms. The sizes of the parameters α_i and β_j determine the short run dynamics of the resulting volatility process. The non-negativity of α_i and β_j ensure that σ_t^2 is strictly positive. The innovation e_t is an independent and identically distributed process with zero mean and unity variance and strict stationarity is ensured by $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

Large ARCH error coefficients α_i imply that volatility reacts significantly to market movements. Large GARCH coefficients β_j indicate that shocks to the conditional variance take long time to die out. High α_i coefficients, relative to β_j indicate that volatility tends to be more extreme. Since σ_t^2 is the one-period ahead forecast volatility based on the past information, it is called conditional volatility and it is specified as a function of three terms: unconditional volatility ω, news about volatility from the previous period measured as the lag of the squared residuals from the mean equation r_{t-1}^2 (ARCH term) and last period volatility σ_{t-1}^2 (GARCH term).

The most common form of the GARCH models uses in financial data is the GARCH (1, 1) model. By examining the behavior of GARCH model described in (3.4) it is clear that the impact of sign is not taken into account since to estimate the variance of

today, the model has to consider the impact of yesterday as a squared value. For this reason the GARCH model is referred to as a symmetric model.

3.3.1   Properties of GARCH Model

The r_t process is called GARCH (p, q) if the following properties are satisfied:

1) $E(r_t|\Omega_{t-1})=0$

2) $Var(r_t|\Omega_{t-1}) = \sigma_t^2$

Where $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$

3) $e_t = {r_t}/{\sigma_t}$ are independent and identically distributed and independent of

$r_{t-1}, r_{t-2}, \ldots$

The symbol Ω_{t-1} is omega contains past information of r_t and σ_t up to time $t-1$.

3.4   Tests

In this study various tests on exchange rate returns and residuals have been performed. Jarque Bera test for normality has been used to test whether the exchange returns series follow normal distribution or not. ARCH effect in residuals series have been tested using Lagrange Mutliplier test. It is also very important in financial time series data to test for stationarity. Augmented Dickey Fuller test is applied to check the stationarity of our data. The Box-Ljung test introduced by Ljung and Box (1978) for autocorrelation testing in residuals series after fitting the proposed model to the data is also used.

3.4.1 Test for Normality

In many empirical studies with time series data Jarque and Bera test is the most popular for normality testing. Jarque Bera test is better for sample size range between 50 and 5000 observations, Jushan (2005). Since the sample size is 2758 observations, this test is useful in this work for normality testing.

Jarque Bera test can be applied using the method of moments, suppose X is a continuous random variable. The first moment μ_x measures the central location of the distribution. Second moment σ_x^2 measures the variability of a continuous random variable. The first two moments of a random variable determine a normal distribution. The third and fourth moments measure the symmetry and tail behavior of X respectively. In statistics, skewness and kurtosis are often used to determine the degree of asymmetry and fat tailedness of a distribution under study.

Mathematically, the *skewness* of a continuous random variable X are defined as follow.

$$S(x) = E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right], \tag{3.5}$$

Under the normality assumption, the estimated skewness and kurtosis (as defined in equation 2.2) are distributed asymptotically as normal with zero mean and variance $6/T$ and $24/T$ respectively, Snedecor and Cochran (1980). Jarque and Bera (1987) combined these two tests for normality testing as follows.

$$JB = \frac{\hat{S}^2(r_i)}{6/T} + \frac{(\hat{K}(r_i)-3)^2}{24/T} \tag{3.6}$$

Where$\{r_1, r_{2,} \ldots, r_T\}$, are the returns series with T observations. The null hypothesis of normality can be rejected only if the $p-$value of the JB statistic is less than the significance level. Therefore we conclude that the distribution is not normally distributed.

3.4.2 Test for Stationarity of the data

In Autoregressive time series model the presence of unit root causes a violation of the assumptions of classical linear regressions. Recall the classical linear assumptions; where we consider the following seven assumptions of classical linear regression models:

i. The dependent variable is linearly related to the coefficients of the model and the model is correctly specified.

ii. The independent variables are not correlated with the equation error term.

iii. The mean of error term is zero.

iv. The error term has a constant variance (homoscedastic error) or no heteroskedasticity.

v. The error terms are uncorrelated with each other. $i.e$ No autocorrelation or series correlation.

vi. No perfect multicolinearity. $i.e.$ Non-independent variable has a perfect linear relationship with any of the other independent variables.

vii. The error term is normally distributed (optional assumption for hypothesis testing).

The presence of the unit root indicates that the given time series data is non-stationary. When non-stationary time series are used in a regression model one may obtain apparently significant relationships from uncorrelated variables. This phenomenon is called spurious regression. The most popular unit root test is the Augmented Dickey Fuller test. The reason is that the standard Dickey Fuller test is only able to test unit root for first order Autoregressive model. For the standard Dickey Fuller test the following equation can be applied:

$$\Delta r_t = (\delta - 1)r_{t-1} + e_t$$

The case where $\delta = 1$, then we have the random walk which is non-stationary. The Dickey fuller test whether $\delta = 1$ or not. This t-statistic does not converge to the normal distribution but instead to the distribution of functional of wiener process. The Augmented Dickey Fuller test builds correlation of parameters for higher order correction by adding lag differences of the time series, when the time series is correlated at higher lags,.

$$\Delta r_t = (\delta - 1)r_{t-1} + \sum_{i=1}^{q} \alpha_i \Delta r_{t-i} + e_t \tag{3.7}$$

The order of q could be chosen by minimizing information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

Null hypothesis (H_0): Time series data is not stationary and

Alternative (H_a): the data is stationary.

3.4.3 Test for Autocorrelation in the series

Ljung and Box (1978) proposed a diagnostic tool (*Box-Ljung test*) to test the lack of fit of financial time series models. It analyzed autocorrelations of the GARCH innovations. In general if the autocorrelations are very small we can conclude that the model does not exhibit significant lack of fit. The Box-Ljung test for autocorrelation can be defined as: H_0: The model does not exhibit lack of fit (there is no autocorrelation), H_1: the model exhibits lack of fit (there is autocorrelation or dependence in data). Test statistic Q of length T is defined as:

$$Q = T(T+2) \sum_{i=1}^{k} \frac{\hat{\rho}_i}{T-i} \tag{3.8}$$

Where $\hat{\rho}_i$ is the estimated autocorrelation of the series at lag i, and k is the number of lags being tested. The Box-Ljung test rejects the null hypothesis if $Q > \chi^2_{1-\varphi,d}$ where $\chi^2_{1-\varphi,d}$ is the chi-square distribution table value with d degree of freedom and significance level φ. This indicates that the model has significant lack of fit. In this study, we have used this test to the residuals of exchange rate returns after fitting the GARCH model for testing autocorrelation in residuals series.

3.4.4 ARCH Effects Testing

One of the most important issues before applying the GARCH model for financial time series data is to test for the presence of ARCH effect in the residuals. If residuals do not exhibit ARCH effects presence, the GARCH model is unnecessary. Using mean equation in (2.5), we can get the residuals $\epsilon_t = r_t - \mu_t$ of the mean equation;

26

the squared series ϵ_t^2 can be applied to test ARCH effects. Lagrange Multiplier (LM) test proposed by Engle (1982) was applied in this work for ARCH effects testing.

In summary the test procedure is performed as follows: First, obtain the residuals ϵ_t from the ordinary least squares regression of the conditional mean equation. Secondly, regress the squared residual ϵ_t on a constant and p lags in the following equation:

$$\epsilon_t^2 = a_0 + a_1\epsilon_{t-1}^2 + \cdots + a_p\epsilon_{t-p}^2 + e_t \tag{3.9}$$

The null hypothesis that there is no ARCH effect up to order p can be formulated as: $H_0: a_1 = a_2 = a_3 = \cdots = a_p = 0$ Against alternative $H_1: a_i \neq 0$, for at least one, $i = 1,2,\ldots,p$. The test statistic $LM = T.R^2 \sim \chi_d^2$ where T is the sample size and R^2 is computed from the regression of (3.9) using the estimated residuals. The null hypothesis is rejected if p-value is less than the conventional level. If the LM test for ARCH effects is significant for a time series then we could proceed to estimate a GARCH model and obtain estimates of the conditional volatility.

3.5   GARCH Model selection

Model selection is an important part of any statistical analysis and it is interpreted as a decision problem through which a statistic model is selected in order to perform statistical analysis, such as policy analysis, forecasting, estimation and testing. The choice of a good model in the application of financial time series data analysis is crucial since in financial modeling, one of the main challenges is to select a suitable

model to characterize the underlying time series process. The best one can be selected based on diagnostics such as, Akaike Information Criterion (A.I.C) or the Schwartz Criterion (S.C), F-test and Q-test. This study has applied the most popular diagnostics AIC and SC tests to select a good model for the exchange rates data.

3.5.1 Autocorrelation Function and Partial Autocorrelation function

Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) are measures of correlation between current and past series values and show which past series values are most useful in predicting future values. Using this knowledge we came up with the order of the processes in GARCH model. Specifically, ACF can be defined as a set of correlation coefficients between the series and the lags of itself over time. The lag at which the ACF cuts off is the indicated number of GARCH term or conditional Variance. In the same way, PACF can also be defined as partial correlation coefficients between the series and lag of itself over time. The lag at which the PACF cuts off is the indicated number of Autoregressive term or ARCH term. A positive correlation indicates that large current values correspond with large values at the specified lag whereas a negative correlation indicates that large current values correspond with small values at the specified lag. The absolute value of a correlation is a measure of strength of the association, with large absolute values indicating stronger relationships (wang *et al*., 2005).

3.5.2 Akaike and Schwartz Information Criterions

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

$$AIC = -2(\ln(likelihood)) + 2k$$

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

$$AIC = -2log(Maximized\ Likelihood\ ) + 2(no.of\ fitted\ parameters)$$

The first term on the right hand side of AIC equation is a measure of the lack of fit of the chosen model while the second term on the right hand side measures the increased number of model parameters.

The Schwartz Information Criterion proposed by Schwartz (1978) is another model selection criterion based on information theory in Bayesian context called BIC. BIC is an estimate of a function of a future probability of a model being true under a certain Bayesian setup. A lower BIC means that a model is considerably more likely to be the true model. Mathematically, BIC can be defined as follows.

$$BIC = -\frac{2Ln}{T} + [K(1 + \ln(T)]/T$$

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

3.6   Estimation of Exchange rate Volatility

We consider the process that describes the exchange rate returns under the model in (2.5) and we redefine it as follows;

$$\begin{cases} r_t = \mu_t + \epsilon_t, \epsilon_t = \sigma_t\, e_t \\ \quad \sigma_t^2 = var(\,r_t|\mathcal{F}_{t-1}) \end{cases} \tag{3.10}$$

In order to estimate the conditional volatility residuals are substituted by sample residuals. The residuals of the returns can be given as $\epsilon_t = r_t - \mu_t$, since e_t is standardized returns (i.e. independently and identically distributed random variable with $E(e_t) = 0$ and $E(e_t^2) = 1$). Residuals may be estimated through sample residuals as follows.

$$\hat{\epsilon}_t = r_t - \hat{\mu}_t$$

$$\hat{\sigma}_t\, \hat{e}_t = r_t - \hat{\mu}_t$$

Where

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i\, \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i\, \sigma_{t-i}^2 \tag{3.11}$$

is GARCH (p, q) model. We can fit GARCH (p, q) model to the negative return exchange rate data using Quasi-Maximum likelihood Estimation procedure to get the

current volatility. From equation (3.11) of general framework of GARCH (p, q) model, the volatility estimator can be obtained as follows.

$$\hat{\sigma}_t = \sqrt{\hat{\omega} + \sum_{i=1}^{p} \hat{\alpha}_i \, \epsilon_{t-i}^2 + \sum_{i=1}^{q} \hat{\beta}_i \, \sigma_{t-i}^2} \qquad \text{(3.12)}$$

Where $\hat{\sigma}$ is the estimated volatility and its asymptotic consistency and asymptotic normality were investigated in the next section.

3.7   Quasi-Maximum Likelihood Method

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

To get the Quasi-Likelihood function, we consider the situation where the true probability distribution $f_0(r_t, \theta_0)$ of the exchange rate at time t and incorrect probability distribution given by $f(r_t, \theta)$ are used to build the likelihood function.

Now model (2.5) can be reformulated by letting $\{r_t\}$ to be a sequence with the true model giving

$$\begin{cases} r_t = \epsilon_{0t}, \ \epsilon_{0t} = \sigma_{0t}\, e_t \\ \sigma_{0t}^2 = Var(r_t|\mathcal{F}_{t-1}) = E[(r_t)^2|\mathcal{F}_{t-1}] \end{cases} \qquad (3.13)$$

Where $\epsilon_{0t} \sim N(0, \sigma_{0t}^2)$, $\qquad E(\epsilon_{0t}|\mathcal{F}_{t-1}) = 0 \qquad$ *almost* $\qquad$ *sure* $\qquad$ *(a.s)* $\qquad$ and $\mathcal{F}_t = \sigma(\epsilon_{0t}, \epsilon_{0t-1}, \epsilon_{0t-2}, \dots)$, the conditional variance can be defined as $E(\epsilon_{0t}^2|\mathcal{F}_{t-1}) = \sigma_{0t}^2$ (the subscript 0 indicates the true values of parameters). We also assume $r_t = \epsilon_t = \sigma_t \varepsilon_t$, $\epsilon_t \sim N(0, \sigma_t^2)$ to be the model for the unknown parameters (misspecified model). Hence the true and misspecified distributions are;

$$f_0(r_t) = \frac{1}{\sigma_{0t}\sqrt{2\pi}} exp\left[-\frac{(\epsilon_{0t})^2}{2\sigma_{0t}^2}\right] \qquad (3.14)$$

$$f(r_t) = \frac{1}{\sigma_t\sqrt{2\pi}} exp\left[-\frac{(\epsilon_t)^2}{2\sigma_t^2}\right] \qquad (3.15)$$

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

$$\sigma_{0t}^2 = \omega_0(1 - \beta_0) + \alpha_0 \epsilon_{0t-1}^2 + \beta_0 \sigma_{0t-1}^2 \ a.s$$

An equivalent expression for the conditional variance can be derived as:

$$\sigma_{0t}^2 = \omega_0 + \alpha_0 \sum_{k=0}^{\infty} \beta_0^k \epsilon_{t-1-k}^2 \ a.s,$$

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

$$\theta_0 = [\omega_0, \alpha_0, \beta_0]' \qquad (3.16)$$

and for the model with the unknown parameters,

$$\sigma_t^2(\theta) = \omega(1 - \beta) + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad t = 2,3,\dots,T$$

with the setup or initial condition $\sigma_1^2(\theta) = \omega$, this gives the convenient expression for the conditional variance process

$$\sigma_t^2 = \omega + \sum_{k=0}^{t-2} \beta^k \epsilon_{t-1-k}^2$$

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

$$\theta = [\omega, \alpha, \beta]', \text{ with} |\beta| < 1 \tag{3.17}$$

Now define the compact parameter space Θ, in the following way.

$$\Theta \equiv \{\theta: 0 < \omega_l \le \omega \le \omega_u; \ 0 < \alpha_l \le \alpha \le \alpha_u; \ 0 < \beta_l \le \beta \le \beta_u < 1\} \tag{3.18}$$

Where subscript l and u indicate lower and upper limits respectively. We assume that the true parameter $\theta_0 \in \Theta$, this implies that $\alpha_0 > 0$, $\beta_0 > 0$, which means that ϵ_t is strictly a GARCH process. We can also define standardized residuals $e_t = {\epsilon_t}/{\sigma_t}$ by constructing $E(e_t|\mathcal{F}_{t-1}) = 0 \ a.s$ and $E(e_t^2|\mathcal{F}_{t-1}) = 1 \ a.s$ frequently, estimation of GARCH models is done under the assumption that $e_t \sim N(0,1)$ so that the likelihood is easily specified. The maximum likelihood estimators of the parameters of the misspecified distribution are obtained by maximizing the log-likelihood function

$$lnL(\theta) = \sum_{t=1}^n lnf(r_t; \theta) \tag{3.19}$$

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

$$l(\theta) = \frac{\partial lnL}{\partial \theta} = \sum_{t=1}^n \frac{\partial lnf(r_t; \theta)}{\partial \theta} \tag{3.20}$$

to zero. Let take expectations of the gradient vector in (3.20) with respect to the true probability distribution $f_0(r_t; \theta_0)$.

$$E_0[l(\theta)] = \int_{-\infty}^{\infty} l(\theta) f_0(r_t; \theta_0) dr_t$$

$$= \int_{-\infty}^{\infty} \sum_{t=1}^{n} \frac{\partial \ln f(r_t; \theta)}{\partial \theta} f_0(r_t; \theta_0) dr_t$$

$$= \int_{-\infty}^{\infty} \sum_{t=1}^{n} \frac{\partial f(r_t; \theta)}{\partial \theta} \frac{f_0(r_t; \theta_0) dr_t}{f(r_t; \theta)} \tag{3.21}$$

Where $E_0[.]$ means that the expectation taken with respect to the true distribution, this expression is not guaranteed to equal zero except in the case where the distribution is specified correctly $f(r_t; \theta) = f_0(r_t; \theta_0)$. In this case, (3.21) may be simplified by exchanging the integration and differentiation operators and using the property of a probability distribution to give

$$E_0[l(\theta)] = \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(r_t; \theta) dr_t$$

$$= \sum_{t=1}^{n} \frac{\partial}{\partial \theta} 1$$

$$E_0[l(\theta)] = 0 \tag{3.22}$$

Thus sufficient condition for (3.22) to hold is that the model is specified correctly. There are however, some important cases where $E_0[l(\theta)] = 0$ even when the distribution is misspecified. Let us assume that the Gaussian Likelihood is applied to form the estimator. Then, the Log Likelihood takes the form

$$L_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} l_t(\theta), \tag{3.23}$$

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

$$\hat{\theta}_n = arg\,\max_{\theta \in \Theta_l} L_n(\theta) \tag{3.24}$$

We also need to investigate the asymptotic consistency and asymptotic normality properties of the quasi-maximum estimator $\hat{\theta}_n$ of the GARCH process.

3.7.1 Asymptotic Consistency

An estimator say $\hat{\theta}_n$ is consistency to the actual parameter θ_n means that when sample size is sufficiently large the estimator $\hat{\theta}_n$ will be very likely to be very close to the actual parameter value θ_n. When an estimator converges in probability to the true value as the sample size increases, then, the estimator is asymptotically consistent. Suppose that we observe the daily exchange rate returns data $r_{-p+1}, \ldots r_0, r_1, \ldots, r_n$ generated by the model (3.13) with θ_0 as the parameter. Assume that the data up to r_0 are available to us and the process r_0 is described with true parameters in the vector form given as in (3.16). One can define

$$\hat{\sigma}_{*t}^2(\theta) = \omega(1-\beta) + \alpha\hat{\epsilon}_{t-1}^2 + \beta\hat{\sigma}_{*t-1}^2 \quad t = 1,2,..T$$

Together with initialization $\hat{\sigma}_{0t}^2(\theta) \geq 0$ this means that the log-likelihood of $(r_1, \ldots, r_n)'$ given $(r_0, \sigma_0)'$ under $e_t \sim N(0,1)$ is approximately equal to

$$\hat{L}_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\ln\hat{\sigma}_{*t}^2(\theta) + \frac{r_t^2}{\hat{\sigma}_{*t}^2(\theta)}\right)$$

The Quasi-Maximum Likelihood Estimator $\hat{\theta}_n$ is the parameter value which maximizes $\hat{L}_n$ on subspace Θ_1, since Θ_1 is an approximately chosen compact subset of the parameter space Θ. Then Quasi-MLE $\hat{\theta}_n$ is strongly consistency if the following conditions on the random variable e_t are satisfied:

D_1. e_t is sequence of independent and identically distributed random variables such that $\mathrm{E}(e_t)=0$

D_2. The vector parameter θ_0 is in the interior of compact set Θ.

D_3. For some a>0 there exists a constant $b < \infty$ such that $E[e_t^{2+a}] \leq b < \infty$

D_4. e_t^{2+a} is non degenerate

D_5. $E[ln(\beta_0 + \alpha_0 e_t^2)] < 0$

D_6. If for some t holds $\sigma_{0t}^2 = \omega_0 + \sum_{k=1}^{\infty}\omega_k\,\epsilon_{t-k}^2$ and $\sigma_{0t}^2 = \omega_0^* + \sum_{k=1}^{\infty}\omega_k^*\,\epsilon_{t-k}^2$

Therefore $\omega_j = \omega_j^*$ for every $1 \leq j < \infty$

If these conditions are satisfied we can conclude the consistency of Quasi-MLE in the following theorem.

Theorem

Under the conditions $D_1 - D_6$ above, the quasi-maximum likelihood estimate $\hat{\theta}_n$ is strongly consistent that is $\hat{\theta}_n \overset{a.s}{\to} \theta_0$, $n \to \infty$. (3.25)

3.7.2 Asymptotic normality

The distribution of estimators is said to be asymptotically normal if, as the sample size increases, the distribution of the estimators approaches a normal distribution. To show that our estimators are asymptotically normal we need the following additional assumptions.

D_7. σ_{*t}^2 is twice continuously differentiable on subspace Θ_1

D_8. The following moment conditions hold: $E(e_0^4) < \infty$, $E\left[\frac{|\nabla\sigma_{*0}^2(\theta_0)|^2}{\sigma_0^4}\right] < \infty$, $E\|\nabla l_n\|_{\Theta_1} < \infty$ and $E\|\nabla^2 l_n\|_{\Theta_1} < \infty$.

If the conditions $D_1 - D_8$ hold, the following theorem can be stated for the asymptotic normality of the quasi-Maximum likelihood estimator.

Theorem

Under the conditions $D_1 - D_8$, the Quasi-Maximum Likelihood Estimator $\hat{\theta}_n$ is strongly consistent and asymptotically normal, that is

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, \ V_0) \ as \ n \to \infty. \tag{3.26}$$

Where V_0 is the asymptotic variance of the estimator $\hat{\theta}_n$. Under asymptotic normality, the estimator $\hat{\theta}_n$ not only converges to the unknown parameter, but it converges fast enough, at a rate $1/\sqrt{n}$. For more details see (Francq and Zakoïan, 2004) and (Posedel, 2005).

3.8   Conclusion

In this section we estimated conditional volatility in GARCH (p, q) model using Quasi-Maximum Likelihood procedure assuming that the innovations are normally distributed. This chapter described some tests need to be done for foreign exchange rate series such as Normality testing, unity root test, test for Autocorrelation in series and also test for presence of ARCH effects in residuals. This chapter also investigated consistency and asymptotic normality of estimated parameters.

CHAPTER FOUR

4.0 QUANTILES ESTIMATION AND VALUE AT RISK

4.1   Background of Extreme Value Theory

Extreme Value Theory (EVT) is a well developed theory in the field of probability that studies the distribution of extreme realizations of a given process, satisfying certain assumptions. The foundations of the theory were set by Fisher & Tippett (1928) and Gnedenko (1943), who proved that the distribution of the extreme values of an independent and identically distributed ($i.i.d$) sample from a cumulative distribution function F, when adequately rescaled, can converge to one out of only three possible distributions; Fréchet family, Weibull family and Gumbel distribution (Embrechts $et\ al$.,1999).

There are two main approaches when identifying extremes in real data. The first approach is *Block Maxima* which considers the maximum (minimum) values that a variable takes over successive periods of same length (blocks). The second approach is known as *Peaks over Threshold* (POT) which focuses on the realizations exceeding a given (high) Threshold.

For financial time series, POT method is employed to modeling extreme events. This approach is considered to follow a Generalized Pareto Distribution (GPD). Extreme Value Theory provides possibility to concentrate on each one of the two tails of the distribution independently, thus allowing a flexible approach which can take skewness of the underlying distribution into account.

4.2   Distribution of Maxima

The limit law for the block maxima, denoted M_n with n the size of the subgroup is given by:

Theorem1 (*Fisher and Tippet, 1928)*

Consider a sequence (x_n) of independent and identically random variables. If there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ and some non-degenerate distribution function $\mathcal{H}$ such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} \mathcal{H}, \text{ as } t \longrightarrow \infty \text{ or } \lim_{t \to \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathcal{H}(x) \tag{4.1}$$

then $\mathcal{H}$ belongs to one of the three standard extreme value distributions.

$$\text{Frechet: } \Phi_\delta(x) = \begin{cases} 0, & x \leq 0 \\ exp(-x^{-\delta}), & x > 0 \end{cases}, \delta > 0 \tag{4.2}$$

$$\text{Weibull: } \Psi_\delta(x) = \begin{cases} exp - (-x)^\delta, & x \leq 0 \\ 1, & x > 0 \end{cases}, \delta > 0 \tag{4.3}$$

Gumbel: $\Lambda(x) = exp\left(-exp[-x]\right), x \in \mathbb{R}$ (4.4)

Where δ is referred to as the tail index of the distribution. Jenkinson (1955) and Von Mises (1954) generalized these distributions above in the following one shape parameter ζ representation.

$$\mathcal{H}_\zeta(x) = \begin{cases} exp\{-(1 + \zeta x)^{-1/\zeta}\} & if \quad \zeta \neq 0 \\ exp(-exp[-x]) & if \quad \zeta = 0 \end{cases} \tag{4.5}$$

Where x is such that $1 + \zeta x > 0$, ζ is the shape parameter, $\mathcal{H}_\zeta(x)$ is known as the Generalized Extreme Value (GEV) distribution. An important concept for application of extreme value theory to extreme quantile estimation is the maximum domain of attraction($\mathcal{D}$). In simple terms, a random variable x_n is said to belong to $\mathcal{D}$ $i.e.$, $(x_n) \in \mathcal{D}(\mathcal{H})$, if and only if the Fisher-Tippet theorem holds for x_n. The shape parameter ζ is very important in determining the class of generalized extreme value distribution. For $\zeta > 0$, derivers the Fréchet family, $i.e.$ the distribution is in $\mathcal{D}(\mathcal{H}_\zeta, \zeta > 0)$ and is heavy tailed. Examples of heavy tailed distributions are, *Pareto, Log-gamma, Cauchy and t-distribution.* For $\zeta \to 0$ $or(\zeta = 0)$, the distribution is said to belong to $\mathcal{D}$ of Gumbel distribution $\mathcal{D}(\mathcal{H}_0)$ and they are characterized by median tails. Examples are, *gamma, normal, lognormal.* For $\zeta < 0$, the distribution is said to belong to $\mathcal{D}$ of Weibull family $\mathcal{D}(\mathcal{H}_\zeta, \zeta < 0)$ and are short tailed or bounded. Examples are *uniform and beta distributions*.

In practice, the true distribution of the returns is not known, therefore we do not have any idea about the constants a_n and b_n, for this reason we may use the parameter specification

$$\mathcal{H}_{\zeta,\ell,\psi}(x) = \mathcal{H}_\zeta\left(\frac{x-\ell}{\psi}\right) \begin{cases} exp\left\{-\left(1 + \zeta\frac{x-\ell}{\psi}\right)^{-\frac{1}{\zeta}}\right\} \\ exp\left(-exp\left[-\frac{x-\ell}{\psi}\right]\right) \\ x \in \mathcal{D}, \end{cases} , \mathcal{D} = \begin{cases}]-\infty, \ell - \frac{\psi}{\zeta}[& \zeta < 0 \\ ]-\infty, \infty[& \zeta = 0 \quad (4.6) \\ ]\ell - \frac{\psi}{\zeta}, \infty[& \zeta > 0 \end{cases}$$

of the GEV, which is the limiting distribution of the not normalizes Maxima. Where ℓ and ψ are the location and the scale parameters representing the unknown constants a_n and b_n, for a location parameter $\ell \in \mathbb{R}$ and a scale parameter $\psi > 0$. In our case,

we are not only focusing on the parameters themselves, but the quantiles also called return levels of estimated generalized extreme value.

4.3   Distribution of Exceedances

Since Peaks over Threshold focuses on the realizations exceeding a given (high) threshold and the threshold method uses data more efficiently, we concentrate on peak over threshold approach. Consider x_n to be independent and identically distributed $(i.i.d)$ random variables and $v_1, v_2, \ldots, v_\tau$ to be a series of exceedances over threshold τ. We assume that the excesses are $i.i.d$ with conditional distribution function F_τ, and threshold τ is less than endpoint $x_F \leq \infty$. The distribution function F_τ is called the conditional excess distribution function and is defined as follows:

$$F_\tau(v) = P(x_n - \tau \leq v | x_n > \tau), \ o < v \leq x_F - \tau$$

$$= \frac{P[(x_n \leq v + \tau) \cap (x_n > \tau)]}{P(x_n > \tau)}$$

$$= \frac{P[(x_n \leq v + \tau) - (x_n \leq \tau)]}{P(x_n > \tau)}$$

$$F_\tau(v) = \frac{F(v+\tau) - F(\tau)}{1 - F(\tau)} \tag{4.7}$$

This is the excess distribution. When rearranged we obtain the tail distribution of the random variable $x_t \ i.e.$

$$1 - F_\tau(v) = 1 - \frac{F(v+\tau) - F(\tau)}{1 - F(\tau)}$$

$$= \frac{1 - F(\tau) - F(v+\tau) + F(\tau)}{1 - F(\tau)}$$

42

$$= \frac{1-F(v+\tau)}{1-F(\tau)}$$

$$\bar{F}_\tau(v) = \frac{\bar{F}(v+\tau)}{\bar{F}(\tau)} \Longrightarrow \bar{F}(v+\tau) = \bar{F}_\tau(v)\bar{F}(\tau) \tag{4.8}$$

Peak over threshold method according to Todorovic & Zelenhasic (1970) gives the framework of estimating the distribution function F_τ of the value of x_t excesses over certain threshold τ, which identifies the starting of the tail.

Theorem *(limiting distribution of $\bar{F}_\tau(v)$), Pickands (1975), Balkema and Haan (1974)*

For a large class of underlying distribution function F the conditional excess distribution function $F_\tau(v)$ given an appropriately high threshold τ is approximated by a Generalised Pareto Distribution function i.e

$$\lim_{\tau \to x_F}\left\{\mathrm{Sup}_{0 \leq v \leq x_F - \tau}\,|F_\tau(v) - G_{\zeta,\psi(\tau)}(v)|\right\} = 0 \tag{4.9}$$

4.3.1 Generalized Pareto Distribution

For a given a series $v_1, v_2, \ldots, v_\tau$ of exceedances over threshold τ, the generalized Pareto distribution function $G_{\zeta,\psi_\tau}(v)$ can be defined as follows:

$$G_{\zeta,\psi_\tau}(v) = \begin{cases} 1 - (1 + \zeta v/\psi_\tau)^{-1/\zeta} & if\ \zeta \neq 0 \\ 1 - \exp\left(-\frac{v}{\psi_\tau}\right) & if\ \zeta = 0 \end{cases} \tag{4.10}$$

With a random variable $x_t - \tau \geq v \geq 0$ and ζ the shape parameter is independent of τ and is the same as for GEV distribution and ψ the scale parameter, $\psi_\tau = \psi(\tau) > 0$. The distribution is said to be generalized because it contains other distributions under

43

common parametric form. The shape parameter ζ controls the tails behavior of the distribution and the tendency to produce heavy extremes while the scale parameter stretches or contracts the distribution. When $\zeta > 0$, we have a reparametrized type of the usual Pareto distribution, if $\zeta = 0$, gives the exponential distribution, and if $\zeta < 0$, gives uniform distribution. In general, one cannot fix an upper bound for financial losses; only distributions with the positive shape parameter are suited to model financial return distributions. From equation (4.9) the distribution function F_τ tends to become $G_{\zeta,\psi(\tau)}$

$$F_\tau(v) \longrightarrow G_{\zeta,\psi(\tau)}(v) \tag{4.11}$$

By setting $x = v + \tau$ and combining expression (4.9) and (4.7) we see that the model can be also written as

$$\bar{F}(x) = \bar{F}(\tau)\bar{G}_{\zeta,\psi(\tau)}(v) \tag{4.12}$$

Equation (4.12) shows that we may interpret the model in terms of the tail of the underlying distribution $F(x)$ for $x > \tau$. The main steps of Peak over Threshold implementation are;

a) Test for independent and identically distributed. Hypothesis Data should be a sequence of independent and identically distributed random variables.

b) Select an appropriate threshold level.

c) Estimate the parameters using the most appropriate method for the considered excesses dataset.

4.3.2 Threshold selection Methods

By choosing a low threshold the risk is to introduce some central observations in the series of extremes. The tail index (shape) in this case is more accurate (less variance but biased). Setting a too high threshold would lead to a reduction of the number of extreme observations and hence increase in the variance. The high threshold implies a less biased but less robust tail index. Here the major problem is to find the optimal threshold for Generalized Pareto Distribution.

A considerable amount of researchers such as Neves & Alves (2008), Smith (1985) and Thompson, *et al*. (2009) proposed and applied various methods to detect the appropriate threshold. Some of these approaches are graphical, numerical and others combine both graphical and numerical approaches. Three techniques of threshold determination are discussed; *Mean Residual Life Plot, Hill Plot, and Square Error Method.*

i.   Mean Residual Life Plot

The Mean Residual Life Plot also known as *Mean Excess plot* is one of the most commonly used graphical methods. The theoretical reasons behind this approach reside in the fact that the distribution of exceedances over the threshold τ is a Generalized Pareto Distribution of exceedances over any threshold $\tau_1 > \tau$ is also a GPD with the same shape parameter ζ and scale parameter $\psi_{\tau_1} = \psi_\tau - \zeta(\tau_1 - \tau)$. The Mean Residual Life Plot is a representation of the empirical estimate of conditional expectation $E(X - \tau | X > \tau)$ as a function of τ. For an optimal threshold τ^*, the underlying distribution function of the excedances is a GPD and the conditional mean excess is given by

$$E(X - \tau | X > \tau) = \frac{\psi_\tau}{1+\zeta} = \frac{\psi_{\tau^*} - \zeta(\tau - \tau^*)}{1+\zeta}, \text{ for } \tau > \tau^* \tag{4.13}$$

Hence, a good Generalized Pareto Distribution fit occurs when the Mean Residual Life Plot is roughly linear.

ii.   Hill Plot Approach

The Hill Plot technique is done by ordering the data with respect to their values as

$X_{1,n}, X_{2,t}, \ldots, X_{t,n}$ where $X_{1,n} \geq X_{2,n} \geq \ldots \geq X_{t,n}$. The Hill estimator of the tail index,

$\delta = 1/\zeta > 0$, is given by

$$\hat{\delta} = \left(\frac{1}{k} \sum_{i=1}^{k} \ln X_{i,n} - \ln X_{k,n} \right)^{-1} \tag{4.14}$$

Where $k \to \infty$ is upper order statistics (the same number of excedances), n is the sample size, $\delta = 1/\zeta$ is tail index and ζ is shape parameter. The Hill Plot is constructed by plotting the estimate of tail index as a function of k −upper order statistics or threshold $\tau = X_{k,n}$. The threshold is selected from the plot where the shape parameter $\hat{\zeta}$ or tail index $\hat{\delta}$ is fairly stable.

iii.   Square error method

The square Error method was proposed by Beirlant *et al*. (1996) with the purpose of choosing the threshold that minimizes the Mean square error of the tail index Hill estimator. The square error method is based on mathematical criteria, so it helps the user in choosing an adequate threshold on a quite objective consideration basis. In this work, we propose an algorithm in the line of Beirlant's work. It is useful in

comparing different estimators, especially when one of them is biased. The square error method is therefore natural to take as optimal threshold that minimizes the mean square error of an estimator based on exceedances (Xiangxian & Wenlei, 2009).

4.4   Tails Estimation

Our interest is to build tail estimator that can be used to obtain quantiles. The method of the nonparametric such as historical simulation may be used to estimate F_τ as $\hat{F}(\tau) = \frac{N - N_\tau}{N}$, where N is the total number of observations and N_τ are the number of the observations above the threshold τ. The MLE of the generalized Pareto distribution parameters give rise to the tail estimator formula

$$\hat{F}(x) = \frac{N_\tau}{N}\left[1 - \left(1 + \hat{\zeta}(x - \tau)/\hat{\psi}_\tau\right)^{-1/\hat{\zeta}}\right] + \left(1 - \frac{N_\tau}{N}\right) \tag{4.15}$$

This simplifies to

$$\hat{F}(x) = 1 - \frac{N_\tau}{N}\left[\left(1 + \hat{\zeta}(x - \tau)/\hat{\psi}_\tau\right)^{-1/\hat{\zeta}}\right] \tag{4.16}$$

Where $\hat{\zeta}$ and $\hat{\psi}_\tau$ are the estimates of ζ and $\psi(\tau)$ shape and scale parameters respectively.

4.5   Maximum Likelihood Method

Maximum likelihood estimation method is a general method for estimating the parameters of an econometric model. In this section it has been used to estimate Generalized Pareto Distribution (GPD) parameters and it may be expressed as

follows. Let $\hat{G}_{\hat{\zeta},\hat{\psi}_\tau}(v)$ be the Estimated Generalized Pareto Distribution with $\hat{\zeta}$ and $\hat{\psi}_\tau$ the estimates of ζ and $\psi(\tau)$ shape and scale parameters respectively then

$$\hat{G}_{\hat{\zeta},\hat{\psi}_\tau}(v) = \begin{cases} 1 - \left(1 + \hat{\zeta}v/\hat{\psi}_\tau\right)^{-1/\hat{\zeta}} & if\ \hat{\zeta} \neq 0 \\ 1 - \exp\left(-\dfrac{v}{\hat{\psi}_\tau}\right) & if\ \hat{\zeta} = 0 \end{cases}$$

Consider a random variable r_t with probability density function $f(r_1, r_2, \ldots, r_T; \Phi)$ where the form of f is known, but the parameter vector $\Phi = (\zeta, \psi(\tau))$ is not known. In its principle one can choose values of the parameters that give the greatest probability of giving rise to the observed sample of data. The following conditions below must be satisfied in deriving the maximum likelihood estimator of a GPD approach;

1)   The distribution of the observed random variable r_t must be known

2)   The likelihood function must be tractable in the sense that it can be evaluated for all admissible values Φ. The joint probability density function is given by

$$f(r_1, r_2, \ldots, r_T; \Phi_1, \Phi_2, \ldots, \Phi_T) \tag{4.17}$$

Where $\Phi_1 = \Phi_2 = \cdots = \Phi_T = \Phi$ is a vector parameter that is constant over time. The standard interpretation of the probability density function in (4.17) above is that f is interpreted as a function of r_t for given parameters Φ in defining maximum likelihood estimators this interpretation is reversed, so that f is taken as a function of Φ for

given r_t. The reason behind this change in the interpretation of the arguments of the probability density function is to consider $(r_1, r_2, \ldots, r_T)$ as a realized exchange rates data set which is no longer random. Therefore, the maximum likelihood estimator is obtained by finding the value of Φ which is most likely to have generalized the observed data. The likelihood is expressed as

$$L(\Phi) = f(r_1, r_2, \ldots, r_T; \Phi_1, \Phi_2, \ldots, \Phi_T) \tag{4.18}$$

It is crucial to remember that the likelihood function is a redefinition of the joint probability density function. The maximum likelihood estimate of Φ is therefore defined as that value of Φ that maximizes the likelihood function in (4.18).

Most of the problems often work with the log-likelihood function. The only reason to use Log-Likelihood instead of the plain old likelihood is mathematical convenience because it lets you turn multiplication into addition. The plain old likelihood is P (parameter\data) i.e assuming data is fixed and vary the parameters of the model. Maximizing this is one way to do parameter estimation is known as maximum likelihood.

$$lnL(\Phi) = lnf(r_1, r_2, \ldots, r_T; \Phi_1, \Phi_2, \ldots, \Phi_T) \tag{4.19}$$

The likelihood functions of the case of independent and identically distributed $(i.i.d)$ is

$$lnL(\Phi) = \sum_{t=1}^{T} lnf(r_t; \Phi) \tag{4.20}$$

Since the objective of maximum likelihood estimation is to find the value of Φ that maximizes the log-likelihood function, a natural way to do this is to use the rules of

calculus. We compute the first derivatives (gradient) and second derivatives (Hessian) of the log-likelihood function with respect to the parameter Φ.

$$l(\Phi) = \frac{\partial lnL\ (\Phi)}{\partial \Phi} \tag{4.21}$$

is known as the score. In the $i.i.d$ case where Φ is a fixed $(kx1)$ vector of the parameters, the score is

$$l(\Phi) = \begin{bmatrix} \frac{\partial lnL\ (\Phi)}{\partial \Phi_1} \\ \vdots \\ \frac{\partial lnL\ (\Phi)}{\partial \Phi_k} \end{bmatrix} \tag{4.22}$$

The maximum likelihood estimate of Φ, namely $\widehat{\Phi}$ is obtained by solving the set of the first order conditions for a maximum obtained by setting the score equal to zero.

Thus, $\widehat{\Phi}$ satisfies

$$l\left(\widehat{\Phi}\right) = \frac{\partial lnL\ (\Phi)}{\partial \Phi}\bigg|_{\Phi=\widehat{\Phi}} = 0 \tag{4.23}$$

The second derivatives of log-likelihood function with respect to the parameter vector Φ is known as the Hessian

$$H(\Phi) = \frac{\partial^2 lnL\ (\Phi)}{\partial \Phi \partial \Phi'} \tag{4.24}$$

The hessian plays two important roles in the maximum likelihood framework. First, the Hessian is used to establish that a maximum for log-likelihood function has been achieved. The maximum of a function is obtained by solving the first order condition

obtained by setting the gradient to zero and checking to see if the second derivative of the function at optimum is negative.

$$H(\widehat{\Phi}) = \frac{\partial^2 lnL(\Phi)}{\partial\Phi\partial\Phi'}\bigg|_{\Phi=\widehat{\Phi}} \tag{4.25}$$

is negative definite matrix. A matrix H is negative definite if and only if $a'Ha < 0$ for all non-zero vectors a. The second crucial role of hessian is that, the hessian plays a role in determining the precision of the maximum likelihood estimator. We need to investigate the consistency and asymptotic normality of the estimated parameters of Generalized Pareto Distribution.

4.5.1   Consistency of Maximum likelihood estimator

To derive the asymptotic properties of maximum likelihood estimators, we assume that $\widehat{\Phi}$ is the maximum likelihood estimator of the parameter vector Φ and the true value is Φ_0. A minimum requirement of estimator to be consistency is that, as the sample size increases the estimate approaches the true population parameter value Φ_0, that is,

$$plim(\widehat{\Phi}) = \Phi_0 \tag{4.26}$$

A result which requires that any finite sample bias and the variance of the estimator both tend to zero as $t \to \infty$. Given the regularity conditions all maximum likelihood estimators are consistent.

4.5.2   Asymptotic Normality

The distribution of estimators are said to be asymptotically normal if, as the sample size increases, the distribution of the estimators approaches a normal distribution. Mathematically the sampling distribution of the maximum likelihood estimator is

$$\sqrt{T}\left(\widehat{\Phi} - \Phi_0\right) \xrightarrow{d} N(0, I_t(\Phi_0)^{-1}) \tag{4.27}$$

The square roots of the diagonal elements of $I_t(\Phi_0)^{-1}$ represent the standard errors while $\xrightarrow{d}$ means that the estimator converge in distribution. T represents the total number of observations. The parameters are obtained by using MLE where the parameter values are chosen to maximize joint probability density of observations. Maximum Likelihood Estimate of GPD parameters are consistent and asymptotically normal as $N_\tau \longrightarrow x_F$ (Smith R. , 1987).

4.6   Estimation of Extreme Quantiles

Consider a random variable X and $\varphi \in (0,1)$ be a given probability level, a quantile of random variable X at probability level φ is any real number e_φ satisfying the following inequalities.

$$P(X \le e_\varphi) \ge \varphi \tag{4.28}$$

Now, defining the quantile, e_φ, of distribution function F as generalized inverse of that distribution at a given probability level $\varphi \in (0,1)$ close to one, $i.e\ F(e_\varphi)$ close to unity.

$$\varphi = \hat{F}(e_\varphi) = 1 - \frac{N_\tau}{N}\left[\left(1 + \hat{\zeta}(e_\varphi - \tau)/\hat{\psi}_\tau\right)^{-1/\hat{\zeta}}\right] \tag{4.29}$$

We obtain the quantile estimate of an underlying distribution by simple inverting the

(4.29) $i.e$ $\hat{e}_\varphi(\tau) = \left[\hat{F}(\varphi)\right]^{-1}$ yield

$$\hat{e}_\varphi(\tau) = \tau + \frac{\hat{\psi}_\tau}{\hat{\zeta}}\left(\left(\frac{N(1-\varphi)}{N_\tau}\right)^{-\hat{\zeta}} - 1\right) \tag{4.30}$$

Where $\hat{\zeta}$ and $\hat{\psi}_\tau$ are the estimates of ζ and $\psi(\tau)$ shape and scale parameters respectively. For extreme quantiles (when φ closes to 1) the empirical quantiles are not efficient estimates of the theoretical quantiles. Given the risk horizon and confidence level $\varphi \geq 0.95$, we can obtain the unconditional Value at Risk estimate $\widehat{UVaR}_\varphi$ which is equal to the quantile at confidence interval $\varphi \geq 0.95$.

$$\widehat{UVaR}_{\varphi \geq 0.95} = \tau + \frac{\hat{\psi}_\tau}{\hat{\zeta}}\left(\left(\frac{N(1-\varphi)}{N_\tau}\right)^{-\hat{\zeta}} - 1\right) \tag{4.31}$$

4.7   Conditional Value at Risk

The conditional volatility provided by GARCH model and extreme quantile estimates are combined to obtain conditional Value at Risk ($CVaR_\varphi^t$). For extreme quantiles, when φ closes to unity the empirical quantiles are not efficient estimates of the theoretical quantiles. The conditional value at risk is given by

$$CVaR_\varphi^t = \sigma_t e_\varphi \tag{4.32}$$

and Conditional Value at Risk estimate is

$$\widehat{CVaR}_\varphi^t = \hat{\sigma}_t \left(\tau + \frac{\hat{\psi}_\tau}{\hat{\zeta}} \left(\left(\frac{N(1-\varphi)}{N_\tau} \right)^{-\hat{\zeta}} - 1 \right) \right)$$ (4.33)

Conditional Value at Risk $\widehat{CVaR}_\varphi^t$ is defined as the φ −conditional quantile of returns at $\varphi \in (0.95, 0.99)$, (Gourieroux and Jasiak, 2009). The conditional VaR estimate is also consistency since it is composed of consistent estimates.

4.8   Conditional Expected Shortfall

The conditional value at risk is not sub-additive risk measure. $i.e$ Let ϑ be a generic measure of risk that maps the riskness of a portfolio to an amount of required reserves to cover losses that regularly occur and let W_1 and W_2 be portfolios of assets. For sub-additivity property, the required reserves for the combination of two portfolios are less than the required reserves for each treated separately. $i.e$

$$\vartheta(W_1 + W_2) \leq \vartheta(W_1) + \vartheta(W_2)$$ (4.37)

To overcome these shortcomings the conditional Expected Shortfall which has better theoretical properties is applied.

Expected Shortfall also called average value at risk or the tail conditional expectation or expected tail loss; can be defined as the conditional expectation of the return given that it falls above the Value at Risk. Estimation of conditional Expected Shortfall, under extreme conditions, requires estimation of volatility σ_t and using appropriate extreme value distribution to obtain quantiles. The estimator for the conditional Expected Shortfall becomes:

$$\widehat{CES}^t_\varphi = \hat{\sigma}_t \left(\frac{\widehat{UVaR}_\varphi}{1-\hat{\xi}} + \frac{\hat{\beta}+\hat{\xi}\tau}{1-\hat{\xi}} \right), 1 > \xi > 0 \tag{4.38}$$

Where $\hat{\beta}$ and $\hat{\xi}$ are the scale and shape parameters respectively of the GPD distribution and τ is threshold. The conditional ES estimate is also consistency since it comprised of consistency parameters.

4.9   Conclusion

In this chapter extreme quantiles have been estimated using Generalized Pareto Distribution under the assumption that the distribution is unknown. We have combined the conditional volatility obtained by fitting GARCH (p, q) model with the extreme quantiles based on independent and identically distributed excesses to estimate conditional Value-at-Risk and conditional Expected Shortfall.

$$\widehat{CES}^t_\varphi = \hat{\sigma}_t \left(\frac{\widehat{UVaR}_\varphi}{1-\hat{\xi}} + \frac{\hat{\beta}+\hat{\xi}\tau}{1-\hat{\xi}} \right), 1 > \xi > 0$$

CHAPTER FIVE

5.0 EMPIRICAL ANALYSIS AND RESULTS DISCUSSIONS

5.1   Introduction

This chapter presents empirical results on the estimation of extreme risk in exchange

rates using volatility and Extreme value theory. The analysis has been done using

opening daily exchange rates for the following currency pairs: Rwanda Francs versus

Kenya Shillings (Frw/Ksh), Rwanda Francs against US Dollars (Frw/USD), Rwanda

Francs against Euros (Frw/Euro) and Rwanda Francs versus Sterling GBP

(Frw/GBP). The choice of these currencies was based on their relative proportions, in

the Bank's foreign exchange investment portfolio and based also on their currency

composition of the Rwanda imports. 2758 daily observations covering the period

from January 1st 2002 to December 31st 2012 were used. The data were obtained from

National Bank of Rwanda.

5.2   Estimation of Volatility

5.2.1   Data Exploration

The exchange rate data were plotted to see the behavior of the data. The plots in

Figure 5.1 below show the daily fluctuations of Exchange rate series of the Rwandan

Francs versus Kenya Shillings, US Dollars, Euros and GB Pounds respectively. The

plot *5.1.b* shows that the Rwanda Francs against US Dollars exchange rate data

exhibit very low volatility since the graph is almost smooth for some period as

indicated on plots. The Plots in Figures 5.1 reveal general trends with high

uncertainty in the exchange rates of all currencies between the end of 2003 and the

beginning of 2004, and relative stability thereafter. This high depreciation of Rwandan Francs is maybe due to the supply of banknotes and coins, the issue of new banknotes and the distribution of notes unfit for circulation with a view to ensure sound management of money in circulation. The expenses associated with the end of political transitional period and the current external deficit deteriorated contributing sharply to the depreciation of Rwandan franc against foreign currencies in that period.
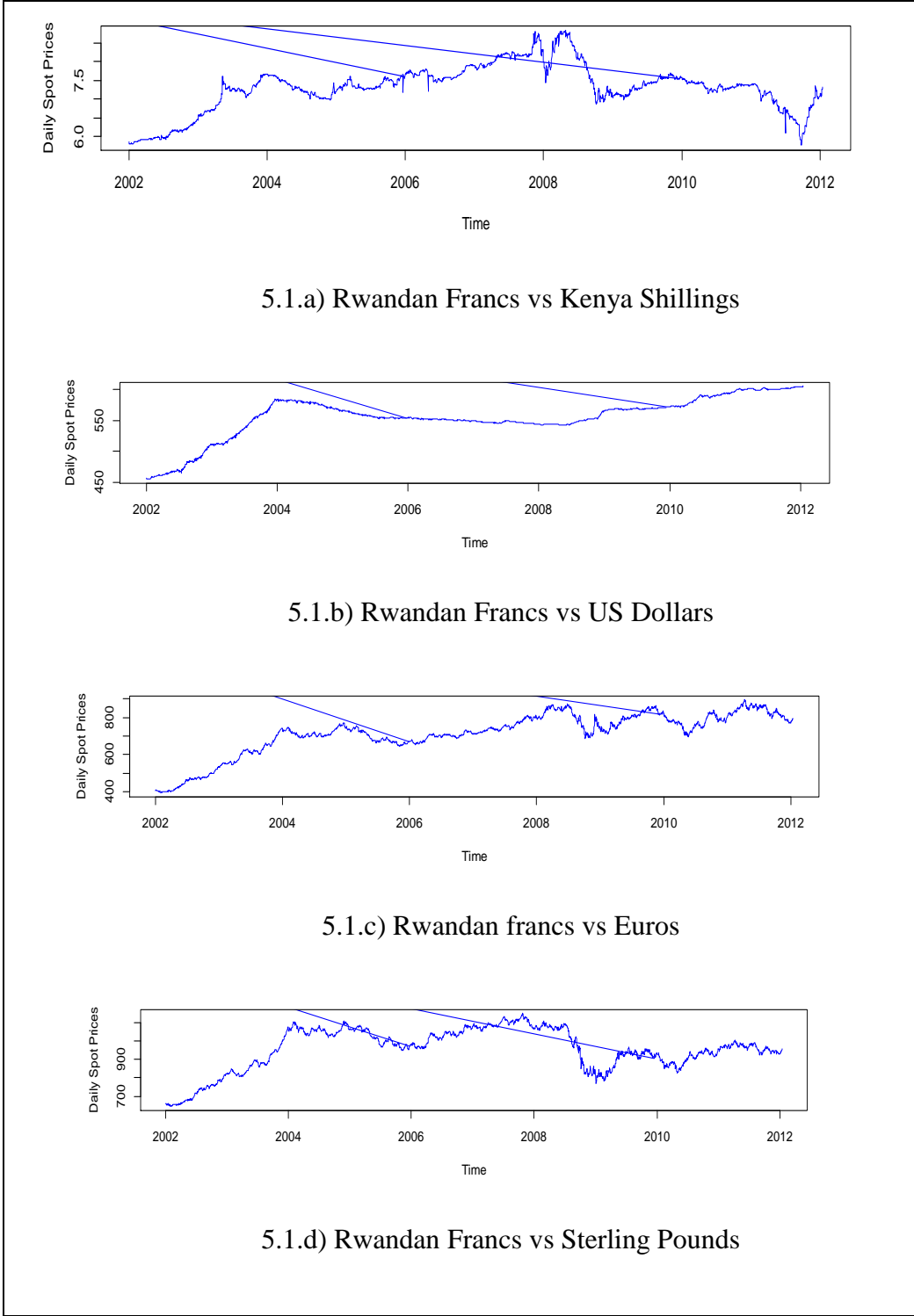
5.1.a) Rwandan Francs vs Kenya Shillings

5.1.b) Rwandan Francs vs US Dollars

5.1.c) Rwandan francs vs Euros

5.1.d) Rwandan Francs vs Sterling Pounds

Figure5.1: Trends in the Daily Exchange Rate series

The daily exchange rate series has a significant difference between its maximum and minimum specifically, in plot *5.1.c* Rwanda Francs against Euros (Frw/Euro) as well as in plot *5.1.d* Francs versus Sterling GBP (Frw/GBP). However, the Rwanda Francs against Kenya Shilling (Frw/Ksh) exhibits low difference as well as Francs versus US Dollars (Frw/USD). The standard deviation of Frw/Ksh series is 8.20% of its mean and that of Frw/USD is insignificant 6.47% of its mean.

Table5.1: Basic statistics of Exchange rate series

Statistics	Frw/Ksh	Frw/USD	Frw/Euro	Frw/GBP
Mean	7.317	560.20	799.7	956.9
Median	7.378	560.10	719.7	967.1
St.dev	0.603	36.25	110.30	110.66
Maximum	8.829	631.50	893.5	1147.0
Minimum	5.759	455.50	394.9	645.6
Kurtosis	3.674	4.021	4.171	3.295
Skewness	-0.437	-0.85635	-1.23	-0.849
observations	2759	2759	2759	2759

5.2.2   Daily Exchange Rate Returns

Most financial time series data are probably decomposed into exponentially growing trend. In financial econometric, it is important to transform data in log-returns because log-returns have good properties such as; it is very simple to aggregate the log-returns over time and in error correlation models, there is assumption that proportions are more stable than absolute differences. In order to estimate the volatility in exchange rates, we have used logarithm exchange rates returns.
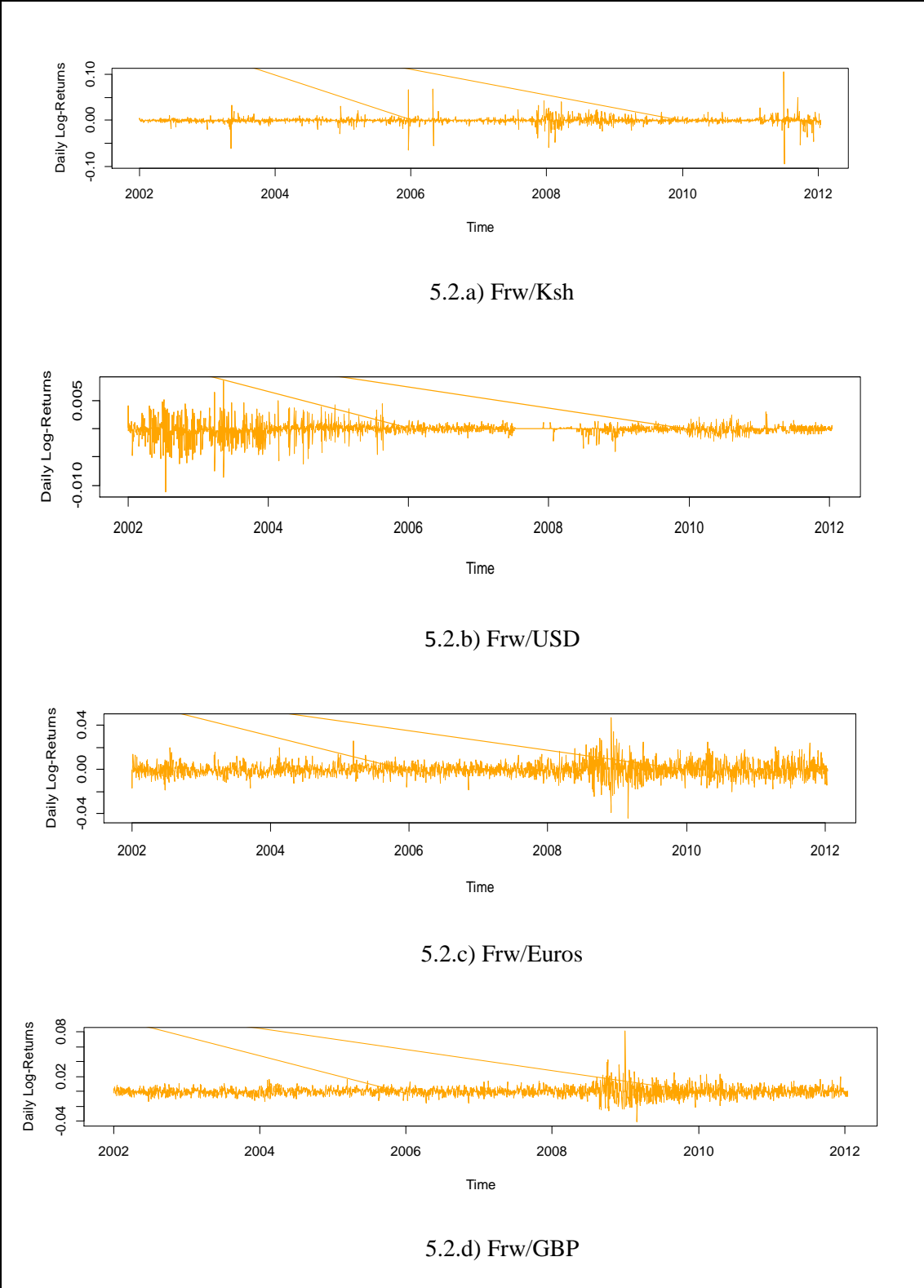
5.2.a) Frw/Ksh

5.2.b) Frw/USD

5.2.c) Frw/Euros

5.2.d) Frw/GBP

Figure5.2: Daily Exchange Rate Returns

The log-returns plots in Figures 5.2 show that the data appear to be stationary in mean after logarithm transformation. These plots also reveal that the returns exhibit dependence structure where period of high returns tend to be followed by high returns and period of low returns tend to be followed by low returns. This is evidence of short-range dependence (volatility clustering in data), which must cast doubt on the assumption of independent and identically distributed ($i.i.d$) data. The clustering of exchange rate returns data indicates presence of stochastic volatility in exchange rate series. Plots in Figures 5.2 allow identifying the most extreme losses and their occurrence.

Descriptive statistics for the exchange rate log returns are presented in Table 5.2. The mean of the exchange rate returns range from -0.026% on Euro to -0.008% on Kenya shillings which are negligible for all currencies. The distribution of returns in plots *5.2.a* of Frw/Ksh and *5.2.b* of Frw/USD exhibit negative skewness (means frequent small gains and few extreme losses). This indicates that they have what statisticians call a long left tail, which for investors can mean a greater change of extremely negative outcomes. The returns series in plot *5.2.c* of Frw/Euro and that of *5.2.d* of Frw/GBP have positive skewness coefficients. The positive skewness coefficients indicate that the distributions of the returns in both currencies are slightly right skewed. This implies that depreciations in the exchange rates occur slightly more often than appreciation. This indicates that investors can have frequent small negative outcomes and few extreme gains.

Table5.2: summary statistics for exchange rate log-returns

Statistics	Frw/Ksh	Frw/USD	Frw/Euro	Frw/GBP
Mean	-8.905e-05	-0.000116	-0.000264617	-0.00015700
Median	-4.329e-05	0.0000000	-0.0003855	-0.0001134
Minimum	-0.09532464	-0.011194	-0.04449859	-0.0413000
Maximum	0.1048363	0.008339	0.04643833	0.0808800
Standard dev.	0.006702908	0.0011167	0.006160647	0.006094035
Skewness	-0.1100568	-1.065407	0.2640723	1.15312
Kurtosis	63.08347	15.5501	7.642768	18.3189
JB.(p-value)	<2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
ADF.Statistics(14)	-12.5243	-9.2241	-12.02	-11.9874
ADF(p-value)	0.01	0.01	0.01	0.01
Sign. level	0.05	0.05	0.05	0.05
Observations	2759	2759	2759	2759

The kurtosis coefficients for log-returns of all currencies are much greater than three for normal distribution. This indicates that the underlying distributions of exchange rate returns have tails which are heavier than that of the normal distribution for log-returns of selected currencies. Jarque Bera test for normality rejects null hypothesis that the distribution is normally distributed since $p - value$ is too small number compare to 5% probability level for all the currencies. Augmented Dickey Fuller (ADF) test has been applied for stationarity testing, the results revealed that the null hypothesis of exchange rate returns series is not stationary has been rejected for all currencies since the p-value is less than 5% level of confidence and more negative indicates the stronger the rejection of the null hypothesis.

5.2.3 GARCH model selection

The Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) were applied to obtain the lags in the GARCH (p, q) model. These functions help us to know which past series values are most useful in predicting future values. The length of past conditional variance (q) was determined by ACF where the lag at which the ACF cuts off is the indicated number of GARCH term (q). The PACF determine the length of past squared innovations (p) where the lag at which PACF cuts off is the indicated the number of ARCH term (p).

5.3.a) Frw/Ksh

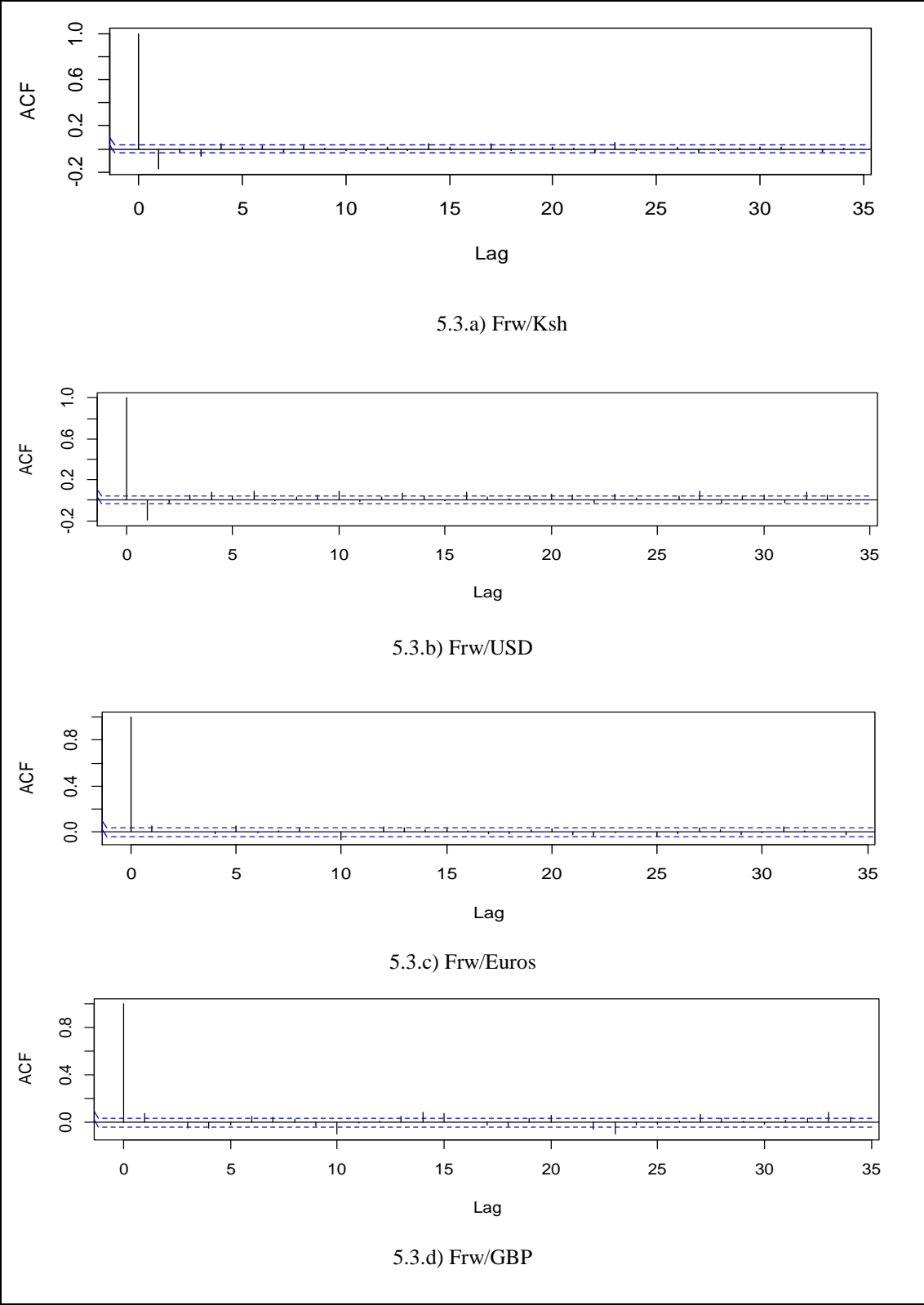5.3.b) Frw/USD

5.3.c) Frw/Euros

5.3.d) Frw/GBP

Figure5.3: Autocorrelation Function for returns series

5.4.a)Frw/Ksh

5.4.b)Frw/USD

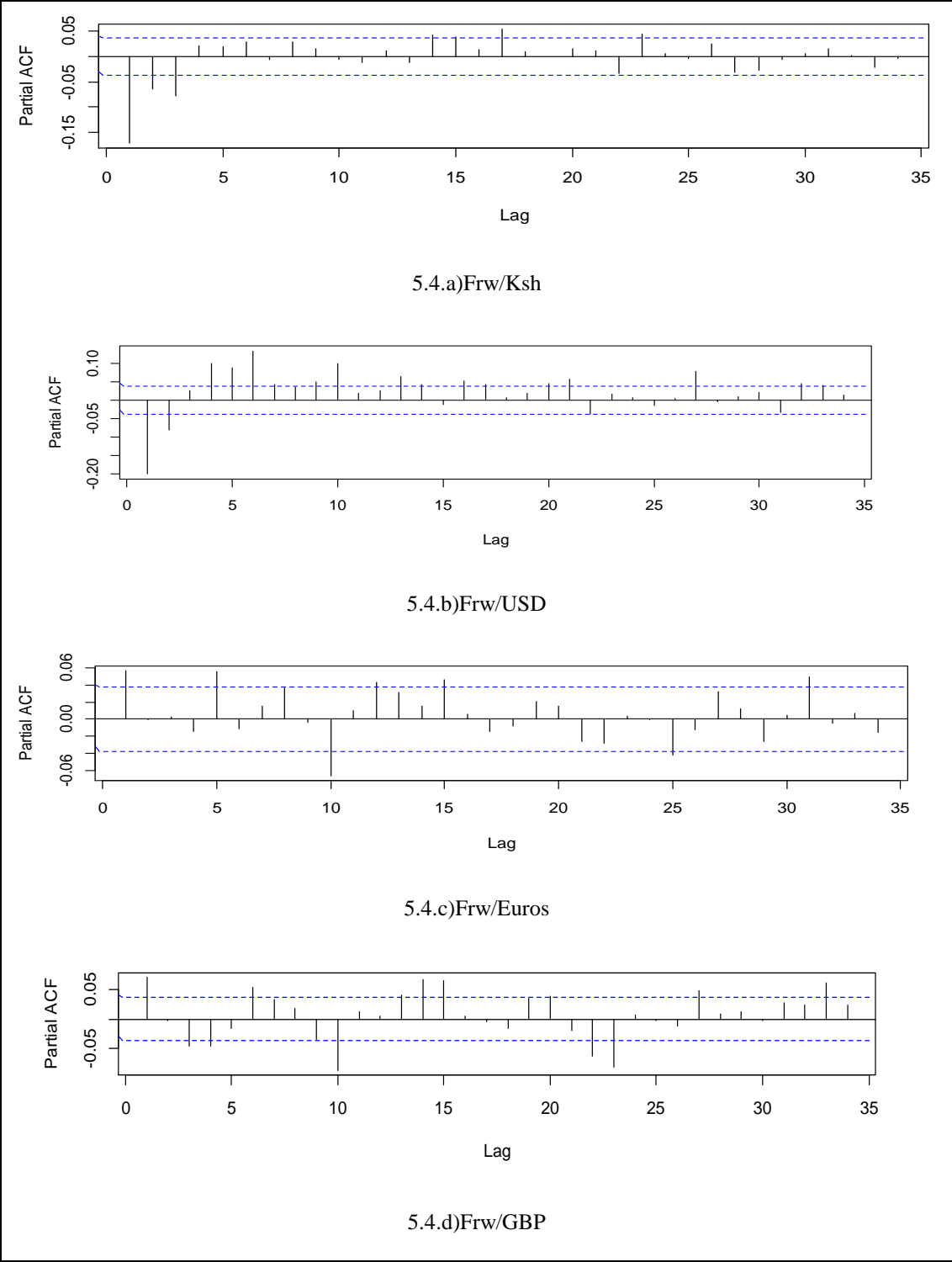5.4.c)Frw/Euros

5.4.d)Frw/GBP

Figure5.4: Partial Autocorrelation Function for returns series

The Akaike information (AIC) criterion and Bayesian information criterion (BIC) tests were used to select the best model for each currency. As suggested by Akaike (1973) and Schwarz (1978) that the best model for financial data is the one that minimize the AIC and BIC respectively. The results are as presented in Table 5.3 below.

Table5.3: BIC and AIC for GARCH model selection

a)   Frw/Ksh

Frw/Ksh	GARCH(1,1)	GARCH(2,1)	GARCH(3,1)
omega	7.42e-07	7.67e-07	9.52e-07
Alpha1	0.115	9.81e-02	7.63e-02
Alpha2		2.57e-02	1.00e-08
Alpha3			1.29e-01
Betha	0.898	0.892	8.44e-01
AIC	-7.597	-7.613	**-7.616**
BIC	-7.589	-7.600	**-7.604**

b)   Frw/USD

Frw/USD	GARCH(1,1)	GARCH(2,1)	GARCH(3,1)
omega	3.09e-09	3.10e-09	3.08e-09
Alpha1	0.112	0.113	0.112
Alpha2		1.00e-08	1.00e-08
Alpha3			1.00e-08
Betha	0.900	0.899	0.900
AIC	-11.49346	**-11.51255**	-11.48645
BIC	-11.48487	**-11.50181**	-11.47956

c)   Frw/Euro

Frw/Euro	GARCH(1,1)	GARCH(2,1)	GARCH(3,1)
omega	1.827e-07	1.83e-07	1.83e-07
Alpha1	3.063e-02	3.070e-02	2.719e-02
Alpha2		1.00e-08	1.00e-08
Alpha3			4.122e-03
Betha	0.963	0.963	0.337
AIC	-7.546322	-7.546272	-7.5460
BIC	-7.537732	-7.535536	-7.5330

d)   Frw/GBP

Frw/GBP	GARCH(1,1)	GARCH(2,1)	GARCH(3,1)
omega	1.31e-07	1.37e-07	1.44e-07
Alpha1	3.364e-02	2.339e-02	2.312e-02
Alpha2		1.127e-02	1.00e-08
Alpha3			1.297e-02
Betha	0.9619e-01	0.9606	0.9590
AIC	-7.701987		-7.701069
BIC	-7.693398		-7.690333

As it can be seen in Table 5.3 from (a) to (d) the results reveal that the GARCH (3,1) is the best GARCH model for Frw/Ksh, GARCH (2,1) for Frw/USD and GARCH (1, 1) is the best model for Frw/Euro and Frw/GBP.

5.2.4   Conditional Volatility estimation

After getting appropriate GARCH model for each currency now we need to estimate the model parameters using Quasi-Maximum Likelihood Procedure as listed in

chap.3 section7 (3.7). The results of estimated models are summarized in the Table 5.4 below.

Table5.4: Summary statistics of the selected GARCH models

	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\beta}$	$\sum_{i=1}\widehat{\alpha}_i + \widehat{\beta}$	Observations
Ksh	9.5e-07	0.076	1.0e-08	0.129	0.844	1.049	High persistent
USD	3.10e-09	0.113	1.0e-08		0.998	1.111	High persistent
Euro	1.82e-07	0.030			0.963	0.993	Low persistent
GBP	1.31e-7	0.033			0.961	0.994	Low persistent

The results in Table 5.4 show that the sums of the ARCH and GARCH coefficients for Frw/Ksh as well as Frw/USD exceed unity. This indicates that the conditional variance is high persistence to the shocks in the volatility of these currencies, so, memories of shocks for these currencies are remembered in the exchange rates markets. The sum of ARCH term and GARCH term coefficients for Frw/Euro and for Frw/GBP are below unity. This indicates that the variance is relatively less persistent to the shocks in volatility (Bollerslev, 1986). The coefficients of ARCH terms (in Table 5.4 above) for variance equation are positive in all currencies and that of GARCH terms are also positive.

The daily exchange rate series indicates volatility clustering characteristics. The values of Q-statistics, ACF and PACF suggest the presence of autocorrelation, for example see plots in Figures 5.3 and 5.4 above. These values continue to decrease with the increase of the number of lags. As can be seen in Table 5.5 below, the Jarque

Bera (J.B) test whether the residuals of returns are normally distributed rejects the null hypothesis to indicate that the residuals of return series are not normal. A test for the presence of ARCH effects in residuals is computed by regressing the squared residuals on a constant and q lags. Lagrange Multiplier (LM) test for ARCH effects rejects the null hypothesis of no ARCH effects at 12 degrees of freedom for all currencies. This indicates the presence of ARCH effects in residuals of exchange rate returns.

For examples, the autocorrelation in the exchange rate returns series for Frw/Ksh dies out after 435 lags and the ARCH effects in residuals die out after 678 lags. The autocorrelation in the returns of Frw/USD dies out after 1732 and ARCH effects in residuals die out after 916 lags. The autocorrelation in the returns Frw/Euro dies out after 746 and ARCH effects in residuals die out after 906 lags. The autocorrelation of the returns in Frw/GBP dies out after 1643 and ARCH effects in residuals die out after 645 lags (These values we got them by increasing number and compare p-value with 5% significance level and we stop where the null hypothesis is accepted to indicate that there is no ARCH effects and we did the same for Autocorrelation). The plots of volatilities in Figure5.5 below reveal volatility clustering characteristics. Statistically, volatility clustering implies a strong autocorrelation in the exchange rate returns series.

Table 5.5: Summary statistics of GARCH model innovations

Tests	statistics	Frw/Ksh	Frw/USD	Frw/Euro	Frw/GBP
J.B	$\chi^2(2)$	415489.8	18654.85	2515.073	27626.04
	$p-value$	< 2.2e-16	< 2.2e-16	< 2.2e-16	<2.2e-16
LM	$\chi^2(12)$	528.3464	323.4492	376.3599	191.2635
	$p-value$	2.2e-16	2.2e-16	2.2e-16	2.2e-16

5.5.a) Frw/Ksh

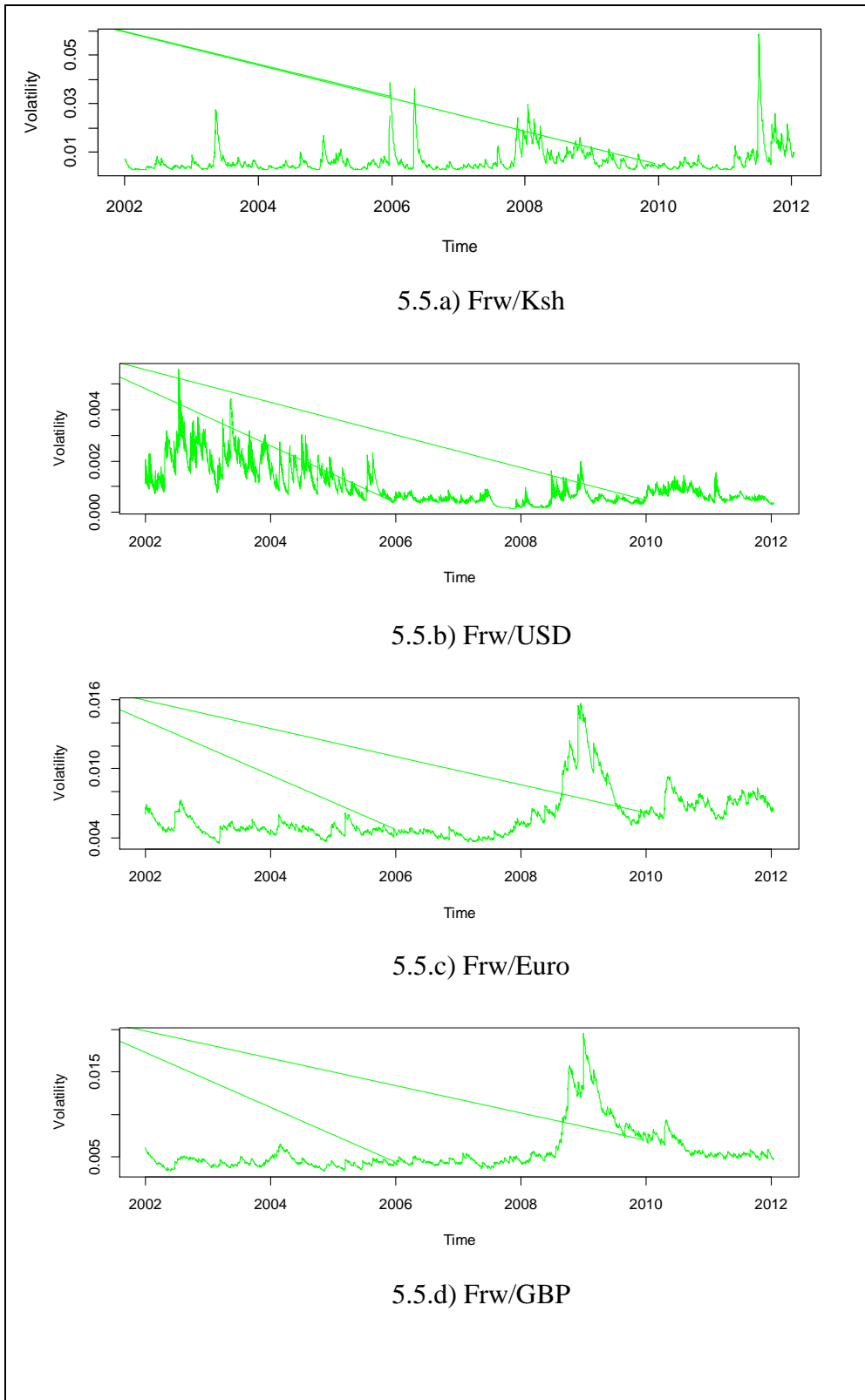5.5.b) Frw/USD

5.5.c) Frw/Euro

5.5.d) Frw/GBP

Figure5.5: Daily Exchange Rate Volatilities

As it is seen on plot *5.5.b* of Frw/USD volatility is high around the year 2003 the reason behind this high volatility is maybe because during this period USD was mostly using by importers of the goods in Rwanda. Therefore new banknotes and coins distribution and political transitional in that period could affect the Frw/USD more than other currencies. The plots *5.5.c* and *5.5.d* of Frw/Euro and Frw/GBP respectively exhibit high volatility around 2009 this is resonable since global crisis (2008) affected Europeans more than the rest of the world. Correlograms of the exchange rate series in Figure 5.5 suggest the evidence of ARCH effects judging from significant autocorrelation coefficients.

The results can be summarized as follows, neither the exchange rate return series nor the residuals series can be considered to be normally distributed since both the series have kurtosis which are greater than 3. This indicates that the curvatures are high in the middle of the distributions and tails are fatter than normal distribution. This means that even if we assumed that e_t is independent and identically distributed standard normal we still get returns which have fat tail behaviors. Therefore, we can conclude that the assumption of conditional normality is not realistic for these data.

5.3   Estimation of Extreme Quantiles

Our aim is to estimate the extreme quantiles in exchange rate returns series using extreme value theory (EVT). The randomness in the model comes through the random variables e_t, which are referred to as noise variables or the innovations of the process and assumed to be independent and identically distributed with unknown

72

distribution function $F(e)$. We begin this stage of estimation of extreme quantile by standardizing the residuals.

$$\hat{e}_t = {r_t}/{\hat{\sigma}_t}, t = 1, 2, \ldots, 2759 \tag{5.1}$$

where r_t is returns series and $\hat{\sigma}_t$ is estimated volatility in the returns. ACF and PACF for squared residuals are plotted in Figures 5.6 and 5.7 below to show that standardized residuals are not autocorrelated.
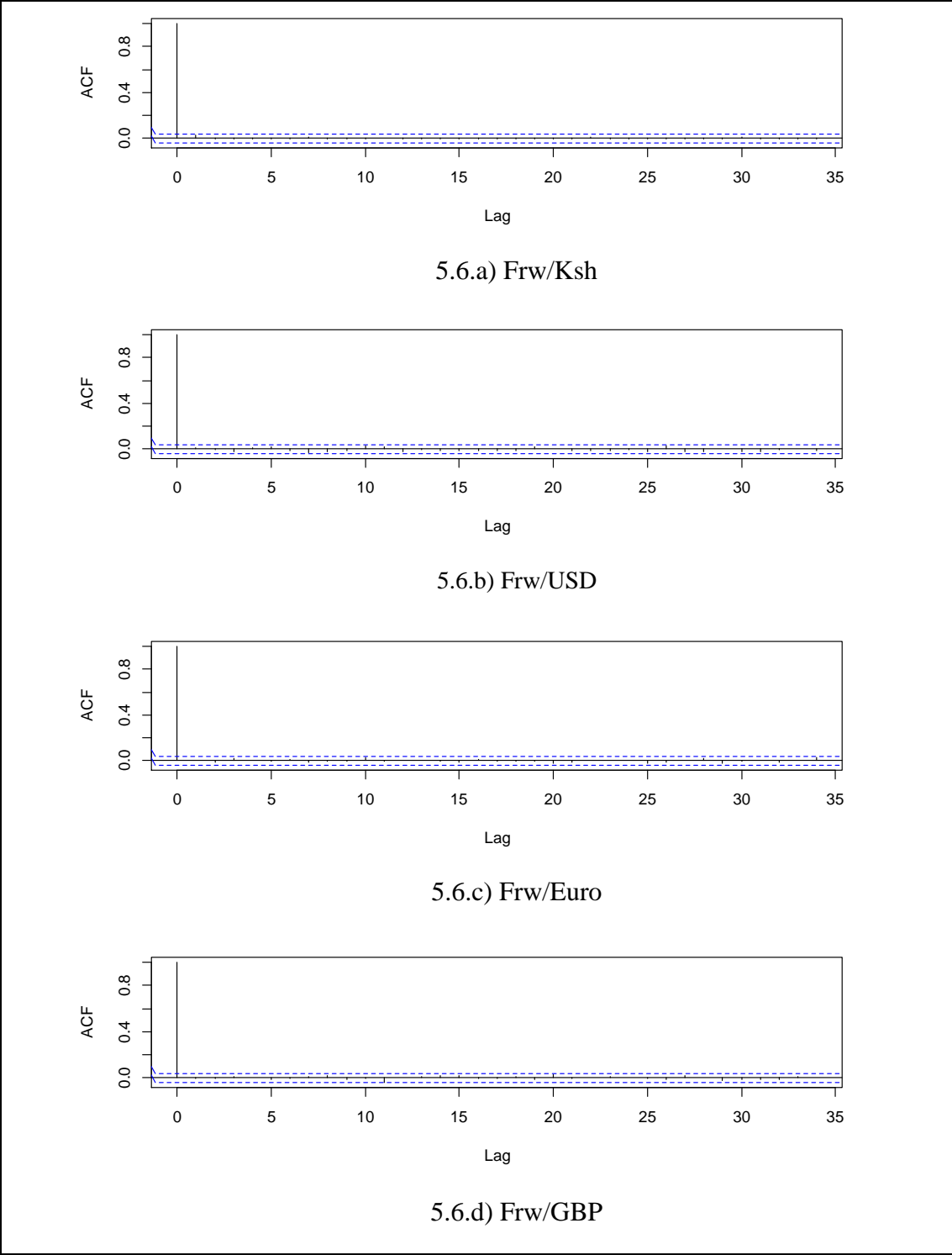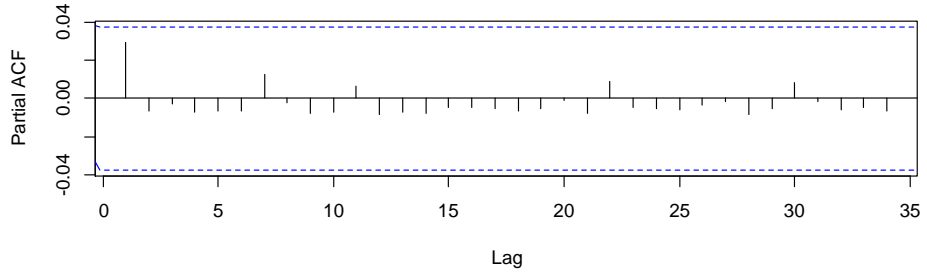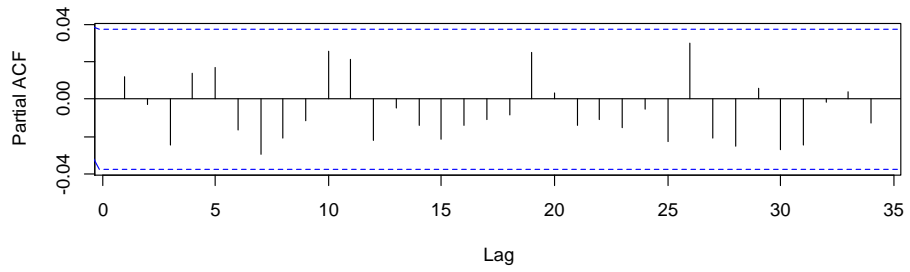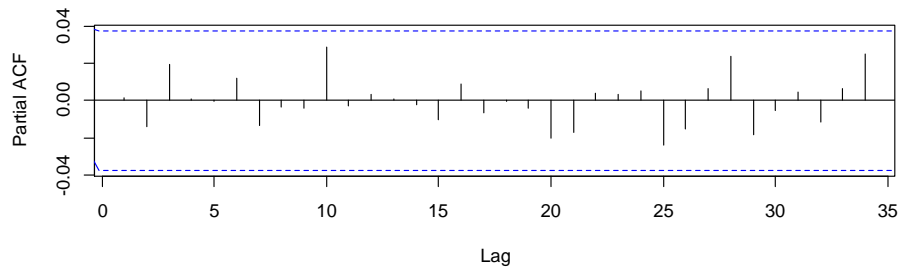
5.6.a) Frw/Ksh

5.6.b) Frw/USD

5.6.c) Frw/Euro

5.6.d) Frw/GBP

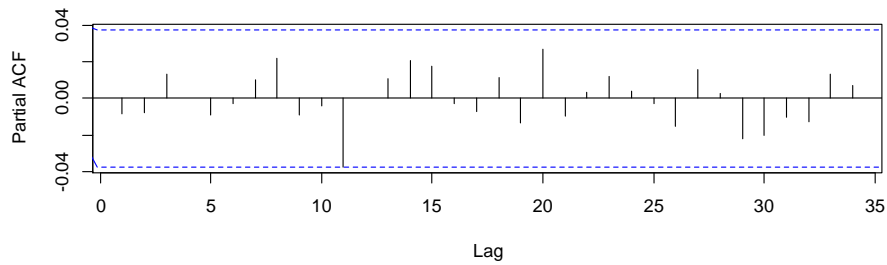**Figure 4.6 ACF of squared residuals for Daily exchange rates**

5.7.a) Frw/Ksh

5.7.b) Frw/USD

5.7.c) Frw/Euro

5.7.d) Frw/GBP

Figure5.7: PACF of squared Residuals

The plots in Figures 5.6 and 5.7 reveal that the residuals exhibit no autocorrelation in all lags for Frw/Euro as well as Frw/GBP but exhibit low autocorrelation in Frw/Ksh for only first lag and for Frw/USD up to 4th lag. Since the standardized residuals exhibit insignificant autocorrelation.

Table5.6: Summary statistics of squared standardized residuals

Tests	Statistics	Frw/Ksh	Frw/USD	Frw/Euro	Frw/GBP
JB	$\chi^2(2)$	7428.669	5415.314	51183.32	40820.44
	$p-value$	< 2.2e-16	<2.2e-16	< 2.2e-16	<2.2e-16
LM	$\chi^2(12)$	16.8744	31.5687	2.2033	3.8016
	$p-value$	0.1544	0.001612	0.999	0.9868

The summary statistics of squared residuals presented in Table 5.6 above where JB test for normality with 2 degree of freedom rejects the null hypothesis to mean that the standardized residuals are not normally distributed. LM test for ARCH effects rejects null hypothesis at lag 12 for Frw/Ksh as well as for Frw/USD and accepts null hypothesis for Frw/Euro and for Frw/GBP. This indicates that up to lag 12 standardized residuals series of Frw/Ksh and that of Frw/USD exhibit ARCH effects whereas standardized residuals series of Frw/Euro and that of Frw/GBP do not present ARCH effects at lag12. Since there is evidence that standardized residuals are not normally distributed and therefore EVT is needed to estimate tails of the exchange rates distributions data.

5.3.1 Threshold selection techniques

1)   Quantile-Quantile Plots

The QQ-plots known as Quantile-Quantile plots are needed for two reasons, first, it completes the results obtained using JB test for normality. This means that JB test showed that the squared residuals are not normal while QQ-plots in Figure5.8 below revealed that standardized residuals are fat tailed. This is the reason of using Extreme value theory to estimate the tails of innovations. Secondly, QQ-plots may also be applied to check if the data points satisfy the generalized Pareto Distribution. Picklands (1975) and Balkema & de Haan (1974) showed that if the empirical plots seem to follow a reasonably straight line with a positive gradient above a certain threshold, therefore these indicate that the exchange rates data follows a Generalized Pareto Distribution with scale and shape parameters. It is possible to choose the threshold where an approximation by the GPD is reasonable by detecting an area with a linear shape on the plot.
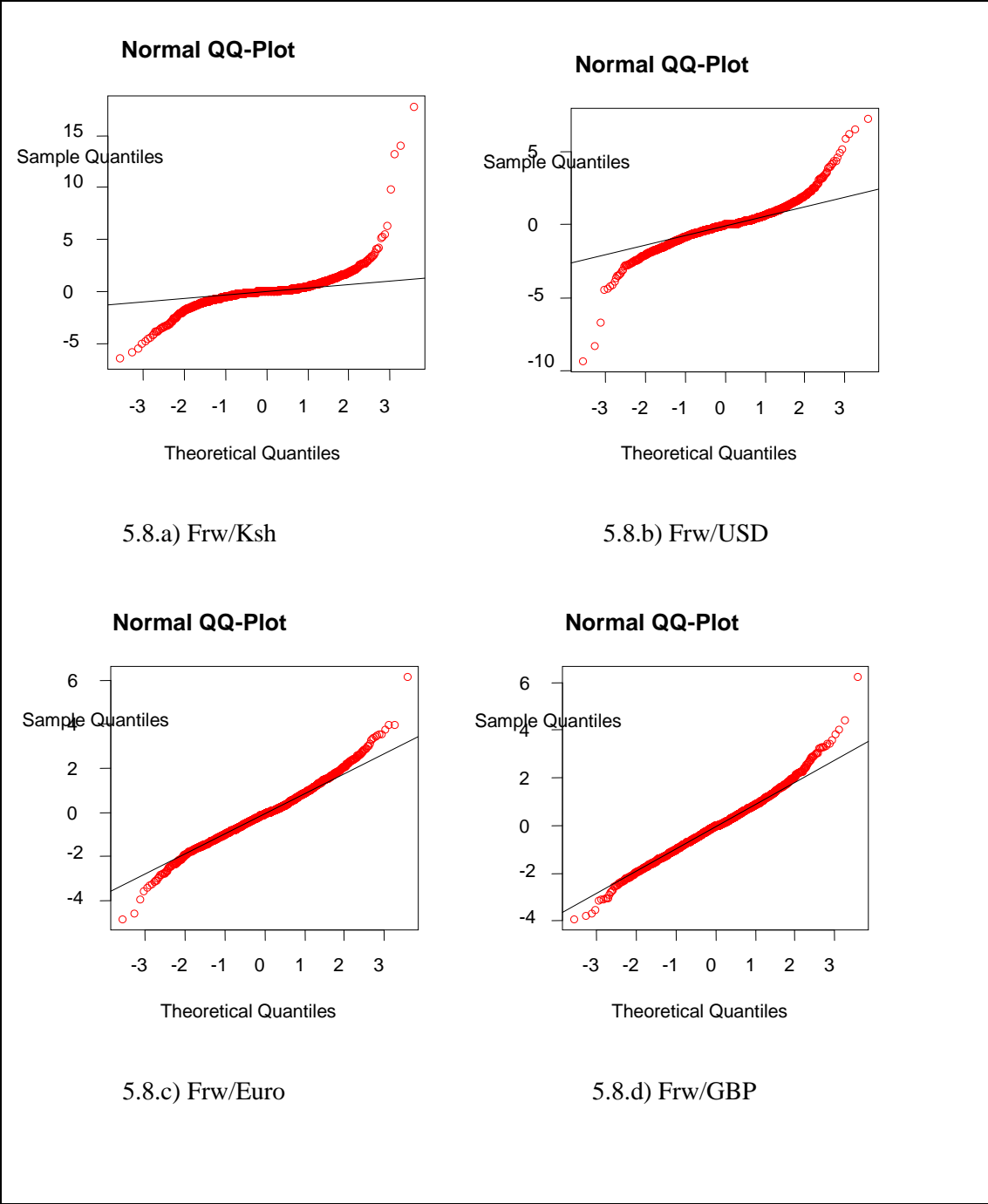
Figure5.8: Quantile-Quantile plots of residuals against the normal distribution

As it can be seen in plots from Figure 5.8, Quantile-Quantile plots of residuals against normal distribution confirm that the standardized residuals for all currencies have a

fat tail since each plot curve down the left and up the right (concave curve). Hence the assumption of conditional normality is unrealistic.

2)   **Mean Residual Life Plot**

The Mean Residual Life Plot (Mean Excess plot) is one of the most common used graphical methods. The reasons behind is that the distribution of exceedances over the threshold τ is a Generalized Pareto Distribution (GPD) of excedances over any threshold $\tau_1 > \tau$.
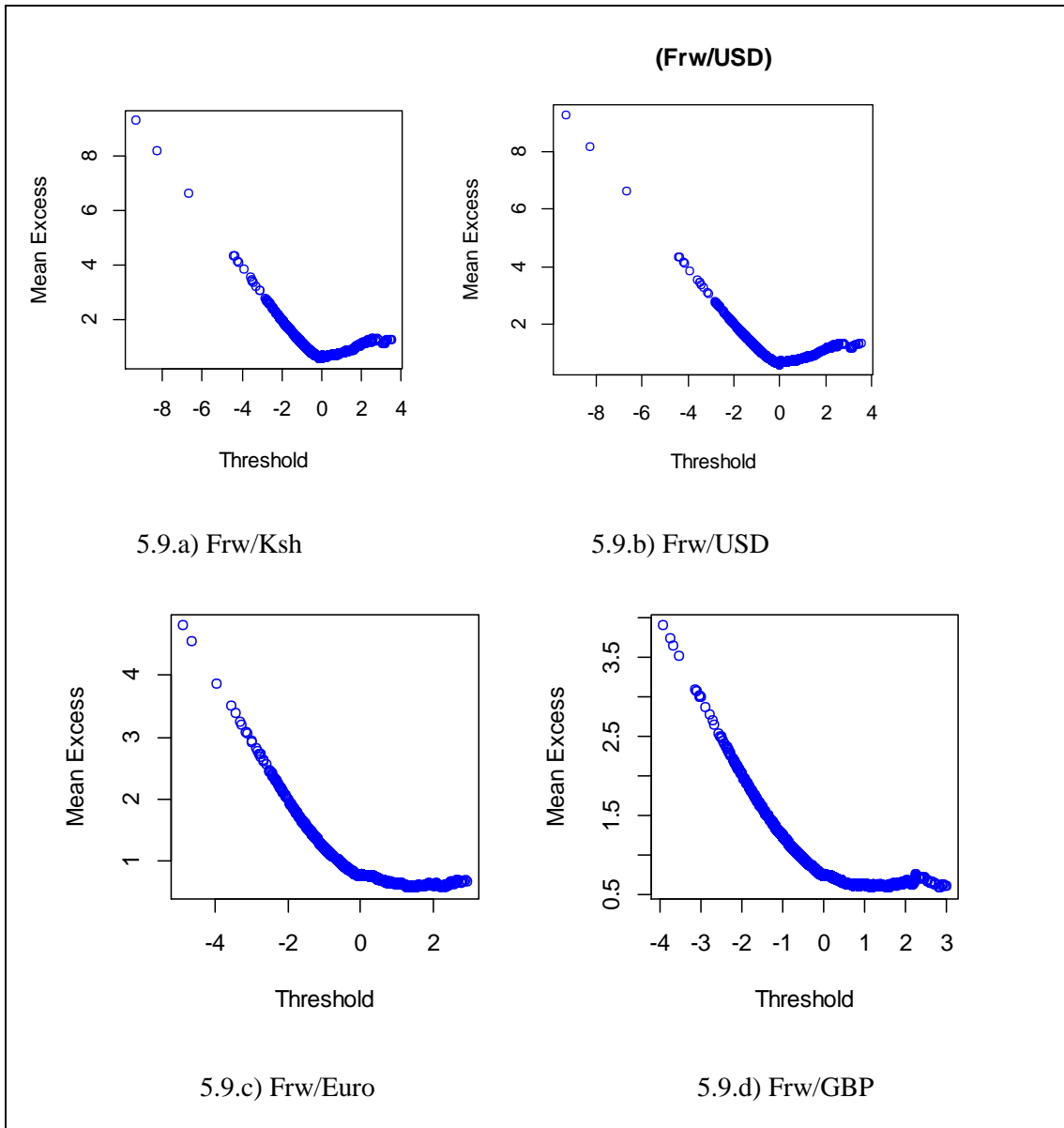
Figure5.9: Mean excess function against threshold

It is observed that mean excess plots in Figure 5.9 shows an upward trend for each currency, which indicates heavy tail behavior. Particularly, since the plot seems to follow a straight line with positive gradient above a certain value of threshold, this is evidence that our data follow a GPD with a positive shape parameter in the tail area above a certain threshold.
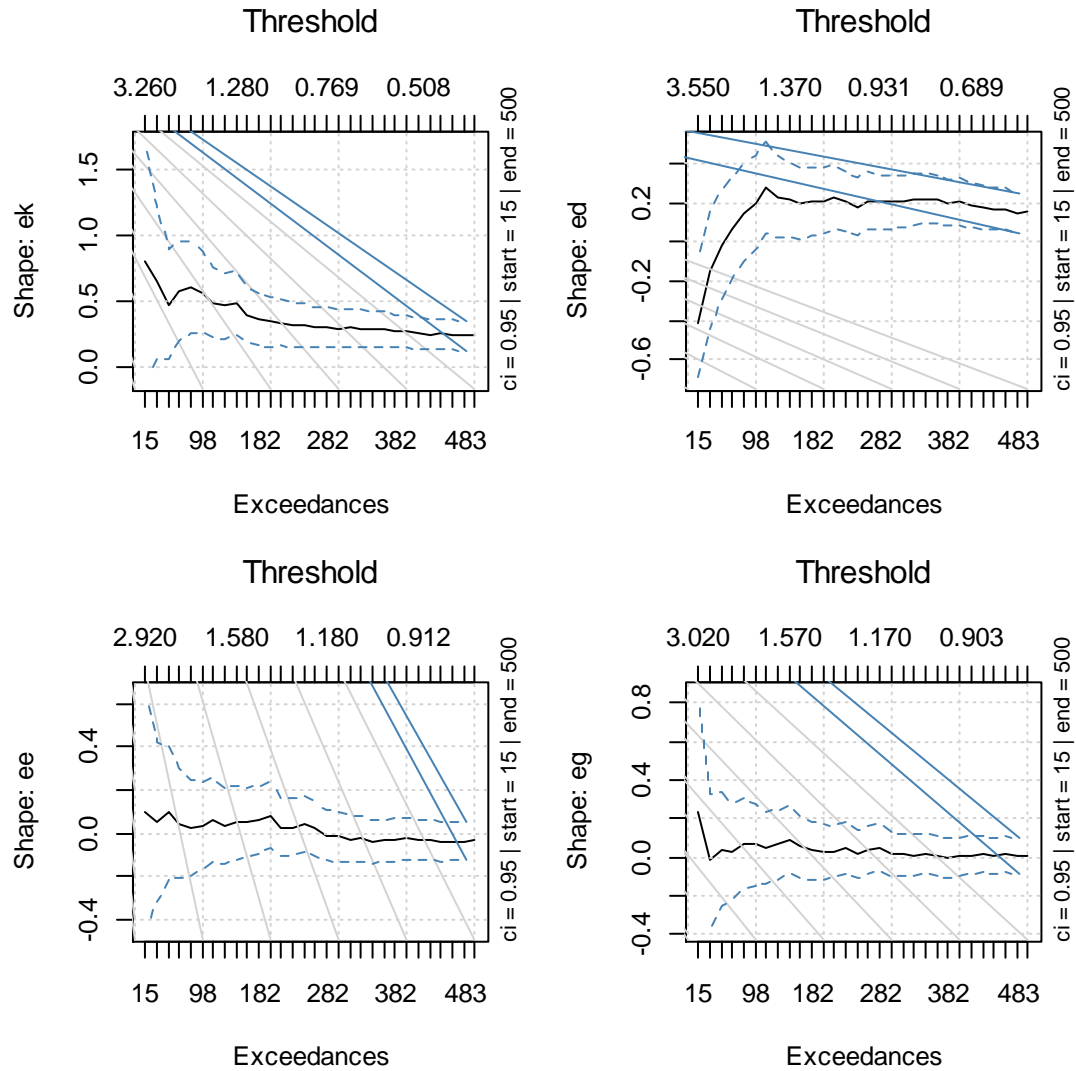
Figure5.10: Shapes against exceedances plots

The shapes against exceedances plots from Figure5.10 above are helpful in threshold selection, where threshold is chosen where the line seems to be horizontal. We also pay attention on the number of observations which require appearing above threshold. Since if low threshold is chosen the number of observations (exceedances) increase and the estimation becomes smoother and also it introduces some observations from the centre of the distribution and estimation becomes biased. If high threshold is

chosen, the estimates based on few largest observations are highly sensitive with large variability.

5.3.2   Estimated parameters of Generalized Pareto Distribution

After identifying the threshold for each currency, the observations in excess of the thresholds are used to determine the Generalized Pareto Distribution parameters which are shape and scale parameters. The statistic results of the estimated parameters of Generalized Pareto Distribution are presented in Table5.7 below;

Table5.7: Generalized Pareto Distribution parameter estimates

Frw/	Threshold(τ)	$\hat{\zeta}$	$\hat{\psi}$	N_τ
Ksh	0.815	0.307503	0.6205089	249
USD	0.975	0.2108428	0.6324132	286
Euro	1.31	0.01203404	0.62319310	225
GBP	1.36	0.0335017	0.6004739	212

$\hat{\zeta}$ represents the shape parameter which determines the type of the distribution, it is positive for all currencies. This indicates that the distributions of selected currencies belong to maximum domain of attraction of Frechet distribution which is heavy tailed. $\hat{\psi}$ represents the scale parameter of underlying distribution.

5.3.3   Extreme Quantiles Estimates

Using the shape and scale parameter estimates obtained above, we can obtain the quantiles at extreme probability values for independent and identically distributed standardized residuals. Let $\hat{e}_\varphi$ be the quantiles estimate of innovations at

probability φ. Typically, the probability φ is such that $0.95 \leq \varphi < 1$. Recall that the quantile estimate is defined in (4.30) as follows.

$$\hat{e}_{\varphi}(\tau) = \tau + \frac{\hat{\psi}_{\tau}}{\hat{\zeta}}\left(\left(\frac{N(1-\varphi)}{N_{\tau}}\right)^{-\hat{\zeta}} - 1\right) \tag{5.2}$$

Where $\hat{\zeta}$ and $\hat{\psi}_{\tau}$ are the estimates of ζ and $\psi(\tau)$ shape and scale parameter respectively. N represents number of observations and N_{τ} represents number of observations over threshold τ. In this work we choose $\varphi = 0.95, 0.99$ and 0.995 the results are presented in table 5.8 below

Table5.8: summary statistics of extreme quantiles estimates

Frw/	$\hat{e}_{0.95}$	$\hat{e}_{0.99}$	$\hat{e}_{0.995}$	$\widehat{UES}_{0.95}$	$\widehat{UES}_{0.99}$	$\widehat{UES}_{0.995}$
Ksh	1.217101	2.766735	3.709776	3.0155	5.253249	6.615046
USD	1.473689	2.887003	3.65989	2.929295	4.72021	5.699593
Euro	1.614201	2.601706	3.021144	2.199196	3.174958	3.589409
GBP	1.620108	2.627472	3.078336	2.33135	3.373632	3.840125

5.4 Conditional VaR and Conditional ES Estimates

The extreme quantiles presented in Table 5.8 are unconditional. To obtain conditional quantile e_{φ}^{t} we combine conditional volatility estimate with the estimated unconditional quantile. Thus conditional quantile e_{φ}^{t}, assuming the mean is negligible, is given as

$e_\varphi^t = \hat\sigma_t \hat e_\varphi, t = 1,2,\ldots,2759$

Using the results obtained in Table 5.8 we can apply the following formula (Eq.5.3) to obtain conditional VaR estimate at $\varphi = 0.995$.

$$\widehat{CVaR}_\varphi^t = \hat\sigma_t\left(\tau + \frac{\widehat\psi_\tau}{\widehat\zeta}\left(\left(\frac{N(1-\varphi)}{N_\tau}\right)^{-\widehat\zeta} - 1\right)\right) \tag{5.3}$$

The results of conditional Value at Risk estimates are presented graphically in Figure5.11 below.

5.11.a) Frw/Ksh

5.11.b) Frw/USD

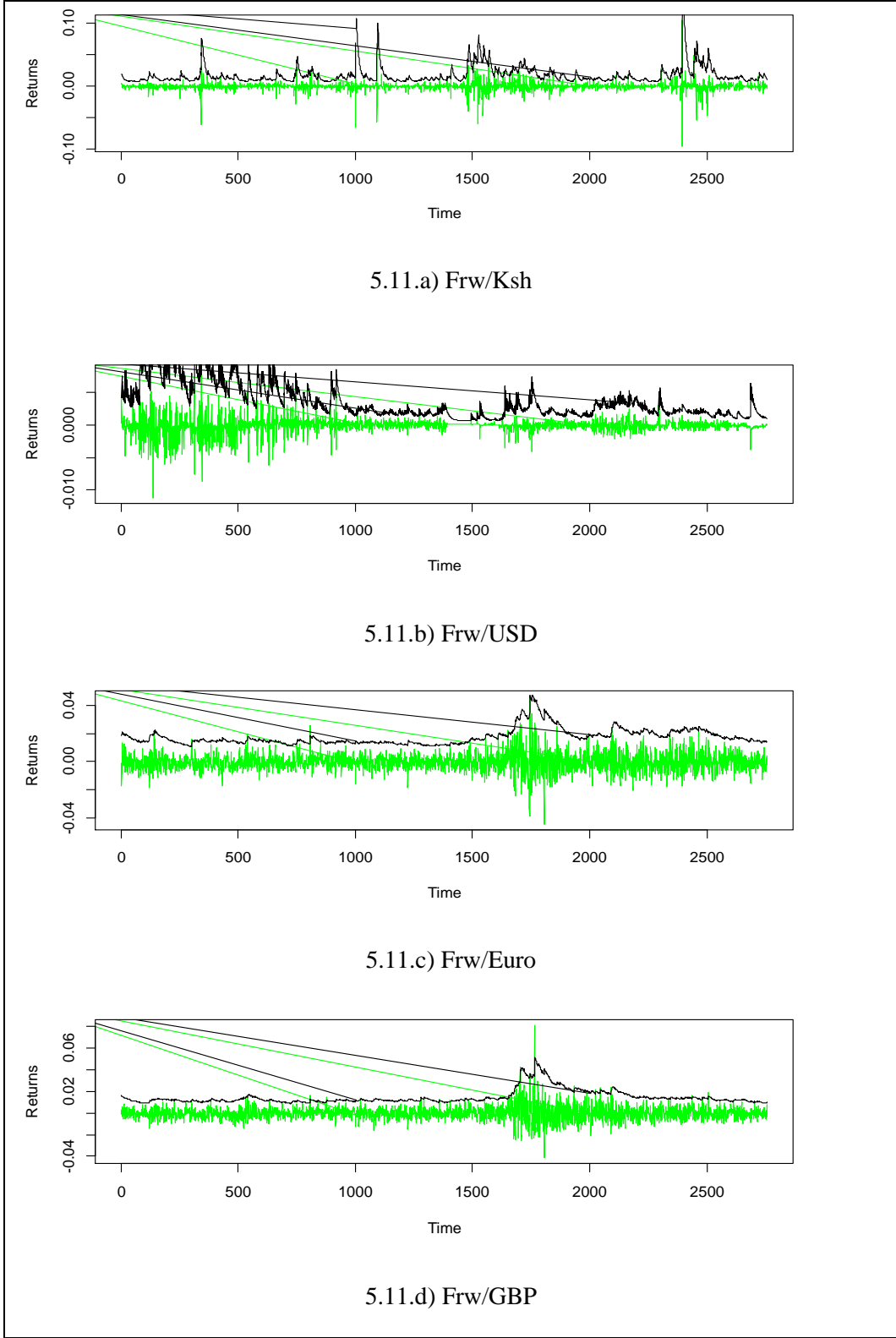5.11.c) Frw/Euro

5.11.d) Frw/GBP

Figure5.11: Exchange Rate Returns with Conditional Value at Risk

The plots in Figure 5.11 show the conditional Value at Risk estimated at probability level $\varphi = 0.995$. The plots in black colour give the extreme quantiles on the daily returns in blue colour. The plots look reasonable compare to the formulae used. Quantiles estimates indicate the scale of losses that could be received if the threshold were to be exceeded. These are referred to as conditional Value at Risk which risk managers have to monitor regularly. If they fall above a certain level the management should be able to advice the institution accordingly.

As it is described some limitations of VaR in previous chapters, we need to plot conditional Expected Shortfall estimate to overcome these shortcomings of VaR. Estimation of conditional Expected Shortfall, under extreme conditions, requires estimation of volatility σ_t and using appropriate extreme value distribution to get quantiles. The plots in Figure 5.12 below show the conditional Expected Shortfall estimated at probability level $\varphi = 0.995$. Recall the equation of estimated conditional is;

$$\widehat{CES}_\varphi^t = \hat{\sigma}_t \left(\frac{\widehat{VaR}_\varphi}{1-\hat{\xi}} + \frac{\hat{\psi}_\tau + \hat{\xi}\tau}{1-\hat{\xi}} \right) \tag{5.4}$$

Where $\widehat{UVaR}_\varphi$ is unconditional VaR which can be compared as $\hat{e}_{\varphi \geq 0.95}$. The Conditional Expected Shortfall estimate is presented in plots below.
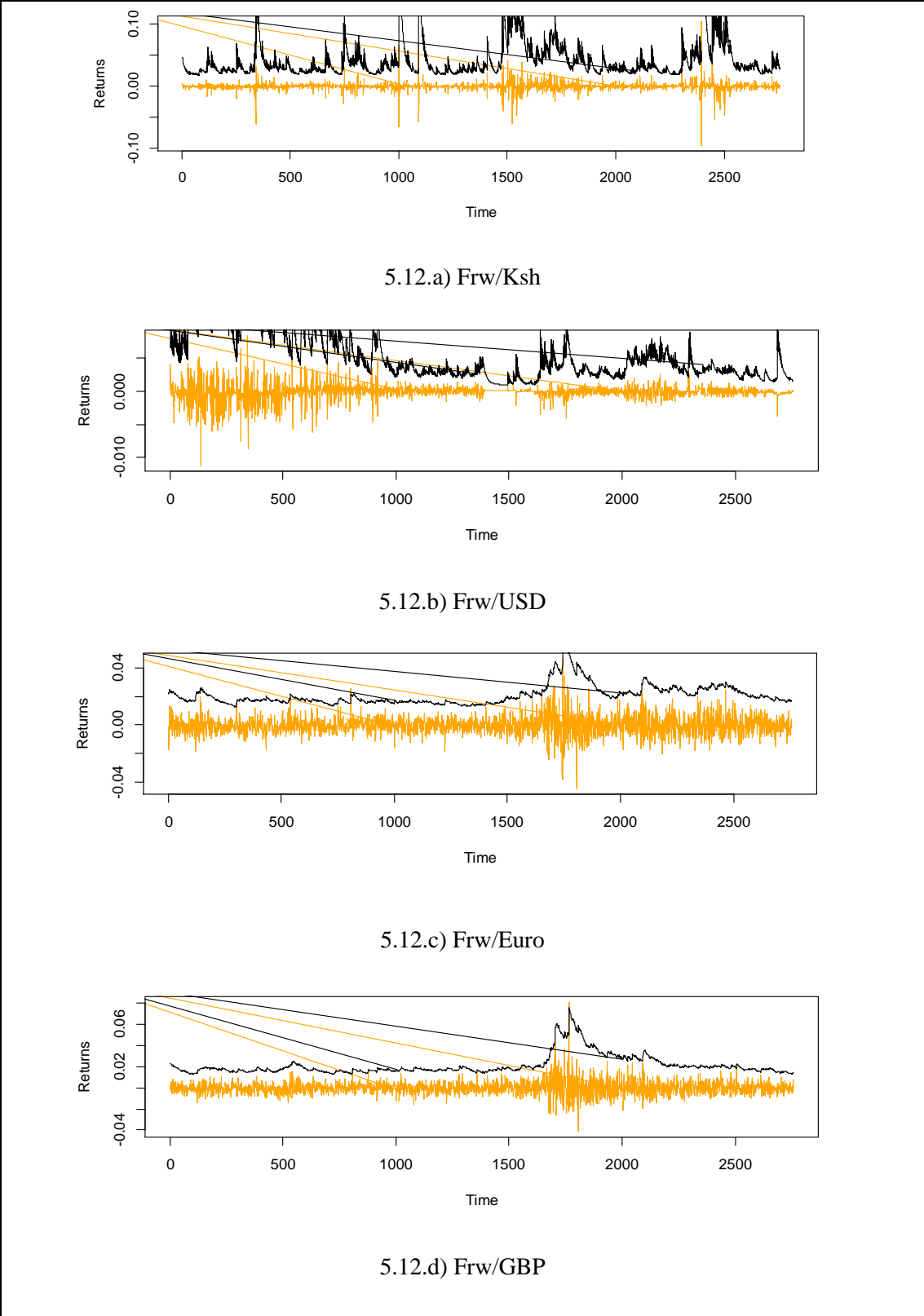
5.12.a) Frw/Ksh

5.12.b) Frw/USD

5.12.c) Frw/Euro

5.12.d) Frw/GBP

Figure5.52: Exchange Rate Returns with Conditional Expected Shortfall

87

From Figure 5.12, the plots in black color give the Expected Shortfall on the daily returns (orange). When you compare Plots in Figure 5.12 of Expected Shortfall and that in Figure5.11 of VaR it is clear that Expected Shortfall contains some information beyond VaR. The CVaR and CES values change dynamically to reflect exchange rate markets conditions in periods of extreme changes and when these values increase the exchange rate markets makers should be careful.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1   Conclusions

In this work, the estimate extreme risk in Rwanda exchange rate series using conditional Value at Risk and conditional Expected Shortfall estimates has been given. These estimates are obtained by combining GARCH model in the estimation of volatility and concepts from Extreme Value Theory. Quasi-Maximum likelihood procedure is used to estimate parameters in GARCH. The estimators are found to be consistent and asymptotically normal. The exploratory analysis showed that exchange rates data are not normally distributed and exhibit leptokurtosis. The distributions of the returns in Frw/Ksh and in Frw/USD exhibit negative skewness this implies that investors can have frequent small gain and few extreme losses. The returns series of Frw/Euro and that of Frw/GBP have positive skewness coefficients this implies that depreciations in the exchange rates occur slightly more often than appreciation. This indicates that investors can have frequent small negative outcomes and few extreme gains. Lagrange Multiplier test showed presence of ARCH effects in both returns series and residuals.

Generalized Pareto Distribution was fitted to the standardized residuals and then, the estimated distribution inverted to obtain extreme quantiles at 99% and at 99.5% probability levels. The Maximum Likelihood estimator of parameters was found to be consistent and asymptotically normal. The conditional value at risk and conditional Expected Shortfall are obtained by combining the two consistent estimators.

A robust risk measure of exchange rate data has been provided. Estimating the uncertainty of value at risk and Expected Shortfall is significant since it allows policy makers and risk managers to make good decisions about direction of portfolio. When compare these models to other single modeling methods for financial data estimation it is clear that dynamic method such as GARCH model with normal distribution assumption provides good estimates as well as the extreme value theory with independent and identically distribution residuals assumption. However, both tend to be violated more often because they do not take into account the leptokurtosis of the residuals. Finally, market makers, risk practitioners, traders, investors and risk managers should understand well the development of GARCH model and Extreme Value Theory approach to estimate conditional value at risk and conditional Expected Shortfall which are crucial in decision making.

6.2   Recommendations

Nelson (1991) introduced Exponential Generalized Autoregressive Conditional Heteroscedasticity (EGARCH) model, listed three shortcomings with the GARCH models; the lack of symmetry in the response of shocks, restrictions imposed to GARCH models to ensure that the conditional variance is positive and the difficulty in measuring persistence using standard GARCH models. From these drawbacks it is recommended that future research should focus on asymmetric models to see whether these shortcomings have significance impact to extreme risk estimation. The threshold level selection is another challenge, since a low threshold value estimator results in biased estimator and setting a too high threshold leads to a reduction of the number of extreme observations and hence increased variance. Therefore the challenge to find optimal threshold forms part of future problem in this area.

References

Acerbi, C., Srtori, C., & Nordio, C. (2001). as a Tool for Financial Risk Management. *Working Paper* .

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov and F. Cs'aki eds. *Akad' emia Kiad'o, Budapest* , 267-281.

Ammann, M., & Reich, C. (2000). Value at Risk for Nonlinear Financial Instruments. *Working paper* .

Andersen, T., & Sorensen, B. (1996). GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business and Economics Statistics* , 329-352.

Angelidis, T., & Degiannakis, S. (2004). Modeling Risk for long and short trading positions. *Journal of Risk Finance Vol.6, No.3* , 226-238.

Artzner, P., Delbaen, F., & Heath, D. (1998). Coherent measures of Risk. *Mathematical Finance Vol.9* , 203-228.

Balkema, A., & De Haan, L. (1974). Residuals life time at great age . *Annals Probability Vol2* , 792-804.

Beirlant, J., Vynckier, P., & Teugels, J. (1996). Tail index estimation, Pareto quantile plots and regression diagnostics. *Journal of American Statistics Association Vol.91* , 1659-1667.

Blum, P., & Dacorogna, M. (2002). Extreme moves in daily Foreign Exchange rates and risk limit setting. *Department of Mathematics General Guisan Quasi Vol26* , 8092Z'urich.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of EconometricsVol.31* , 307-327.

Dekkers, A. L., Einmahl, J., & De Haan, L. (1989). A Moment Estimator for the Index of an Extreme-Value Distribution. *The Annals of Statistics Vol.17* , 1833-1855.

Delbaen's, F. (2002). Coherent risk measures on general Probability soaces. In: Advancees in Finance and Stochastics. *Essays in Honour of Dieter Sondermann. Springer- Verlag Berlin* , 1-37.

Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). Modelling Extremal Events for insurence and Finance. *Springer-Verlag, Berlin* .

Embrechts, P., Sidney, I., Resnick, & Samorodnitsky, G. (1999). Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal Vol.3* .

Engle, F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation Vol.50. *Econometrica* , 50-62.

Engle, R. (2004). Risk and Volatility: Econometric Models and Financial Practice. *American Economic Review Vol.94* , 405-420.

Fisher, R., & Tippett, H. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a Sample. *Proceedings of the Camblidge Philosophical Society Vol.24* , 180-190.

Francq, C., & Zakoian, J. (2004). Maximam Likelihood Estimationof Pure GARCH and ARMA-GARCH Processes. *Bernoulli Vol.10* , 605-637.

Ghysels, E., Harvey, A., & Renault, E. (1996). Stochastic Volatility. In Handbook of Statistics. *Statistical Methods in Finance, Maddala, G.S. Ed. North- Holland, Amsterdam Vol.14* , 119-91.

Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relation between the Expected Value and the Volatility of nominal excess return on stocks. *Journal of Finance Vol.48* , 1779-1801.

Gnedenko, B. V. (1943). Sur la Distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics Vol.44* , 423-453.

Gourieroux, C., Jasiak, J., & Sufana, R. (2009). The Wishart Autoregressive Progress of Multivariate Stochastic Volatility . *Journal of Econometrics Vol150* , 167-181.

Gravril, A., & Altar, M. (2009). Exchange Rate Risk: Heads or Tails. *Working Paper* .

Hendricks, D. (1996). Evaluation of Value at Risk Models Using Historical data. *Econometric Policy Review, Federal Reserve Bank of New York, New York, NY* , 39-69.

Hill, B. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *The Annals of Statistics Vol.3* , 1163-1174.

Hols, M. A., & de Vries, C. (1991). The Limiting Distribution of Extremal Exchange Rate Returns. *Journal of Applied Econometrics V0l.6 N0.3* , 287-302.

Huisman, R., Koedijk, K., Kool, C., & Palm, F. (1998). The Fat-Tailedness of Exchange rate Returns. *Working Paper at University of Maastricht, Department of Economics Center for Economic studies and Institute for Economic Research, Maastricht.*

Hull, J., & White, A. (1998). Value at Risk When Changes in Market Variables are not Normally Distributed. *Journal of Derivatives Vol.5* , 9-19.

Jarque, C., & Bera, A. (1987). A Test for Normality of Observations and Regression Residuals. *International Review Vol.55 No.2* , 163-172.

Jenkinson, A. (1955). The frequency distribution of the annual maximum (or Minimum) Values of the Meteorogical elements. *Quart.J.Roy. Vol.81* , 158-171.

Jorian, P. (2007). *Value at Risk, the new benchmark for managing financial risk.* Third Edition McGraw-Hill.

Juan-Juan, C., Einmahl, J. H., & DE Haan, L. (2011). Estimation of Extreme risk Regions under Multivariate Regular Variation. *Journal of Annals Statistics Vol39, No.3* , 1803-1826.

Jushan, B. (2005). Test for skewness, kurtosis and Normality for time series data. *American Statistics Association Journal of Business and Economics statistics Vol.23, No.1* .

Ljung, G., & Box, G. (1978). Measure of lack of fit in time series models. *Biometrika Vol65* , 297-303.

Longin, F. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business Vol.69* , 383-408.

Lumsdaine, R. (1996). Consistency and asymptotic normality of the Quasi-Maximum Likelihood estimator in GARCH (1,1) and covariance stationary GARCH (1,1) Models. *Econometrica Vol64* , 575-596.

Maana, I., Mwita, N., & Odhiambo, R. (2010). Modelling the Volatility of Exchange rates in the Kenya Market. *African Journal of Business Management Vol.4 No.7* , 1401-1408.

MacNeil, A. (1999). Extreme Value Theory for Risk Managers. *ETH preprint (www.math.ethz.ch/~m neil)* .

Madura, J. (1989). *International Financial Management 2nd edution.* St.Paul, Minnesta: West Publishing Company.

Manganelli, S., & Engle, R. (2001). Value at Risk Models in Finance,European Central Bank Working Paper Series, Frankfurt. Modelling. *Mathematical Finance Vol.14* , 75-102.

McNeil, A. (1998). *Calculating Quantile Risk Measures for Financial Return Series Using Extreme Value Theory.* ETH preprint(www.math.ethz.ch/~mcneil).

McNeil, A. (1997a). Estimating the tails of loss severity Distributions Using Extreme Value Theory. *ASTIN Bulletin Vol.27* , 117-137.

McNeil, A. (1999). *Extreme Value Theory for Risk Managers.* ETH preprint (www.math.ethz.ch/~mcneil).

McNeil, A. (1997b). *The Peaks over Threshold Method for Estimating High Quantiles of Loss Distributions.* ETH preprint(www.math.ethz.ch/~mcneil).

McNeil, A., & Frey, R. (2000). Estimation of the tail related risk measures for heteroscedastic Finance time series: An extreme value approach. *Journal of Empirical Finance Vol.7* , 271-300.

Neftci, S. (2000). Value at Risk Calculations Extreme Events, and Tail Estimation. *Journal of Derivatives vol.7 No.3* , 23-38.

Nelson, D. (1991). Conditional Heteroscedasticity in Asset Returns: A new approach. *Econometrica Vol.59* , 347-370.

Neves, C., & Fraga Alves, M. (2008). Testing Extreme Value Conditions- an overview and recent approaches. *Statistical Journal* , 83-100.

Papaioannou, G. M. (2006). Exchange Rate Risk Measurement and Management: Issues and Approaches for Firms . *South-Eastern Europe Journal of Economics Vol.2* , 129-146.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics Vol.3* , 119-131.

Posedel, P. (2005). Properties and Estimation of GARCH(1,1) Model . *Metodolo S Ki Zvezki Vol.2* , 243-257.

Resnick, S. (2007). Heavy-Tail Phenomena-Probability and Statistical Modelling . *Springer, New York* , 73-114.

Ruey, S. (2005). *Analysis of Financial Time Series 2nd Edition.* Wley Interscience A JOHN Wiley and Sons INC., Publication.

Sandmann, G., & Koopman, S. (1998). Estimation of Stochastics Volatility Models Via Monte Carlo Maximum Likelihood . *Journal of Econometrics Vol.87* , 271-301.

Schwartz, G. (1978). Estimating the Dimension of a Model . *Annals of Statistics Vol6.* , 461-464.

Smith, F. (1994). *Writing and the write 2nd Edition Hillsdale.* NJ: Lawrence Erlbaum Assocates.

Smith, R. (1987). Estimation of Tails of Probability Distributions . *Annals Statistics Vol15* , 1174-1207.

Smith, R. (1985). Maximum Likelihood Estimation in a Class of non-regular cases. *Biometrika Vol.72* , 67-90.

Smith, W., Clifford, C., & Stykes, W. (1990). Managing Financial Risk. *New York, Harper and Row.*

Snedecor, G., & Cochram, W. (1980). Statistical Methods, 7th Edition. *Iowa State University Press, Ames, IA.*

Soltane, H. B., Karaa, A., & Makram, B. (2012). Conditional VaR Using GARCH-EVT approach: Forcasting Volatility in Tunisian Financial Market . *Journal of Computations &Modelling Vol.2, No.2* , 95-115.

Straumann, D., & Mikosch, T. (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A stochastic Recurrence Equations Approach. *The Annals of Statistics Vol.34 No.5* , 2449-2495.

Thompson, P., Cai, Y., Reeve, D., & Stander, J. (2009). Automated Threshold Selection Methods for Extreme wave analysis. *Coastal Engineering Vol.56* , 1013-1021.

Todorovic, P., & Zelenhasic, E. (1970). A Stochastic Model for flood analysis . *Water Resour. Res. Vol.6 No.6* , 1641-1648.

VonMises, R. (1936). La Distribution de la plus Grande de n Values . *Reprinted in Selected Paper Vol.2 American Mathematical Society Providence,R.I* , 271-294.

Wagner, N., & Marsh, T. (2003). Measuring Tail Thickness under GARCH and Application to Extreme Exchange Rate Changes. *Working Paper at Haas School of Business University of Calfornia Berkeley, California.*

Wang, W., Van Gelder, J., Vrijling, J., & Ma, J. (2005). Testing and Modelling Autoregressive Conditional Heteroskedastity of Streamflow Processes . *Nonlinear Processes in Geophysics Vol.12* , 55-66.

Xiangxian, Z., & Wenlei, G. (2009). A New Method to Choose the Threshold in the POT Model. *First International Conferece on Information Science and Engineering ICISE Vol.9* , 750-753.