KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI

DEPARTMENT OF MATHEMATICS

FACULTY OF PHYSICAL SCIENCES

COLLEGE OF SCIENCE

APPLICATION OF EXTREME VALUE THEORY FOR ESTIMATING DAILY BRENT CRUDE OIL PRICES

BY

ABDUL-AZIZ IBN MUSAH

FEBRUARY 2010

APPLICATION OF EXTREME VALUE THEORY FOR ESTIMATING DAILY BRENT CRUDE OIL PRICES

BY

ABDUL-AZIZ IBN MUSAH

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN

PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER OF SCIENCE

FACULTY OF PHYSICAL SCIENCES

COLLEGE OF SCIENCE

FEBRUARY, 2010

DECLARATION

I hereby declare that this submission is my own work towards the M.Sc. and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

Abdul-Aziz Ibn Musah (PG 9670306) ……………………….. …………

(Student/ID) **(Signature)** **(Date)**

Certified by:

Dr. F.T Oduro ....…………………………. ………………..

(Supervisor) **(Signature)** **(Date)**

Mr. S.K. Appiah ……..…………

(Supervisor) **(Signature)** **(Date)**

Dr. S.K. Amponsah ...…………………………… …………….

(Head of Department) **(Signature)** **(Date)**

DEDICATION

This work is dedicated to:

- MY LATE MOTHER OF BLESSED MEMORY AND ANY OTHER FAMILY MEMBER
- THE NEEDY IN THE SOCIETY

ACKNOWLEDGEMENTS

It is quite natural that no academic and research work of this nature can be successfully accomplished without the least form of assistance from one quarter or the other. I have realized the logic in this assertion and the success of this assignment relied on certain personalities who need to be commended.

I thank the Almighty God for the divine guidance throughout the period of study. The successful completion of this work was made possible by the support I received from my friends particularly Mr. Salman Ibn Faris; a research student of Department of Mathematics, KNUST, Bashiru Imoro Ibn Saeed, Lecturer at Kumasi Polytechnic and Mr. Alhassan Abdul-Rasheed Akeji, Lecturer at Tamale Polytechnic. I again extend my most heartfelt thanks to Mr. S.K. Appiah, under whose dedicated supervision; this work has come to reality despite his heavy schedule. I am equally grateful to Mr. Umar Gibreel, a brother and friend.

I am particularly grateful to my family whose prayers and sacrifices have seen me through my education all the years. The last but not the least include: Mr. Adam Sayibu Wilberforce (Headmaster of Northern School of Business Senior Secondary School), Mr. Ambrose of Kalipohin Senior Secondary School and Hajia Mariama Mohammed for their efforts and encouragement to make this work come true.

As much as I appreciate the immense contribution of the above mentioned personalities to the successful completion of this work, I am solely responsible for any shortcomings and inadequacies in this work.

ABSTRACT

Assessing the probability of rare and extreme events is an important issue in the risk management of financial portfolios. Crude oil markets are highly volatile and risky. Extreme Value Theory (EVT), an approach to modelling and estimating risks under rare events, has seen a more prominent role in risk management in recent years. This thesis presents an application of EVT to the daily returns of Brent crude oil prices in the spot market between 1987 and 2009. We focus on the peak over threshold method by analysing the generalized Pareto distributed exceedances over some high threshold. This method provides an effective means for estimating tail risk measures specifically, Value-at-Risk (VaR) and Expected Shortfall (ES). The estimates of these risk measures computed under high quantile (99^{th} percentile) provides estimates of VaR as 8.1% and 8.0% for daily positive and negative returns, respectively. The estimates for expected shortfall are 12.3% and 10.7% for daily positive and negative returns, respectively.

TABLE OF CONTENTS

CHAPTER FOUR

FINANCIAL-RISK MEASURES AND MODELLING APPROACHES

CHAPTER FIVE

MODELLING OF DAILY BRENT CRUDE OIL PRICES

CHAPTER : SIX CONCLUSION AND RECOMMENDATIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND INFORMATION

The oil market analyst and investors speak daily about the risks that affect the movement of their world. They search tirelessly to understand how these risks arise, the exact effect they have, and how they act. The past years and even now, have been characterised by significant instability in financial markets worldwide. This has led to numerous criticisms about the existing risk management systems and motivated the search for more appropriate methodologies to cope with rare events that have heavy consequences. The typical question one would like to answer is: “If things go wrong, how wrong can they go?” The problem is then how to model the rare phenomena that lie outside the range of availably observations. In such a situation it seems essential to rely on well founded methodology. The Extreme Value Theory (EVT) provides a firm theoretical foundation on which we can build statistical models describing extreme events.

In many fields of modern science, engineering and insurance, EVT is well established; Embrechts et al., (1997and1999), Reiss and Thomas, (1997)). Recently, numerous research studies have analyzed the extreme variations that financial markets are subject to, mostly because of currency crises, stock market crashes and large credit defaults. The tail behaviour of financial series has, among others, been discussed in Koedijk et al. (1990), Dacorogna et al. (1995), Loretan and Phillips (1994), Longin (2001), Danielsson and de Vries (2000), Kuan and Webber (1998), Straetmans (1998), McNeil (1997), Jondeau and Rockinger (1999), Rootzµen and

KlÄuppelberg (1999), Neftci (2000), McNeil and Frey (2000) and Gencay et al. (2003a). An interesting discussion about the potential of EVT in risk management is given in Diebold et al. (1998).

This thesis deals with the behaviour of the tails of Brent crude oil pricing. More specifically, the focus is on the use of EVT to compute tail risk measures and the related confidence intervals.

1.2 STATEMENT OF PROBLEM

Crude oil markets can be volatile and risky. The world crude oil prices have risen dramatically during the past decade. However, oil prices did not sustain a constant rise – rather, they showed high volatility, reflecting market conditions such as political turmoil, supply disruptions, unexpected high demand and speculation. Research conducted by Acadian Asset Management Inc. shows that the daily returns on crude oil (West Texas Intermediate) have a wider range than that of gold, copper and major U.S. stock market indices through the period of 1990-2006.

The percentage change in oil price over one trade day was as high as 36.12% and as low as -18.13% between 1987 and 2009. For instance, following the September 11, 2001 attacks, the price of oil plummeted as oil traders believed that weakened economies in the U.S. and elsewhere would use less oil. In particular, on September 24, the price went by 24%, its biggest one-day drop through that period. In contrast, on March 23, 1998, the crude oil price increased sharply by 17% over one day because of the news that three of the world's biggest oil producers agreed to cut supply. Relative to the average positive daily return of 1.74% and the average negative daily return of -1.69% during that period, these cases provide examples of extreme events.

Moreover, volatile oil prices may lead to price variability of other energy commodities and can have wide-spread impacts on the international economy. The bulk of Ghana's crude oil is imported from the Europe, Nigeria, U.S.A., Saudi Arabia and other oil producing countries. So, volatility in the price of world's crude oil impacts significantly on the global economy. Specific examples of impacts include the obvious example of gasoline, and hence transportation costs. Other examples include the stock market and exchange rates, which can be affected substantially by the price of oil (Nandha and Faff, 2007), as well as the chemical industry. If relevant risk management organizations and investors in these markets cannot predict and capture the risks appropriately, their losses could be huge. The highly volatile behaviour of crude oil prices and the substantial impacts of this volatility motivate us to undertake research on modelling oil price fluctuations and providing an effective instrument to measure energy price risks.

In order to better disclose the nature of the risks under extreme situations, and finally avoid the risks in the most degree, we need certain risk measures. EVT, a theory for assessing the asymptotic probability distribution of extreme values, models the tail part of the distribution where the risk exists. This theory is playing an increasingly important role in dealing with modelling rare events. While application of EVT is not foolproof, it provides a relatively safe method for extrapolating beyond what has been observed (Embrechts et al., 1997). On successfully modelling tail-related risks, we then need to find suitable instruments to measure these risks. Two popular measures are Value-at-Risk (VaR) and Expected Shortfall (ES). VaR is the maximum loss of a portfolio such that the likelihood of experiencing a loss exceeding that amount over a specified risk horizon is equal to a pre-specified tolerance level. Expected Shortfall measures the mean of the losses that are equal to or greater than VaR. In particular, VaR has become one of the most commonly used techniques in risk management. In order to capture the effect of market behaviour under extreme events, EVT has been widely adopted in VaR estimation in recent years. Because extreme value methods are derived from a sound statistical theory, and provide a parametric form for the tail of a distribution; these methods are attractive when dealing with measuring risks. There is a large literature that studies EVT for risk measures in areas where extreme observations may appear, such as finance, insurance, hydrology, climatology, engineering and modern science. Specifically, numerous studies in finance and insurance have been conducted, including Embrechts et al. (1997), Reiss and Thomas (1997), Danielsson and de Vires (2000), McNeil and Frey (2000), Gencay et al. (2003) and Gilli and Këllezi (2006). However, to the best of our knowledge, there is only limited discussion of the application of EVT to markets for crude oil, which is a crucial commodity to the world economy. Among the studies on Value-at-Risk(VaR) estimation on energy markets with EVT approaches is Krehbiel and Adkins (2005) who examined the price risk in the New York Mercantile Exchange (NYMEX) energy complex. This study constructed risk statistics for unconditional distributions of daily price changes and applies the conditional extreme value method for estimating VaR and related risk statistics. Another research undertaken by Marimoutou et al. (2006) explored the daily spot Brent oil prices and compared the performances of unconditional and conditional EVT models with that of conventional models such as Generalised AutoRegressive Conditional Heteroskesdasticity (GARCH) and historical simulation.

1.3 JUSTIFICATION AND OBJECTIVES OF STUDY

First and foremost, nowadays, statistics admittedly holds an important place in all the fields of our lives. Almost everything is quantified and, most often, averaged. Indeed, ‘averaging’ is the statistical notion most easily understood and most widely used. On the other hand, not few are the cases where the central tendency, as this is captured by an average measure, is not suitable to ‘statistically’ describe the situations and their impacts.

Two important dates in the history of risk management are February 1, 1953 and January 28, 1986. During the night of February 1, 1953 at various locations the sea-dykes in the Netherlands collapsed during a severe storm, causing major flooding in large parts of coastal Holland and killing over 1,800 people. The second date corresponds to the explosion of the space shuttle Challenger. According to Dalal et al. (1989), a probable cause was the insufficient functioning of the so-called O-rings due to the exceptionally low temperature the night before launching. In both of these cases, an extremal event caused a protective system to breakdown. One could argue that statistical concepts, such as averages, also ‘broke-down’, since not only they offer no help but they can also be misleading, if used. In such cases, it is the examination of extremes that provides us with insight of the situations.

The primary goal of this research is to use EVT to assess the size of extreme/rare events of Brent crude oil. In particular:

- To model related risk measures such as value-at-Risk, expected shortfall and return levels applying it to daily log-returns on Brent crude oil pricing.
- To compare the methods of Extreme Value Theory.

Natural or man-made disasters, crashes on the stock market or other extremal events form part of society. The systematic study of extremes may be useful in contributing towards a scientific explanation of these. As will be made clearer in the sequel, analysis of extreme values is an aspect of statistical science that has much to offer too many fields of human activity. Secondly, most of the world's markets including Ghana are affected by the fluctuations of Brent Crude oil pricing and therefore the need to model the extremal events particularly at the tail of the distribution. Finally, Ghana will be going to oil production next year and therefore the need to estimate the risk measures associated with the world oil pricing.

1.4 METHODOLOGY OF STUDY

To achieve these objectives of the study, both qualitative and quantitative approaches are employed. Data analysis will be primarily on secondary data from http://tonto.eia.doe.gov/dnav/pet/hist/rbrted.htm. A starting point for modelling the extremes of a process is based on distributional models derived from asymptotic theory. The p*arametric approach* to modelling extremes is based on the assumption that the data in hand ($X_1, X_2, ..., X_n$) form an independent identical distribution (i.i.d) sample from an exact General Extreme Value (GEV) distribution function. In this case, standard statistical methodology from parametric estimation theory can be utilised in order to derive estimates of the parameters. In practice, this approach is adopted whenever our dataset is consisted of maxima of independent samples.

The *Block maxima/minima* and *Peaks Over Threshold* will be the main methods for sampling the data for analysis. The application of the method of block maxima goes through the following steps : divide the sample in n blocks of equal length, collect the maximum value in each block, fit the GEV distribution to the set of maxima and, finally, compute point and interval estimates for return level at a period.

The delicate point of this method is the appropriate choice of the periods defining the blocks. The calendar naturally suggests periods like months, quarters, etc. In order to avoid seasonal effects, we choose yearly periods which are likely to be sufficiently large for Fisher and Tippett (1928), Gnedenko (1943) to hold; to be discussed. The data sample can be divided into say, 30 non-overlapping sub-samples, each of them containing the daily returns of the successive calendar years. Therefore not all our blocks are of exactly the same length. The maximum return in each of the blocks constitutes the data points for the sample of maxima M which is used to estimate the generalized extreme value distribution (GEV). However, this approach may seem restrictive and not very realistic since the grouping of data into maxima is sometimes rather arbitrary, while by using only the block maxima, we may lose important information (some blocks may contain several among the largest observations, while other blocks may contain none). Moreover, in the case that we have few data, block maxima cannot be actually implemented.

The implementation of the peak over threshold method involves the following steps: select the threshold *u*, fit the GPD function to the exceedances over *u* and then compute point and interval estimates for Value-at-Risk and the expected shortfall. The selection of the threshold u by theory (Pickands (1975), Balkema and de Haan (1974)) tells us that *u* should be high, but the higher the threshold the less observations than are left for the estimation of the parameters of the tail distribution function. So far, no automatic algorithm with satisfactory performance for the selection of the threshold *u* is available. The issue of determining the fraction of data belonging to the tail is treated by Danielsson et al. (2001), Danielsson and de Vries (1997) and Dupuis (1998) among others. However these references do not provide a clear answer to the question of which method should be used. A graphical tool that is very helpful for preliminary data analysis and the selection of the threshold *u* is the use of R or matlab functions.

This thesis focuses on the univariate case; the approach is not easily extended to the multivariate case, because there is no concept of order in a multidimensional space and it is difficult to define the extremes in the multivariate case.

1.5 DEFINITION OF SOME TERMS

Risk: Uncertainty about a situation can often indicate risk, which is the possibility of loss, damage, or any other undesirable event. Most people desire low risk, which would translate to a high probability of success, profit, or some form of gain.

Value -at-Risk (VaR): A forecast of a given percentile, usually in the lower tail of the distribution of returns on a portfolio over some period: similar in principle to an estimate of the expected return on a portfolio which is a forecast of the 50th percentile. Value-at-risk is can also be defined as the loss level that will not be exceeded with a certain confidence level during a certain period of time.

Expected Shortfall(ES): A measure that produces better incentives for traders than VaR is expected shortfall(ES). This is also sometimes referred to as conditional VaR, or tail loss. Where VaR asks the question 'how bad can things get?' expected shortfall asks 'if things do get bad, what is our expected loss?'

Market Risk: The possibility of loss in an investment or speculation due to movements in market forces

Extreme Value Theory (EVT): EVT is the theory of modelling and measuring events which occur with very small probability.

Portfolio: The collection of financial and real assets-bank deposits, treasury bills, government bonds, ordinary shares of industrial companies, gold, work of art-which the financial investor's wealth is held.

Volatility: A trading conditions likely to change suddenly or sharply; unstable: volatile stock-markets, exchange rates, interest rates etc.

1.6 ORGANISATION OF THESIS

This thesis deals with the behaviour of the tails of brent crude oil pricing. More specifically, the focus is on the use of extreme value theory to assess tail related risk; it thus aims at providing a modelling tool for modern risk management.

Chapter 1 proves a general introduction, whiles in chapter 2, the review of related literature is presented for practical applications. Chapter 3 reviews the fundamental results of extreme value theory used to model the distributions underlying the risk measures. In chapter 4, some of the questions concerning the concepts of financial risk measures are answered. Chapter 5 looks at how practical application is presented, where the observations of twenty-two years of daily returns on an index representing the brent crude oil prices are analysed. Both tails are modelled and point and interval estimates computed. Finally, a brief analysis of the risks estimates with conclusions and recommendations are presented in chapter 6.

CHAPTER TWO

LITERATURE REVIEW

2.1 DEVELOPMENT OF EXTREME VALUE THEORY (EVT)

The EVT is a blend of a variety of applications concerning natural phenomena such as rainfall, floods, wind gusts, air pollution and corrosion and sophisticated mathematical results on point processes and regular varying functions. So, engineers and hydrologists on the one hand and theoretical probabilists on the other were the first to be interested in the development of EVT. It is only recently that this theory attracted mainstream statisticians. Indeed, the founders of probability and statistical theory Laplace, Pascal, Fermat, Gauss, and so on were too occupied with the general behaviour of statistical masses to be interested in rare extreme values.

Historically, work on extreme value problems may be dated back to as early as 1709 when N. Bernoulli discussed the mean largest distance from the origin when *n* points lie at random on a straight line of length *t* (Johnson et al., 1995). A century later Fourier stated that, in the Gaussian case, the probability of a deviation being more than three times the square root of two standard deviations from the mean is about 1 in 50,000, and consequently could be omitted (Kinnison, 1985). This seems to be the origin of the common, though erroneous, statistical rule that plus or minus three standard deviations from the mean can be regarded as the maximum range of valid sample values from a Gaussian distribution.

The first to investigate extreme value statistics were early astronomers who were faced with the problem of utilizing or rejecting suspected observations that appeared to differ greatly from the rest of a data-set. Still, systematic study and exploration of

EVT started in Germany in 1922. At that time a paper by Bortkiewicz (1922) appeared which dealt with the distribution of the range of random samples from the Gaussian distribution. The contribution of Bortkiewicz is that he was the one to introduce the concept of '*distribution of largest values'*. A year later another German, von Mises, introduced the concept of "expected value of the largest member of a sample of observations" from the Gaussian distribution (Mises, 1923). Essentially, he initiated the study of the asymptotic distribution of extreme values in samples from Gaussian distribution. At the same time, Dodd (1923) studied largest values from distributions other than the normal.

Indeed, a major first step occurred in 1925, when Tippet (1925) presented tables of the largest values and corresponding probabilities for various sample sizes from a Gaussian distribution, as well as the mean range of such samples. The first paper where asymptotic distributions of largest values (from a class of individual distributions) were considered appeared in 1927 by Frechet (1927). A year later, Fisher and Tippet (1928) published the paper that is now considered the foundation of the *asymptotic theory of extreme value distributions*. Independently, they found Frechet's asymptotic distribution and constructed two others. These three distributions have been found adequate to describe the extreme value distributions of all statistical distributions. We will explore further this result in subsequent chapter. Moreover, they showed the extremely slow convergence of the distribution of the largest value from Gaussian samples toward asymptote, which has been the main reason for the difficulties encountered by prior researchers. Indeed, the use of the Gaussian distribution as starting point has hampered the development of the theory, because none of the fundamental extreme value theorems is related in a simple way to the Gaussian distribution. Some simple and useful sufficient conditions for the weak convergence of the largest order statistic to each of the three types of limit distributions were given by Von Mises (1936). A few years later Gnedenko (1943) provided a rigorous foundation for the EVT and necessary and sufficient conditions for the weak convergence of the extreme order statistics. The theoretical developments at the 1920s and mid 1930s were followed in the late 1930s and 1940s by a number of publications concerning applications of extreme value statistics. Gumbel (1941) was the first to study the application of EVT. His first application was to old age, the consideration of the largest duration of life. In the sequel he showed that the statistical distribution of floods could be understood by the use of EVT. Extreme value procedures have also been applied extensively to other meteorological phenomena (such as rainfall analysis), to stress and breaking strength of structural materials and to the statistical problem of outlying observations.

The applications mentioned above all refer to the early development of statistical analysis of extremes from a theoretical as well as practical point of view. Gumbel's book of 1958 Gumbel, (1958) contains a very extensive bibliography of the developed literature up to that point of time. Of course, since then many more refinements of the original ideas and further theoretical developments and fields of applications have emerged. Some of these recent developments will be further discussed in the chapters to follow. Still, while this extensive literature serves as a testimony to the validity and applicability of the extreme value distributions and processes, it also reflects the lack of co-ordination between researchers and the inevitable duplication of results appearing in a wide range of publications.

2.2 FIELDS OF APPLICATIONS OF EVT

We have so far gained an idea of the diversity of fields where extreme-value analysis can be applied. Now in this sequel, we give only a short review of the most important areas where EVT has already been successfully implemented.

2.2.1 Hydrology – Environmental Data

As we have mentioned hydrologists were of the first to use EVT in practice. Here, the ultimate interest is the estimation of the T-year flood discharge, which is the level once exceeded on the average in a period of T years. Under standard conditions, the T-year level is a high-quantile of the distribution of discharges. Thus, one is primarily interested in a quantity determined by the upper tail of the distribution. Since, usually the time span T is larger than the observation period, some additional assumptions on the underlying distribution of data have to be made. If the statistical inference is based on annual maxima of discharges, then hydrologists favoured model is the Generalised extreme-value (GEV) model. Alternatively, if the inference is based on a partial duration series, which is the series of exceedances over a certain high threshold, the standard model for the flood magnitudes is the generalized Pareto model.

There is a large literature of extreme-value analyses applied to hydrological data. Hosking et al. (1985) and Hosking and Wallis (1987) apply their proposed estimation method(GEV) to river Nidd data (to 35 annual maxima floods of the river Nidd, at Hunsingore, Yorkshire, England). Davison and Smith (1990) apply the generalized Pareto distribution to more detailed data of the same river, taking into account both seasonality and serial dependence of data. Dekkers and de Haan (1989) were concerned with the high tide water levels in one of the islands at the Dutch coast. The increasing need to exploit coast areas combined with the concern about the greenhouse effect has resulted in a demand for the height of sea defences to be estimated accurately, and to be such that the risk of the sea-dyke being exceeded is small and pre-specified.

So, more elaborate techniques have recently been developed. Tawn (1992) performed extreme-value analysis to hourly sea levels by taking into account the fact that the series of observations is not a stationary sequence (due to astronomical tidal component). Barão and Tawn (1999) utilized bivariate extreme value distributions to model data of sea-levels at two UK east coast sites, while de Haan and de Ronder (1998) modelled wind and sea data of Netherlands using bivariate extreme value distribution function. Extreme low sea levels are of independent interest in applications to shipping and harbour developments and for the design of nuclear power station cooling water intakes. With simple adaptations most methods can be applied to sea-level minima to solve such problems.

Another related issue is that of rainfalls. The design of large-scale hydrological structures requires estimates to be made of the extremal behaviour of the rainfall process within a designated catchment region. It is common to simulate extreme events (rainfalls) and then to access the consequent effect on hydrological models of reservoirs, river flood networks and drainage systems. Coles and Tawn (1996) exploited extreme value characterizations to develop an explicit model for extremes of spatially aggregated rainfall over fixed durations for a heterogeneous spatial rainfall process. Furthermore, in ecology, higher concentration of certain ecological quantities, like concentration of ozone, acid rain or SO_2 in the air are of great interest due to their negative response on humans and generally, on the biological system. For example, Smith (1989) performs extreme value analysis in ground-level ozone data, taking into account phenomena common in environmental time series, such as seasonality and clustering of extremes. Similar is the subject dealt with in Küchenhoff and Thamerus (1996).

2.2.2 Insurance

Estimating loss severity distributions (i.e. distributions of individual claim sizes) from historical data is an important actuarial activity in insurance. In the context of reinsurance, where we are required to choose or price a high-excess layer, we are specifically interested in estimating the tails of loss severity distributions. In this situation it is essential to find a good statistical model for the largest observed historical losses; it is less important that the model explains smaller losses. In fact, a model chosen for its overall fit to all historical losses may not provide a particularly good fit to the large losses. Such a model may not be suitable for pricing a high-excess layer. It is obvious that EVT is the most appropriate tool for this job, either by using GEV to model large claims or generalized Pareto distribution to model exceedances over a high threshold.

The applicability of EVT to insurance is discussed by Beirlant et al.(1994), Mikosch (1997), McNeil (1997), McNeil and Saladin (1997) with application to Danish data on large fire insurance losses, Rootzen and Tajvidi (1997) with application to Swedish windstorm insurance claims.

2.2.3 Finance – Risk Management

Finance and, even more general, risk management are areas where only recently EVT has gained ground. Insurance and financial data can both be investigated from the viewpoint of risk analysis. Therefore, the insight gained from insurance data can also be helpful for the understanding of financial risks. Mainly due to the increase in volume and complexity of financial instruments traded, risk management has become a key issue in any financial institution or corporation of some importance. Globally accepted rules are put into place aimed at monitoring and managing the full diversity of risk. Extreme event risk is present in all areas of risk management. Whether we are concerned with credit, market or insurance risk, one of the greatest challenges to the risk manager is to implement risk management models which allow for rare but damaging events, and permit the measurement of their consequences. In market risk, we might be concerned with the day-to-day determination of the VaR for the losses we incur on a trading book due to adverse market movements.

In credit or operational risk management our goal might be the determination of the risk capital we require as a cushion against irregular losses from credit downgrading and defaults or unforeseen operational problems. No discussion has perhaps been more heated than the one on VaR. The biggest problem with VaR is the main assumption in the conventional models, i.e. that portfolio returns are normally distributed.

In summary the main points of risk management are the followings:

- Risk management is interested in estimating tail probabilities and quantiles of profit/loss distributions, and indeed of general financial data
- Extremes do matter.
- We want to have methods for estimating conditional probabilities concerning tail events: ‘Given that we incur a loss beyond VaR, how far do we expect the excess to go?’
- Financial data show fat tails.

EVT is a subject whose motivations match the four points highlighted above. It has a very important role to play in some of the more technical discussions related to risk management issues. The usefulness of EVT to risk management is stressed by Danielsson and de Vries (1997), McNeil (1998 and 1999), Embrechts et al. (1998, 1999), Embrechts (1999).

2.2.4 Teletraffic Engineering

Classical queuing and network stochastic models contain simplifying assumptions guaranteeing the Markov property and insuring analytical tractability. Frequently, inter-arrival and service times are assumed to be i.i.d. and typically underlying distributions are derived from operations on exponential distributions. At a minimum, underlying distributions are usually assumed nice enough that moments are finite. Increasing instrumentation of teletraffic networks has made possible the acquisition of large amounts of data. Analysis of this data is disturbing since there is strong evidence that the classical queuing assumption of thin tails and independence are inappropriate for these data. Such phenomena as file lengths, Central Processing Unit (CPU) time to complete a job, call holding times; inter-arrival times between packets in a network and length of on/off cycles appear to be generated by distributions which have heavy tails. Resnick and Stărică (1995), Kratz and Resnick (1996), and Resnick (1997a) here dealt with such kind of data.

Other areas where extreme-value analysis has found application are engineering strength of materials (Harter, 1978 provides a detailed literature on this subject); earthquake size distribution (Kagan, 1997); athletic records (Strand and Boes, 1998, Barão and Tawn, 1999); city-sizes, corrosion analysis, exploitation of diamond deposits, demography, geology and meteorology among others.

CHAPTER THREE

EXTREME VALUE THEORY (EVT)

3.1 INTRODUCTION

The aforementioned schemes work on the quantiles of the random variables using information from the whole data set. However, due to the fact that quantiles at 1% or 5% are ultimate values of the distribution, it is natural to emphasize the representation of the tails directly as an alternative to attempt the whole structure of the distribution. EVT, then, provides a theoretical justification to such procedures, as it plays the same essential role as the Central Limit Theorem performs when modelling sums of random variables. This section will supply some basic notions indispensable for the rest of the thesis. When modelling the maxima of a random variable, EVT plays the same fundamental role as the central limit theorem plays when modelling sums of random variables. In both cases, the theory tells us what the limiting distributions are. Generally, there are two related ways of identifying extremes in real data. Let us consider a random variable representing daily losses or returns. The first approach considers the maximum the variable takes in successive periods, for example months or years. These selected observations constitute the extreme events, also called block (or per period) maxima.

In Figure 3.1 below, the observations X_2, X_5, X_7 and X_{11} represent the block maxima for four periods of three observations each.

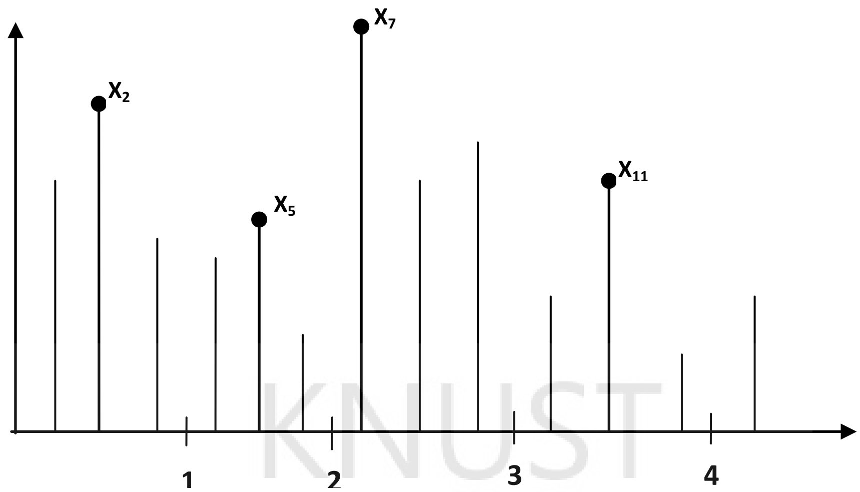

Figure 3.1: Block-maxima

The second approach focuses on the realizations exceeding a given (high) threshold. The observations X_1, X_2, X_7, X_8, X_9 and X_{11} in Figure 3.2, all exceed the threshold u and constitute extreme events.

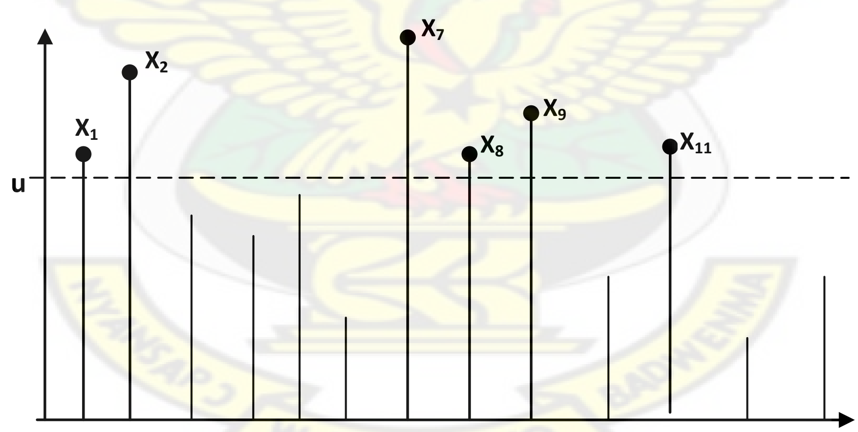

Figure 3.2: Excesses over Threshold u

The block maxima method is the traditional method used to analyze data with seasonality as for instance hydrological data. However, the threshold method uses data more efficiently and, for that reason, seems to become the method of choice in recent applications.

In the following subsections, the fundamental theoretical results underlying the block maxima and the threshold method are presented.

3.2 DISTRIBUTION OF MAXIMA

The limit law for the block maxima, which we denote by M_n, with *n* the size of the sub-sample (block), is given by the following theorem:

Theorem 3.1: *(Fisher and Tippett (1928), Gnedenko (1943)).*

Let (X_n) be a sequence of i.i.d (independent and identical distribution) random variables. If there exist constants $c_n > 0$, $d_n \in R$ and some non-degenerate distribution function H such that $\frac{M_n - d_n}{c_n} \xrightarrow{d} H$ *then H belongs to one of the three standard extreme value distributions:*

$$\left.\begin{array}{lll} Fre'chet: & \Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} & \alpha > 0 \\ Weibull: & \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \le 0 \\ 1, & x > 0 \end{cases} & \alpha > 0 \\ Gumbel: & \Lambda(x) = e^{-e^{-x}}, \quad x \in R & \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 3.1$$

The shapes of the probability density functions for the standard Frechet, Weibull and Gumbel distributions are given in Figure 3.3, 3.4 and 3.5.

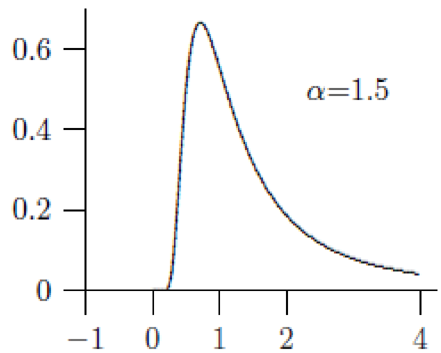

Figure 3.3. Density function for Frechet

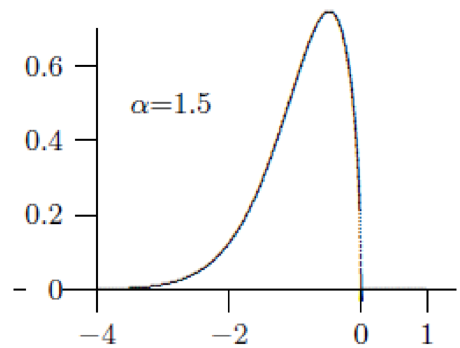

Figure 3.4. Density function for Weibull

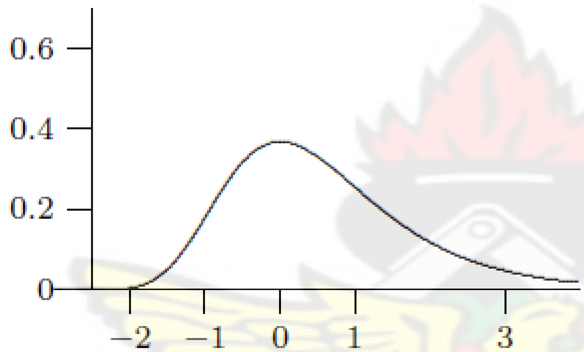

Figure 3.5. Density function for Gumbel

We observe that the Frechet distribution has a polynomially decaying tail and therefore suits well heavy tailed distributions. The exponentially decaying tails of the Gumbel distribution characterize thin tailed distributions. Finally, the Weibull distribution is the asymptotic distribution of finite endpoint distributions. Jenkinson (1955) and von Mises (1954) suggested the following one-parameter representation:

$$H_{\xi}(x) = \begin{cases} {}^{-(1+\gamma x)^{-1/\gamma}} & if \quad \gamma \neq 0 \\ e^{-e^{-x}} & if \quad \gamma = 0 \end{cases} \qquad \text{..............................} \ 3.2$$

As representation of these three standard distributions, with x such that $1+\gamma x>0$. This generalization, known as the ***generalized extreme value*** (GEV) distribution, is obtained by setting $\gamma=\alpha^{-1}$ for the Frechet distribution, $\gamma=-\alpha^{-1}$ for the Weibull distribution, and by interpreting the Gumbel distribution as the limit case for $\gamma=0$.

As in general we do not know in advance the type of limiting distribution of the sample maxima, the generalized representation is particularly useful when maximum likelihood estimates have to be computed. Moreover the standard GEV defined in equation (2.2) is the limiting distribution of normalized extrema. Given that in practice we do not know the true distribution of the returns and, as a result, we do not have any idea about the norming constants c_n and d_n, we use the three parameter specification

$$H_{\gamma,\sigma,\mu(x)}=H_{\gamma}\left(\frac{x-\mu}{\sigma}\right),\ x\in D\ ,\quad D=\begin{cases}\left]-\infty,\mu-\frac{\sigma}{\gamma}\right[& \gamma<0\\ \left]-\infty,\infty\right[& \gamma<0\\ \left]\mu-\frac{\sigma}{\gamma},\infty\right[& \gamma>0\end{cases} \qquad \text{..............................3.3}$$

of the GEV, which is the limiting distribution of the un-normalized maxima. The two additional parameters μ and σ are the location and the scale parameters representing the unknown norming constants. The quantities of interest are not the parameters themselves, but the quantiles, also called return levels, of the estimated GEV:

$$R^k=H^{-1}_{\gamma,\sigma,\mu}\left(1-\frac{1}{k}\right) \qquad \text{..3.4}$$

Substituting the parameters γ, σ and μ by their estimates $\hat{\gamma}$, $\hat{\sigma}$ and $\hat{\mu}$, we get

$$\hat{R}^k = \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left(1 - \left(-\log\left(1 - \dfrac{1}{k}\right)\right)^{-\hat{\gamma}}\right) & \hat{\gamma} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left(-\log\left(1 - \dfrac{1}{k}\right)\right) & \hat{\gamma} = 0 \end{cases} \quad \text{...................................3.5}$$

A value of $\hat{R}^{10}$ of 8 means that the maximum loss/gain observed during a period of one year will exceed 8% once in ten years on average.

3.3 DISTRIBUTION OF EXCEEDANCES

An alternative approach, called the *peak over threshold* (POT) method, is to consider the distribution of exceedances over a certain threshold. Our problem is illustrated in Figure 3.6 and 3.7 where we consider an (unknown) distribution function F of a random variable X. We are interested in estimating the distribution function F_u of values of x above a certain threshold u.

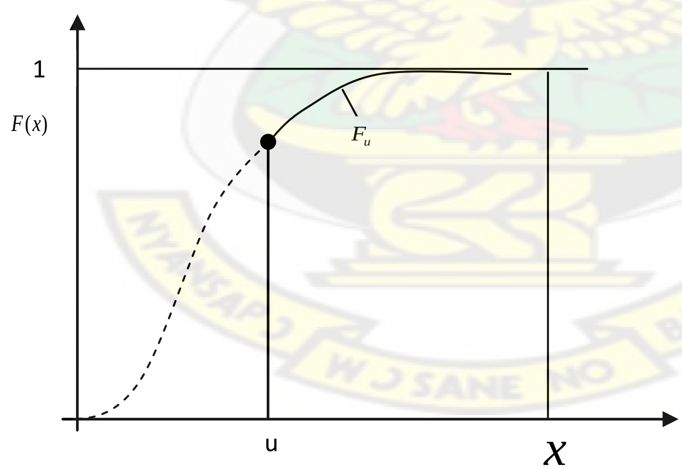

Figure 3.6 : Distribution function F(x)

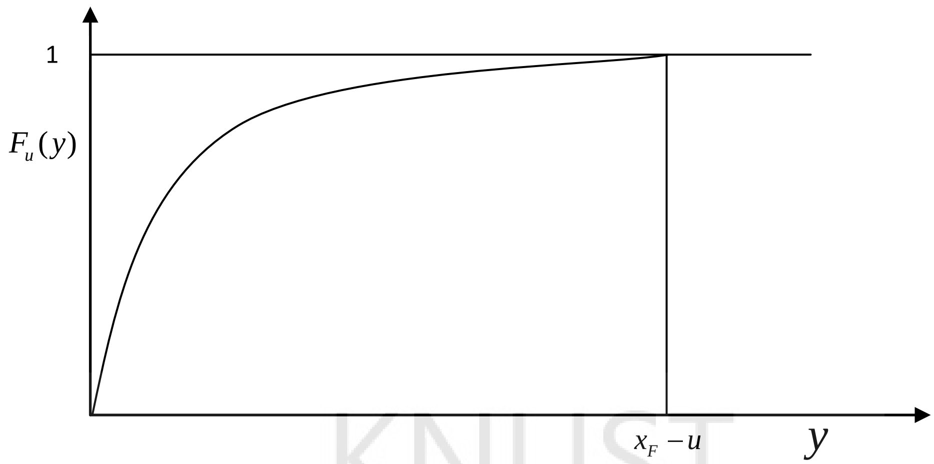

Figure 3.7: Conditional Distribution function F_u

The distribution function *Fu* is called the *conditional excess distribution function* and is defined as $F_u\left(y\right) = P\left(X - u \le y \middle| X > u\right), \qquad 0 \le y \le x_F - u$ 3.6

where *X* is a random variable, *u* is a given threshold, $y = x - u$ are the excesses and $x_F \le \infty$ is the right endpoint of *F*. We verify that *Fu* can be written in terms of *F*, i.e.

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)} \quad ....3.7$$

The realizations of the random variable *X* lie mainly between 0 and *u* and therefore the estimation of *F* in this interval generally pose no problem. The estimation of the portion *Fu* however might be difficult as we have in general very little observations in this area. At this point EVT can prove very helpful as it provides us with a powerful result about the conditional excess distribution function which is stated in the following theorem:

Theorem 3.2 *(Pickands (1975), Balkema and de Haan (1974))*

For a large class of underlying distribution functions F the conditional excess distribution function Fu(y), for u large, is well approximated by $F_u(y) \approx G_{\gamma,\sigma}(y)$, $u \to \infty$, where

$$G_{\gamma,\sigma}(y) = \begin{cases} 1-\left(1+\frac{\gamma}{\sigma}y\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1-e^{-y/\sigma} & \text{if } \gamma = 0 \end{cases} \quad \text{.................................3.8}$$

for $y \in \left[0, \left(x_F - u\right)\right]$ if $\gamma \geq 0$ and $y \in \left[0, -\frac{\sigma}{\gamma}\right]$ if $\gamma < 0$.

$G_{\gamma,\sigma}$ is *the so called generalized Pareto distribution (GPD).*

If *x* is defined as *x* = *u* + *y*, the GPD can also be expressed as a function of *x*, i.e.

$$G_{\gamma,\sigma}(x) = 1-(1+\gamma(x-u)/\sigma)^{-1/\gamma} \quad \text{...3.9}$$

Figure 3.8 illustrates the shape of the generalized Pareto distribution $G_{\gamma,\sigma}$ when γ called the *shape parameter* or *tail index*, takes a negative, figure 3.9 a positive and figure 4.0, a zero value. The scaling parameter σ is kept equal to one.

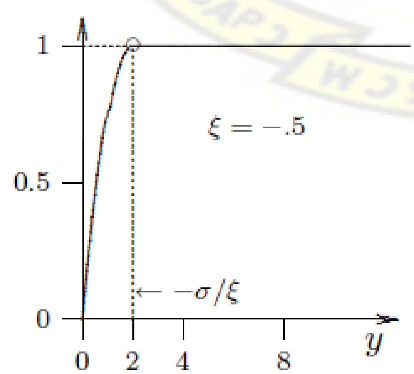

Figure 3.8 Shape of pareto distribution
$\gamma = -0.5$

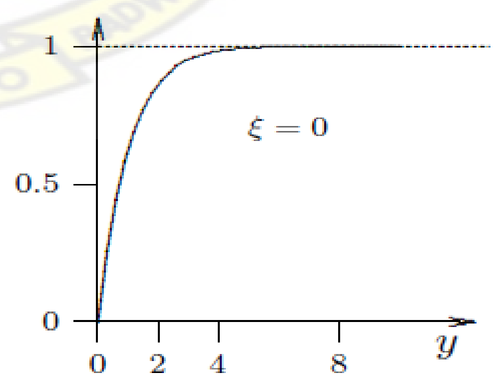

Figure 3.9 Shape of pareto distribution
$\gamma = 0$

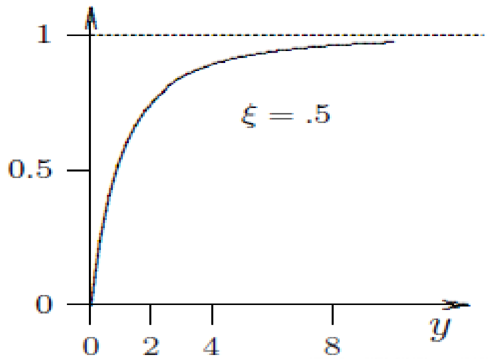

Figure 3 .10 : shape Pareto distribution $\gamma = 0.5$

The tail index γ gives an indication of the heaviness of the tail, the larger γ, the heavier the tail. As, in general, one cannot fix an upper bound for financial losses, only distributions with shape parameter $\gamma \geq 0$ are suited to model financial return distributions.

Assuming a GPD function for the tail distribution, analytical expressions for VaR_p and ES_p can be defined as a function of GPD parameters. Isolating F(x) from equation (3.7), $F(x) = (1 - F(u))F_u(y) + F(u)$ and replacing F_u by the GPD and $F(u)$ by the estimate $(n - N_u)/n$, where n is the total number of observations and N_u the number of observations above the threshold u, we obtain

$$\hat{F}(x) = \frac{N_u}{n}\left(1 - (1 + \frac{\gamma}{\hat{\sigma}}(x - \mu))^{-\frac{1}{\hat{\gamma}}}\right) + \left(1 - \frac{N_u}{n}\right) \qquad \text{...3.9}$$

which simplifies to

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 - \left(1 + \frac{\hat{\gamma}}{\hat{\sigma}}\left(x - \mu\right)\right)\right)^{-\frac{1}{\hat{\gamma}}} \qquad \text{..3.10}$$

inverting (2.10) for a given probability p gives

$$\hat{VaR}_p = u + \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}p\right)^{-\hat{\gamma}} - 1\right) \quad \ldots\ldots 3.11$$

Let us rewrite the expected shortfall as

$$\hat{ES}_p = \hat{VaR}_p + E(X - \hat{VaR}_p \mid X > \hat{VaR}_p) \quad \ldots\ldots 3.12$$

where the second term on the right is the expected value of the exceedances over the threshold VaRp. It is known that the *mean excess function* for the GPD with parameter $\gamma < 1$ is

$$e(z) = E(X - z \mid X > z) = \frac{\sigma + \gamma z}{1-\gamma}$$

$$\sigma + \gamma z > 0 \quad \ldots\ldots 3.13$$

This function gives the average of the excesses of *X* over varying values of a threshold *z*. Another important result concerning the existence of moments is that if *X* follows a GPD then, for all integers $r < 1/\gamma$ such that, the *r* first moments exist. Similarly, given the definition for the expected shortfall and using expression (3.13), for $z = VaR_p - u$ and *X* representing the excesses *y* over *u* we obtain

$$\hat{ES}_p = \hat{VaR}_p + \frac{\hat{\sigma} + \hat{\gamma}\,(\hat{VaR}_p - u)}{1-\hat{\gamma}} = \frac{\hat{VaR}_p}{1-\hat{\gamma}} + \frac{\hat{\sigma} - \hat{\gamma}u}{1-\hat{\gamma}} \quad \ldots\ldots 3.14$$

3.4 EXTREME-VALUE INDEX ESTIMATORS

In this section, we give the most prominent answers to the issue of parameter estimation. We mainly concentrate on the estimation of the shape parameter γ due to its (already stressed) importance. The setting on which we are working is:

Suppose that we have a sample of i.i.d random variable (r.v's) $X_1, X_2, ..., X_n$ (where $X_{1:n} \geq X_{2:n} \geq ... \geq X_{n:n}$ are the corresponding descending order statistics) from an unknown *continuous* distribution function. F. According to EVT, the normalized maximum of such a sample follows asymptotically a GEV distribution function. $H_{\gamma:\mu,\sigma}$ i.e. $F \in MDA(H)\gamma; \mu, \sigma$. In the remaining of this section, we give the most prominent answers to the above question of estimation of extreme-value index *γ*. We describe the most known suggestions, ranging from the first contributions, of 1975, in the area to very recent modifications and new developments.

3.4.1 Pickands Estimator

The Pickands estimator (Pickands, 1975), is the first suggested estimator for the parameter $\gamma \in \Re$ of GEV distribution function and is given by the formula

$$\hat{\gamma}_p = \frac{1}{\ln 2} \ln \left(\frac{X_{(k/4):n} - X_{(k/2):n}}{X_{(k/2):n} - X_{k:n}} \right) \qquad 3.15$$

The original justification of Pickands estimator was based on adopting a percentile estimation method for the differences among the upper-order statistics. A more formal justification is provided by Embrechts et al. (1997). A particular characteristic of Pickands estimator is the fact that the largest observation is not explicitly used in the estimation. One can argue that this makes sense since the largest observation may add too much uncertainty.

The properties of Pickands estimator were mainly explored by Deckers and de Haan (1989). They proved, under certain conditions, weak and strong consistency, as well as asymptotic normality. Consistency depends only on the behaviour of k, while asymptotic normality requires more delicate conditions (2nd order conditions) on the underlying distribution function. F, which are difficult to verify in practice. Still, Deckers and de Haan (1989) have shown that these conditions hold for various known and widely-used distribution function's (normal, gamma, GEV, exponential, uniform, Cauchy).

3.4.2 Hill Estimator

The most popular tail index estimator is the Hill estimator, (Hill, 1975), which, though, is

restricted to the Fréchet case $\gamma > 0$. The Hill estimator is provided by the formula

$$\hat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k} \ln X_{i:n} - \ln X_{k+1:n} \qquad \text{.................................} \; 3.16$$

The *original derivation* of the Hill estimator relied on the notion of conditional maximum likelihood estimation method. The statistical behaviour and properties of the Hill estimator have been studied by many authors separately, and under diverse conditions. Weak and strong consistency as well as asymptotic normality of the Hill estimator hold under the assumption of i.i.d. data (Embrechts et al., 1997). Similar (or slightly modified) results have been derived for data with several types of dependence or some other specific structures (see for example Hsing, 1991 as well as Resnick and Stărică, 1995, 1996 and 1998).

Note that, the conditions on k and distribution function, F, that ensure the consistency and asymptotic normality of the Hill estimator are the same as those imposed for Pickands estimator. Such conditions have been discussed by many authors, such as Davis and Resnick (1984), Haeusler and Teugels (1985), de Haan and Resnick (1998).

Though the Hill estimator has the apparent disadvantage that is restricted to the case γ>0, it has been widely used in practice and extensively studied by statisticians. Its popularity is partly due to its simplicity and partly due to the fact that in most of the cases where extreme-value analysis is called for, we have long-tailed distribution function.'s (i.e. γ>0).

3.4.3 Adapted Hill Estimator

The popularity of the Hill estimator generated a tempting problem to try to extend the Hill estimator (with its simplicity and good properties) to the general case $\gamma \in \Re$. Such an attempt, led Beirlant et al. (1996) to the so-called adapted the Hill estimator, which is applicable for any γ in the range of real numbers :

$$\left.\begin{aligned} \hat{\gamma}_{adH} &= \frac{1}{k}\sum_{i=1}^{k} \ln(U_i) - \ln(U_{k+1}) \\ where \quad U_i &= X_{(i+1):n}\left(\frac{1}{i}\sum_{j=1}^{i} \ln X_{j:n} - \ln X_{(i+1):n}\right) \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.17$$

3.4.4 Moment Estimator

Another estimator that can be considered as an adaptation of the Hill estimator, in order to obtain consistency for all $\gamma \in \Re$, has been proposed by Deckers et al. (1989). This is the moment estimator, given by

$$\left.\begin{aligned} \hat{\gamma}_M &= M_1 + 1 - \frac{1}{2}\left(1 - \frac{(M_1)^2}{M_2}\right)^{-1} \\ where \quad M_j &= \frac{1}{k}\sum_{i=1}^{k}(\ln X_{i:n} - \ln X_{(k+1):n})^j \quad j = 1, 2. \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 3.18$$

Weak and strong consistency, as well as asymptotic normality of the moment estimator have been proven by Dekkers et al. (1989).

3.4.5 QQ – Estimator

One of the approaches concerning Hill's derivation is the 'QQ-plot' approach (Beirlant et al., 1996). According to this, the Hill estimator is approximately the slope of the line fitted to the upper tail of Pareto QQ plot. A more precise estimator, under this approach, has been suggested by Kratz and Resnick (1996), who derived the following estimator of γ:

$$\hat{\gamma}_{pq} = \frac{\sum_{i=1}^{k}\ln\frac{i}{k+1}\left\{\sum_{j=1}^{k}\ln X_{j:n} - k\ln X_{i:n}\right\}}{k\sum_{i=1}^{k}\left(\ln\frac{i}{k+1}\right)^2 - \left(\sum_{i=1}^{k}\ln\frac{i}{k+1}\right)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 3.19$$

They proved weak consistency and asymptotic normality of QQ-estimator (under conditions similar to the ones imposed for the Hill estimator). However, the asymptotic variance of QQ-estimator is twice the asymptotic variance of the Hill estimator, while similar conclusions are drawn from simulations of small samples. On the other hand, one of the advantages of QQ-estimator over the Hill estimator is that the residuals (of the Pareto plot) contain information which potentially can be utilised to confront the bias in the estimates when the approximation is not exactly valid.

3.4.6 Moments Ratio Estimator

Concentrating on cases where $\gamma > 0$, the main disadvantage of the Hill estimator is that it can be severely biased, depending on the second order behaviour of the underlying distribution function, F. Based on an asymptotic second order expansion of the distribution function. F, from which one gets the bias of the Hill estimator, Danielsson et al. (1996) proposed the moments ratio estimator:

$$\hat{\gamma}_{MR} = \frac{1}{2}\frac{M_2}{M_1} \quad \text{..3.20}$$

They proved that $\hat{\gamma}_{MR}$ has lower asymptotic square bias than the Hill estimator (when evaluated at the same threshold, i.e. for the same k), though the convergence rates are the same.

3.4.7 Peng's and W estimators

An estimator related to the moment estimator $\hat{\gamma}_M$ is Peng's estimator, suggested by Deheuvels et al. (1997):

$$\hat{\gamma}_W = 1 - \frac{1}{2}\left(1 - \frac{(L_1)^2}{L_2}\right)^{-1} \quad \textit{and} \quad \hat{\gamma}_L = \frac{M_2}{2M_1} + 1 - \frac{1}{2}\left(1 - \frac{(M_1)^2}{M_2}\right)^{-1} \text{..................3.21}$$

These estimator has been designed to somewhat reduce the bias of the moment estimator.

Another related estimator suggested by the same authors is the W estimator, where

$$L_j = \frac{1}{k}\sum_{i=1}^{k}\left(X_{i:n} - X_{(k+1):n}\right)^j \quad j=1,2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.22$$

As Deheuvels et al. (1997) mentioned, $\hat{\gamma}_L$ is consistent for any $\gamma \in \Re$ (under the usual conditions), while $\hat{\gamma}_W$ is consistent only for $\gamma < 1/2$. Moreover, under appropriate conditions on F and $k(n)$, $\hat{\gamma}_L$ is asymptotically normal. Normality holds for $\hat{\gamma}_W$ only for $\gamma < 1/4$.

3.5 USE OF MEAN EXCESS PLOT

A graphical tool for assessing the behaviour of a distribution function, F is the mean excess function (MEF). The limit behaviour of MEF of a distribution gives important information on the tail of that distribution function (Beirlant et al., 1995). MEF's and the corresponding mean excess plots (MEP's) are widely used in the first exploratory step of extreme-value analysis, while they also play an important role in the more systematic steps of tail index and large quantiles estimation. MEF is essentially the expected value of excesses over a threshold value u. The formal definition of MEF (Beirlant et al., 1996) is as follows:

Let X be a positive random variable X with distribution function F and with finite first moment. Then MEF of X is

$$e(u) = E(X - u | X > u) = \frac{1}{\overline{F}(u)} \int_u^{x_F} \overline{F}(y) dy \quad \text{..................3.23}$$

for all u > 0.

The corresponding MEP is the plot of points {u, $e(u)$, for all u > 0}. The empirical counterpart of MEF based on sample $(X_1, X_2, ..., X_n)$ is

$$\hat{e}(u) = \frac{\sum_{i=1}^{n} (X_i - u) 1_{(u,\infty)}(X_i)}{\sum_{i=1}^{n} 1_{(u,\infty)}(X_i)} \quad \text{..................................3.24}$$

where $1_{(u,\infty)}(x) = 1$ if x > u, 0 otherwise.

Usually, the MEP is evaluated at the points. In that case, MEF takes the form

$$E_k = \hat{e}(X_{(k+1):n}) = \frac{1}{k} \sum_{i=1}^{k} X_{i:n} - X_{(k+1):n} \quad k = 1,2,...,n \quad \text{........................3.25}$$

If $X \in MDA(H)\ \gamma$, γ > 0, then it s easy to show that

$$e_{\ln X}(\ln u) = E(\ln X - \ln u | X > u) \rightarrow \gamma \quad \text{as } u \rightarrow \infty \quad \text{.............................. 3.26}$$

Intuitively, this suggests that if the MEF of the logarithmic-transformed data is ultimately constant, then $X \in MDA(H)\ \gamma$, γ > 0 and the values of MEF converge to the true value γ. Replacing u, in the above relation, by a high quantile $Q\left(1 - \frac{k}{n}\right)$, or empirically by $X_{(K+1):n}$, we find that the estimator $e_{\ln X}(X_{(K+1):n})$ will be a consistent estimator of γ in case $X \in MDA(H)\ \gamma$, γ > 0. This holds when $k/n \rightarrow 0$ as

$n \to \infty$. Notice that the empirical counterpart $e_{\ln X}(X_{(K+1):n})$ of is the well-known the Hill estimator.

3.6 ESTIMATION OF PARAMETERS

3.6.1 Introduction

There are different methods available for performing parameter estimation these include: Method of Moments Estimation (MME), Probability Weighted Moments (PWM) or equivalently L-Moments (LM), Maximum Likelihood Estimation (MLE), and Bayesian methods. For smaller sample sizes (n < 50), the MLE is unstable and can give unrealistic estimates for the shape parameter (e.g., Hosking and Wallis (1997), Coles and Dixon (1999), and Martins and Stedinger (2000,2001)). Madsen et al (1997) argue that the MME quantile estimators have smaller root mean square error when the true value of the shape parameter is within a narrow range around zero. For weather and climate applications, enough data are typically available to expect that MLE would be comparable in performance, especially when blocks smaller than years are used. Additionally, MLE allows one to easily incorporate covariate information into parameter estimates.

Furthermore, it is more straightforward to obtain error bounds for parameter estimates with MLE compared with most alternative methods. Although work on Bayesian estimation with respect to extreme-value analysis has been sparse in the literature, good examples are available (Stephenson and Tawn (2004) and the references therein, Coles (2001, Section 9.1), and Cooley et al. (2005a, 2005b)). Obviously, one will never select the Gumbel when fitting data to a GEV because the Gumbel is reduced to a single point in a continuous parameter space. A common approach is to perform an initial hypothesis test to determine which of the three extremal types is appropriate, and then fit data only to that type. However, this approach does not account for the uncertainty of the choice of extremal type on the subsequent inference, which can be quite large. Stephenson and Tawn (2004) suggested a Bayesian approach to estimate these parameters that allows for the Gumbel to be achieved with positive probability; though results can be highly sensitive to choice of prior distributions.

Suppose we have a random variable X which has probability density function, $f(x,\theta)$ where θ is either a real number or a vector of real numbers. Assume that $\theta \in \Omega$ which is a subset of R^p, for $p \geq 1$. For example, θ could be the vector (μ, σ^2) when X has a $N(\mu, \sigma^2)$ distribution or θ could be the probability of success p when X has a binomial distribution. Our information about θ comes from the sample $X_1, X_2, \ldots, X_n$. We often assume that this is a random sample which means that the random variables $X_1, X_2, \ldots, X_n$ are independent and have the same distribution as *X*; that is, $X_1, X_2, \ldots, X_n$ are independently and identically distributed (*iid*). We may use the statistic $T = T(X_1, X_2, \ldots, X_n)$, a function of the sample to estimate θ and say that *T* is a point estimator for θ. For example, suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and variance, σ^2. Then the statistics $\overline{X}$ and S^2 (the sample mean and variance of this random sample) are point estimators of μ and σ^2 respectively.

In order to decide which point estimator of a parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators and also gives the most economical information. Ideally, we want an estimator which generates estimates that can be expected to be close in value to the parameter. These properties are presented in the following Section:

3.6.2 PROPERTIES OF POINT ESTIMATORS

The desirable properties: unbiasedness, efficiency, consistency, and sufficiency, which need be satisfied by point estimator, are presented as follows:

3.6.2.1 Unbiasedness and efficiency

Let $X_1, X_2, \ldots, X_n$ be a random sample from the random variable X with the probability distribution function, $f(x, \theta)$, where $\theta \in \Omega$. Let $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ be a statistic. We say T is an unbiased estimator for θ if $E(\hat{\theta}) = \theta$, for all $\theta \in \Omega$. If $\hat{\theta}$ is not unbiased (that is, $E(\hat{\theta}) \neq \theta$), we say that T is a biased estimator for θ.

Selecting among several unbiased estimators is to choose the one with minimum variance, minimum variance unbiased estimator (*MVUE*). The lower bound for the variance of an unbiased estimator can be established if the appropriate derivatives exist and can pass under the integration sign. A remarkable inequality, called the *Rao-Cramer Lower Bound* (*CRLB*), gives a lower bound on the variance of any unbiased estimator. We then show that under the following regularity conditions, the variances of the maximum likelihood estimates achieve this lower bound asymptotically.

4 *Assumptions* (*Regularity Conditions*):

(i) The probability density functions ($pdf's$) are distinct.

(ii) The $pdf's$ have common support for all θ.

(iii) The point θ_0 (the true value of) is an interior point in, the parameter space.

(iv) The pdf, $f(x,\theta)$ is twice differentiable as a function of θ.

(v) The integral $\int_{Rx} f(x,\theta)dx$ can be differentiated twice under the integral sign as a function of θ.

Theorem 2.3 (*Rao-Cramer Lower Bound*):

Let $x_1, x_2, \ldots, x_n$ be independent and identically distribution (iid) with common pdf $f(x,\theta)$, for $\theta \in \Omega$ and T be an unbiased estimator for θ. then under the regularity conditions, $Var(T) \geq [nI(\theta)]^{-1}$,

where $I(\theta) = E\left[\left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}\right]$ 3.27

Proof: The proof for the continuous pdf $f(x,\theta)$ is given as follows: $1 = \int_{-\infty}^{\infty} f(x;\theta)dx$, since $f(x,\theta)$ is probability density function. Taking the derivative with respect to θ results in, $0=\int_{-\infty}^{\infty}\frac{\partial f(x;\theta)}{\partial\theta}dx$. The latter expression can be rewritten as

$$0=\int_{-\infty}^{\infty}\frac{\partial f(x;\theta)/\partial\theta}{f(x;\theta)}f(x;\theta)dx$$

Or, equivalently,

$$0=\int_{-\infty}^{\infty}\frac{\partial\log f(x;\theta)}{\partial\theta}f(x;\theta)dx \quad ....3.28$$

Writing this last equation as expectation, we have established,

$$E\left[\frac{\partial\log f(X;\theta)}{\partial\theta}\right]=0, \quad ....3.29$$

that is, the mean of the random variable $\frac{\partial\log f(X;\theta)}{\partial\theta}$ is zero(0). If we differentiate equation (2.28) again, it follows that

$$0=\int_{-\infty}^{\infty}\frac{\partial^2\log f(x;\theta)}{\partial\theta^2}f(x;\theta)dx+\int_{-\infty}^{\infty}\frac{\partial\log f(x;\theta)}{\partial\theta}.\frac{\partial\log f(x;\theta)}{\partial\theta}f(x;\theta)dx \quad ....3.30$$

(3.18)

The second term of the right hand side of this equation (2.30) can be written as an expectation. This is *Fisher information*, denoted it by

$$I(\theta)=\int_{-\infty}^{\infty}\left(\frac{\partial\log f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)dx=E\left[\left(\frac{\partial\log f(x;\theta)}{\partial\theta^2}\right)^2\right] \quad ....3.31$$

Clearly, from equations (2.32) and (2.33), we have

$$I(\theta) = -\int \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} f(x;\theta)dx = Var\left(\frac{\partial \log f(X,\theta)}{\partial \theta}\right) \quad \text{..3.32}$$

which means that the Fisher information is the variance of $\frac{\partial \log f(X;\theta)}{\partial \theta}$.

Note that the important function, $\frac{\partial \log f(x;\theta)}{\partial \theta}$ is called the *score function*. Recall that it determines the estimating equations for the maximum likelihood estimator, that is, the maximum likelihood estimator, $\hat{\theta}$ solves equation:

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta} = 0 \text{ for } \theta.$$

Now let $T = h(x_1, x_2, \ldots, x_n)$ be the unbiased estimator for θ, $Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}$,

and define $T = g(T,\theta)$, where and define $E(T) = \int Tg(T,\theta)dT$. Then

$$E(Z) = E\left(\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta}\right) = 0 \text{ and } Var(Z) = -n\sum_{i=1}^{n} \frac{\partial^2 \log f(x_i,\theta)}{\partial \theta^2}$$

Under the regularity conditions and differentiation under the integral sign:

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} \log f(x_i,\theta)dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} \log f(x_i,\theta)dx$$

$$Cov(T,Z) = E(TZ) - E(T)E(Z) \quad \text{..3.33}$$

$$\rho = Corr = \frac{Cov(T,Z)}{\sigma_T \sigma_Z} \quad \text{and} \quad \rho\sigma_T\sigma_Z + E(T)E(Z) = 1$$

$\rho\sigma_T\sigma_Z = 1$, and $\rho^2\sigma^2{}_T\sigma^2{}_Z \leq 1$, since $E(Z) = 0$ and $\rho^2 \leq 1$ 3.34

$$\sigma^2{}_T \geq \frac{1}{\sigma_Z^2} = \frac{1}{nE\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right]} = \frac{1}{-nE\left[\left(\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2}\right)\right]} = \text{CRLB}$$

Definitions: The following are the consequences of the CRLB inequality:

(i) An unbiased estimator T is said to be efficient if it attains the CRLB.

(ii) For any unbiased estimator T, its efficiency is defined by $eff(T) = \dfrac{CRLB}{Var(T)}$.

(iii) Often $T = t(x_1, x_2, x_3, ..., x_n)$ is efficient at $n \to \infty$. Specifically, T is said to be asymptotically efficient.

For example, if X has the Bernoulli distribution, $B(1,\theta)$, we obtain $I(\theta)$ as follows:

$$\log f(x;\theta) = x\log\theta + (1-x)\log(1-\theta),$$

$$\frac{\partial \log f(x;\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta},$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} = -\frac{x}{\theta} - \frac{1-x}{(1-\theta)^2}$$

$$I(\theta) = -E\left[\frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1}{\theta(1-\theta)}$$

3.6.2.2 Consistency and sufficiency

Let $X_1, X_2, \ldots, X_n$ be a random sample from the random variable X with cumulative distribution function, $F(x,\theta),\ \theta \in \Omega$.. The statistic $\hat{\theta}_n$ is said to be consistent if it converges to θ. In other words, the estimator θ_n is said to be a consistent estimator of θ if, for any positive ε,

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \le \varepsilon\right) = 1, \text{ or equivalently, } \lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = 0.$$

This means that an unbiased estimator $\hat{\theta}_n$ for θ is a consistent estimator of θ if

$\lim_{x\to\infty} Var(\hat{\theta}_n) \to 0$

The property of sufficiency provides methods for finding statistics that in a sense summarize all the information in a sample about a target parameter. Such statistics are said to be sufficient. Often, sufficient statistics are used to develop estimators that have minimum variance among all unbiased estimators.

Definition 2.2: The statistic $T(x_1, x_2, ..., x_n)$ is said to be sufficient for θ if and only if, for each value of t, the conditional distribution of the random sample $(x_1, x_2, ..., x_n)$ given $T = t$ does not depend on θ.

$$f(x_1, x_2, ..., x_n \mid T = t) = \frac{f(x_1, x_2, ..., x_n, \theta)}{h(t,\theta)} = g(x_1, x_2, ..., x_n) \quad ....3.33$$

The following serves as an illustrative example: Let $x_1, x_2, ..., x_n$ be a sequence of independent random sample drawn from the Bernoulli distribution, $B(1,\theta)$, where $P(x_i = 1) = \theta$ and $f(x_i,\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $x_i = 0,1.$ To show that $T = \sum x_i$ is sufficient for θ we have the following:

The statistic $T = \sum x_i \ B(n,\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$

$$f(x_1, x_2, ..., x_n, \theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = \theta^t(1-\theta)^{n-t}$$

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | T = t)$$

$$= \frac{f(x_1, x_2, ..., x_n \mid T = t)}{P(T = t)} = \frac{f(x_1, x_2, ..., x_n, \theta)}{h(t,\theta)}$$

$$= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}, \text{ which is independent of } \theta$$

Theorem 2.4(Neyman Factorisation): This theorem which states as follows may also provide a convenient means of identifying sufficient statistics:

A necessary and sufficient condition for the statistic $T(x_1, x_2, ..., x_n)$ to be sufficient for θ is that the joint probability distribution function of the random sample $x_1, x_2, ..., x_n$ can be factorised into two non-negative functions:

$$f(x_1, x_2, ..., x_n, \theta) = h(T(x_1, x_2, ..., x_n), \theta) \cdot g(x_1, x_2, ..., x_n) \quad \text{...3.34}$$

where $g(x_1, x_2, ..., x_n)$ is independent of θ.

The Proof: Let $X = (X_1, X_2, ..., X_n)$ and $x = (x_1, x_2, ..., x_n)$

$$P(T = t) = \sum_{T(x)=t} P(X = x) = h(t, \theta) \sum_{T(x)=t} g(x)$$

$$P(X = x \mid T = t) = \frac{P(X = x, T = t)}{P(T = t)} = \frac{h(x)}{\sum_{T(x)=t} h(x)}$$ which does not depend on θ. To show that the conclusion holds in other direction, suppose that the conditional distribution of X given T is independent of θ. Let $h(t, \theta) = P(T = t, \theta)$ and $g(x) = P(X = x \mid T = t)$

$$P(X = x|\theta) = P(T = t, \theta).P(X = x \mid T = t = h(t, \theta).g(x)$$

The most commonly used approaches to the statistical estimation of parameters are the least squares method and method of maximum likelihood. The later is usually used for generalized linear models, where the estimates are usually obtained numerically by an iterative procedure which turns out to be closely related to weighted least squares estimation.

3.7 METHODS OF POINT ESTIMATION

3.7.1 Methods of maximum likelihood

Suppose that $x_1, x_2, \ldots, x_n$ are independent and identically distributed (iid) random variables with common probability density function, $f(x;\theta), \theta \in \Omega.$. The basis of our inferential procedures is the likelihood function given by,

$$L(\theta, x) = \prod_{i=1}^{n} f(x_i;\theta), \;\; \theta \in \Omega \quad \text{...............................3.35}$$

where $X = (x_1, x_2, \ldots, x_n)^T$, and the likelihood function which is a function of θ is simply denoted by $L(\theta)$. The maximum likelihood estimator $(\hat{\theta})$ of the parameter θ is obtained by maximizing $L(\theta)$. Usually, for mathematical convenience, we rather work with $l(\theta) = \log L(\theta)$, which interestingly gives us no loss of information in using $l(\theta)$ because the log is a one-to-one function. Thus

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta), \;\; \theta \in \Omega \text{ ....3.36}$$

Illustrative example 2.1:

(a) Given the *iid* random sample, $x_1, x_2, \ldots, x_n$ from the logistic density ,

$$f(x,\theta) = \frac{\exp\{-(x-\theta)\}}{(1+\exp\{-(x-\theta)\})^2}, \qquad -\infty < x < \infty, -\infty < \theta < \infty.$$

the log of the likelihood simplifies to

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) = n\theta - n\overline{x} - 2\sum_{i=1}^{n} \log(1 + \exp\{-(x_i - \theta)\})$$

Using this, the first partial derivative becomes:

$$l^1(\theta) = n - 2\sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{1 + \exp\{-(x_1 - \theta)\}}.$$

Setting this equation to 0 and rearranging terms results in the equation:

$$\sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{1 + \exp\{-(x_i - \theta)\}} = \frac{n}{2}$$

The above has a unique solution. The derivative of the left side of the above equation simplifies to,

$$\frac{\partial}{\partial\theta}\left(\sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{1 + \exp\{-(x_i - \theta)\}}\right) = \sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{(1 + \exp\{-(x_i - \theta)\})^2} > 0.$$

Thus the left side of equation $\sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{1 + \exp\{-(x_i - \theta)\}} = \frac{n}{2}$ is a strictly increasing function of θ. Finally, the left side of the same equation approaches 0 as $\theta \to -\infty$ and approaches n as $\theta \to \infty$. Thus the equation $\sum_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{1 + \exp\{-(x_i - \theta)\}} = \frac{n}{2}$ has a unique solution. Also the second derivative of $l(\theta)$ is strictly negative for all θ; so the solution is a maximum.

(b) Also, suppose we have observations $Y_1, \ldots, Y_N$ which are data for which the GEV distribution is appropriate. For example, perhaps we take one year as the unit of time, and $Y_1, \ldots, Y_N$ represent annual maximum values for each of N years. The corresponding log likelihood is

$$\ell_Y(\mu,\sigma,\gamma) = -N\log\sigma - \left(\frac{1}{\gamma}+1\right)\sum_i\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right) - \sum_i\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right)^{-1/\gamma}$$

provided $1+\gamma(Y_i-\mu)/\sigma > 0$ for each i.

For the Poisson-GPD model discussed above, suppose we have a total of N observations above a threshold in time T, and the excesses over the threshold are $Y_1, \ldots, Y_N$. Suppose the expected number of Poisson exceedances is μT, and the GPD parameters are σ and γ. Then the log likelihood is

$$\ell_{Y,N}(\mu,\sigma,\gamma) = N\log\mu - \mu T - N\log\sigma - \left(1+\frac{1}{\gamma}\right)\sum_i^N \log\left(1+\gamma\frac{Y_i}{\sigma}\right)$$

provided $1+\gamma Y_i/\sigma > 0$ for all i. Similar log likelihoods may be constructed from the joint densities and for the r largest order statistics approach and the point process approach.

The maximum likelihood estimators are the values of the unknown parameters that maximize the log likelihood. In practice these are local maxima found by nonlinear optimization. The standard asymptotic results of consistency, asymptotic efficiency, and asymptotic normality hold for these distributions if $\gamma > -1/2$ (Smith 1985). In particular, the elements of the Hessian matrix of − (the matrix of second-order partial derivatives, evaluated at the maximum likelihood estimators) are known as the observed information matrix, and the inverse of this matrix is a widely used approximation for the variance-covariance matrix of the maximum likelihood estimators.

The square roots of the diagonal entries of this inverse matrix are estimates of the standard deviations of the three parameter estimates, widely known as the standard errors of those estimates. All these results are asymptotic approximations valid for large sample sizes, but in practice they are widely used even when the sample sizes are fairly small.

3.7.2 L- and M-estimators

L-estimators are constructed primarily for estimating the location parameter. Formally, θ will be called a location parameter if $f(x,\theta) = h(x-\theta)$ for some probability density (or probability mass function) h. Similarly, θ is called the scale parameter if $f(x,\theta) = \frac{1}{\theta} h\left(\frac{x}{\theta}\right)$ for some probability density (or probability mass function) h. The mean of a distribution may not be its location parameter. For instance, it is so in case of the normal distribution. However, while in case of a general gamma distribution the mean is neither a scale nor a location parameter.

Let $X_{1,\dots,}X_n$ be a random sample from the distribution $f(x,\theta)$, and let $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ denote the order statistics of the sample. By an L-estimator we mean a statistic from the form $T = \sum_{k=1}^{n} \gamma_{n,k}$, where $\gamma_{n,k}$, $k = 1, \dots, n$ is a double array of coefficients. Thus, L-estimators are simple linear combinations of order statistics. The class of L-estimators contains many well-known estimators: Choosing

$\gamma_{n,k} = {}^{1}/_{n}$ for $k = 1, \dots, n$ gives $= \bar{X}$. The choices $\gamma_{n,1} = 1$, $\gamma_{n,k} = 0$ for $k \geq 2$ or $\gamma_{n,n} = 1$ $\gamma_{n,k} = 0$ for $k < n$ gives two extreme order statistics,

$Y_{1:n} = \min\,(X_1, \dots, X_n)$ and $Y_{n:n} = \max\,(X_1, \dots, X_n)$. In a similar way one can obtain any sample quantile. Choosing $\gamma_{n,[n/4]+1} = -1$, $\gamma_{n,k} = 0$ for the remaining k , one can obtain the sample inter-quartile range, and so on.

Perhaps the most important L-estimators are the trimmed and windsorized means, defined as follows,

Definition: Let $0 < \alpha < \frac{1}{2}$. Then the α-trimmed mean is

$$U_n = \frac{1}{n-2[n\alpha]} \sum_{k=[n\alpha]+1}^{n-[n\alpha]} Y_{k:n}$$

while the α-windsorized mean is

$$V_n = \frac{1}{n}\left([n\alpha]Y_{[n\alpha]+1:n} + \sum_{k=[n\alpha]+1}^{n-[n\alpha]} Y_{k:n} + [n\alpha]Y_{n-[n\alpha]:n}\right)$$

Thus, α-trimming consists of rejecting from the sample the fraction α of lowest and the fraction α of largest observations, and taking the average of the remaining ones(the middle $1-2\alpha$ fraction of observations). On the other hand, α-windsorizing consists of replacing each observation in the lower α and $1-\alpha$, respectively. The windsorized mean is then calculated as the mean of the windsorized sample. It is clear that the purpose of trimming (or windsorizing) is to eliminate (or decrease) the effect of outliers in the sample.

The main objective of L-estimates, apart from establishing their asymptotic properties, is to define the notion of optimality and then to determine the optimal level of α at which the mean should be trimmed or windsorized.

Another class of estimators is M-estimator obtained as follows: Let $h(x,u)$ be a function of two arguments. Given a ample $x_1, x_2, \ldots, x_n$ from the distribution $f(x,\theta)$, one takes as an estimator of θ the solution of the equation

$$\sum_{k=1}^{n} h(\, x_k, u) = 0 \quad \ldots\ldots 3.37$$

Such estimators are most often obtained by solving an approximate minimisation problem. Suppose that we have a "distance" of some sort (not necessarily satisfying any conditions for metric), say $H(x\,,u)$. As an estimator of θ we choose a point u^* that minimise the sum

$$\sum_{k=1}^{n} H(\, x_k, u) \quad \ldots\ldots 3.38$$

interpreted as the sum of distances from u to all sample points. In a sense, u^* is the point closest to the sample, with closeness being expressed by function H. differentiating (2.37) with respect to u and setting the derivative equal to zero(0), we obtain equation (2.38) with $h(x,u) = \frac{\delta}{\delta u} H(x,u)$. This formation comprises two important classes of estimators, MLE's and least squares estimators. Indeed, if we define the $H(x\,,u)$ as $-logf(x,u)$, then $h(x,u) = -\frac{\delta}{\delta u} \log \mathrm{f(x,u)}$ and the M-estimator corresponding to this choice is the maximum likelihood estimator (the minus sign is connected with the fact that the problem is now formulated as a minimisation problem).

If we take appropriate functions H and h, we can also obtain different variants of least squares estimators. Similarly, trimmed or windsorised means can be obtained by appropriate choice of the function H and h. For example, we may take a special form of function $H(x, u)$, namely $H(x, u) = H(x - u)$, for some function H of one argument. The M-estimator then minimises the sum $\sum_{i=1}^{n} H(X_i - u)$. For $H(x) = x^2$ we have the simplest least squares estimator; if $H(x) = x^2$ for $|x| \leq k$ and $H(x) = k^2$ for $|x| > k$, we obtain a form of windsorised mean.

As with L-estimators, the main direction of research is to study the asymptotic properties of M-estimators under some general assumptions on functions H or h and distributions of X_i.

3.7.3 The bayesian estimation principle

Bayesian estimation is another likelihood–based method besides the maximum likelihood (ML) method. In addition to the statistical model, the statistician must specify a prior distribution. In this section, we introduce Bayes estimators within a decision–theoretical framework.We estimate a real–valued functional parameter $T(\theta)$, $E\left(\left(\hat{T}(X)-T(\theta)\right)^2|\theta\right) := E\left(\hat{T}(X)-T(\theta)\right)^2$ where $\theta = (\theta_1, \theta_2, \theta_3, ..., \theta_m)$ is a parameter vector. This includes, e.g., the estimation of the jth component θ_j of θ if $T(\theta)=\theta_j$, and of the mean $T(\theta)=\int x dF_\theta(x)$ of the underlying df F_θ represented by θ .

T he Prior Density, Minimizing the Bayes Risk: Let $\hat{T}$(X) be an estimator of the functional parameter $T(\theta)$, where X is a random variable having a distribution represented by θ . For example, $X = (X_1, X_2, X_3, ..., X_n)$ is a sample of size *n*, $T(\theta)$ is the mean of the common distribution and $\hat{T}$(X) is the sample mean X. The performance of $\hat{T}$(X) as an estimator of can be measured by the mean squared error (MSE), $E\left(\left(\hat{T}(X)-T(\theta)\right)^2|\theta\right):=E\left(\hat{T}(X)-T(\theta)\right)^2$ 3.39

where the left–hand side emphasizes the fact that the expectation is taken under the parameter θ. The performance of the estimator can be made independent of a special parameter vector θ by means of a "prior" probability density $P(\theta)$ which may be regarded as a weight function. Some prior knowledge about the parameter of interest is included in the statistical modeling by means of the prior density $P(\theta)$. The Bayes risk of the estimator $\hat{T}$ with respect to the prior $P(\theta)$ is the integrated MSE

$$\mathrm{R(p,\hat{T})= E(\hat{T}(X)\text{-}T(}\theta))^2 \ | \ \theta)\mathrm{P}(\theta)\mathrm{d}\theta_1....d\theta_m \quad ....3.40$$

An estimator $\hat{T}$, which minimizes the Bayes risk, is called Bayes estimator.

Computing the Bayes Estimator, the Posterior Density : We introduce the posterior density, which is determined by the prior density $P(\theta)$ and the likelihood function, and deduce an explicit representation of the Bayes estimator by means of the posterior density. Let $X=(X_1,X_2,X_3,...,X_n)$ be a vector of iid random variables with common distribution function F_θ and density f_θ, where again $\theta=(\theta_1,\theta_2,\theta_3,...,\theta_m)$ is the parameter vector.

Let $L\left(x|\theta\right)=\prod_{i\le n} f_\theta(x_i)$ be the likelihood function given the sample vector $X=(X_1,X_2,X_3,...,X_n)$. Using the likelihood function, one gets the representation

$$E((\hat{T}(X)-T(\theta))^2|\theta)=\int E((\hat{T}(X)-T(\theta))^2|\theta)\,L(x|\theta)dx_1.....dx_n \quad ....3.41$$

of the MSE. We verify that the Bayes estimate can be written

$$T^*(x) = \int T(\theta)P(\theta|x)d\theta_1.....d\theta_m \quad ....3.42$$

where the function:

$$P(\theta|x) = \frac{L(x|\theta)P(\theta)}{\int L(x|\theta)P(\theta)d\theta_1.....d\theta_m} \quad ....3.43$$

is the "posterior" density for a given sample vector x, whenever the denominator is larger than zero. If ϑ is a one–dimensional parameter and $T(\theta)=\theta$, then the Bayes estimate is the mean $\int \theta P(\theta|x)d\theta$ of the posterior distribution according to (2.42). Thus, by means of the prior density $P(\theta)$ and the likelihood function $L(x|\theta)$ one gets the posterior density $P(\theta|x)$. Notice that the posterior density in (2.42) and the Bayes estimate can be computed whenever $L(x|\theta)P(\theta)$, as a function in θ, is known up to a constant. Writing $g(\theta) \propto f(\theta)$, when functions g and f are proportional, we have $p(\theta|x) \propto L(x|\theta)p(\theta)$.

To prove the representation (2.42) of the Bayes estimate, combine (2.40) and (2.41) and interchange the order of the integration. The Bayes risk with respect to the prior $P(\theta)$ can be written

$$\begin{aligned} R(p,\hat{T}) &= \int E(\hat{T}(X)-T(\theta))^2 \mid \theta)L(x|\theta)p(\theta)dx_1....dx_n d\theta_1........d\theta_m \\ &= \int\left(\int (\hat{T}(X)-T(\theta))^2 p(\theta|x)d\theta_1.....d\theta_n\right)f(x)dx_1.......dx_n \end{aligned} \quad ....(3.44)$$

where $f(x) = \int L(x|\theta)P(\theta)d\theta_1........d\theta_m$. To get the Bayes estimate, compute for every **x** the value z which minimizes the integral $\int\left(z-\hat{T}(\theta)\right)^2 p(\theta|x)d\theta_1..........d\theta_m$. This value z can be computed by showing that the derivative

$$\frac{\partial}{\partial z}\int\left(z-\hat{T}(\theta)\right)^2 p(\theta|x)d\theta_1..........d\theta_m = \int 2\left(z-\hat{T}(\theta)\right)p(\theta|x)d\theta_1..........d\theta_m$$

$$= 2(z - T^*(x))$$

is equal to zero, if z is equal to the value $T^*(x)$ in (2.42). In those cases where the posterior density $P(\theta|x)$ is of the same type as the prior density $P(\theta)$ one speaks of a conjugate prior.

3.7.4 The Maximum likelihood Test

Let $X_1, X_2, \ldots, X_n$ be *iid* with *pdf* $f(X,\theta)$ for $\theta \in \Omega$, where θ is a scalar. Consider the two-sided hypotheses $H_o : \theta = \theta_o$ versus $H_1 : \theta \neq \theta_o$, where θ_o is a specified value. The maximum likelihood function and its logarithm are given by:

$$L(\theta) = \prod_{i=1}^{n} f(x_i.,\theta) \quad \text{and} \quad l(\theta) = \sum_{i=1}^{n} \log f(x_i.,\theta) \text{ ....3.45 .}$$

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ. To motivate the test, if θ_o is the true value of θ then, asymptotically, $L(\theta_o)$ is the maximum value of $L(\theta)$. Consider the ratio of two likelihood functions, namely

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} \qquad \text{...3.46}$$

Note that $\Lambda \leq 1$, but if H_0 is true Λ should be large (close to 1); while H_1 is true, Λ should be smaller. For a specified significance level α, this leads to the intuitive decision rule, Reject H_0 in favour of H_1 if $\Lambda \leq c$ where c is such that $\alpha = P_{\theta_0}[\Lambda \leq c]$. This test is called the *likelihood ratio test.*

In the multi-parameter case, hypotheses of interest often specify θ to be in a sub-region of the space. Suppose X has a $N(\mu,\sigma^2)$ distribution. The full space is $\Omega = \{(\mu,\sigma^2) : \sigma^2 > 0, -\infty < \mu < \infty\}$. This is a two-dimensional space. We may be interested through in testing that $\mu = \mu_0$, where μ_0 is a specified value. Under H_0, the parameter space is the one-dimensional space $\omega = \{(\mu,\sigma^2) : \sigma^2 > 0\}$. We say that H_0, is defined in terms of one constraint on the space Ω.

In general, let $x_1,\ldots\ldots,x_n$ be *iid* with *pdf* $f(x;\theta)$ for $\theta \in \Omega \subset R^p$. The hypotheses of interest here are, $H_0 : \theta \in \omega \; versus \; H_1 : \theta \in \Omega \cap \omega^c$, where $\omega \subset \Omega$ is defined in terms of $q, 0 < q \leq p,$ independent constraints of the form, $g_1(\theta) = a_1, \ldots., g_p(\theta) = a_q$. The functions $g_1, g_2, \ldots, g_q$ must be continuously differentiable. This implies ω is a $p - q$ dimensional space. Based on Theorem;

$L(\theta, x) = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Omega,$, the true parameter maximizes the likelihood function, so an intuitive test statistic is given by the likelihood ratio.

$$\Lambda = \frac{\max_{\theta\in\omega} L(\theta)}{\max_{\theta\in\Omega} L(\theta)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.47$$

Large values (close to one) of Λ suggests that H_0 is true, while small values indicate H_1 is true. For a specified level $\alpha, 0 < \alpha < 1,$ this suggests the decision rule reject H_0 in favour of H_1, if $\Lambda \le C,$ where C is such that $\alpha = \max_{\theta \in \omega} P[\Lambda \le C].$ As in the scale case, this test often has optimal properties; To determine $C,$ we need to determine the distribution of Λ or a function of Λ when H is true.

Let $\hat{\theta}$ denote the maximum likelihood estimator when the parameter space is the full space Ω and let $\hat{\theta}_0$ denote the maximum likelihood estimator when the parameter space is the reduced space ω. For convenience, define $L(\hat{\Omega}) = L(\hat{\theta})$ and $L(\hat{\omega}) = L(\hat{\theta}_0).$ Then we can write the test statistic as $\Lambda = \dfrac{L(\hat{\omega})}{L(\hat{\Omega})}.$.(Mckean, et al 2005)

3.7.5 Profile Likelihoods for Quantiles

Suppose we are interested in the *n*-year return level y_n, i.e., the $(1-1/n)$-quantile of the annual maximum distribution. This is given by solving the equation

$\exp\left\{-\left(1+\gamma\dfrac{y_n-\mu}{\sigma}\right)^{-1/\gamma}\right\} = 1-\dfrac{1}{n}$. Exploiting the approximation

$1-\dfrac{1}{n} \approx \exp\left(-\dfrac{1}{n}\right)$, this simplifies to $\left(1+\gamma\dfrac{y_n-\mu}{\sigma}\right)^{-1/\gamma} = \dfrac{1}{n}$, and hence $y_n = \mu + \sigma\dfrac{n^{\gamma}-1}{\gamma}$.

One approach to the estimation of y_n is simply to substitute the maximum likelihood estimates $\hat{\mu}, \hat{\sigma}, \hat{\gamma}$ for the unknown parameters μ, σ, γ respectively, thus creating a maximum likelihood estimator $\hat{y}_n$. The variance of $\hat{y}_n$ may be estimated through a standard delta function approximation.

CHAPTER FOUR

FINANCIAL-RISK MEASURES AND MODELLING APPROACHES

4.1 CONCEPTS OF FINANCIAL-RISK MEASURES

Some of the most frequent questions concerning risk management in finance involve extreme quantile estimation. This corresponds to the determination of the value a given variable exceeds with a given (low) probability. A typical example of such measures is the Value-at-Risk (VaR). Other less frequently used measures are the expected shortfall (ES) and the return level. Hereafter, we define the risk measures we focus on in the following section.

4.1.1 Extreme Market Risk

In the market risk interpretation of our random variables

$$X_t = -\left(\log P_t - \log P_{t-1}\right)*100 \approx 100*\left(P_{t-1} - P_t\right)/P_{t-1} \quad \text{..4.1}$$

represents the loss on a portfolio of traded assets on day *t*, where P*t* is the closing value of the portfolio on that day. We change to subscript *t* to emphasize the temporal indexing of our risks. As shown above, the loss may be defined as a relative or logarithmic difference, both definitions giving very similar values.

In calculating daily VaR estimates for such risks, there is now a general recognition that the calculation should take into account *volatility* of market instruments. An extreme value in a period of high volatility appears less extreme than the same value in a period of low volatility. Various authors have acknowledged the need to *scale* VaR estimates by current volatility in some way (see, for example, Hull & White (1998)). Any approach which achieves this call for a *dynamic* risk measurement procedure. In this thesis we focus on how the dynamic measurement of market risks can be further enhanced with EVT to take into account the extreme risk over and above the volatility risk.

Most market return series show a great deal of common structure. This suggests that more sophisticated modelling is both possible and necessary; it is not sufficient to assume they are independent and identically distributed. Various stylized facts of empirical finance argue against this. While the correlation of market returns is low, the serial correlation of absolute or squared returns is high; returns show *volatility clustering* - the tendency of large values to be followed by other large values, although not necessarily of the same sign.

4.1.2 Basel I and II

Paul Embrechts referred to modelling and analysis of extremes in econometric data as "extremes in economics". This has been illustrated by Longin, F. (2001), one of the pioneers of EVT in finance. In this paper, the author, for instance, estimates the probability of exceedance and waiting time period for the ten largest daily return price movements in the U.S. equity market (S&P 500) over the period July 1962 to December 1999. This "hit parade" ranges from −18.35% on October 19, 1987 to −3.29% on October 9, 1979. Other, perhaps less well known applied econometric work on extremes in economics concerns spill-over events; see for instance Hartmann et al.(2001) . In this case, EVT in a multidimensional setup appears.

"Economics of Extremes," concentrates on the crucial question: given the econometric evidence on quantifiable extremal events in finance (and insurance), how can we handle these extremes from an economic point of view?. Some concrete questions could be: How can one devise prudent regulatory rules aiming at market stability? Here the Basel Committee enters.

Measure and more importantly price the time-dimension of system-wide risk. For these questions, see for instance, Crockett, A. (2000) and Borio et al. (2001). An interesting review on systemic risk, an area where EVT as a quantitative tool has a lot to offer, is De Brandt, O and Hartmann, P (2000). Important in these problems is finding typically macroeconomic structures that help the economy/market to dampen (hopefully avoid) the more negative consequences of extremal events. For most of the more mathematically minded extreme value theorists, working in risk management is equivalent to estimating Value-at-Risk (VaR) for ever more complicated stochastic models. In their terminology, VaR is "just" a quantile of some underlying process or distribution. However, VaR is to finance what body temperature is to a patient; an indicator of bad health but not an instrument telling us what is wrong and far less a clue on how to get the patient (system) healthy again. Let us look at some of the main issues in risk management (RM) from the perspective of the regulator as personified by the famous Basel Committee. Details underlying the summary below are to be found on the homepage www.bis.org of the Bank of International Settlements in Basel.

The Basel Committee was established by the Central-Bank Governors of the Group of Ten at the end of 1974. The committee does not possess any formal supranational supervisory authority, and hence its conclusions do not have legal force. Rather, it formulates broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements — statutory or otherwise — that are best suited to their own national system. In 1988, the committee introduced a capital measurement system, commonly referred to as the Basel Capital Accord (also called Basel I). This system provided for the implementation of a credit risk measurement framework with a minimum capital standard of 8% (a so-called haircut) by the end of 1992. From the start, banks criticized the lack of risk sensitivity in this approach.

On the credit risk side, this led to the New Capital Adequacy (so-called Basel II) framework of June 1999. The latter is now under discussion with the industry and was planned to become operational from 2006. Besides these key developments within the credit risk area, already around 1994 we saw various amendments to Basel I catering to market risk, in particular for derivative positions. The 1996 report on the amendment to the Capital Accord to incorporate market risks opened the floodgates for the VaR–modellers. Through this amendment, a direct link between the quantitative VaR measure for market risk and regulatory capital was established. The exact form of the link very much depends on the statistical qualities of the underlying market risk models through backtesting.

4.1.3 Value-at-Risk (VaR)

In today's financial world, the Value-at-Risk (VaR) has become the benchmark risk measure. Following the *Basle Accord on Market Risk* (1988, 1995, 1996) every bank in more than 100 countries around the world has to calculate its risk exposure for every individual trading desk. The standard method (the normal model) prescribes in the following steps:

- estimate the *p*-quantile of the profit/loss distribution for the next 10 days
- set $p = 1\%$ (or $p = 5\%$) based on observations of at least one year (220 trading days).
- Finally, multiply the estimated quantile by 3. This number is negative and its modulus is called Value-at-Risk (VaR).

The factor 3 is supposed to account for certain observed effects, also due to the model risk; it is based on backtesting procedures and can be increased by the regulatory authorities, if the backtesting proves the factor 3 to be insufficient. The importance of VaR is undebated since regulators accept this model as a basis for setting capital requirements for market risk exposure. A textbook treatment of VaR is given in Joriot (1997). Interesting articles on risk management are collected in Embrechts,et al(1997).

There were always discussions about the classical risk measure, which has traditionally been the variance, and alternatives have been suggested. They are typically based on the notion of downside risk concepts such as lower partial moments.

Value-at-Risk is generally defined as the capital sufficient to cover, in most instances, losses from a portfolio over a holding period of a fixed number of days. Suppose a random variable *X* with continuous distribution function *F* models losses or negative returns on a certain financial instrument over a certain time horizon. VaR_p can then be defined as the *p*-th quantile of the distribution *F*, that is $VaR_p = F^{-1}(1-p)$

where F^{-1} is the so called *quantile function* defined as the inverse of the distribution function *F*. For internal risk control purposes, most of the financial firms compute a 5% VaR over a one-day holding period. The Basel accord proposed that VaR for the next 10 days and p = 1%, based on a historical observation period of at least 1 year of daily data, should be computed and then multiplied by the `safety factor' 3. The safety factor was introduced because the normal hypothesis for the profit and loss distribution is widely recognized as unrealistic.

More generally a quantile function is defined as the generalized inverse of *F*:

$$F^{\leftarrow}(p) = \inf \{x \in R : F(x) \geq p\} \qquad \text{4.2}$$

4.1.4 Conditional Value - at – Risk (CVaR)

As it can be easily seen, VaR is a risk measure that only takes account of the probability of losses, and not of their sizes. Moreover, VaR is usually based on the assumption of normal asset returns and has to be carefully evaluated when there are extreme price fluctuations. Furthermore, VaR may be convex for some probability distributions. Due to these deficiencies, other risk measures have been proposed. Among them, the *Expected Shortfall (ES)* as defined in Acerbi et al. (2001), also called *Conditional Value-at-Risk (CVaR)*. Note that in Acerbi and Tasche (2002), several risk measures related to ES are considered and the coherence of ES is proved.

4.1.5 The Benefits of using VaR

The previous sections have shown that VaR is not unproblematic to use, and is not a coherent risk measure, however its estimate is downward biased and subject to large errors. These shortcomings which can be exploited by individuals within the company as well as the company as a whole do not necessarily imply that VaR is not a useful tool in risk management. The obvious benefit of VaR is that it is easily and intuitively understood by non-specialists. It can therefore be well communicated within a company as well as between the company and regulators, investors, or other stakeholders. Furthermore, it can address all types of risks in a single framework, which not only allows the aggregation of risks but also further facilitates communication.

With the help of VaR we can address most problems arising from risks. It can be used to set risk limits for individual traders, divisions, and the entire company; it facilitates performance measurement and can be used as the basis for performance-related pay; it facilitates decisions on the allocation of capital and gives an indication of the total capital requirement; finally, it can help to decide which risks to reduce, if necessary. No other risk management system developed thus far addresses all these aspects in a single framework, while still being accessible to managers and a wide range of employees. Therefore VaR has proved to be a very useful tool that is readily accepted, despite its shortcomings. But it is exactly these shortcomings that limit the extent to which VaR can be used. The VaR estimate should not be taken as a precise number, but it provides an indication as to how much risk is involved. It also aids in detecting any trends in the behavior of individuals, divisions, or the company as a whole. Properly used, VaR is a powerful but still simple tool in risk management. On the other hand, overreliance on its results and justifying important decisions solely on its basis are likely to be counterproductive. No risk management system can replace the sound judgment of managers, and those using it should be aware of its limits. The benefits of the simplicity of VaR cannot be underestimated. A much more precise and improved method that is not understood by decision makers is of much less value than an easily understood method, even if it gives only a rough estimate of the risks involved, provided these limits are understood. Results based on systems that are not understood are either ignored or used without the necessary precautions. In both cases decisions are likely to be inferior.

4.1.6 Expected shortfall and return level

Another informative measure of risk is the *expected shortfall* (ES) or the *tail conditional expectation* which estimates the potential size of the loss exceeding VaR. The expected shortfall is defined as the expected size of a loss that exceeds VaR_p .

$$ES_p = E\left(X \middle| X > VaR_p\right) \qquad \text{...4.3}$$

Artzner et al. (1999) argue that expected shortfall, as opposed to Value-at-Risk, is a coherent risk measure.

If H is the distribution of the maxima observed over successive non overlapping periods of equal length, the *return level* $R_n^k = H^{-1}\left(1-\frac{1}{k}\right)$ is the level expected to be exceeded in one out of k periods of length n. The return level can be used as a measure of the maximum loss of a portfolio, a rather more conservative measure than the Value-at-Risk.

4.2 SOME MODELLING APPROACHES

4.2.1 Block maxima approach (GEV)

A starting point for modelling the extremes of a process is based on distributional models derived from asymptotic theory. The *parametric approach* to modelling extremes is based on the assumption that the data in hand ($X_1, X_2, X_3, ...X_n$) form an i.i.d. sample from an exact GEV distribution function. In this case, standard statistical methodology from parametric estimation theory can be utilised in order to derive estimates of the parameters θ In practice, this approach is adopted whenever our dataset consist of maxima of independent samples (e.g. in hydrology we have disjoint time periods). This method is often called method of *block maxima*. Such techniques are discussed in DuMouchel (1983), Hosking (1985), Hosking et al. (1985), Smith (1985), Scarf (1992), Embrechts et al. (1997) and Coles and Dixon (1999). However, this approach may seem restrictive and not very realistic since the grouping of data into maximum/minimum is sometimes rather arbitrary, while by using only the block maxima, we may loose important information (some blocks may contain several among the largest observations, while other blocks may contain none). Moreover, in the case that we have few data, block maxima cannot be actually implemented.

In this thesis we deal with another widely used approach, the so-called 'Maximum Domain of Attraction Approach' by (Embrechts et al., 1997), or Non-Parametric. In the present context we prefer the term '*semi-parametric*' since this term reflects the fact that we make only partly assumptions about the unknown distribution function, *F*. So, essentially, we are interested in the distribution of the maximum (or minimum) value. Here is the point where EVT gets involved. According to the Fisher-Tippet theorem, the limiting distribution function of the (normalized) maximum value (if it exists) is the GEV distribution function. $H_{\theta} = H_{\gamma;\mu,\sigma}$. So, without making any assumptions about the unknown distribution function, *F* (apart from some continuity conditions which ensure the existence of the limiting distribution function.), EVT provides us with a fairly sufficient tool for describing the behaviour of extremes of the distribution that the data in hand stem from. The only issue that remains to be resolved is the estimation of the parameters of the GEV distribution function. $\theta = (\gamma, \mu, \sigma)$.

Of these parameters, the *shape parameter* γ (also called *tail index* or *extreme-value index*) is the one that attracts most of the attention, since this is the parameter that determines, in general terms, the behaviour of extremes. According to EVT these are the parameters of the GEV distribution function that the maximum value follows asymptotically. Of course, in reality, we only have a finite sample and, in any case, we cannot use only the largest observation for inference. So, the procedure followed in practice is that we assume that the asymptotic approximation is achieved for the largest k observations (where k is large but not as large as the sample size n), which we subsequently use for the estimation of the parameters. However, the choice of k is not an easy task. On the contrary, it is a very controversial issue. Many authors have suggested alternative methods for choosing k, but no method has been universally accepted.

One approach to working with extreme value data is to group the data into blocks of equal length and fit the data to the maximums of each block; for example, monthly maxima of daily oil pricing. The choice of block size can be critical as blocks that are too small can lead to bias and blocks that are too large generate too few block maxima, which leads to large estimation variance Coles, Stuart(2001). The block maxima approach is closely associated with the use of the GEV family. Note that all parameters are always estimated by maximum likelihood estimation (MLE), which requires iterative numerical optimization techniques.

4.2.2 The R- largest order statistics model

An extension of the maximum approach is to use the *r* largest observations in each fixed time period (say, one year), where *r* > 1. The mathematical result on which this relies is that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le x\right\} = F\left(a_n x + b_n\right)^n \to H(x) \quad \text{...4.5}$$

which is easily extended to the joint distribution of the *r* largest order statistics, as $n \to \infty$ for a fixed $r > 1$, and this may therefore be used as a basis for statistical inference. A practical caution is that the *r*-largest result is more vulnerable to departures from the i.i.d. assumption (say, if there is seasonal variation in the distribution of observations, or if observations are dependent) than the classical results about extremes.

The main result is as follows: if $Y_{n,1} \ge Y_{n,2} \ge Y_{n,3} \ge \ldots \ge Y_{n,r}$ are r largest order statistics of i.i.d. sample of size *n*, and a_n and b_n are the normalising constants in equation (3.5), then $\left(\frac{Y_{n,1}-b_n}{a_n},\ldots,\frac{Y_{n,r}-b_n}{a_n}\right)$ converges in distribution to a limiting random vector $\left(X_1,\ldots,X_r\right)$, whose density is

$$h(x_1,\ldots,x_r)=\sigma^{-r}exp\left\{-\left(1+\gamma\frac{x_r-\mu}{\sigma}\right)^{-1/\xi}-\left(1+\frac{1}{\gamma}\right)\sum_{j=1}^{r}\log\left(1+\gamma\frac{x_j-\mu}{\sigma}\right)\right\}\text{..........4.6}$$

Some examples using this approach are the papers of Smith (1986) and Tawn (1988) on hydrological extremes, and Robinson and Tawn (1995) and Smith (1997) for a novel application to the analysis of athletic records.

4.2.3 Point process approach

This was introduced as a statistical approach by Smith (1989), though the basic probability theory from which it derives had been developed by a number of earlier authors. In particular, the books by Leadbetter et al. (1983) and Resnick (1987) have much information on point-process viewpoints of EVT. In this approach, instead of considering the times at which high-threshold exceedances occur and the excess values over the threshold as two separate processes, they are combined into one process based on a two-dimensional plot of exceedance times and exceedance values. The asymptotic theory of threshold exceedances shows that under suitable normalisation, this process behaves like a non homogeneous Poisson process.

4.2.4 Peaks over threshold (pot)/point process (pp) approach

The GPD model looks at exceedances over a threshold and those values are fitted to a generalized Pareto distribution. A more theoretically appealing way to analyze extreme values is to use a point process characterization. This approach is consistent with a Poisson process for the occurrence of exceedances of a high threshold and the GPD for excesses over this threshold. Inferences made from such a characterization can be obtained using other appropriate models from above (Coles, Stuart (2001)). However, there are good reasons to consider this approach. Namely, it provides a nice interpretation of extremes that unifies all of the previously discussed models. For example, the parameters associated with the point process model can be converted to those of the GEV parameterization. In fact, the point process approach can be viewed as an indirect way of fitting data to the GEV distribution that makes use of more information about the upper tail of the distribution than does the block maxima approach (Coles, Stuart(2001)).

CHAPTER FIVE

MODELLING OF DAILY BRENT CRUDE OIL PRICES

5.1 INTRODUCTION

In this chapter our primary objective is to analyze the volatility of daily crude oil prices by applying the concepts EVT to model the tails of the distribution for daily returns. We describe the historical data for Brent crude oil prices on worlds market, the preliminary tests undertaken on the data and exploratory techniques, the determination of thresholds, the fitting of the GPD, and the examination of tail modelling. The empirical analysis has been undertaken by writing program code that was executed using the R package. We have used GEV distribution fitting to the extreme value data and investigate the likelihood-based confidence intervals for quantiles of the fitted distribution. We will also use Generalised Pareto Distribution (GPD) and fit the extreme value data.

Fitting a parametric distribution to the daily Oil price (Brent crude) sometimes results in a model that agrees well with the data in high density regions, but poorly in areas of low density. For unimodal distributions, such as the distribution of oil pricing, these low density regions are known as the "tails" of the distribution. One reason why a parameter model might fit poorly in the tails is that by definition, there are fewer data in the tails on which to base a choice of model, and so models are often chosen based on their ability to fit data near the mode. Another reason might be that the distribution of real data is often more complicated than the usual parametric models. However, in many applications, fitting the data in the tail is the main concern. The GPD was developed as a distribution that can model tails of a wide variety of distributions, based on theoretical arguments. One approach to distribution fitting that involves the GPD is to use a non-parametric fit (for example the empirical cumulative distribution function) in regions where there are many observations, and to fit the GPD to the tail(s) of the data.

We will also demonstrate how to fit the GPD to tail data, using functions in the Statistics software of R and Matlab for fitting this distribution and estimating parameters by maximum likelihood method. The Generalized Pareto (GP) is a right-skewed distribution, parameterized with a shape parameter, γ, and a scale parameter σ, sigma. The shape parameter, γ is also known as the "tail index" parameter, and can be positive, zero, or negative.

5.2 PRELIMINARY ANALYSIS OF DATA

The descriptive analysis performed in this section gives the important features of the data obtained. Table 4.1 presented below give quick descriptive summary of all the parameters under study. The data used in this study is the daily closing prices of Brent crude oil over the period extending from 21 May 1987 to 18 May 2009 making total observations of 5594 excluding public holidays. The data are obtained from the website http://tonto.eia.doe.gov/dnav/pet/hist/rbrted.htm source Energy Information Administration

Table 5.1: Descriptive statistics for the Oil pricing returns

Mean	30.91842	Skewness	2.126262
Maximum	143.9500	Kurtosis	7.681817
Minimum	9.100000	Jarque-Bera	9324.118
Std Dev.	22.77285	Probability	0.000000

The returns at time *t* are defined in the natural logarithm of the oil price (p), that is,

$$X_t = -\left(\log P_t - \log P_{t-1}\right)*100 \approx \ 100*\left(P_{t-1} - P_t\right)/P_{t-1} \quad ....5.1$$

Generally the index has a large difference between its maximum and minimum returns. The standard deviation is also high indicating a high level of fluctuations of the daily Oil price. There is also evidence of positive skewness, which means that the right tail is particularly extreme, an indication that the Oil pricing has non-symmetric returns. Spot Oil price are leptokurtic or fat-tailed, given its large kurtosis statistics in Table 5.1. The kurtosis exceeds the normal value of 3. The series is non-normal according to the Jarque-Bera test, which rejects normality at the 1% level for each series.

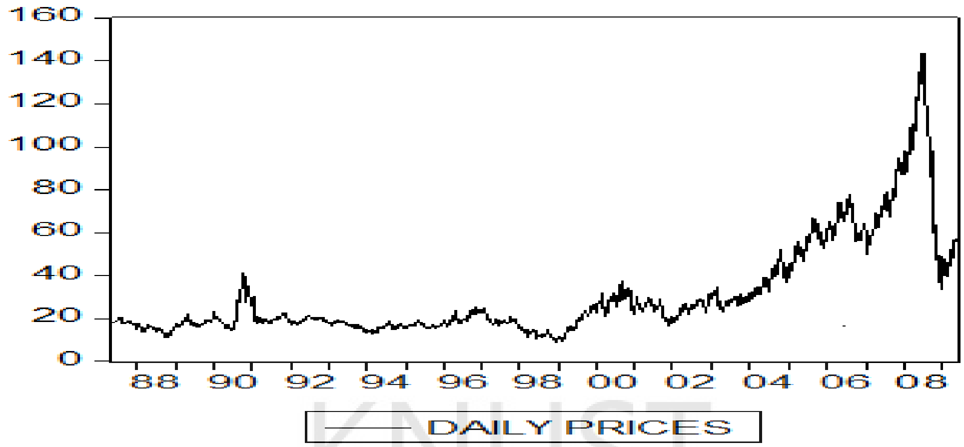

Figure 5.1: Plots of Daily oil prices

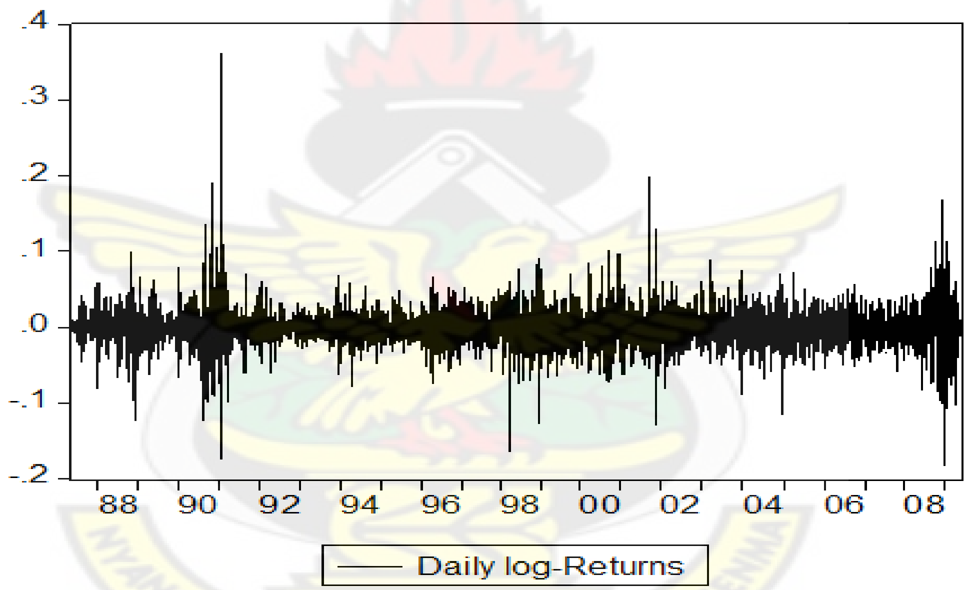

Figure 5.2: Plots of daily logarithmic returns.

There are 5,593 adjusted observations, including 2,655 observations of gains and 2,793 observations of losses. The plot of the daily crude prices (Figure 4.1a) shows a substantial increase since 2002 with lots of fluctuations, and the graph of daily returns (Figure 4.1b) confirms the volatility of the crude oil market.

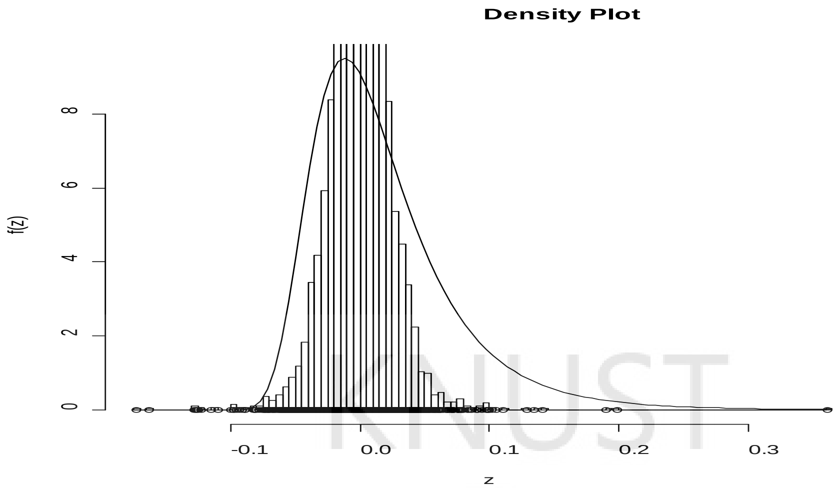

Figure 5.3: Histogram of standardized returns

In Figure 5.3 above, we reproduce the histogram of the standardized returns and we observe that in the tails the data are far from normal.

5.2.1 Distribution of Gains and Losses

Some studies, including Krehbiel and Adkins (2005), claim that the upper and lower tails behave differently, and thus should be treated separately while estimating risk measures. Evidences from our empirical study show the small difference in risk statistics on both tails, implying that the thickness of two tails is likely to be similar.

In Figure 5.4 below, the empirical distribution of both tails of the distribution are shown. The gains distribution (top) shows a higher extremal observation than that of the losses distribution (bottom). The losses on the other hand, show many extremal observations than the gains distribution.

The distributions of the gains and losses observations are shown in figure 5.4. Thus the figure gives the empirical distribution of both tails of the gains and losses.

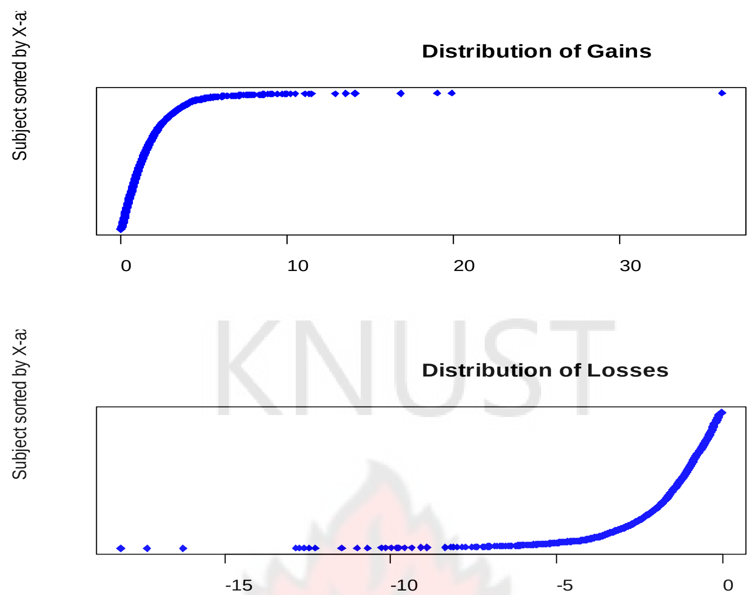

Figure 5.4: Distribution of gains (top) and losses(bottom)

5.3 MODELLING OF DATA

5.3.1 Block maxima approach (GEV)

We now apply the block maxima method to our daily return data for the gains (right tail) distribution. Similar method will be applied for losses (left tail). For this method the delicate point is the appropriate choice of the periods defining the blocks. The suggested periods is block size of 30 observations, this is due to the volatile nature of the returns and also for sufficiently large data for *Theorem 3.1* in chapter 3 to hold. For the gains, the data used over the period under study (ie 2,655 observations) is divided into 30 non-overlapping sub-samples and the maximum observed value picked. This gives a total sample size of 89 observations, each of them containing the daily returns of the successive month.

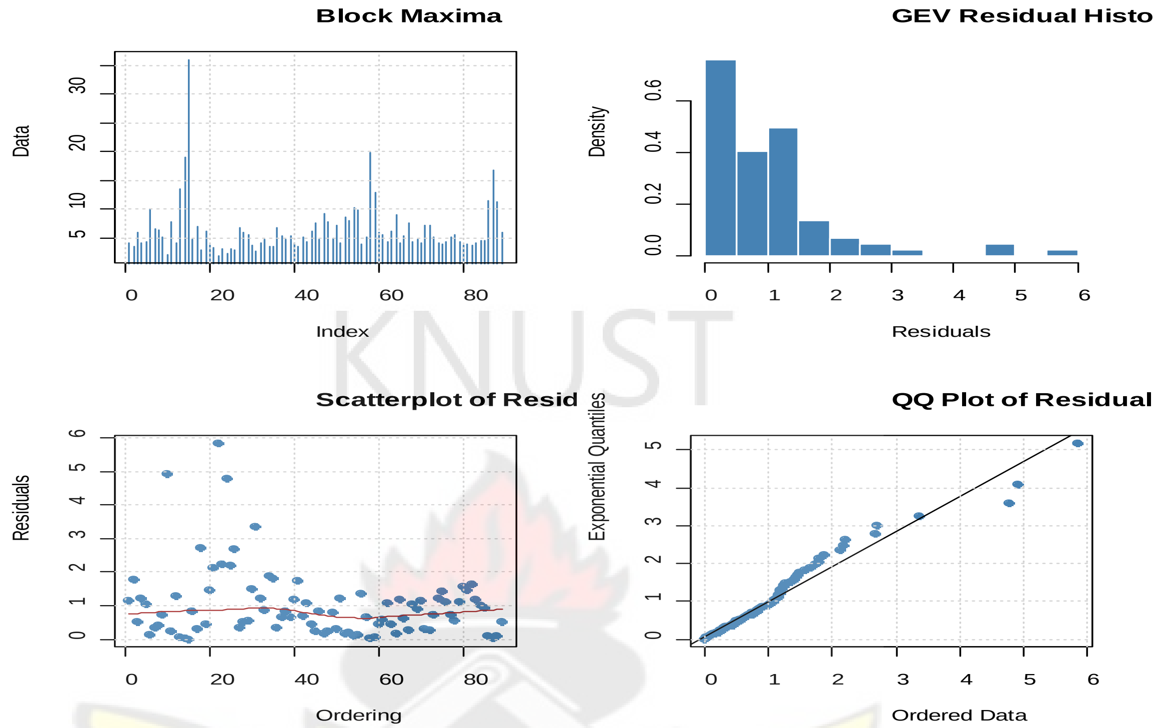

Figure 5.5: Plots of the monthly maxima of the daily returns of GEV fit. (gains)

In Figure 5.5 the two top plots; gives the sample plot of block maxima (left) and a histogram showing the GEV of residuals (right). In the bottom is a scatter plot of the residuals showing time of block maxima. The idea here is to observe a possible time trend in the observations. To aid in judging this, a simple fitted curve (using the function of fExtremes package in R) is superimposed and there is evidence of a systematic trend. The solid line is the smooth of scattered residuals obtained by a spline method. The QQ-plot of the residuals is shown in the bottom right. The QQ-plot of the fitted model does not deviate much from the straight line; the plot is based on the GEV fitted to all 89 block maxima.

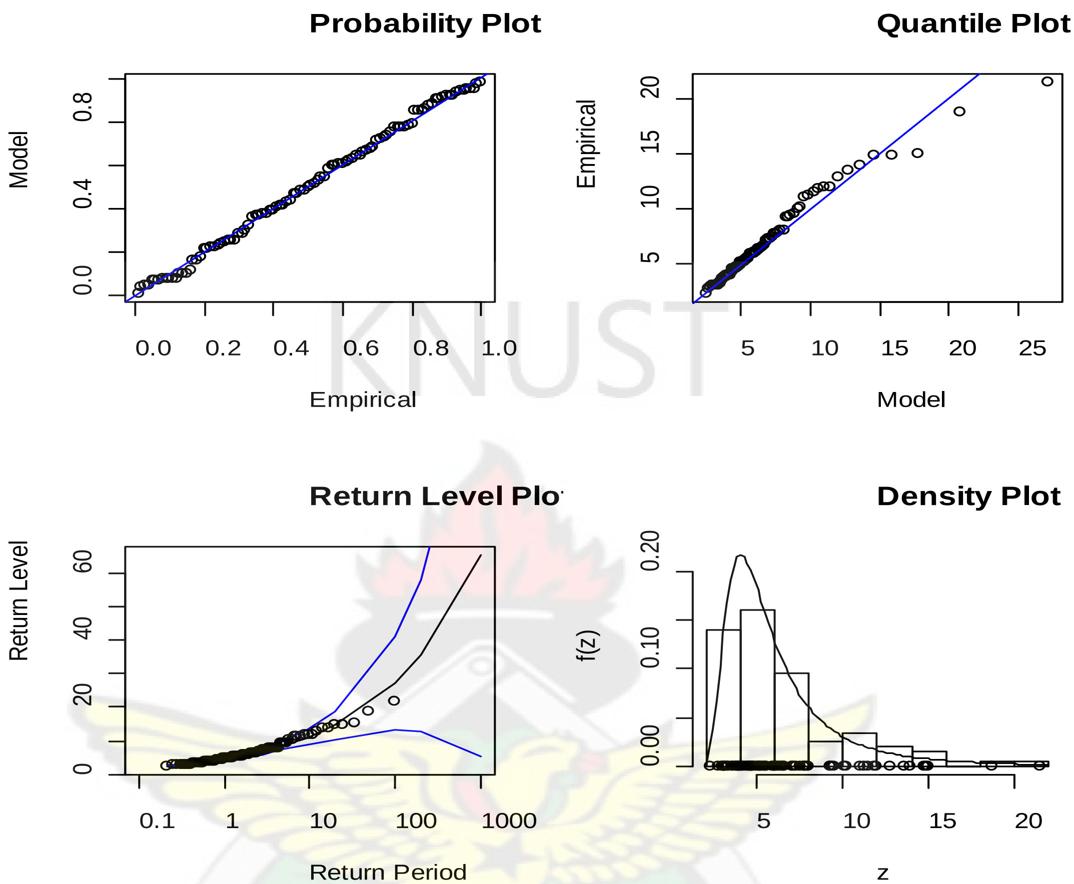

Figure 5.6: Diagnostic plots for GEV fit (Gains)

Figure 5.6 displays the diagnostic plots for GEV fit to block maximum data of sample size 89. The top two plots do not deviate much from the straight line and the histogram match up with the curve. The return level plot gives an idea of the expected return level for each return period. The maximum likelihood estimates (MLE) for the parameters of the fit shown in Figure 4.2 were found to be $\hat{\mu}$ = 4.4489(0.0789),

$\hat{\sigma}$ =1.8113 (0.2130) and $\hat{\gamma}$= 0.2940(0.1777) with a negative log-likelihood value for this model of approximately 208.5651 and standard errors in parentheses. Similarly, for the left tail the maximum likelihood.

5.3.2 Estimation of Shape Parameters

We mainly concentrate on the estimation of the shape parameter γ due to its (already stressed) importance in this section.

Figure 5.7 shows Hill-plot of the oil pricing data with a 0.95 confidence interval. In this plot, estimated parameters γ are plotted against upper order statistics (number of excedances). Alternatively, estimated parameters may be plotted against different thresholds. A threshold is selected from the plot where the shape parameter γ is fairly stable. The number of upper order statistics or thresholds can be restricted to investigate the stable part of the Hill-plot. The function hill Plot investigates the shape parameter and plots the Hill estimate of the tail index of heavy-tailed data, or of an associated quantile estimate. This plot is usually calculated from the alpha perspective. For a generalized Pareto analysis of heavy-tailed data using the Generalised Pareto Distribution (GPD) fit function, it helps to plot the Hill estimates for γ (the shape parameter).

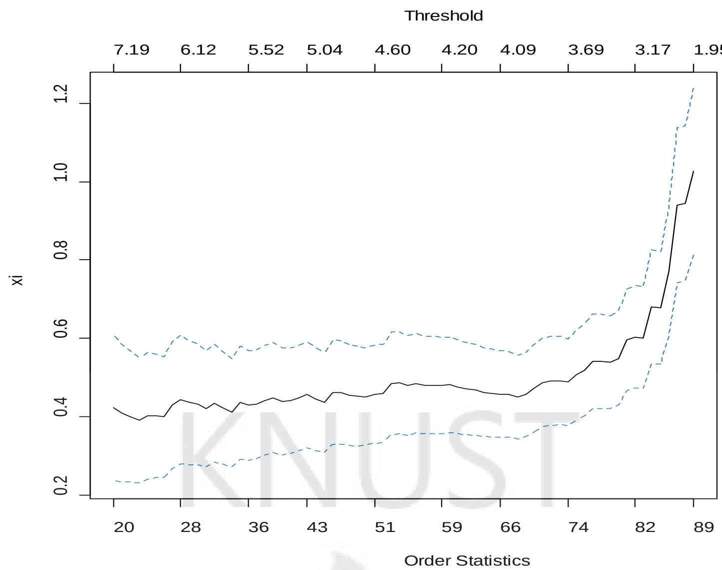

Figure 5.7: Hill plot of oil price data

From figure 5.8 below, the Pickands estimate of the shape parameter is, $\hat{\gamma} = 0.3131$

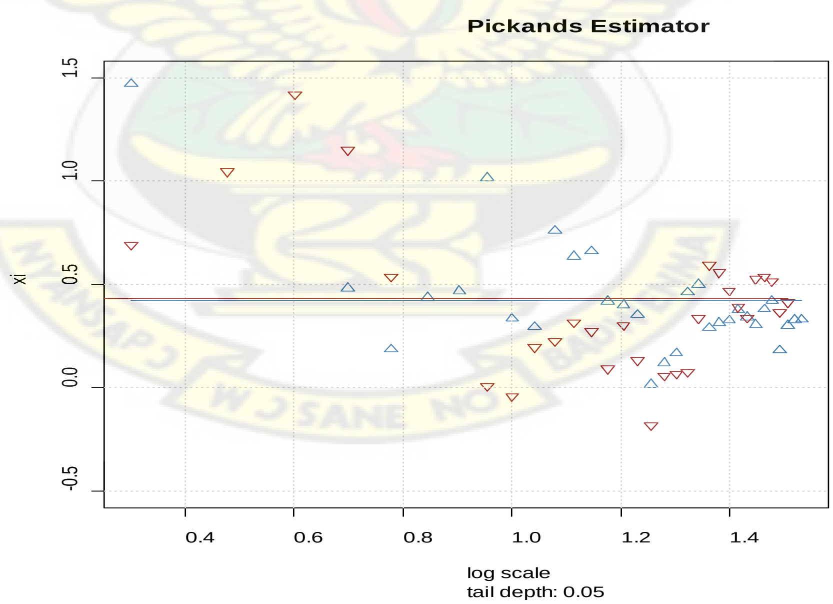

Figure 5.8: Plot of Pickands Estimator

From Figure 5.9 below, the hill estimate of the shape parameter is, $\hat{\gamma} = 0.2861$

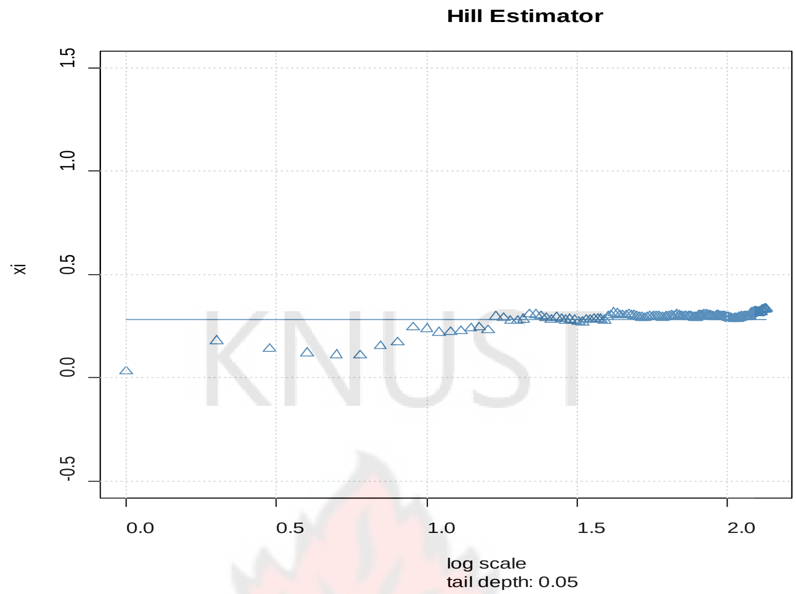

Figure 5.9: Plot of Hill Estimator

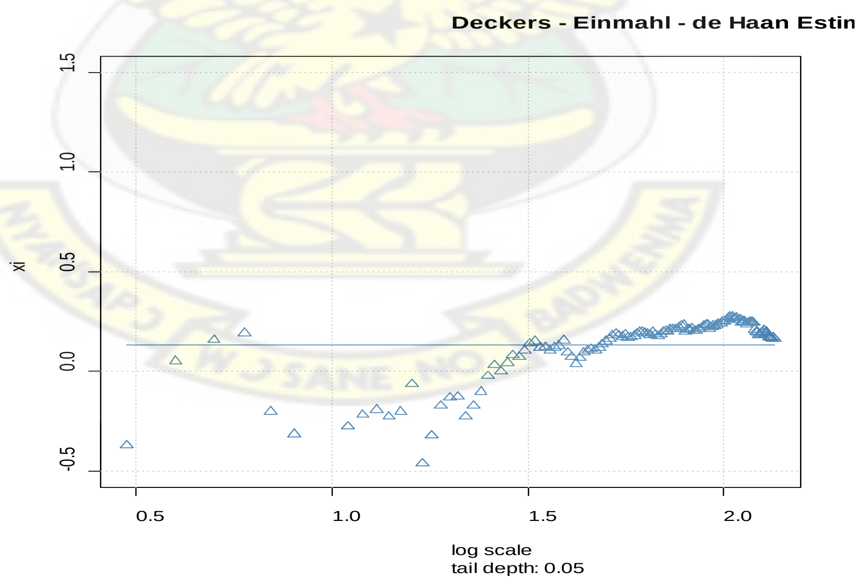

Figure 5.10: Plot of Decker – Einmahl – de Haan Estimator with shape parameter is, $\hat{\gamma} = \mathbf{0.1192}$

5.3.3 Return Level Plot for Positive and Negative Returns using GEV Fit

The return level plot shown in Figure 5.11 shows the return level, and an estimated 95% confidence interval. The return level is the level we expect to be exceeded only once every k (time period): This is level is (in this case percentage) that is expected to be exceeded, on average, once every m time points (in this case months). The return period is the amount of time expected to wait for the exceedance of a particular return level.

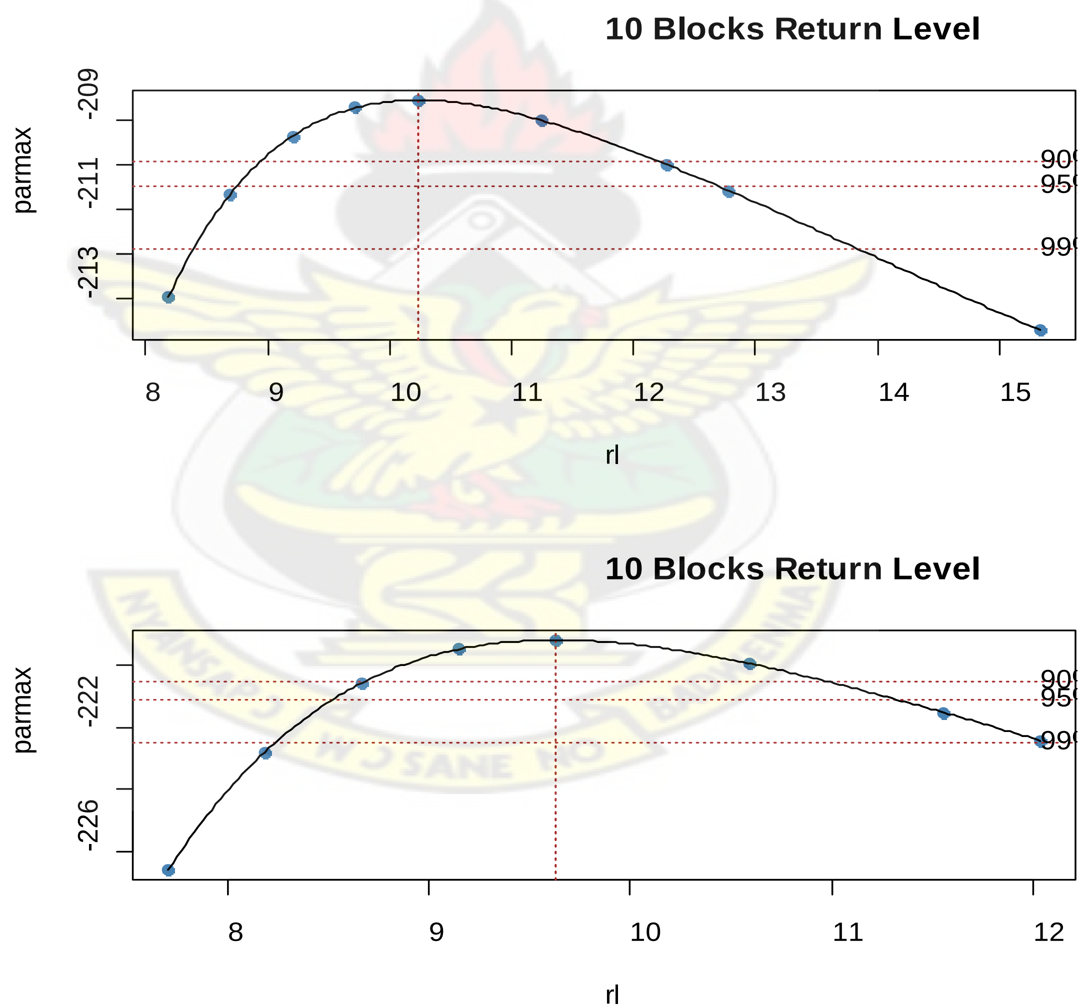

Figure 5.11 : Profile likelihood plots for the one-month return level and 95% confidence (top positive returns) and (bottom negative returns)

In figure 5.11, the top graph is the profile likelihood plots for the ten-month return level and 95% confidence interval for return value 10.227 (8.973 , 12.17525) of the right tail with confidence interval in parentheses. The left tail (shown bottom) gave return level value 9.63 (8.6838, 10.9704) with confidence interval in parentheses. A return value of 10.227 means that the maximum gain (positive returns) observed during a period of 12 mouths will exceeds 10.23% (8.97%, 12.17%) in one out of one year on an average with 95% confidence interval in parenthesis. Similarly, for the left tail, a return value of 9.63 means that the maximum loss observed during a period of 12 months will exceeds 9.23% (8.68% ,10.97%) in one out of one on an average with 95% confidence interval in parenthesis.

In table 5.2 below, the point estimate of the shape parameter is greater than zero an indication of fat-tailness of the distribution. The confidence interval also does not include zero confirming the heavy-tail distribution at the right tail. The scale parameter does not include one(1) an indication of extremal events.

Table 5.2: Generalised Extreme Value Parameter Estimates for Gains

PARAMETER	POINT ESTIMATE	CONFIDENCE INTERVAL
Shape parameter	0.2940	[0.1394 0.4489]
Location parameter	4.4493	[4.0313 4.8662]
Scale parameter	1.8113	[1.4943 2.1949]

In table 5.3 below, the point estimate of the shape parameter is greater than zero an indication of fat-tailness of the distribution. The confidence interval includes zero confirming the thin -tail distribution at the left tail. The scale parameter does not include one(1) an indication of extremal events.

Table 5.3: Generalised Extreme Value Parameter Estimates for Losses

PARAMETER	POINT ESTIMATE	CONFIDENCE INTERVAL
Shape parameter	0.1244	[-0.0125 0.2613]
Location parameter	4.4853	[4.0413 4.9292]
Scale parameter	1.9802	[1.6727 2.3443]

5.3.4 The threshold selection

Choosing some suitable threshold is critical in order to adopt the POT method to model the tails of the distribution of daily returns. So far, no automatic algorithm with satisfactory performance for the selection of the threshold u is available. The issue of determining the fraction of data belonging to the tail is treated by Danielsson et al. (2001), Danielsson and de Vries (1997) and Dupuis (1998) among others. However, these references do not provide a clear answer to the question of which method should be used. It is desired to find a threshold that is high enough that the underlying theoretical development is valid, but low enough that there is sufficient data with which to make an accurate fit. That is, selection of a threshold that is too low will give biased parameter estimates, but a threshold that is too high will result in large variance of the parameter estimates. Some useful descriptive tools for threshold selection includes, the mean excess, or mean residual life, plot and another method involving the fitting of data to a GPD several times using a range of different thresholds. The ME (mean excess) plot is helpful in detecting graphically the quantile above which the Pareto's relationship is valid.

Chapter three details the empirical mean excess plot is approximately linear in the threshold u given that the underlying distribution of sample data is a GPD. More specifically, the ME plot of the data can be used to distinguish between light- and heavy-tailed models. The plot of a heavy-tailed distribution shows an upward trend, a medium tail shows a horizontal line, and the plot is downward-sloped for light-tailed data. A common ground in our sample data is that both the ME plots of positive and negative returns have an upward-trend part followed by an irregular portion in the far end. The initial and small part of the gain plot is downward-sloping until $u \approx 0$ to 4.5, followed by a roughly upward-sloping, where upon it varies sharply. The case of losses shows an approximate linearity with slightly upward trend in the threshold from $u \approx 1$ to $u \approx 5$. Therefore, there is some evidence to choose thresholds from 4.5 to 6 for the right tail and from 4.5 to 5.5, for the left tail based on the criterion of linearity in the ME plots shown in Figure 5.12 below.

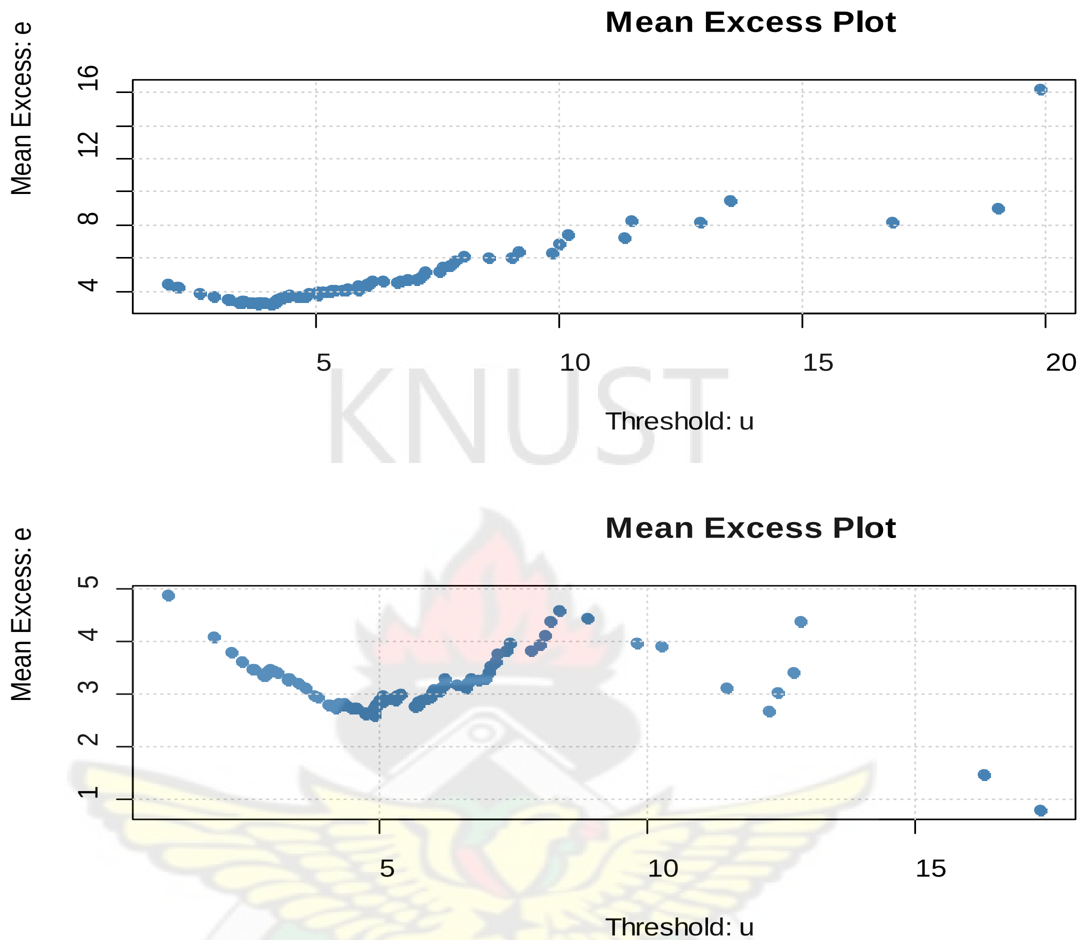

Figure 5.12: Mean excess plot for positive returns (top) and negative returns (Bottom)

The empirical distribution function is given in figure 5.13 on a double logarithmic scale. This scale is used to highlight the tail region. Here an exact Pareto distribution corresponds to linear plot.

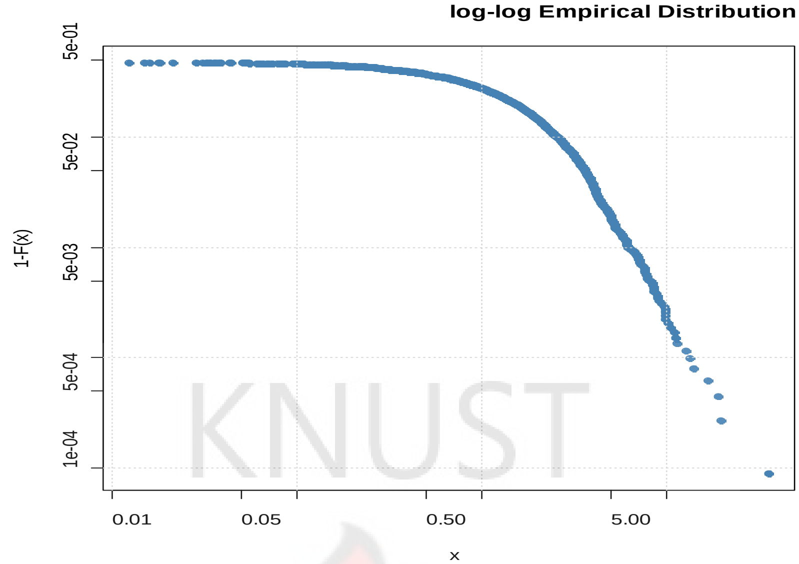

Figure 5.13: log-log Empirical distribution

The sample mean excess function, which is an estimate of the mean excess function $e(u)$ defined in equation 2.23, should be linear. This property can be used as a criterion for the selection of u. Figure 5.14 shows the sample mean excess plots corresponding to the oil pricing data. From a closer inspection of the plots we suggest the values u = 5.0 for the threshold of the right tail. This value is located at the beginning of a portion of the sample mean excess plot that is roughly linear, leaving respectively 110 observations in the tails.

A straight line with positive slope above a given threshold u = 5.0 is a sign of the GPD

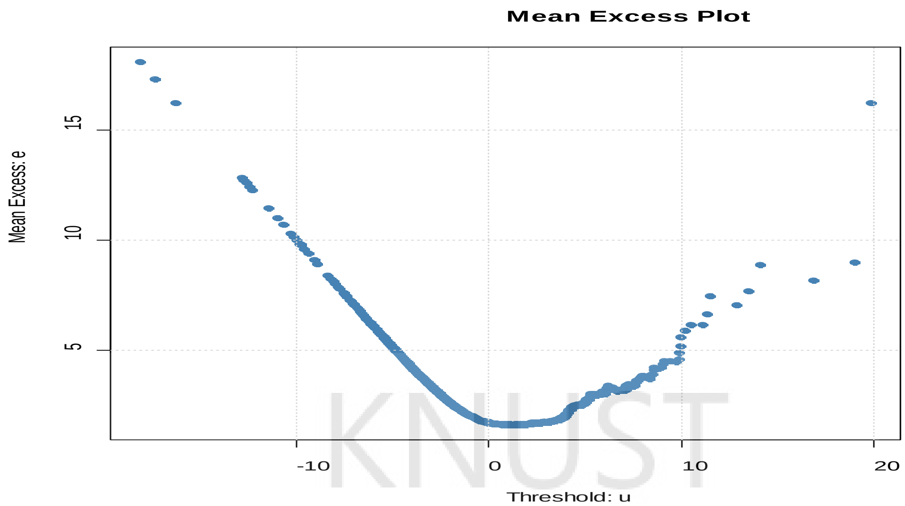

Figure 5.14: Mean excess plot

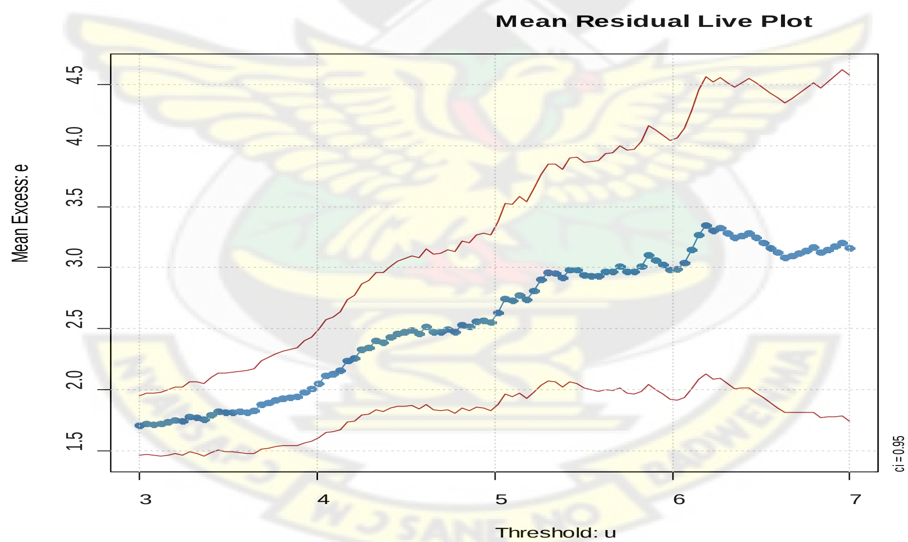

Figure 5.15: Mean Residual Plot

Figure 5.1 shows the mean excess plot, with confidence bands, for the daily oil price data, based on all exceedances over thresholds u = 5.0. The continuous dotted line (middle) is the estimated theoretical mean excess assuming the GPD at threshold u, the dotted lines above and below is the estimated confidence bands. The plot lie nearly everywhere inside the confidence bands, but plot appears to show more systematic departure from a straight line, adding to the evidence that threshold 5.0 is a better.

As a means of threshold selection, the ME plots may be difficult to interpret, and the results can be treated as preliminary conclusions. A further step is to apply the GPD fitting and look for stability of shape parameter estimates. We fit the exceedances of daily returns beyond the associated threshold in each tail to the GPD. Because the maximum likelihood estimator of the shape parameter is asymptotically normal, we can calculate the associated approximate standard errors and construct confidence intervals for this parameter. The plots of the shape parameter estimates against different threshold levels are shown in Figure 5.16 for gains and figure 5.17 for losses. The upper and lower dashed lines constitute confidence intervals at an approximate 95% level.

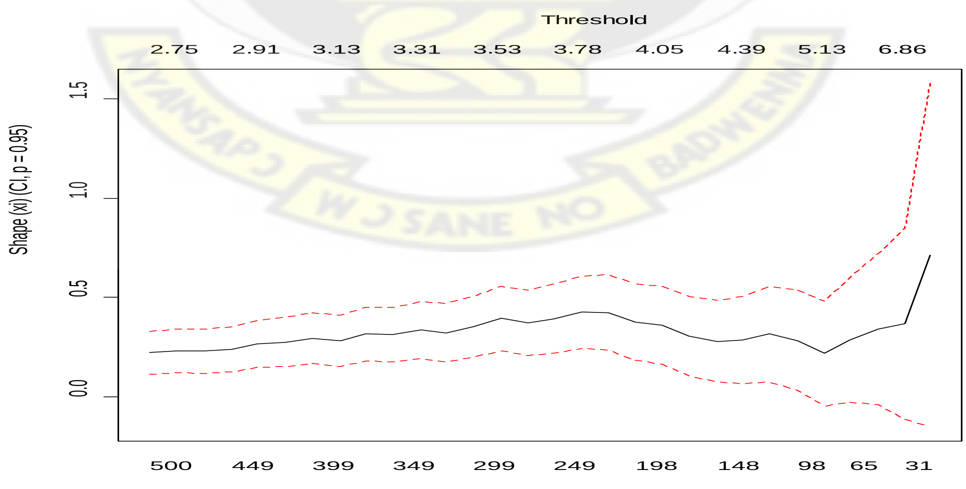

Figure 5.16: Estimates for shape parameter (positive returns)

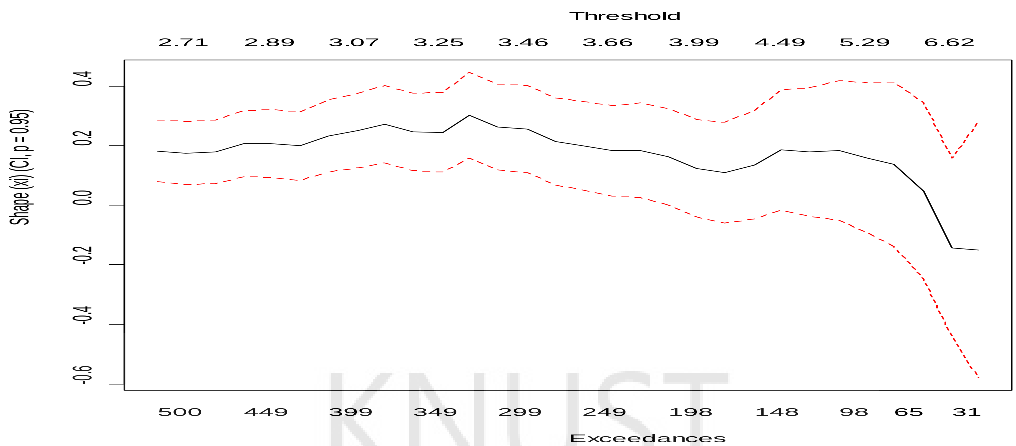

Figure 5.17: Estimates for shape parameter (negative returns)

The estimated shape parameter $\hat{\gamma}$ and scale parameter $\hat{\sigma}$ as well as their associated standard errors under different thresholds for both tails are listed in Table 5.4

Table 5.4: Maximum likelihood parameter estimation under different thresholds for both returns

	Parameter estimates for positive returns			**Parameter estimates for negative returns**		
	u= 4.8827	u = 5.1293	u=5.4067	u= 4.7641	u = 4.9889	u=5.2912
$\hat{\gamma}$	0.3151	0.2815	0.2164	0.1867	0.1807	0.1853
(s.e.)	0.1226	0.1294	0.1352	(0.1033)	(0.1101)	(0.1197)

In order to apply EVT, the threshold should be sufficiently large so that only the tail of the distribution can be analyzed. When the threshold is close to zero, there are too many observations included. Practical experience suggests it is reasonable including observations up to roughly one-fifth of the total number of observations for both positive returns and negative returns. This is somewhat arbitrary, but provides a reasonable compromise. Combining this restriction and results from the ME plots and the shape parameter plots, we choose the range of the threshold from 4.5 to 5.5 for positive returns and from 4.6 to 5.4 for negative returns.

5.3.5 Generalized Pareto Distribution(GPD)

Sometimes using only block maximum can be wasteful if it ignores much of the data. It is often more useful to look at exceedances over a given threshold instead of simply the maximum (or minimum) of the data. EVT provides for fitting data to GPD models as well as threshold selection.

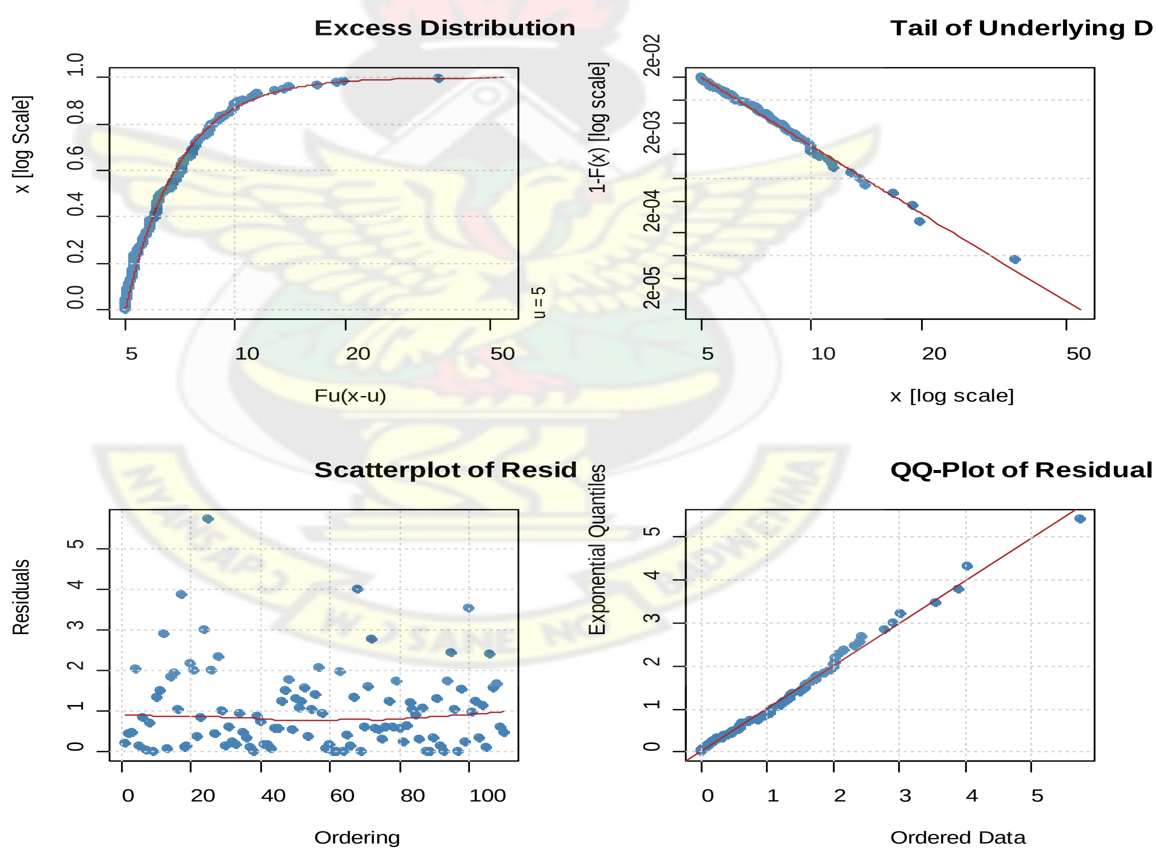

Figure 5.18: Diagnostic plots for the GPD fit of the daily oil price returns using a threshold of 5.0(Gains)

In Figure 5.18 (top left) the estimated GPD model for the excess oil pricing is plotted as a curve while the actual daily oil pricing thresholds are shown in circles (in logarithmic scale). With threshold of u = 5.0, the estimated parameters are $\hat{\gamma} = 0.351$ and $\hat{\sigma} = 1.671$. The number of exceedances is 110 from oil pricing data sets. The top right is a plot of the empirical distribution of the data vector on logarithmic scaled axes. The bottom left of figure 5.18 shows scatter plot of the residuals indicating time of exceedance. The idea here is to observe a possible time trend in the observations. To aid in judging this, a simple fitted curve (using the function of fExtremes package in R) is superimposed and there is evidence of a systematic trend. The solid line is the smooth of scattered residuals obtained by a spine method. The QQ-plot of the residuals is shown in the bottom right. The plot of the fitted model does not deviate much from the straight line; it is based on the GPD fitted to all exceedances above threshold 5.0. The main point here is that although there are reasons for treating the largest observations as outliers, they are not in fact very far from the straight line in other words; the data are in fact consistent with the fitted GPD, which in this case is extremely fat-tailed. If this interpretation is accepted, there is no reason to treat those observations as outliers.

In order to start this procedure, a threshold value u = 5.0 and 4.5 for the right and left tail respectively, has to be chosen as estimate depend on the excesses over this threshold. The estimates of the key shape parameter γ and their corresponding standard deviations for both tails as a function of u (alternatively, as a function of the number of order statistics used) are given in Figure 5.19 and Figure 5.20.

Table 5.5 : GPD estimate of positive tail- index

PARAMETER	GPD ESTIMATE	STANDARD DEVIATION
Shape parameter	0.351	0.1386
Scale parameter	1.683	0.2775

Table 5.6: GPD estimate of negative tail- index

PARAMETER	GPD ESTIMATE	STANDARD DEVIATION
Shape parameter	0.118	0.0941
Scale parameter	1.929	0.2413

Table 5.7 : Risk Measures of right-tail distribution (GPD-fit)

PROBABILITY	VALUE-AT-RISK(VaR)	EXPECTED SHORTFALL(ES)
0.9500	4.694	7.122
0.9900	8.1023	12.374
0.9950	10.2776	15.725
0.9990	17.9258	27.509
0.9995	22.8069	35.031
0.9999	39.9687	61.474

Table 5.8 : Risk Measures of left-tail distribution (GPD-fit)

PROBABILITY	VALUE-AT-RISK(VaR)	EXPECTED SHORTFALL(ES)
0.9500	4.5726	6.771
0.9900	8.0079	10.668
0.9950	9.7018	12.589
0.9990	14.2158	17.712
0.9995	16.4417	20.236
0.9999	22.3735	26.966

The results in Table 5.7 indicate that, with probability 0.01 that is 99% confidence interval, the tomorrow's gain will not exceed the value 8.10% and that the corresponding expected gain, that is the average gain in situations where the gains exceed 8.10%, is 12.37%. Similarly, results in Table 5.8 indicate that, with probability 0.01 that is 99% confidence interval, the tomorrow's gain on a short position will loss the value 8.00% and that the corresponding expected loss, that is the average loss in situations where the losses exceed 8.10%, is 10.67%.

For higher quantiles that is 99.99% confidence interval, in table 5.7 the tomorrow's gain on a long position will exceed the value 39.96% and that the corresponding expected gain, that is the average gain in situations where the gains exceed 39.96%, is 61.47%. Similarly, results in table 5.8 indicate that, with 99.99% confidence interval, the tomorrow's loss on a short position will lose the value 22.37% and that the corresponding expected loss, that is the average loss in situations where the losses exceed 22.37%, is 26.96%. It is interesting to note that the upper bound of the confidence interval for the shape parameter is such that the first order moment is finite (1/0.68 > 1). This guarantees that the estimated expected shortfall, which is a conditional first moment, exists for both tails as shown in figure 5.19

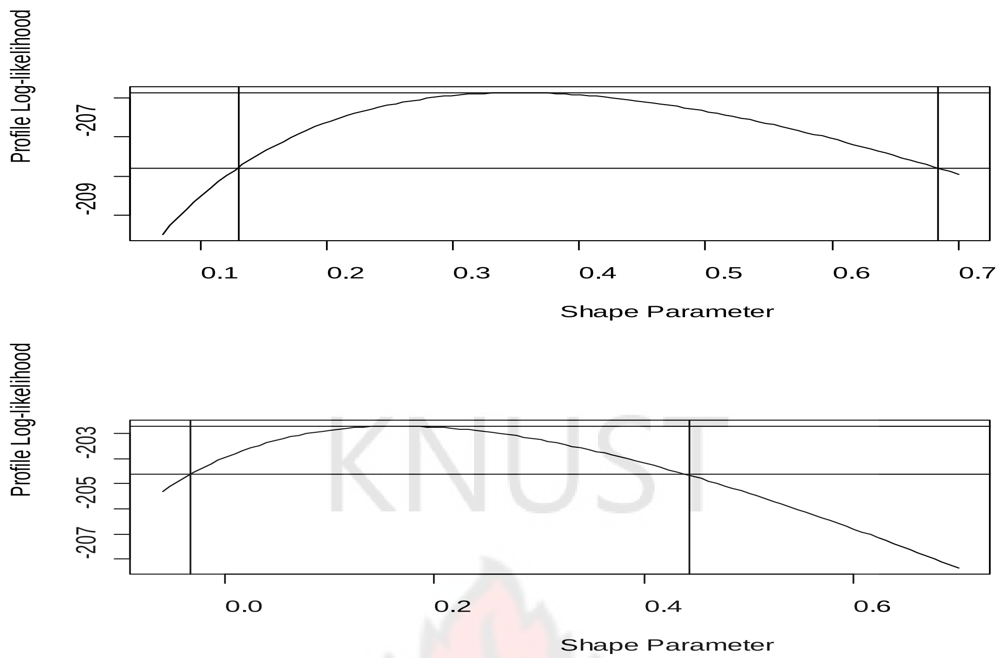

Figure 5.19: The profile log-likelihood and confidence intervals for shape parameter. Gains (top) and losses(bottom)

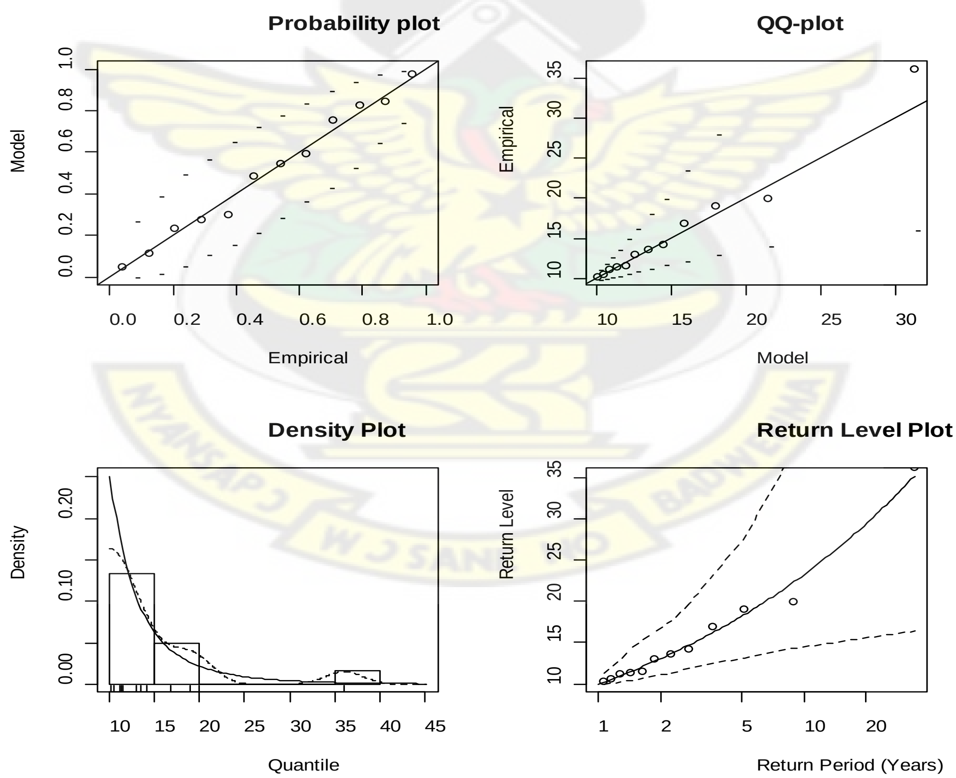

Figure 5.20: Graphical diagnostic for a fitted POT model (univariate case)

Figure 5.20 above shows graphic diagnostics for the fitted model. It can be seen that the fitted model with maximum likelihood estimates seems to be appropriate.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 INTRODUCTION

The study was carried out to explore the application of EVT in relation to world's oil pricing of Brent crude. The main objective of this study was to find out the long-term effect on investing in Brent Crude oil industry in Ghana. This chapter presents a summary of the various analyses performed, the findings obtained and what they imply with regard to the research objectives. The chapter is concluded with recommendations including the further work to be done.

6.2 SUMMARY OF FINDINGS

In this study, Value-at-Risk measures the best/worst case scenario on the market value of the Brent crude oil over one trade day. We first consider the cases of point estimates under the thresholds for both tails (u=5 for the right tail and 4.5 for the left tail) with statistics shown in Table 4.6 and 4.7. For example, we calculate VaR as u = 5.0 at the 99th percentile for the right tail. That is, given usual conditions, we expect a daily change in the value of crude oil market would not increase by more than 8.10%. In other words, the market value, with a probability of 1%, would be expected to gain by $81,000 or more if we have an investment of $1 million in that market. On the other hand, VaR is estimated as u=4.5 at the 1st percentile for the left tail. This implies that, for the lowest 1% negative daily returns, the worst daily loss in the market value may exceed 8.00% in expectation. Put differently, if we invest $1 million in crude oil, we are 99% confident that our daily loss at worst will not exceed

$80,000 during one trade day. Similarly, at a lower quantile of 95-level, the estimated VaR is 4.69% for gains and 4.57% for losses. We can state that, with 95% confidence, the expected market value of crude oil would not gain by more than 7.12% for the best case scenario or lose more than 6.77% for the worst case scenario within one-day duration. Under the higher threshold for both tails (5.5 for the right tail and 5.8 for the left tail), the estimates of VaR are very close to their corresponding values under the lower threshold, and the estimates may or may not be larger than that under a lower threshold. These estimates can be used in different ways. For example, the VaR results in table 4.7 and 4.8 imply that, given the same amount of investment the possibility of loss for an investment in the oil market is relatively lower than the possibility of gain. In addition, the difference between the VaR and ES for the positive returns is bigger than that for the negative returns. This means that the expected gain over the VaR under the situation of gain is more than the expected loss over the VaR under the situation of loss.

The risk measures could also help oil producers to forecast the required number of barrel of oil to produce.

Researchers have conducted sound studies on the tail distribution modelling by applying some methods of univariate EVT, especially in the financial field. An important argument is that the EVT approach well captures the features of the innovation distribution and can provide more accurate estimates of risk measures compared with other approaches (for example, McNeil, 1997; Gencay *et al*., 2003; Fernandez, 2005), and one can obtain better estimates with the application of the GPD fitting of the excess distribution based on threshold models (for example, Coles, 2001; Gilli and Këllezi, 2006; Marimoutou et al., 2006). This confirms our belief of choosing the POT method to apply the EVT. Overall, the assessment of our results shows that the point and estimations are stable and reliable, implying that this approach of modelling extreme values can be used to further application of extreme events.

6.3 CONCLUSION

The high volatility of prices in oil markets requires the implementation of effective risk management. EVT is a powerful tool to estimate the effects of extreme events in risky markets based on sound statistical methodology. This study exhibits how EVT can be used to model tail-related risk measures such as Value-at-Risk and Expected Shortfall by applying it to the daily returns of world's crude oil prices market. Our application captures the heavy-tailed behaviour in daily returns and the asymmetric characteristics in distributions, suggesting us to treat positive and negative returns separately. An unconditional approach is favoured as no evidence indicates the existence of conditional heteroskedasticity in our sample data. In the context of applying EVT, the peak over threshold method provides a simple and effective means to choose thresholds and estimate parameters. By assessing empirical excess distribution functions and survival functions with associated theoretical distribution simulations, we find the goodness of fit in tail modelling. Furthermore, as we increase the threshold, the fit is gets less precise for both gains and losses; at an either lower or higher threshold level, positive daily return series fits a GPD slightly better than negative one does.

The EVT-based Value-at-Risk approach adopted in this study provides quantitative information for analysing the extent of potential extreme risks in oil markets, particularly the crude oil markets.

Our conclusion is that EVT can be useful for assessing the size of extreme events. From a practical point of view this problem can be approached in different ways, depending on data availability and frequency, the desired time horizon and the level of complexity one is willing to introduce in the model. One can choose to use a conditional or an unconditional approach, the Block Maxima Method (BMM) or the Peaks over Threshold (POT) method.

In our application, the POT method proved superior as it better exploits the information in the data obtain. Being interested in long term behaviour rather than in short term forecasting, we favoured an unconditional approach.

6.4 RECOMMENDATIONS

- Interested organizations and corporations could employ this technique as one of the means of risk management. For those who invest in the Ghana crude oil market, our estimates of VaR and ES provide quantitative indicators for their investment decisions.
- Government should employ all necessary legislature and business plans to speed up the exploitation of oil as there is an indication of a substantial rise in oil price for the next five (5) month.
- Government is encouraged to speed up the passage of freedom to information bill to enable student of research to have access and usage of public data.

- Students of research are encouraged to look into other areas of EVT applications, for instance, extreme climate (rainfall) in Ghana.
- The Department of Mathematics should established good faith with companies to unable them used their data.
- Statistical Software packages should be taught in relation to the courses undertaken at the Postgraduate level in the Mathematics department.

REFERENCES

Acerbi, C., Nordio, C., and Sirtori, C., (2001): Expected shortfall as a tool for financial risk management, Working paper, http://www.gloriamundi.org.

Acerbi, C. and Tasche, D., (2002): On the coherence of expected shortfall, *Journal of Banking and Finance*, 26, 1487-1503.

Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1997), `Thinking coherently', *RISK* 10(11), 68-71.

Barão, M.I. and Tawn, J.A. (1999). Extremal Analysis of Short Series with Outliers:Sea-Levels and Athletic Records. *Applied Statistics* 48 (4), 469-487.

Balkema, A.A. and L. de Haan (1974), Residual life time at great age, *Annals of Probability*, 2, 792-804.

Beirlant, J., Teugels, J. & Vynckier, P. (1996), *Practical analysis of extreme values*, Leuven University Press, Leuven.

Borio, C., Furline, C. and Lowe, P. (2001) Procyclicality of the financial system and financial stability; issues and policy options. Working Paper, BIS, Basel.

Bortkiewicz, L., von (1922). Variationsbreite und mittlerer Fehler. *Sitzungsberichte derBerliner Mathematischen Gesellschaft* 21, 3-11. (in Johnson et al. 1995)

Coles, S. (2001), An Introduction to Statistical Modelling of Extreme Values, Springer-Verlag, London.

Coles, S. and Dixon, M. (1999) Likelihood-Based Inference for Extreme Value Models. Extremes 2(1):5–23.

Coles, S.G. and Tawn, J.A. (1996). Modelling Extremes of the Areal Rainfall Process. *Journal of the Royal Statistical Society B* 58(2), 329-347.

Cooley D., Nychka D., Naveau P., (2005a): Bayesian Spatial Modelling of Extreme Precipitation Return Levels, (submitted).

Cooley D., Naveau P., and Jomelli, V., (2005b): A Bayesian Hierarchical Extreme Value Model for Lichenometry. Environmetrics, (in press).

Crockett, A. (2000) Marrying the micro- and macro-prudential dimensions of financial stability. Working Paper, BIS, Basel.

Dacorogna, M. M., MÄuller, U. A., Pictet, O. V., and de Vries, C. G. (1995). The distribution of extremal foreign exchange rate returns in extremely large data sets. Preprint, O&A Research Group.

Dalal, S.R., Fowlkes, E.B. and Hoadley, B. (1989). Risk Analysis of the Space Shuttle:Pre-Challenger Prediction of Failure. *Journal of the American Statistical* Association 84, 945-957.

Danielsson, J. & de Vries, C. (1997), `Tail index and quantile estimation with very high frequency data', *Journal of Empirical Finance* 4, 241- 257.

Danielsson, J., Hartmann, P. & de Vries, C. (1998), `The cost of conservatism', *RISK* 11(1), 101-103.

Danielsson, J., de Vries, C., (2000). Value-at-Risk and extreme returns. Annales d'Economie et de Statistique 60, 239-270.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society B* 52 (3), 393-442.

De Brandt, O. and Hartmann, P. (2000) Systemic risk: a survey. Working Paper No. 35, European Central Bank, Frankfurt.

Deheuvels, P., de Haan, L., Peng, L. and Pereira, T.T. (1997). NEPTUNE T400: EUR- 09, Comparison of Extreme Value Index Estimators.

Dekkers, A.L.M. and de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation. *The Annals of Statistics* 17 (4), 1795 - 1832.

Diebold, F., Schuermann, T. & Stroughair, J. (1999), Pitfalls and opportunities in the use of EVT in risk management, *in* `Advances in Computational Finance', Kluwer Academic Publishers, Amsterdam. To appear.

Dodd, E.L. (1923). The Greatest and Least Variate under General Laws of Error. *Transactions of the American Mathematical Society* 25, 525-539. (in Johnson et al.1995)

DuMouchel, W.H. (1983). Estimating the Stable Index α in Order to Measure Tail Thickness: a Critique. *The Annals of Statistics* 11, 1019-1031.

Dupuis, D. J. (1998). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(3):251- 261.

Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer, Berlin.

Embrechts, P., McNeil, A. & Straumann, D. (1999), `Correlation and dependency in risk management: properties and pitfalls', preprint, ETH Zurich.

Embrechts, P., Resnick, S. & Samorodnitsky, G. (1998), `Living on the edge', *RISK Magazine* 11(1), 96-100.

Fernandez, V., (2005). Risk management under extreme events. International Review of financial analysis 14, 113-148

Financial Toolbox User's Guide COPYRIGHT (1995–2000) by The MathWorks, Inc.

Fisher, R. A., Tippett, L. H. C., (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 24, 180-190.

Frechet, M. (1927). Sur la loi de Probabilité de l' Écart Maximum. *Annales de la Société Polonaise de Mathématique, Cracovie* 6, 93-116. (in Johnson et al. 1995)

Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York. (in Kinnison,1985)

Haan, L. de and de Ronde, J. (1998). Sea and Wind: Multivariate Extremes at Work. *Extremes* 1 (1), 7-45.

Hartmann, P., Straetmans, S. and de Vries, C.G. (2001) Asset market linkages in crisis periods. Working Paper, No. 71, European Central Bank, Frankfurt.

Haeusler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation. *The Annals of Statistics* 13 (2), 743-756.

Gencay, R., Selcuk, F., Ulugülyagci, A., (2003). High volatility, thick tails and EVT in value-at-risk estimation. Insurance: Mathematics and Economics 33, 337-356.

Gencay, R., Selcuk, F., and UlugÄulya·gci, A. (2003a). EVIM: a software package for extreme value analysis in Matlab. *Studies in Nonlinear Dynamics and Econometrics*, 5:213-239.

Gnedenko, B. V., (1943). Sur la distribution limite du terme d'une série aléatoire. Annals of Mathematics 44, 423-453.

Grimshaw, S. (1993). Computing the Maximum Likelihood Estimates for the Generalized Pareto Distribution to Data. *Technometrics*, 35(2):185-191.

Harter, H.L. (1978). A Bibliography of EVT. *International Statistical Review* 46, 279-306.

Hartmann, P., Straetmans, S. and de Vries, C.G. (2001) Asset market linkages in crisis periods. Working Paper, No. 71, European Central Bank, Frankfurt.

Holton, G. A., (2002). History of value-at-risk: 1922-1988. Contingency Analysis Working Paper.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalised Pareto distribution. *Technometrics*, 29:339-349.

Hsing, T. (1991). On Tail Index Estimation Using Dependent Data. *Annals of Statistics*, 19 (3), 1547-1569.

Hull, J. & White, A. (1998), `Incorporating volatility updating into the historical simulation method for value at risk', *Journal of Risk* **1**(1).

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (minimum) values of meteorological events. *Quarterly Journal of the Royal*

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Jondeau, E. and Rockinger, M. (1999). The tail behavior of stock returns: Emerging versus mature markets. Mimeo, HEC and Banque de France.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, *Volume 2*, 2nd Edition. Wiley, New York.

Joriot, P. (1997) *Value at Risk: the New Benchmark for Controlling Market Risk.* Irwin, Chicago.

Kagan, Y.Y. (1997). Earthquake Size Distribution and Earthquake Insurance. *Communication in Statistics – Stochastic Models* 13 (4), 775-797.

Kinnison, R.R.(1985). *Applied Extreme Value Statistics*. MacMillan, New York.

Koedijk, K. G., Schafgans, M., and de Vries, C.(1990). The Tail Index of Exchange Rate Returns. *Journal of International Economics*, 29:93-108.

Kratz, M. and Resnick, S.I. (1996). The QQ Estimator and Heavy Tails.

Communication in Statistics – Stochastic Models 12 (4), 699-724.

Krehbiel, T., Adkins, L.C., (2005). Price risk in the NYMEX energy complex: an extreme value approach**.** Journal of Futures Markets 25, 309-337.

Kuan, C. H. and Webber, N. (1998). Valuing Interest Rate Derivatives Consistent with a Volatility Smile. Working Paper, University of Warwick.

Küchenhoff, H. and Thamerus, M. (1996). Extreme Value Analysis of Munich Air Pollution Data. *Environmental and Ecological Statistics* 3(2), 127-141.

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983), *Extremes and Related Properties of Random Sequences and Series.* Springer Verlag, New York.

Longin, F. (2001) Stock market crashes: Some quantitative results based on EVT. *Derivatives Use, Trading and Regulation*, 7, 197–205.

Loretan, M. and Phillips, P. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1(2):211-248.

Madsen, H., Rasmussen, P.F., and Rosbjerg, D., (1997): Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events, 1, At-site modelling, Water Resour. Res. 33(4):747–758.

Martins, E.S. and Stedinger, J.R., (2000): Generalized maximum-likelihood generalized extreme value quantile estimators for hydrologic data, Water Resour. Res. 36(3):737–744.

Martins, E.S. and Stedinger, J.R., (2001): Generalized maximum likelihood Pareto-Poisson estimators for partial duration series, Water Resour. Res. 37(10):2551–2557.

Marimoutou, V., Raggad, B., Trabelsi, A., (2006). EVT and value at risk: Application to oil market, GREQAM Working Paper 2006-38.

McNeil, A. (1997), `Estimating the tails of loss severity distributions using EVT', *ASTIN Bulletin* 27, 117-137.

McNeil, A. (1998), `History repeating', *Risk* 11(1), 99.

McNeil, A. & Frey, R. (1998), `Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', preprint, ETH Zurich.

McNeil, A.J. and Saladin, T. (1997). The Peaks Over Thresholds Method for Estimating High Quantiles of Loss Distributions. *Proceedings of 28th International ASTIN Colloqium.*

McNeil, A. J., (1999). Extreme value theory for risk managers. British Bankers' Association, Internal Modelling and CAD II: Qualifying and Quantifying Risk within a Financial Institution 93-113. RISK Books, London.

McNeil, A. J., Frey, R., (2000). Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. Journal of Empirical Finance 7, 271-300.

Mikosch, T. (1997). Heavy-Tailed Modelling in Insurance. *Communication in Statistics– Stochastic Models* 13 (4), 799-815.

Mises, R., von (1923). Über die Variations breite einer Beobachtunggsreihe. *Sitzungsberichte der Berliner Mathematischen Gesellschaft* 22, 3-8. (in Johnson et al.1995)

Nandha, M., Faff, R., (2007). Does oil move equity prices? A global view. Energy Economics.

Neftci, S. N. (2000). Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives*, pages 23-37.

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer, New York.

Paper. Jarque, C. M., Bera, A. K., (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6, 255–259

Pickands, J. (1975). Statistical Inference Using Extreme order statistics. *The Annals of Statistics* 3 (1), 119-131.

Reiss, R. & Thomas, M.(1997), *Statistical Analysis of Extreme Values*, Birkhauser, Basel.

Resnick, S. (1987), *Extreme Values, Point Processes, and Regular Variation.* Springer Verlag, New York.

Resnick, S. (1997a). Heavy Tailed Modelling and Teletraffic Data. *The Annals of Statistics* 25 (5), 1805-1869.

Resnick, S. and Stărică, C. (1995). Consistency of Hill's Estimator for Dependent Data. *Journal of Applied Probability* 32, 139-167.

Robert B, Magdalena N. B, 1996, Probability and Statistical Inference, John Wiley and Sons, Inc. New York.

Robert V. Hogg, Joseph W. Mckean & Allen T. Craig, (2005).Introduction to Mathematical Statistics,6th Edition

Robinson, M. E. and Tawn, J. A. (1995), Statistics for exceptional athletics records. *Applied Statistics* 44, 499–511.

Rootzµen, H. and KlÄuppelberg, C. (1999). A single number can't hedge against economic catastrophes. *Ambio*, 28(6):550-555. Preprint.

Rootzén, H. and Tajvidi, N. (1997). Extreme Value Statistics and Wind Storm Losses: A Case Study. *Scandinavian Actuarial Journal*, 70-94.

Smith, R. (1994), Multivariate threshold methods, *in* J. Galambos, ed., `EVT and Applications', Kluwer Academic Publishers, pp. 225-248.

Stephenson, A.G., Tawn, J. A.: Bayesian inference for extremes: accounting for the three extremal types. Extremes 7, 291–307 (2004)

Straetmans, S. (1998). Extreme financial returns and their co movements. Ph.D. Thesis, Tinbergen Institute Research Series, Erasmus University Rotterdam.

Strand, M. and Boes, D. (1998). Modelling Road Racing Times of Competitive recreational Runners Using EVT. *The American Statistician* 52 (3), 205-210.

Tawn, J. A. (1992). Estimating Probabilities of Extreme Sea-Levels. *Applied Statistics* 41 (1), 77-93.

Tippet, L.H.C. (1925). On the Extreme Individuals and the Range of Samples Taken from a Normal Distribution. *Biometrika*. 17(3-4), 364-387. (in Kinnison, 1985)

APPENDIX

A. SOME CODES USED IN *R* SOFTWARE

A.1 Generalised Extreme Value Models

```
par(mfcol = c(2, 1)); m = blockMaxima(Gains,30,doplot="TRUE");
n = blockMaxima(-Losses, 30,doplot="TRUE");
  #  plot of heavy-tailed data
 p = 0.01*(1:10); q=0.1*(1:10)
hillPlot(m, start = 20, ci = 0.95, doplot = TRUE, plottype = c("xi"), labels = TRUE)
hillPlot(n, start = 5, ci = 0.95, doplot = TRUE, plottype = c("xi"), labels = TRUE)
d=shaparmPlot(x, p = 0.01*(1:10), xiRange = NULL, alphaRange = NULL,doplot =
TRUE, plottype = c("upper"))
shaparmPickands(x, p = 0.05, xiRange = NULL,doplot = TRUE, plottype =
c("both"), labels = TRUE)
shaparmHill(x, p = 0.05, xiRange = NULL,doplot = TRUE, plottype = c("upper"),
labels = TRUE)
y <- pgev(q, xi=0.2092864, mu=0.03428153, beta =0.01540)
qgev(p, xi=0.2092864, mu=0.03428153, beta=0.01540)
dgev(y, xi=0.2092864, mu=0.03428153, beta=0.01540)
# Sample mean excess plot of heavy-tailed
 par(mfcol = c(2, 1));mePR<-mePlot(m); mePL<-mePlot(n)
   # Fit GEV to monthly Block Maxima:
 fit = gevFit(as.vector(m));print(fit); par(mfcol = c(2, 2)); summary(fit)
 fit1 = gevFit(as.vector(n));print(fit1); par(mfcol = c(2, 2)); summary(fit1)
```

A.2 GENERALISED PARETO DISTRIBUTION

```
#    Explorative Data Analysis

emdPlot(x, doplot = TRUE, plottype = "xy",labels = TRUE);

mePlot(Losses, doplot = TRUE, labels = TRUE)

qqparetoPlot(x, xi=0.32241, trim = NULL, threshold = 5.0, doplot = TRUE,labels =
TRUE);

mrlPlot(Gains, ci = 0.95, umin = 3, umax = 7, nint = 100, doplot = TRUE, plottype =
c("autoscale"), labels = TRUE)

mxfPlot(Gains, u = 5.0, doplot = TRUE, labels = TRUE)

recordsPlot(m, ci = 0.95, doplot = TRUE, labels = TRUE)
```

exindexPlot(x, block = 30, start = 5, end = NA, doplot = TRUE, plottype = "thresh", labels = TRUE)

exindex(x, block = 30, quantiles = c(0.99, 0.995, 0.999, 0.9995, 0.9999), length = 10, doplot = TRUE, labels = TRUE)

A.3 PEAKS OVER THRESHOLD (POT)

gs = shape(Gains,30, 15,500, TRUE, labels = TRUE)

shape(-Losses, models = 30, start = 5, end = 100, reverse = TRUE,auto.scale = TRUE, labels = TRUE)

mrlplot(x, u.range = c(1, quantile(x, probs = 0.995)), col = c("green", "black", "green"), nt = 200)

G <- fitgpd(Gains,5, est = "mle")

L <- fitgpd(-Losses,4.5,est="mle")

mle <- fitgpd(Gains, 4.5:5.5, est = "mle");

gpd(Gains,110, "expected")

B. PORTION OF BRENT OIL PRICES OBTAINED (1987 -2009)

Date	Spot Prices	Date	Spot prices	Date	Spot prices	Date	Spot prices
May 20, 1987	18.63	Aug 04, 1987	20.65	Oct 10, 1990	40.2	Apr 30, 2009	50.3
May 21, 1987	18.45	Aug 05, 1987	19.8	Oct 11, 1990	41.15	May 01, 2009	51.75
May 22, 1987	18.55	Aug 06, 1987	19.75	Oct 12, 1990	39.9	May 04, 2009	53.26
May 25, 1987	18.6	Aug 07, 1987	19.65	Oct 15, 1990	38.28	May 05, 2009	53.16
May 26, 1987	18.63	Aug 10, 1987	19.43	Oct 16, 1990	38.93	May 06, 2009	55.07
May 27, 1987	18.6	Aug 11, 1987	19.45	Oct 17, 1990	35.33	May 07, 2009	56.63
May 28, 1987	18.6	Aug 12, 1987	19.5	Oct 18, 1990	35.65	May 08, 2009	56.02
May 29, 1987	18.58	Aug 13, 1987	19.4	Oct 19, 1990	33.2	May 11, 2009	55.99

Jun 01, 1987	18.65	Aug 14, 1987	19.25	Oct 22, 1990	27.45	May 12, 2009	56.52
Jun 02, 1987	18.68	Aug 17, 1987	18.85	Oct 23, 1990	28.95	May 13, 2009	56.84
Jun 03, 1987	18.75	Aug 18, 1987	18.75	Oct 24, 1990	30.1	May 14, 2009	56.25
Jun 04, 1987	18.78	Aug 19, 1987	18.5	Oct 25, 1990	32.9	May 15, 2009	56.33
Jun 05, 1987	18.65	Aug 20, 1987	18.3	Oct 26, 1990	33.73	May 18, 2009	56.51

C. EXTREME VALUES OIL PRICES USED FOR ANALYSIS

C.1 Extreme Values of Transformed Data for Gains (Block Maxima method)

4.203338	3.485457	5.867122	4.115807	4.401689	9.977115	6.663438
5.043085	2.141409	7.867457	4.035130	13.511242	19.01837	36.12143
6.359620	4.766467	7.059432	2.887592	6.164416	3.790152	3.230094
1.953187	3.166491	2.176364	3.181960	2.902493	6.860632	5.861190
5.610139	3.780828	2.602377	4.086649	4.729832	3.411183	3.468556
6.748230	5.299538	4.817686	5.307656	4.154600	3.512925	5.140434
4.330768	6.057139	7.596326	4.796198	9.145378	7.764039	4.860770
7.194008	4.133633	8.534919	8.036722	10.184082	9.859426	3.943464
5.232134	19.89064	12.88262	6.036974	5.523340	4.311481	6.117130
9.000286	4.160712	5.406722	7.527378	4.390796	4.660776	4.191365
7.120778	7.242899	5.040163	4.089635	3.859921	4.244973	5.037134
5.642480	4.274962	3.687136	3.817735	3.612872	4.178705	4.448924
4.604872	11.46205	16.83201	11.32620	5.858422		

C.2 Extreme Values of Transformed Data for Losses (Block Maxima method)

2.456623	4.490750	8.104121	4.095060	6.669137	7.221821	12.44377
6.237033	4.251548	2.876172	6.615867	5.086920	6.991604	12.26979
9.789019	17.333273	3.490287	6.016852	2.955880	4.815430	5.990461
4.591625	2.660911	2.885370	1.917342	3.040133	6.138608	7.835769
4.074522	4.310771	4.748667	3.351563	3.120976	3.303985	2.251388
2.980443	7.455016	4.764109	5.787313	4.914469	2.907785	4.206719
5.851318	6.225397	16.25594	7.378990	7.068766	12.73735	6.706423
6.449469	4.915569	5.310983	5.968551	5.676893	7.036149	7.168992
6.172817	5.080076	4.220388	5.397184	12.85340	7.988946	4.935690
2.934021	2.697172	3.867787	4.919332	5.078315	4.225743	8.898349
4.927410	6.863158	5.732228	11.46876	5.342517	4.988998	4.955381
5.719516	5.291180	3.803921	4.997704	5.209159	4.346279	4.558116
3.311317	2.859573	5.913052	4.369357	3.622088	8.195202	8.355146
18.12974	10.27656	1.073670				

D. LIST OF ABBREVATIONS

Maximum Domain of Attraction (MDA)

Mean Excess Function (MEF)

Mean Excess Plot (MEP)