LAPPEENRANTA UNIVERSITY OF TECHNOLOGY

School of Business

Strategic Finance

Jesse Koponen

Extreme market risk in Gulf markets

Examiners: Professor, Eero Pätäri

Associate Professor Sheraz Ahmed

ABSTRACT

Author:	Koponen, Jesse
Title:	Extreme market risk in Gulf markets
Faculty:	School of Business
Major:	Master of Strategic Finance
Year:	2017
Master's Thesis:	LUT School of Business 79 pages, 5 tables, 26 figures and 1 appendix.
Supervisor:	Professor, Eero Pätäri
2nd examiner:	Associate Professor, Sheraz Ahmed
Keywords:	Extreme value theory, Value at Risk, tail risk, market risk, Peaks Over Threshold, Cornish Fisher, Expected shortfall, Gulf markets

Tail risk is a method for estimating market risk at extreme level. This thesis studies different market risk estimation models in Gulf markets. The study brings insights to time series features of Gulf market data and provides various estimates of extreme market risk. The comparison of results to more developed markets is accompanied.

Different methods for estimating tail risk at Gulf market are compared. The results confirm non-normal features of Gulf markets. Skewness and kurtosis adjusted risk measures outperforms traditional VaR and ES tail risk methods. Developed market shows more normal features of market data than emerging markets. Further developed market fails to account risk by many risk measures that in turn generated adequate tail risk estimate at GULF markets.

TIIVISTELMÄ

Tekijä:	Jesse Koponen
Tutkielman nimi:	Äärimmäiset markkinariskit Gulf markkinoilla
Tiedekunta:	Kauppatieteellinen tiedekunta
Pääaine:	Strategic Finance
Vuosi:	2017
Ohjaaja:	Professori, Eero Pätäri
2. tarkastaja:	Apulaisprofessori, Sheraz Ahmed
Pro Gradu-tutkielma:	Lappeenrannan teknillinen yliopisto
	79 sivua, 5 taulukkoa, 26 kuvaa, 1 liite
Hakusanat:	Ääriarvo teoria, Value at Risk, häntä riski, markkina riski, ylitemenetelmä, Cornish-Fisher, Expected shortfall, Gulf-markkinat

Häntä riski on markkinariskin määrittämistapa äärimmäisellä tasolla. Tämä tutkielma käyttää erilaisia markkinariskin määrittämistapoja Gulf-markkinoitten häntäriskin määrittämiseen. Tutkielma tarjoaa myös näkemystä Gulf-markkinoitten aikasarja ominaisuuksista. Tulokset on vertailtu kehittyneempään markkina-alueeseen.

Tulokset vahvistavat etteivät Gulf-markkinat ole jakautuneet normaali jakauman mukaan. Vinous- ja huipukkuus- oikaistut riskimittarit tuottivat paremmat tulokset häntäriskin määrittämisessä kuin normaalit VaR- ja ES-menetelmät. Aikasarja ominaisuudet kehittyneimmällä markkinalla osoittautuivat enemmän normaalijakautuneemmiksi kuin mitä tulokset Gulf-markkinoilta osoittavat. Häntäriski estimaatit kehittyneemmältä markkinalta eivät olleet riittävän luotettavia, samat häntäriski estimaatit tuottivat kuitenkin hyväksyttävät tulokset Gulf-markkinoilta.

ACKNOWLEDGEMENTS

I would like to thank Professor Eero Pätäri for his guidance and patience during the long time of this thesis. I would also thank the LUT School of Business for opportunity to educate myself through all these years of studies in Lappeenranta.

I am grateful to my parents for always being there for me. I wish to thank all colleagues that I have worked with during these years.

In Leppävirta 20th of May 2017

Jesse Koponen

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. Background and motivation

Recent crises have brought out the questions on tail risk estimation and returns modeling for better adjustment of risk. Extreme events[1] have seen as epitome of crises and naturally interest towards extremes has grown. Unexpected nature of risk is a hard to look for and anticipate. Therefore estimation of extreme events is a natural approach to beware and assess the unforeseen risk. Estimation refers to a probability structure of future unexpected events. Although the exact and true probability is impossible to achieve, it has an opportunistic position.

Tail risk is a broad subject and a discussion can be directed to various perspectives to approach tail risk. Tail risk is approachable by indication of the phenomenon itself. Extreme observations are not only subject of finance but also studied in other fields of academic research and practical work. Extremes are studied in field of hydrology, wind engineering, environmental science and seismology. In financial context tail risk is usually linked to market risk estimation purposes. Thus, it’s not surprise that the number of estimation methods for tail risk have emerged. Tail risk can also be approached by repositioning it to other risk managing methods. Indeed Tail risk is categorized under market risk, where the risk comes from harmful market movements.

Rare events matter, and above all, extreme events are not phenomenon of faraway concern but current and present in our risk taking world. Wording “extreme” refers to rarely present and therefore remote; however, the effects are well covering and existing. Visibility is perceived by incorporating the extreme events into the analysis. Extremes represent the most properties of the totality; nonetheless the economy is omitting the extremes and concerning the means. The salience is in the properties that define bulk of the effect, not an average of it. Therefore, non-normal features such as fat tails are in the essence of market movements.

[1] Here wordings extreme, unforeseen, unfortunate, tail- and rare events refer to the very same nature of phenomenon under the study.

Tail risk in finance is usually defined by non-normal features of data. Non-normality refers opposite of normal and in finance by this implication normal can be seen as standard as the wording literally refers. There are number of non-normal features and the existence of these features cause data to show extreme behavior. Non-normal movements in asset returns are characterized by fat-tails and high-peakedness. Another way to consider tail risk is to characterize it as decrease in market price of an asset. Decrease can be come from different factors of that affect on market prices. These include price changes of stocks and commodities and changes in rates of interest rates and foreign exchange rates for example. Usually the study of extremes does not specify these factors in the model.

Tail risk can be approached directly trough concept of fat tails, which refers to distributional feature. Distribution with higher density than under the assumptions of normal distribution is defined as fat tailed distribution. Distribution that has power decay i.e. infinite endpoint of the density function in the tails is considered a fat tailed distribution, whereas exponential decay i.e. finite endpoint is considered thin tailed, as in the case of normal distribution (Lebaron & Samanta, 2005). Typical for power decay appearance is also referred as polynomial growth while the normal distribution is referred as exponential growth. Furthermore, figure 1 illustrates the fact that distribution with exponential decay goes faster to zero than distribution with power decay. In power law function, the exponent is constant while in the exponential function exponent is variable as in Gaussian normal distribution.

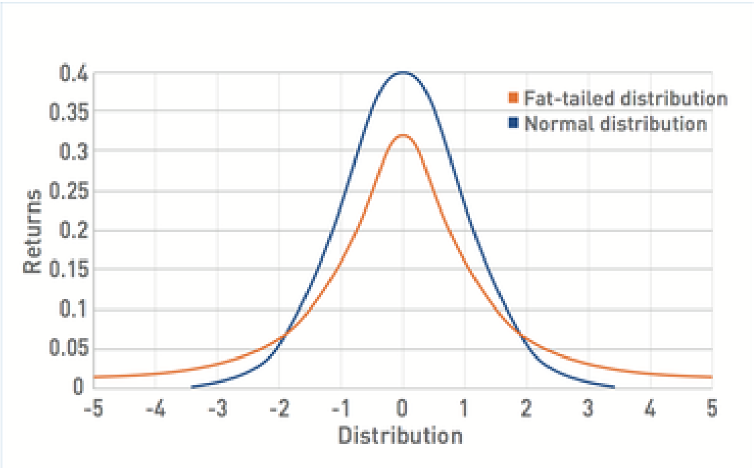

Figure 1: Fat-tailed distribution compared to normal distribution

Tail risk measurement can be approached by predicting the distribution of possible changes, over a given time horizon. For traditional dispersion measures tail risk becomes invisible, and therefore, special tail risk measures are needed. In more detail, Tail risk modelling signifies to incorporate extreme events into the risk analysis.

After the financial crisis of 2008, the questions were raised whether existing models and practices based largely to Gaussian distribution present adequate framework for market risk estimation. Therefore, other methods than Gaussian based have been applied to study tail risk in a more realistic framework. Tail risk has been in the essence of various risk measures which has led to capture higher moments of return distribution and present more realistic contribution to market risk estimation.

1.2. The research problem

This study strives to bring light to the following research questions and present the subject markets in a right position from the point of risk management of extreme market risk.

Which tail risk estimates are most appropriate for market risk estimation in Gulf markets?

Are the Gulf market results generalizable to more developed markets?

The pre-eminence of tail risk estimates is conducted by dividing the data into two sample periods in-sample and out-of-sample. Tail risk is estimated at in-sample period and the performance of tail risk estimates is confirmed in backtesting at out-of-sample period. Thereby the both questions are answered once the comparison of Gulf markets to more developed markets is conducted in the results section.

1.3. Structure of the study

Second section enters into literature review and theoretical framework for study. In literature review empirical studies are discussed regarding to concepts and frameworks of tail risk. Theoretical background provides an introduction to the methods of tail risk modeling by concepts of probability theory and statistics. Third section represents the data and methodology used in this study. Section 4 covers the results where all estimated risk measures are shown with performance evaluation. Section 5 summarizes the study and discusses implications and gives suggestions to further research.

2. Literature review and theoretical background

This section provides introduction to literature of tail risk measures. Section 2.2 introduces theoretical background to risk measurement. The focus is on notion of risk and structure of tail risk measure. The very nature of risk in context of managing tail risk is disclosed. Section 2.2.1 introduces extreme value theory as a framework for tail risk estimation. Section 2.3 shows different tail risk estimates based on the theoretical background introduced in previous sections.

2.1. Literature review

Estimation of unexpected event has been given many attempts and various methods in financial literature. All these methods have similar mission to generate accurate estimate of unexpected event with given information. The focus is on probability structure of events and risk associated within. Risk measures summarize effects of unexpected events and volatility standard deviation is the most simplified risk measure to infer returns of empirical sample and map the unforeseen future risks. However financial crises have laid bare the failure of standard risk models. Albeit the true data generating process is unclear and uncertain, advanced risk measures have been developed and trusted – paradoxically or not.

Early empirical discussion of Mandelbrot (1963) and Fama (1963) on the properties of asset return distributions have lead to rejection on the assumptions of normality and given inspiration to the class of stable Paretian distributions as an alternative to the Gaussian assumptions. The Lévy stable, i.e. α-stable, distribution has been one of the most popular distributions to model the asset returns. Since then number of studies has challenged Gaussian distribution (Longin, 1996; McNeil, Frey & Embrechts, 2005; Bali & Neftci, 2003; Duffie & Pan, 1997) and given support and evidence to fat tailed distributions in characterizing empirical returns both in developed markets (Gettinby et al., 2006; Bali, 2003; Bali, 2007; Byström, 2004; Chan & Grey, 2006) and emerging markets (Gencay, Selcuk & Ulugulyagci, 2003; Mandira, Thomas & Shah, 2003; Zikovic & Aktan, 2009; Makhwiting, Sigauke & Lesaoana, 2014; Tolikas 2011; Onour, 2009). The interest towards stable distributions originates from statistical property of domain of attraction of stable laws, i.e. central limit theorem, and stable distributions belong own domain of attraction of stability. Moreover the popularity has lean towards the properties of stable distribution, which are flexible enough for allowing excess kurtosis and skewness. (Akgiray & Booth, 1988)

In 1970s and 1980s empirical studies reported inconsistencies between stock returns and popular stable distribution assumptions. The discrepancy mainly considers failure of the temporal aggregation property, i.e. tail thickness does not change with return frequency, and assumption of infinite variance of stable law distribution (Stoyanov et al. 2011). In addition, inconsistency was caused by studies that considered symmetric stable distributions while stock returns showed property of asymmetry, i.e. skewness (Akgiray and Booth, 1988). Other classes of distributions were suggested, examples include Student’s t distribution from the work of Blattberg and Gonedes (1974) which has fatter tails than Gaussian distribution. In 1995 Eberlein and Keller considered general class of hyberbolic distributions and later on tempered stable distributions grew attention which relates to stable distribution but overcomes discrepancy of stable distribution with finite variance and treats properly the temporal aggregation property of the distribution (Stoyanov et al. 2011). In 1988 Akgiray and Booth began the discussion of separating tail behavior and the body of the distribution. Later that work was extended by Hols and de Vries (1991).

Risk measures dates back to Roy’s (1952) safety first decision theory of investors minimizing risk rather than maximizing return. Since then other types of risk measures have taken place by simplest Markowitz (1952) mean and variance based. Further Sortino and Price (1994) favored drawdown risk measures over Gaussian based standard deviation and beta. Eftekhari, Pedersen & Satchell (2000) supported alternative risk measures along with Sortino and Price (1994) and were followed by number of others like Pedersen and Satchell (2002) and Hwang and Pedersen (2004). VaR was introduced in 1996 by Committee of Basel. The critic against VaR and discrimination it from coherent risk measures were influenced by Arzner et al. (1999) and crisis of LTCM in 1998. The common features of financial data to exhibit high peakedness gathered around the mean and skewness, i.e. asymmetry (Ang, Chen & Xing 2006; Hyung and de Vries, 2005; Campbell and Kraeussl, 2007; Cheng, 2005; Post and Van Vliet, 2006), have led to incorporate the properties of stylized facts to risk measurement. Li (1999) proposed skewness and kurtosis adjusted VaR, later Favre and Galeano (2002) suggested modified VaR using Cornish-Fisher expansion. More recently Rockafellar and Uryasev (2000) suggested ES as coherent risk measure.

In the beginning of 21th century Extreme Value Theory (EVT) gained attention as more popular and prominent way to model tail behavior. The basic ricochet of extreme value studies can be viewed as rejection of normality assumptions of financial time series and furthermore promote the necessity of EVT as an accurate method for tail estimation. Both of which have been documented in number of studies (Makhwiting, Sigauke & Lesaoana, 2014; Tolikas, 2011; Jeyasreedharan, Alles & Yatawara, 2009; Duffie & Pan, 1997; Huisman, Koedijk & Pownall, 1998). Extreme value theory estimates the possible loss more accurately than normal distribution based methods (Bali, 2007; Gencay, Selcuk & Ulugulyagci, 2003; Onour, 2009). During the last decade the attention has grown towards Copula methods.

Characterization between emerging and developed markets discloses market risk to liquidation, political and economical related risks. Fluctuations of credit rating and exchange rates and financial shocks from internal and external sources are also categorized under market risk. Effect of economic crises is faster in emerging markets than in more developed markets. The exposure from various domains of risks causes unnecessary volatility in emerging markets. Empirical evidence related to emerging markets regularities is provided e.g. Tolikas (2011).

Since Markowitz (1952) risk measurement has emphasized variance as proxy for risk, thereafter attempt to characterize risk has based mainly on the concept of variability. Artzner et al. (1998) structures risk in comprehensive way, however a comprehension of risk might enjoy renewed definition of risk and proceeding to the unexpected events.

2.2. Theory of Tail Risk Measurement

In finance probability theory offers standard description of uncertainty from where risk measurement becomes approachable. Probabilistic models have been estimated from empirical data which requires statistical inference. Probability and statistics goes hand in hand when estimating tail risk[2]. (Rachev, et al., 2007) Risk estimation is usually conducted for the purposes of capital allocation and capital requirements. Tail risk as a part of market risk is estimated mainly for capital requirement purposes. VaR stands as standard tail risk measure for capital requirement purposes. Risk measure such as VaR can be decomposed by probabilistic and statistical notions, which assists structuring risk measure to components for more detailed analysis and comprehension. Tail risk estimation consists of setting time horizon and position value, but also statistics are involved by variability measures of risk factors, i.e. meaning returns series, and probabilistic terms by figuring out factor yield from valid probability distribution. Tail risk measures are therefore constructed by terms of probabilistic and statistical notions.

2.2.1. Distributed returns

Asset returns are treated as variable that exhibit randomness, therefore they are called random variables. In theory random variable can be discrete or continuous but in finance asset returns are considered as outcomes of continuous random variable. Sample space represent a set of all possible outcomes of random variable and event stands for subset of the sample space, i.e. single outcome. Continuous random variable can take any value within the range of outcomes. In general, the number of outcome of continuous variables is so large in the interval that it is not possible to add probabilities to each one of them. Therefore, a concept of probability distribution must be considered. By its nature probability distribution assigns value to every event in the sample space, i.e. probability distribution is used to describe potential outcomes of a random variable.

[2] *Risk and uncertainty is represented as comparable notions in this study albeit risk is usually the one commonly referred incorrectly instead of uncertainty. Risk is by nature a domain with known probability structure. Uncertainty have unknown probability structure and should thus be a domain of interest.*

Probability distribution can be represented mathematical ways and probability density function is the probability distribution of continuous case. (Rachev, Menn & Fabozzi, 2005*)*

Probability density function f(x) for random variable X can be interpreted as follows: realized value near some point x with high probability are shown as large values of density function at point x. The probability of some outcome is achieved by calculating the area between a point on the horizontal axis and the curve. Probability distribution in general can be summarized by various measures: location, dispersion, asymmetry, concentration in tails and quantiles. Probability density function is important as estimation of probabilities of events like summarizing key features of the phenomenon. (Rachev, Menn & Fabozzi, 2005)

For continuous distribution the probability of observing a single outcome is close to zero. In this case the density function must be used to estimate how likely an observation will fall in the neighborhood of estimated value of density function. Therefore a joint density function as a function of unknown parameters, i.e. mean and variance, is used to estimate probability of sample observations. Probability of sample observations is achieved by maximizing the likelihood of unknown parameters. Consequently most likely estimates are produced for the sample observations. In other words, the likelihood is performed by fashion of choosing the estimates for the unknown parameters so that the likelihood given by the joint density function measured at the observed sample is maximal. (Rachev, Menn & Fabozzi, 2005) Once the distribution is approximated by estimated parameters it can be used for statistical analysis, e.g. tail estimation, to describe the occurrence of the phenomenon.

Financial returns are commonly estimated based on normal distribution which relies on central limit theorem. Central limit theorem states that the independent random variables follow normal distribution, i.e. an asymptotic distribution, as observations increase. Central limit theorem relies on law of large numbers, which states a sample average converges in probability to the expected value as sample size grows to infinity. Unfortunately central limit theorem applies to the centre of the distribution. (Dowd, 2002 p.82) Normal distribution is a benchmark in many ways and comparable notion in statistical modeling; nevertheless, financial returns have common features which don’t rely on the normality assumptions. Financial data is observed to have heavy tails, low autocorrelation of return series, high autocorrelation of squared return series, volatility cluster and mean reversion. Stylized facts can be illustrated by probability distributions; where common properties e.g. single peakedness and centralized mass likewise asymmetry are represented.

Tail risk estimation involves a probability of all possible events. Distribution should illustrate tail events and therefore such probability distributions are called non-normal as well as fat tailed. Various distributions have proposed: Mixture of Normal distributions, student-t distribution, extreme value and gamma distributions, for example. Regardless of the number of distributions that have been considered, stable distributions have property of stability that favors stable distribution from others. Stability means that distribution does not depend on time interval over which returns are considered; therefore any time interval can be addressed to stable distributions[3]. Stable distributions include Cauchy, Normal and Levy distributions as a special cases. Pareto which is many times suggested as heavy tailed distribution belongs to the class of Levy distributions. (Rachev et al., 2007)

2.2.2. Stochastic Process

While financial returns are treated as random variables the evolution of financial returns can be regarded as random process, i.e. stochastic process. Stochastic processes are used to model random phenomenon that evolves over time. Stochastic process can be addressed questions of dependence, long term average and extreme events for example. Stochastic process is a function of both time and outcome of process and it is formed by number of cumulative joint distribution functions of random variables.

Process in general can be considered as a sequence of random variables (Billingsley, 1995); Poisson process for example is simply number of events in a given interval. Natural feature of the process is that it generates events in time

[3] Stable distributions can be referred as Levy stable and alpha stable.

(Pelagatti, 2016). Simplest model to account stochastic behavior of asset returns is white noise process, which is a process of i.i.d. realizations with zero mean and constant variance. Random walk, Markov processes, Levy- and Poisson process are reckoned as more advanced processes. Poisson process is counting process which counts the time of the occurrence of events and also time between consecutive events. Poisson process is used in estimating rare events in EVT based model. More complex processes can take account for example clustering of returns, i.e. non-i.i.d. characteristics, by allowing dependence behavior. These processes are ARMA, ARCH and GARCH type of models that can incorporate time-varying conditional volatility (Rachev, Menn & Fabozzi, 2005).

In finance more commonly continuous process has been considered. Two most used continuous processes are Brownian motion, i.e. Wiener process, and Poisson process. Both processes belong to the class of Levy processes. These processes corresponds general process or limiting process. As the number of random variables forms stochastic process in continuous time, the path of the process describes the evolution of one specific realization of random variable through time. In order to draw true path, the value of the process for every real number must be calculated, which is not feasible. Therefore an assumption is made that the processes converge to some limiting process, which stands as a general process when the number of paths are added together. (Rachev, Menn & Fabozzi, 2005)

A peculiar relationship between process and probability distribution is that probability distribution is formed to characterize the process. Probability distributions can be associated to process to describe the process from different aspects; such as distribution of time between events and the distribution of number of events in a time interval, for example.

Return series are considered as a realization of random variable. Therefore, some assumptions regarding return series are addressed. Autocorrelations indicate the relationship between today’s return and historical values of return. Autocorrelation is also referred to as serial correlation. Therefore, assumptions of behavior between occurrences of consecutive observations have been addressed through idea of dependence and stationarity. General assumption of returns supposes that current return is not dependent on the previous returns. Indeed, most techniques for estimation of distribution parameters propose that the return observations are independent (Rachev, Menn & Fabozzi, 2005). Empirical results show that this assumption is violated by most return samples. The violation occurs, for example, as the observations exhibit seasonality, i.e. more volatile and less volatile times, which violates random behavior (Coles, 2001).

In addition to assumption of dependence of the time series, the assumption of stationarity has been brought out. The notion of stationarity is standard assumption which implies a process to be identical when measured over different time periods. Stationary process allows today's returns being dependent of yesterday's return, but the dependence must remain same as the subset is viewed a time period later. Therefore, it responds to real nature of the asset return behavior. (Coles, 2001) Stationarity is time-homogeneity of time series, which is defined as time-invariance for all moments of the probability distribution for strict stationarity and for first two moments for weak stationarity (Pelagatti, 2016). However, the number of extreme observations in financial time series cast doubt that higher moments would exist and then be bounded. Mandelbrot (1963) provided evidence that even the second moment won't exist for fat-tailed data and suggested stable distributions with infinite variance as an alternative to finite variance statistical models. Against Mandelbrot's view Akgiray and Booth (1988) considers that the second moments of most series are finite. Fat tailed distributions allow unbounded moments e.g. Fréchet allow tails to vary with $-\alpha$ which follows power law where the tail index α measures tail thickness and the number of bounded moments (Cotter & Dowd, 2011).

To conclude independence and stationarity are general assumptions regarding to behavior of consecutive observations of financial time series. Together these assumptions form assumption of independent and identically distributed series, i.i.d., which stands for general assumption of white noise series process. Time series analysis mainly concentrates on confirming assumptions of i.i.d.

2.2.3. *Risk measure formation*

Risk measures produce estimates from the sample to describe accurately as possible the phenomenon under the study. At simplest level, risk measure can be considered statistic such as variance (Corelli, 2012). More often than not a risk measure is needed in drastic times and should meet the properties of these situations. Here statistics such as mean and variance solely are insufficient but can be involved as building blocks of the more complex risk measure.

Probability theory provides framework for estimating probability structure of the phenomenon e.g. tail risk. Once the probability structure is known and approximated in appropriate way it can be used to construct statistical measures to better assist understanding the phenomena. The approximated probability distribution can be exploited in constructing risk measure by simple transformation. In a case of normal distribution, this is achieved by transforming approximated distribution to standard one. The resulting value is called z-value measuring the number of standard deviations from mean the observations lies. Here the z-value refers previously mentioned component of risk estimate namely factor which is yield from the probability density function for a chosen confidence level. This factor is used to construct more complex risk measures such as VaR, which exploits the information of probability structure of the phenomenon. Statistical methods can further formulate tail risk measures where the indication of riskiness of the phenomenon is interpreted.

The basic tools for summarizing the distribution of returns are approached by statistical moments. For instance, when quantity has a sufficiently strong tendency to cluster around some particular value it’s better to characterize the value by numbers that are related to its moments. However not all quantities behave like this (Clauset, Shalizi & Newman, 2009). Moment is a statistical quantitative measure representing the shape of a set of observations. Mean is a first moment which estimates the value around which central clustering occurs and location of a probability distribution. Second moment variance characterizes variability or dispersion relative to mean. One alternative to variance is standard deviation and another alternative is mean absolute deviation in the case variance is infinite and does not exist then. Third moment skewness characterizes shape of the distribution in the form of symmetry of distribution around its mean. Distribution with one tail longer than another is skewed, hereby asymmetric. The additional feature of skewness is that dispersion of distribution can be characterized roughly by referring to asymmetric property of distribution. Fourth moment: Kurtosis measures relative peakedness of the distribution to normal distribution and fatness of the tails. Positive kurtosis is also called leptokurtosis and negative platykurtosis, respectively. (Rachev, Stoyanov & Fabozzi, 2008)

Volatility by simplistic can be considered as constant, i.e. it does not change over time. Empirical evidence does not support constant volatility and therefore, better approximation of volatility is achieved by assuming volatility change over time under which recent observations are given more weight than more distant ones. This can be implemented by ARCH and GARCH type of models, among many others.

Quantiles are related to notion of moments to summarize and study tails of the probability distribution; moreover quantiles divides the range of empirical probability distribution. Quantiles can characterize the risk, whereas moments can characterize the distribution. Quantiles are simplest method to summarize the tail risk of the phenomenon. For example the first quantile describes 25 percent of range, second quantile 50 percent and third quantile 75 percent of the range. The 1%, 5%, 95% and 99% quantiles are called percentiles. Risk estimation exploits the very nature of quantiles by taking lowest percentile of the range. Therefore, risk estimates such as VaR are referred as quantile of distribution. (Rachev, Stoyanov & Fabozzi, 2008)

The role of statistics is tolerably to bring out the characteristics of randomness from the process that generates the data. That can be achieved by considering summarizing the behavior of randomness e.g. in simplest manner by estimating the average exposure of the phenomenon for period out of the sample, or extrapolate nature of future maximum values from sample extremes. The objective of statistical modeling of extremes is to quantify the stochastic behavior of a process at extreme level. (Coles, 2001)

2.2.4. Extreme value theory (EVT)

Extreme value theory has suffused in various different fields and applications. Extreme value theory originates from hydrology, environment science, insurance mathematics and wind engineering. Later on EVT has been applied to risk management and telecommunications. Applications include skyscraper building, dam building and mining industry for example. EVT examines extreme observations of random and rare events. The object is to approximate the probability of event that has not occurred. (Coles, 2001) EVT can be characterized as a modelling method that studies the tail of the distribution. Empirical evidence supports the sufficient procedure of EVT in capturing heavy tailed features of asset returns.

EVT centers around two theorems: the Fisher-Tippett-Gnedenko and the Picklands-Balkema-de Hann theorem. Former, the Fisher-Tippett-Gnedenko theorem provides a framework for modeling the maximum of a random variable, while the Picklands-Balkema-de Hann theorem offers an approach based on modeling the largest values over predetermined high threshold. (Da Costa Lewis, 2012)

The Fisher and Tippett theorem states that maxima of independent and identically distributed (i.i.d.) random variables follows one of the three extreme value distributions as an asymptotic i.e. limiting distribution: *Fréchet* distribution, with infinite tail, *Gumbel* distribution, whose tail is also infinite, but lighter than the Fréchet distribution or the *Weibull* distribution with finite tail. (Zoglat et al., 2013) The infiniteness of the tail is natural property of heavy tails.

The maxima of sequence of independent random variables are denoted by (Coles, 2001):

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

Distribution function F represent unknown probability distribution of random variables $X_1, \ldots, X_n$. Approximation of distribution function F can be done by looking approximate distribution of maxima and minima. (Coles, 2001)

The limit distributions for M_n is $\mathrm{G}(z)$ if there exists sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \rightarrow G(z) \quad as\, n \rightarrow \infty$. The G is a non-degenerate distribution function (Coles, 2001):

Gumbel: $\mathrm{G}(z) = \exp(-exp\,[-(\frac{z-b}{a})])\ , \qquad -\infty < z < \infty$

Frechet: $G(z) = \begin{cases} 0, & z \leq b \\ exp^{\{-(\frac{z-b}{a})^{-a}\}}, & z > b \end{cases}$

Weibull: $G(z) = \begin{cases} exp\,\{-\left(\frac{z-b}{a}\right)^{\alpha}\} & z < b \\ 1, & z \geq b \end{cases}$

The Fisher and Tippett (1928) theorem implies that the asymptotic distribution of the sample maxima converges to one of the three distributions above. In other words sample maxima belongs one of the extreme value distribution families known as Gumbel, Fréchet and Weibull respectively. Shape of the distribution is governed as the tail index parameter α goes ∞ and - ∞. (Corelli, 2012; Coles, 2001)

The density of the limiting distribution, G, decays exponentially for the Gumbel distribution and polynomially for the Fréchet distribution, thus indicating relatively different rates of decay in the tails of unknown probability distribution, F. Indeed three different extreme value distributions differs considerably (Coles, 2001)

The Fisher-Tippett- Gnedenko theorem is congruent to central limit theorem (CLT) and uses the tail index to unify the possible characterization of the density function of an extreme value distribution. The CLT is centralized to the normally distributed sum of random variables, whereas Fisher-Tippett-Gnedenko theorem indicates appearance of the distribution in the extreme limits as the sample size reaches to infinity. To be more specific, Fisher-Tippett-Gnedenko states limit law for maximum whereas CLT constitutes limit law for the averages. (Da Costa Lewis, 2012)

The behavior of extreme values differs considerably by the representation of three extreme value distributions. The weakness of adopting one of the three extreme value distributions is culminating to a choice of which distribution to select. Problem is that the selection is considered as accurate one as the empirical data at hand, thus there might exist some uncertainty. (Coles, 2001) Therefore better approach to modeling extreme values is parameterization of single distribution. Extreme value theory has two choices of modeling returns by single distribution parameterization: Generalized Extreme Value (GEV) distribution which relies on Block maxima method and Generalized Pareto (GP) distribution based on threshold model.

2.2.5. Generalized extreme value distribution

The three extreme value distributions can be reformulated and combined to single family model. (Coles, 2001)

$$G(z) = \begin{cases} \exp\ \{-\left[1+\ \xi\left(\dfrac{z-\ \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\}, & if\ \xi \neq 0, \\ \exp[-\exp\ \{-\left(\dfrac{z-\ \mu}{\sigma}\right)\}], & if\ \ \xi = 0 \end{cases}$$

The former corresponds to Fréchet when ξ is positive and Weibull when ξ tends to negative. The latter represents Gumbel family distribution function. The model has three parameters: a location μ, a scale σ and the shape ξ parameter. The variability of shape parameter ξ determines which extreme family is most appropriate for the empirical data. (Coles, 2001)

The tail of distribution function $F(x)$ decays like a power function as Gnedenko (1943) shows; ergo it is domain attraction of the Fréchet distribution. The well known heavy and fat tailed distributions with power decaying tail include Pareto, Cauchy, Student-t and mixture distribution among to Fréchet. (Corelli, 2012)

Generalized extreme value (GEV) distribution is a limiting distribution of block maxima method. Block maxima method divides the data to equally sized consecutive blocks and fit the GEV distribution to maxima of a sample data of vertically divided blocks. Block maxima is parametric method for estimating the unknown random variables $Z_1, \dots, Z_m$. The notion of Z_i represent the maximum values of blocks divided in data. In other words block maxima is the extreme value method which estimates extreme events from the empirical data by picking the maximum value of each equally divided blocks of data. Block size determines the accuracy of the model to represent the data in an appropriate way in tail risk estimation. The block size should not be too small in order to avoid bias in estimation. Too large block size in turn result variance due small number of maxima observations. (Coles, 2001)

Block maxima method for extreme values was strongly criticized because the estimation of the distribution based on extracted blocks of maxima values involves a loss of information. Information loss can occur as width of the block may not be appropriate. For example, one block may consist of more than one extreme observation when the entire sample is concerned, but only the highest value is taken into account. Information loss may also occur as returns of the time series may not come from the one and only distribution but divide into two different volatility sections, low and high. Because the volatility periods are different, the returns are therefore generated by two different distributions (Stoyanov et al., 2011). An alternative to the Block Maxima method is the Peaks-Over-Threshold (POT) model, where the maxima is replaced by modeling the exceedance above predetermined threshold.

2.2.6. Peaks-over-Threshold

The peaks-over-threshold (POT) keeps General Pareto distribution (GPD) as a limiting distribution for tail observations. The POT method is similar to Block Maxima method – here the asymptotic distribution is fitted to extreme data that exceed high threshold. (Stoyanov et al., 2011)

Threshold method is preferred over the block maxima method mainly due to more efficient use of data. Threshold exceedance exploits all the data that exist over high level of threshold whereas the Block Maxima exploits only maxima observations of each block. Moreover the behavior of large observations is more important than the existence of the observation itself (Gencay & Selcuk, 2004).

Sample observations can be noted as $Xt, t = 1,2, \ldots, n$ with a distribution function $F(x) = Pr\{Xt > x\}$. One may exploit high predetermined threshold u to extract extremes from the sample. An exceedance of a threshold u occurs when $Xt > u$ for any t in $t = 1,2, \ldots, n$. An excess over u is defined by $y = Xi - u$. (Gencay & Selcuk, 2004)

The probability distribution of excess values of X over threshold u is defined (Gencay & Selcuk, 2004):

$$F_u(y) = \Pr\{X - u \leq y \mid X > u\}$$

which represents the probability that the value of X exceeds the threshold u by at most an amount y given that X exceeds the threshold u.

According to Balkema and de Haan (1974) and Pickands (1975) theory for sufficiently high threshold u, the distribution of the excess converges to the GPD. The GPD is defined as

$$G\xi,\sigma,\mu\,(x) = \begin{cases} 1 - (1 + \xi\,\dfrac{x-\mu}{\sigma})^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - e^{-\frac{x-\mu}{\sigma}} & if\ \xi = 0, \end{cases}$$

with

$$x \in \begin{cases} [\mu, \infty], & if\ \xi \geq 0 \\ \left[\mu, \mu - \dfrac{\sigma}{\xi}\right], & if\ \xi < 0 \end{cases}$$

where $\xi = \frac{1}{\alpha}$ is the shape parameter, α is the tail index, σ is the scale parameter, and μ is the location parameter. When μ = 0 and σ = 1, the representation corresponds to the standard GPD. The shape parameter ξ is governed by tail index, α, and the tail of the distribution is governed by shape parameter ξ. (Gencay & Selcuk, 2004) The shape parameter ξ determines the upper bound or limit for distribution of excess similar way as GEV model. When the $\xi < 0$ the upper bound exist and can be formalized as $u - \tilde{\sigma} / \xi$; if $\xi > 0$ or $\xi = 0$ there are no upper bound. (Coles, 2001)

The GPD can take a form of number of other distributions. That can be done by changing the value of shape parameter, ξ, when $\xi > 0$, it takes the form of the ordinary Pareto distribution. Pareto represents the heavy tail distribution, which belongs to the class of Lévy distributions**.** (Gencay & Selcuk, 2004)

Exceedance over high threshold is distributed by Poisson process i.e. limiting process of exceedance. Poisson process counts time occurrence of events and time between events, which are exponentially distributed. GPD in turn describes intensity of exeedance i.e. probability of events.

Threshold estimation

A choice of threshold is as important to the threshold method, as a choice of block size to the block maxima method. High threshold results insufficient exceedance and high variance whereas the low threshold violates the asymptotic assumptions of the model leading to bias. Threshold can be chosen by graphical methods such as Mean Excess plot and Zipf plot among others. Graphical methods for threshold selection are based on the mean of the GPD. Graphical methods may suffer from the lack of data which leads to the inaccurate selection of threshold. (Coles, 2001)

Critics against Extreme Value Theory

Extreme Value Theory should be considered critically. Therefore one might consider whether EVT is realistic model for risk measurement purposes. The convergence in EVT relies on law of large numbers (LLN), where sample extremes converge to limiting distribution as sample size, n, goes infinity. The LLN states: no single observation can make difference to the total sample, when n goes to infinity, i.e. as the observations adds up, the asymptotic distribution is achieved. This assumption or presumption is harder to prove in finance. In finance, extremes matter more than in some other domains, because extremes tend to present larger stake of the entire effect than in some other domains. (Coleman, 2012) Another problem arises when returns tend to appear in clusters, which violates the independent and identically distributed (i.i.d.) returns assumption. This restriction leads to demand of large number of observations and better precision in a threshold determination. (Stoyanov et al., 2011)

2.3. Tail Risk Measures

Uncertainty demands risk measuring. However, these two notions are not synonymous. Risks are quantifiable by risk measure procedures while uncertainty requires functionalities of dispersion measures. Risk can be approached as a subjective way by uncertainty and exposure. Risk management exposes to uncertainty, which in turn is realized by risk and loss. Indeed, the main objective of risk management is to identify and manage risk. Risk measuring is utilized into this extent. (Rachev, Stoyanov & Fabozzi, 2011)

Often risk is qualified as an asymmetric measure because it is related to loss only. Uncertainty is symmetric risk measure and can quantify both upside and downside movements, such as variance. Partly due to the lack of appropriate definition of risk, risk measures are imperfect capturing only some part of the risk. (Rachev, Stoyanov & Fabozzi, 2011). That does not necessary mean that the risk measuring tools at hand are worse. Indeed, one is able to extract information even from fragile risk measure under criticisms. One can see if risk is growing or decreasing by re-evaluating the risk. To summarize the relationship and crossroads of uncertainty and risk; there are no risk without uncertainty.

2.3.1. Value at Risk (VaR)

Value at risk (VaR) is a celebrated risk measure of Basel Committee since 1996. The announcement required banks to have sufficient amount of capital to cover losses of next 10 days for 99 % probability. VaR can be defined as quantile of loss, single number and statistical measure. (Da Costa Lewis, 2012; Chen, Giles & Feng, 2012) The capability of VaR to numerate the component of market risk in an easy, accessible and reasonable way has probably endorsed VaR's wide applicability and popularity. Nevertheless there has been clear understanding and message from the point of the introduction of VaR and the manual within that this measure of risk is incomplete and should be treated and applied in knowing that it does not produce the full picture. Yet the world adopted it as a unique measure of risk with fragile results as the economic crises have evidenced.

VaR can be estimated through non-parametric and parametric ways. Non-parametric methods do not make strong assumptions about the distribution of returns. Parametric methods tend to approach risk estimation by fitting the probability curves to the data and work with fitted curve to VaR. In addition to parametric and non-parametric methods, VaR approaches can be divided further to unconditional and conditional classes (Danielsson & De Vries, 2000, p.3). The conditional approach estimates VaR assuming conditional heteroskedasticity on asset returns, which states that estimation are more efficient in conditional on higher weight to recent information. In conditional fitting the reasoning is build on realizing that returns experience stylized facts, such as volatility clustering. Unconditional approach assumes that returns are realizations of i.i.d. (Choi & Min, 2011) Thus, in unconditional fitting the recent market conditions are not taken into account (Danielsson & De Vries, 2000). Extreme Value Theory is one suggested approach if VaR is quantified at extreme confidence levels. (Byström, 2004; Danielsson & De Vries, 2000) Classification of VaR methods are merely expressed to reposition different estimation methods from theoretical aspect.

Calculation of VaR in general requires estimating the volatility of return process and quantiles of standardized returns whether or not parametric or non-parametric techniques are considered (Fan & Gu, 2003). Value at risk (VaR) is defined as (Finkenstädt & Rootzén, 2004):

$$Pr\,\{X_T > x\} = \alpha$$

where X_T is the cumulative loss over a given time horizon T (e.g. 1 day) and α is a given probability (e.g. 0.05).

Non-Parametric methods

The most common non-parametric methods for estimating VaR are Historical simulation and Monte Carlo simulation.

Historical simulation

Historical simulation method calculates probability density for the empirical data (Novak, 2012). The advantage and popularity of historical simulation stems from its easy and fast implementation and from the fact that it won't imply any distributional assumptions. The main problem dealing with historical simulation data is that there might not be enough data available to produce an adequate risk estimate. (Alexander, 2003)

The simulation in historical simulation could be illustrated by return series from historical data by selecting different periods and repeating the process. Moreover, the simulation comes forth as the periods are chosen by minor modifications. (Barone-Adesi & Kostas, 2001) Historical simulation assumes that the most recent observations should be valued more and therefore, given more weight than the distant ones (Christoffersen, 2003). The idea behind proper perception period is that it should be large enough to be significant but at the same time avoid intervening from one to another cluster of volatility (Manganelli & Engle, 2001). VaR is approached by computation relying on the result of discovered paths by historical simulation. (Barone-Adesi & Kostas, 2001)

Monte Carlo simulation

The Monte Carlo (MC) is calculation method for random variables which is utilized in computing of functions (Huynh, Lai & Soumaré, 2008). Monte Carlo simulation method relies on covariance matrix and can produce as accurate estimate of risk as the covariance matrix allows. Method consists of simulating movements of underlying assets and risk factors from current point to future in time. Monte Carlo consists of a set of large scenarios of events for the portfolio in future time. (Alexander, 2003) The average of these events is taken to represent the accurate estimates (Huynh, Lai & Soumaré, 2008).

In the historical simulations approach the risk factor scenarios are directly taken from the past, but in Monte Carlo approach scenarios are simulated. The law of large numbers is employed in Monte Carlo simulation and as a number of trials goes infinity, the estimator converges to the exact value of the true estimate (Huynh, Lai & Soumaré, 2008).

Parametric methods

VaR can be modelled by parametric approaches e.g. Variance-Covariance method and Extreme Value Theory (EVT). Typical for parametric methods is that they make distributional assumptions. The idea behind parametric approaches is to fit a specific probability distribution to the data and approach risk estimation from the fitted curve. Parametric methods allows easy way to conduct risk estimation, although the empirical distribution may not necessarily belong to any specific parametric distribution family that one would suggest. (Novak, 2012) Thus parametric method is sensitive to defects. However, parametric approaches are preferred over non-parametric specifically due to additional distributional assumptions. (Dowd, 2002)

Variance Covariance method

Variance-Covariance method was introduced in 1994 and it represents the traditional way among VaR methodology to assess market risk. It is also referred as delta normal method and can be approached from conditional and unconditional way. Method assumes sample returns to follow normal distribution. Loss distribution requires estimation of mean and covariance matrix. (McNeil, Frey & Embrechts, 2005) Variance-Covariance method for VaR calculation involves covariance matrix of the all assets of the portfolio. In Variance-Covariance method volatility determines portfolio VaR. (Alexander, 2003)

Variance-Covariance VaR formula:

$$VaR_{\alpha,h} = Z_{\alpha}\sigma_t - \mu_t$$

where α refers significance level and h for holding period. As the equation above demonstrates historical data is used to calculate mean and standard deviation for variance-covariance method. The advantage of variance-covariance method is that it is very simple and quick to calculate. The disadvantages of covariance method are its limitedness to linear relationship with respect to risk factors. Furthermore it assumes that returns are normally distributed, which underestimate the tail loss of the distribution (McNeil, Frey & Embrechts, 2005). The assumption to capture all historical data is not realistic. Therefore, the accuracy of the variance-covariance is related to quality of parameters in the model. (Alexander, 2003)

Limitations of VaR

The disadvantages of VaR are its incapability to convey risks further in the tail and relative exposure between assets. First, risk estimate of VaR is strictly restricted to maximum loss of given probability. Thus the extreme events are not incorporated beyond the given confidence level e.g. 99% and VaR gives us no indication of how much the loss might be in that case. This can result a situation that produce small gains in most of circumstances and the occasional large losses that cannot be handled. (Homescu, 2014) Second, in some cases, number of different assets might result similar VaR estimates, but if the tail of distribution is studied more intense the results could show huge variation among observations beyond the estimated VaR confidence level. It must remember that VaR is not designed to account these extreme observations, therefore it is a lack of property of VaR, but as a defense, there were no intention to do so. The relative exposure can be further illustrated by concerning two assets with same VaR estimates for the next day at given 99 % confidence level. The assets risk relations differs from the point that VaR is defined. Beyond that point, the assets are differently exposed to risks, where one might have exceeded the loss of 10% by number of occasions and the other only once or twice. At the same time as the VaR estimates are solely compared, they show within that point they share the same indication of the prevailing risk exposure. (Chen, Giles & Feng, 2012) The portfolio VaR should not be greater than the sum of individual VaR’s of the constituents of the portfolio.

(Rachev, Stoyanov & Fabozzi, 2011) The relative exposure between assets is also known as subadditivity property. To overcome shortcoming related to VaR, expected shortfall (ES) has to be considered.

2.3.2. Expected shortfall

Expected shortfall is referred by different names, such as expected tail loss and conditional VaR. Expected shortfall is an expected value of loss exceeding the VaR. The difference between ES and VaR can be illustrated by the amount that we can expect to lose in a case the tail event occurs (Dowd, 2002). Another parting property of describing the difference between VaR and ES is the share of the loss. VaR is the maximum loss within the threshold, whereas ES is the average or mean loss beyond the threshold. (Homescu, 2014)

Expected shortfall at a confidence level α, is (Zivot & Wang, 2006):

$$ES_a = E\left[X \mid X \geq VaR_a\right]$$

Expected shortfall overcomes the sub-additivity problem of VaR and thus can be characterized as a coherent risk measure. In addition ES requires finite mean of a return distribution. (Novak, 2012)

Both of the risk measures VaR and Expected Shortfall are prone to estimation error of parameters. Thus the accuracy of describing the true tail risk is questionable. Due to the model error and possible insufficiency of the data these risk measures might not able to disclose the true tail risk. The accuracy of VaR and ES tends to be lower at higher confidence levels; in some cases ES might have relatively lower accuracy than VaR (Isogai, 2014).

2.3.3. EVT VaR and ES

Peaks Over Threshold (POT) VaR and ES are estimated based on the Generalized Pareto distribution fitted to tail returns of sample, where the parameter estimates of the sample returns are plugged into the formula.

The POT VaR is calculated as:

$$VaR_{\alpha} = u + \frac{\sigma}{\xi}((\frac{n}{N_u} * a)^{-\xi} - 1)$$

where u is threshold, σ is scale parameter, ξ is shape parameter, n is number of observations, N_u is number of observations above threshold and a is confidence level. (Dowd, 2002)

POT ES is calculated as:

$$ES_{\alpha} = \frac{VaR_{\alpha}}{1 - \xi} + \frac{\sigma - \xi u}{1 - \xi}$$

The assumptions of traditional EVT model do not facilitate the needs of financial data. Volatility clustering and homogenous Poisson process of exceedance above threshold does not apply to characteristics of financial data. One solution is to declustering the data and another is to modeling returns first with GARCH type model and then utilize POT model to the residuals generated by GARCH model. GARCH-EVT model performs better when volatility is stochastic, i.e. returns are not i.i.d. as volatility varies over time.

2.3.4. Tail risk measures other than VaR and ES

The purpose of this section is to introduce some other risk measures besides VaR and ES to facilitate benchmarking. Four additional risk measures are considered: Cornish Fisher VaR (MVaR), Tail Risk based on Cornish Fisher expansion(TR_{CF}), Maximum Drawdown (MDD) and Kataokas's safety first (KSF), respectively.

2.3.5. Cornish Fisher Expansion

Cornish Fisher (CF) expansion is a general method applied for risk measurement in taking account of non-normal features of data. CF is designed to take account of higher moments of distribution of asset returns. Thus, while standard VaR considers variance as a measure of risk, CF can take more sophisticated risk measures into analysis. Cornish Fisher can be viewed as the Taylor expansion around the normal distribution. The Taylor expansion adjusts the critical value $z(a)$

of normal parametric VaR to skewness and kurtosis of the empirical distribution (Gregoriou, 2009)

2.3.6. Modified VaR (MVaR)

VaR formula for normal distributed returns can be replaced by adjusting skewness and kurtosis for fat-tailed data by Cornish-Fisher expansion. The Cornish-Fisher VaR is build by average return, standard deviation of return series and so called critical value or more sophisticated Cornish Fisher factor. CF expansion considers losses and gains separately i.e. an asymmetric way. (Bali & Gokcan, 2004)

MVaR is calculated as:

$MVaR_\alpha = r_p + Z_{CF}^a \sigma_p$, where Z_{CF}^a is calculated on the basis of the fourth order CF expansion. (Pätäri, 2011)

$$Z_{CF}^a = z(a) + \frac{1}{6}(z(a)^2 - 1)S + \frac{1}{24}(z\big(a)^3 - 3z(a)\big)K - \frac{1}{36}\big(2z(a)^3 - 5z(a)\big)S^2$$

2.3.7. Tail Risk based on Cornish Fisher expansion

Tail Risk (TR) is a measure of loss beyond value at risk level (Bali, Demirtas & Levy, 2009). Tail Risk measures variance of loss beyond VaR level, and therefore separates TR from risk estimates of mean loss, such as Expected Shortfall. (Bali, Demirtas & Levy, 2009).

Tail Risk is defined as follows (Bali, Demirtas & Levy, 2009):

$$TR_\Phi(R_t) = E\,[(Rt - E\,(R_t|R_t \le VaR_\Phi(R_t))^2\,|R_t \le VaR_\Phi(R_t)]$$

Tail Risk based on Cornish Fisher expansion and can be calculated as:

$TR_{CF_p} = \sqrt{\frac{(r_p - \overline{r_p})^2}{n}}$, for all $r_p < MVaR$, where n is the number of outcomes for which $r_p < MVaR$. (Pätäri, 2011)

2.3.8. *Maximum Drawdown*

Downside risk measures data back to Safety first measure of risk. Maximum drawdown is maximum loss from peak to bottom of the time series and it is referred as a maximum bound for VaR. Maximum drawdown is among the key risk measures of survival probability, i.e. probability of hitting the stop-loss that causes liquidations. (Lopez de Prado & Peijan, 2004) Drawdown is a downside risk measure given by the sum of sequential negative returns. (Leal & Vaz de Melo Mendes, 2005) Maximum Drawdown is calculated as:

$$MDD = 1 - \frac{P_{Min}}{P_{Max}}$$

2.3.9. *Kataoka safety first*

In principle, Kataoka's Safety First measure is build on an assumption that investors prefer to limit their risk of unfortunate events rather than maximize their utility. Kataoka's safety first maximizes the lower limit in a way that the probability of return less than or equal to lower limit is not greater than preferred value. (Pätäri, 2011)

Kataoka's criterion can be defined as follows:

$\max r_t \ subject\ to\ P_r\ \left(r_p < r_t\right) \leq a,$

where a is a probability of return less than the lower limit. (Pätäri, 2011)

Kataoka safety first can be obtained by:

$$r_p = E(r_t) - \ 1.645\sigma_p$$

Maximizing the lower limit is equivalent to maximizing the lower insured return or expected mean of returns (Ding & Zhang, 2009). $E(r_t)$ is mean return of the portfolio, 1.645 represent the critical value of 95 percentile of normal distribution and σ_p is the standard deviation of the portfolio. Kataoka safety first is the second safety criterion after Roy's (1952). According to KSF, the best performing asset is the one that have highest mean relative to standard deviation. (Elton et al., 2009)

2.4. Backtesting risk estimates

Accuracy of risk measure is confirmed by backtesting at given confidence level. In other words the data from out of the sample period confirms whether the estimated risk measure is exceeded in future. Thus backtesting involves comparing estimated risk measure to future realizations of the returns (Franke, Härdle & Hafner, 2011). However, despite of the ideal procedure of backtesting, violations takes place in rare occasions (Daníelsson, 2011). It takes time that realizations are noticeable.

Backtesting determines whether specific model is appropriate for risk estimation at certain confidence level. Basel Committee on banking Supervision (1996) has outlined framework for backtesting method, in which the magnitude of violations are considered. Capital requirements are increased as the violations increases.

Backtesting framework proposed by McNeil and Frey (2000) is based on violations procedure. The violation occurs as realized risk exceeds the estimated risk. This is an unconditional method and does not take account of time period between violations. Conditional backtesting overcomes this inter-arrival problem between violations. Higher market turbulence causes clustering of events. Thus the model cannot capture volatility and correlations when violations tend to appear in clusters. Accurate model capture coalitions in such a way that today's violations does not depend of violations occurred in previous days. Independence of violations can be tested by likelihood ratio tests. The likelihood ratio test is Chi-square distributed, where accuracy is achieved by lower test statistic than critical value of Chi-square.

3. Data and Methodology

3.1. Data

The data consist of four Gulf region market indexes: Saudi Tadawul All Share index (TASI), Oman Muscat Securities MSM index, Kuwait SE market index and United Arab Emirates ADX General Price index. For comparison purposes, Standard and Poor's 500 index is incorporated. Saudi price index covers the period from 1998 to 2006, Oman Muscat index from 1996 to 2006, Kuwait SE index from 2000 to 2006 and ADX General from 2001 to 2006. SP500 index comprises the period from 1994 to 2006. The data is taken from DataStream and prices are in US dollars.

All index prices are computed to returns by following formula:

$$r_t = ln\,(p_t/p_{t-1})$$

The transformation to negative daily returns $-r_t$ is performed for EVT analysis, where only the positive values are taken under analysis. Therefore losses are possible to study by turning them to positive values. Backtest period is chosen from 2007 to 2014.

3.2. Methodology

This section shows how the theoretical aspects are used in this study. The main goal is to measure tail risk by selected risk measures to provide comparable results. Risk measures being compared are Value at Risk (VaR), Expected Shortfall (ES), Peaks Over Threshold (POT) VaR and ES, Tail Risk based on Cornish Fisher expansion (TR_{CF}), Modified VaR (MVaR), Maximum Drawdown (MDD) and Kataoka Safety First (KSF). Confidence intervals of 95 % and 99% are used for comparison purposes.

Non-parametric VaR and ES are based on historical simulation method, which utilize empirical data. VaR and ES are quantiles from the probability distribution of the returns, which are shown by simple formulas (sections 2.3.1 and 2.3.2). In addition to non-parametric technique, VaR is approached by modified normal distribution to comprise skewness and kurtosis. Here TR_{CF} and MVaR utilize the Cornish-Fisher expansion to incorporate skewness and kurtosis into VaR estimation (sections 2.3.4 and 2.3.5). Extreme value theory restricts the tail from the body of the distribution. EVT is parametric method that fits theoretical heavy tailed GPD distribution to the tail of the data. VaR and ES are then calculated from the Poisson process of exceedance (section 2.3.3). MDD and KSF are also calculated based on simple formulas for additional risk assessment (sections 2.3.6 and 2.3.7). Maximum drawdown does not make any assumptions of returns, while KSF rely on normally distributed returns.

The distributional characteristics such as moments are summarized via basic statistics. Comparison of empirical distributions to theoretical ones is preceded by graphical tools e.g. quantile plots. Return series characterization is carried out by statistics of autocorrelation function to illustrate assumptions of independence. Results of estimated parameters are provided to support an analysis of heavy tails. Graphical methods have been used for proper threshold estimation of EVT-based VaR calculation and volatility is assumed to be constant for simplicity. In this study unconditional backtest procedure is utilized for reason of simplicity. Therefore, the results are also analyzed from the viewpoint of these premises.

4. Results

This section of the study answers to questions that have been addressed beforehand. How do the tail risk measures perform in Gulf markets and are these results comparable to more developed markets? It seems worth to note that indexes do not share the same time period.

Empirical evidence of stylized facts of financial data is concentrated on volatility clustering, asymmetry, high peakedness and fat tails (Stoyanov et al., 2011). Carrying out stylized facts of empirical data begins by distributional moments and basic statistics. Mean, standard deviation, skewness and kurtosis are supported by additional measures for distinctive characterization of stylized facts which are mainly concentrated on comparison to normal distribution. Here the model assumptions are confirmed regarding to independent and identically distributed returns based on normal distributed return series and later, Poisson nature of EVT based model. Once the assumptions are confirmed the risk estimates can be analyzed.

	Oman	UAE	Kuwait	Saudi	SP500
No. Observations	2658	1435	1825	2139	3201
Mean	0,0413 %	0,0754 %	0,1097 %	0,0777 %	0,0351 %
Skewness	1,4413	0,1234	-0,4994	-0,5117	-0,1093
Kurtosis	69,5680	11,6502	10,2170	16,0902	3,8489
Std. dev.	0,0107	0,0117	0,0099	0,0150	0,0105
Minimum	-0,1483	-0,0865	-0,0809	-0,1169	-0,0711
Maximum	0,1985	0,0825	0,0614	0,1622	0,0557

Table 1 Basic Statistics of four Gulf markets and SP500.

Moments provides basic indication and introduction to the market indexes. Central tendency of all markets seems to be near to the value zero, as means appear to be positive it indicates that markets rise on average. Standard deviation represents second moment of the probability distribution is near zero indicating narrow dispersion. Skewness of the Gaussian distribution is zero, consequently negative values of skewness indicate that the data is left skewed and positive values indicate skewness to the right. Empirical evidence of Gulf markets show negative values of skewness for Kuwait, Saudi and SP500 referring losses are more extreme than the returns, whereas Oman and UAE show positive skewness. The Kurtosis is fourth moment of the distribution and for the normal distribution it is three. However kurtosis of three does not necessarily confirm the normality. The kurtosis values for Gulf market indices are positive and very high, above all kurtosis of Oman is relatively high compared to other index returns. High kurtosis as an indicator of fat tails suggests returns are not sampled from the Gaussian distribution, meaning realizations of extremes are more probable. Statistics of distributional moments divides markets to three different groups. First group the most extreme includes Oman and Saudi, second group UAE and Kuwait represent more extreme than third group which consists of SP500 index characterizing nature of developed markets.

4.1. Independent and Identically Distributed (IID) returns

Empirical evidence against independent and identically distributed returns is provided by return plot, extremal index and plot of ACF of returns. Volatility clustering is evident in return plot and also supported in extremal index (Figure 3. and table 5. in appendix) referring violation against assumptions of independence of returns. In return plot volatility is shown by periods of more and less intense fluctuation and extremal index indicates volatility by value lower than 1. Rapid decay of ACF in turn refers independent returns, which is supported by empirical results of sample markets (Figure 25. and Figure 26. in appendix). The assumptions of independence and identical distribution among sample returns turn out ambiguous results.

4.2. Non-normal features

Minimum and maximum values differ a bit as compared to mean. However, this difference gives no indication of the riskiness of indexes. More depth analysis of these down moves shows e.g. extreme observation of Saudi is roughly 11.7% negative, which is 8 standard deviations away from the mean, violating thereby the assumption of Gaussian distributed returns (Carmona, 2014). Abnormal values for Oman show that minimum value is 14 standard deviations away from the mean, whereas the minimum for UAE is 7 standard deviations away from the mean and for Kuwait 8 standard deviations away. SP500 index, as a proxy for more developed markets, shows the maximum negative value of 7 standard deviations below from the mean. These results clearly typify abnormal returns. Furthermore, according to normal distribution 99.7% of all observations should be within 3 standard deviations of the mean. 7 standard deviations from the mean correspond to odds of 1 to 390 billion. That is, under normal assumptions, an event of 7 standard deviations away from mean happens once in 1.07 billion years.

As the analysis of abnormal returns indicates normal distribution is not appropriate distribution for sample returns. This contradiction is also illustrated in figure 10 in appendix, where the empirical returns are standardized. Returns should follow the reference line according to normal assumptions. Contrary to assumptions, returns turn counter-clockwise from the tails indicating heavy tails while the centre part of the distribution behaves like normal distribution. In addition, indications of relative difference between heaviness of positive to negative tails are not observable.

Thus, neither negative nor positive tail shows heavier features of one another. The histograms in Figure 4 in appendix support the implication of return plot and show much quicker decay than normal distribution. Excess kurtosis is also evident which refer from heavy tails. However, otherwise the tails are rather difficult to interpret from histograms.

Distribution features can be further confirmed using finite moments as an indicator of non-normality. Here heavy tails are then again confirmed by Maximum to sum ratio (MS) plot which approaches heavy tails by approximation order of finite moments of the distribution (Figure 8. in appendix). The convergence to zero suggests that the moments of that order may be finite, otherwise the reasoning convey infinite moments. The thin tailed distribution such as normal converges to 0 for all moments. Thin tailed features are presented in markets of Kuwait and Saudi. For Oman and UAE, the higher moments seems to be infinite, albeit the first moment is finite. Here the MS plot confirms both finite mean and infinite fourth moment for Gulf market indexes which in turn indicates existing of fat tails.

4.2.1. Pickands Plot

Pickands estimator estimates extreme value index that is usually expressed as Greek letter gamma γ (Figure 9 in appendix). Extreme value index is also called tail index and it indicates decay of the tail of the distribution. The tail index of $\gamma < 0$ indicates finite endpoint of the distribution, while $\gamma = 0$ represents exponential decay of tails and $\gamma > 0$ indicates polynomial decay of tail referring to heavy tail which in turn convey Pareto behavior. (Beirlant, Dierckx & Guillou, 2005) Pickands plot suggests that the shape parameter is near zero and positive. The horizontal axis represents ordered values in increasing order, where the first value is highest negative return. The plot indicates variation for lowest orders. Assuming this is rather standard behavior Pickands plot indicates positive behavior of shape parameter, as the line fluctuates above zero, referring to polynomial decay i.e. heavy tails for all Gulf markets.

To summarize the results up to this point, empirical evidence from sample markets indicates volatility clustering, i.e. non i.i.d. behavior, high peakedness and abnormal minimum values. Heavy tails as non-normal features of the sample markets has been confirmed first by moment estimates of the probability distribution, abnormal values of return, and then by plots such as QQ plot, histogram, MS plot and Pickands plot. ACF (Figure 25 in appendix) does not indicate violation against assumptions of independent returns, which contradicts with return plot and extremal index statistic.

The characteristics of developed markets show that volatility clustering is evident (Figure 20 in appendix) though ACF of returns does not indicate volatility clustering (Figure 26 in appendix). Standardized Q-Q plot of developed market indicates heavy tails and thus supports the perspective of emerging markets (Figure 20 in appendix). In contrast with emerging markets, MS ratio plot shows that all moments are finite (Figure 21 in appendix). Pickands plot describing the shape parameter turns out to be unreadable, as the fluctuation is extremely high (Figure 22 in appendix). Compared to Gulf markets, the results of developed market show less extreme behavior and the assumptions of normality can be rejected by much clear margin in Gulf markets than in developed market.

After analyzing stylized facts of empirical returns the EVT model is conducted for tail estimation. The EVT method provides indication of heavy tail behavior concentrating on the tail of the distribution only. Therefore, threshold estimation for POT model is a starting point.

4.3. Threshold

Threshold approximation is conducted for the EVT model by graphical methods where Mean Excess plot (ME plot), Zipf plot and Mean Residual Life plot (MRL plot) are in key role. ME-plot and MRL-plot are conventional graphical methods for threshold determination. In addition Zipf-plot shows Pareto behavior of tail where the threshold point comes imperative. All plots utilizes positive values, and therefore, illustration uses transformation of negative returns to study high losses.

Threshold determination must be conducted by avoiding bias, and therefore, the interest should be concentrated on appropriate amount of observations. Threshold is detected from the point by observing the Paretian behavior i.e. upward sloping trend in the ME-plot (Figure 5 in appendix). Regardless of the difficulties in the interpretation of mean excess plot, the appropriate threshold value 0.02 is approximated for all markets. The mean residual life (MRL) plot is similar to mean excess plot but here the confidence bands are used to indicate the fit of the model to the empirical data. MRL plot for all the markets indicates linear behavior up to a point where threshold u > 0.05 (Figure 7 in appendix). However, all the graphs seem to show negative linear relationship beyond threshold value 0.05. The negative slope trend might indicate volatility of the data. The choice of threshold is rather difficult, though some indication can be made. For Oman, linearity seems to hold up to point 0.05, which could facilitate the means of threshold. Similar indications lead to a threshold of 0.05 for UAE, 0.02 for Kuwait and 0.02 for Saudi. Larger amount of data could draw more lucid picture of threshold. The third graph for threshold approximation is Zipf plot; where the Paretianity is conducted as negative linear behavior from the point where a threshold value is recognized (Cirillo, 2013). Thereby Zipf plot indicates quite similar values than the mean excess plot. Here the threshold appears to be close to zero and thus gives no additional light in the issue (Figure 6 in appendix). All three plots (i.e., ME-plot, Zipf-plot and MRL-plot) suggest similar views of threshold selection. Alike to Gulf markets, developed markets indicate increasing feature of ME-and MRL-plots above point 0.02. Therefore, the appropriate threshold could be near that value as long as the number of observations above that point is line with other markets (Figure 21 in appendix).

4.4. MLE estimation of parameters

The Peaks over threshold method approaches VaR calculation using maximum likelihood estimation to determine the parameter values. The maximum likelihood method for estimating parameters is based on numerically maximizing the likelihood function for the given excess data, hence choosing a threshold parameter u is inevitable (Finkenstädt & Rootzén, 2004). Threshold is chosen as $\eta = 2.0\%$ for Oman, $\eta = 1.8$ for UAE and Kuwait and $\eta = 2.5$ for Saudi and SP500 index. The aforementioned threshold selection is based on graphical presentation of data and sufficient amount of exceedance (shown in the Figures 5, 6 and 7 in appendix).

		Threshold	1,50 %	
Market	Exc.	Shape ξ	Scale α	Location β
Oman	103	0.2894 (0.0417)	0.0031 (0.0003)	-0.0080 (0.0013)
UAE	76	0.2851 (0.0579)	0.0038 (0.0006)	-0.0024 (0.0018)
Kuwait	75	0.0723 (0.0672)	0.0087 (0.0021)	-0.0163 (0.0048)
Saudi	152	0.0061 (0.0699)	0.01873 (0.0047)	-0.0349 (0.0084)
SP500	216	0.1169 (0.0330)	0.0045 (0.0004)	0.0007 (0.0008)
		Threshold	2,0% Oman / 1,8 % UAE & Kuwait / 2,5 % Saudi & SP500	
Market	Exc.	Shape ξ	Scale α	Location β
Oman	61	0.3935 (0.0381)	0.0018 (0.000002)	0.0048 (0.0014)
UAE	50	0.0662 (0.0885)	0.0107 (0.0037)	-0.0223 (0.0087)
Kuwait	55	0.0211 (0.0816)	0.0115 (0.0038)	-0.0239 (0.0085)
Saudi	92	0.0163 (0.0841)	0.0177 (0.0060)	-0.0321 (0.0123)
SP500	51	0.3538 (0.0367)	0.0012 (0.000002)	0.0137 (0.0011)
		Threshold	3,00 %	
Market	Exc.	Shape ξ	Scale α	Location β
Oman	20	0.6502 (0.0650)	0.0004 (0.000002)	0.0164 (0.0030)
UAE	22	0.0210 (0.1249)	0.0132 (0.0078)	-0.0276 (0.0206)
Kuwait	25	0.0186 (0.0964)	0.0103 (0.0046)	-0.0160 (0.0123)
Saudi	69	-0.0231 (0.0968)	0.0215 (0.0090)	-0.0409 (0.0189)
SP500	25	0.4714 (0.0501)	0.0005 (0.000002)	0.0198 (0.0016)

Table 2 MLE estimated parameters with three different thresholds

Confirming the good fit of the model is discovered by checking how the parameter estimates behave across a range of thresholds (Finkenstädt & Rootzén, 2004). The parameters do not seem to vary much with different thresholds, that refers to good fit of the model.

The *shape* parameter of Oman seems to rise as the threshold increases, and the same also holds for SP500. For UAE, Kuwait and Saudi, the shape parameter tends to diminish as the threshold grows. Negative value of shape parameter indicates features of normal distribution and positive heavy tailed distributional features. The *scale* parameter of Oman and SP500 declines as the threshold increases, whereas for UAE, Kuwait and Saudi, it increases as the threshold grows. The *location* parameter grows as the threshold increases for Oman and SP500, but for Kuwait, Saudi and UAE, the location parameter falls as the threshold grows. Interesting the parameters of different markets move into different directions as the threshold grows. The variability of parameters doesn't seem to be so huge that it should cause any attention.

The shape parameter shows positive sign for all the markets. This implies that the upper bound is unbounded and higher moments exist, indicating heavy tails. The shape parameter for Saudi turns out to be negative when the threshold is above 2,5%. For example when the threshold is 3,0% the shape parameter turns to negative. That is also the reason behind the value of Saudi threshold, although the higher threshold would be more congruent compared to other markets at least as measured by the number of exceedances over the threshold. The chance of dropping 2.0% or more in a day for Oman occurred with probability 2.3%. Similarly for other indexes the chance of dropping 1.8 % for UAE, 1.8% for Kuwait, 2.5 % for Saudi and 2.5% for SP500 index occurred with probability 3.5%, 3.0%, 4.3% and 1.6%, respectively. The probability corresponds also to the percentage of the observations exceeding the threshold out of total sample.

Remaining parameter values of Generalized Pareto distribution result following; the scale parameter shows positive sign, which supports the assumption of GPD with positive shape parameter. The location parameter shows mainly negative values indicating negative mean. Standard errors look relatively low and don't cause any concern.

4.5. Comparison to Paretian distribution

In Zipf plot the logarithm of the empirical survival function is plotted against the logarithms of the ordered values of returns. Negative linear relationship in the graph implies power law nature (Cirillo, 2013). Correspondingly, lack of negative linearity of the Zipf plot suggests departure from Pareto behavior. Based on these results, there is not much evidence of linear features among subject markets observed herein (Figure 6 in appendix), as empirical survival function shows mainly normal or Weibull, $\alpha > 1$, features. In more detailed examination, the Zipf plot starts to curve and the linear relationship is observable for all markets when $x > 0.01$. In other words, for these values, the plot suggests possible presence of a Paretianity, i.e. heavy tail.

Mean Excess plot shows power law trend is apparent, referring fat tailed distribution, when the returns show increasing positive slope (Figure 5 in appendix). For all markets, power law trend tends to show up values near zero. Oman's mean-excess plot shows the most positive linear trend and the plot is less erratic than the other ME-plots. All markets show features of positive linear slope until they became volatile, which is rather standard behavior. Overall deviance from Pareto behavior can be considered based on results over here. Features of Pareto behavior of developed market is apparent above threshold point, which is supported by Zipf plot. (Figure 21 in appendix)

4.6. Fit of Peaks over Threshold model

The Peaks Over Threshold (POT) method is used to achieve an accurate estimate of tail losses by inferring the tail of the distribution only. The POT method uses positive values of returns and therefore, lower tail values must be converted to positive values to incorporate into analysis. Graphical analysis concentrates on time series features to support assumptions of the Poisson distributed exceedance and Generalized Pareto distributed excess returns.

The number of events in a non-overlapping time interval, i.e. referring frequency or occurrence rate of observations above threshold, should be independent and Poisson distributed while the distribution of time between events, i.e. interarrival time is assumed to be independent and exponentially distributed.

The distribution of exceedances is characterized by their intensity, which is approximated by GPD. Intensity refers to amount of excess observations exceed the threshold. As the excess distribution is considered to characterize GPD nature the exponential distribution is used as reference distribution of residuals, which is special case of GPD when $\xi = 0$. However, this is just one of the distributions that GPD allows. As the tail parameter ξ is non-zero in this case, the excess takes

GPD as a limiting distribution. The model assumptions regarding to Poisson and GPD process of i.i.d. returns are studied at Figures 13 and 16 in appendix.

4.6.1. Poisson nature of exceedance

The Point process graph shows that the exceedances are clustered (see Figure 13 in appendix). This violates the POT model, which assumes independent and identically distributed returns. Indeed, under homogeneous Poisson process, the exceedance times should be independent on exponential distributed random variables. For developed markets clustering of events are also shown (Figure 23 in appendix).

Plots 14 and 15 in appendix assess the Poisson nature of the exceedance process by looking at the scaled gaps, i.e. time, between consecutive exceedances, which should be i.i.d. unit exponentially distributed. Indeed independence of the observations can be examined by analyzing waiting time between events, which should be exponentially distributed. Quantile plot of gaps deviate considerable from the exponential line, giving evidence against Poisson process. Auto Correlation Function (ACF) of gaps suggests that exceedances are clustered, thus supporting Quantile plot of gaps in violating the Poisson assumptions again. For the developed market, the quantile plot of gaps doesn’t show perfect fit, implying that Poisson assumptions cannot be confirmed. However, Auto Correlation of gaps supports Poisson nature (Figure 23 in appendix).

4.6.2. GPD nature of excess

The exponential QQ-plot shows threshold exceedance, and the fit is reasonably good among all markets at the scope (Figure 11 in appendix). Exponential distribution is interpreted by looking features of concave and convex departure from the reference line, indicating heavy tail and thin tail behavior. The QQ-plot is quite linear and does not show much deviation from the exponential line. Oman’s graph shows clear deviation from the diagonal suggesting that exponential distribution is not appropriate for tail of the data. The submission of QQ-plots is to make interpretation of outliers, whether they are result of long-tailed process or simply irregular. Therefore outliers should not immediately be treated by elimination (Finkenstädt & Rootzén, 2004). The observations are ordered from the smallest to the largest, and the largest observations should be given more weight.

The difference between exponential QQ-plot and the QQ-plot of GPD where the reference line is GPD, i.e. when the shape parameter is non-zero, is illustrated in Figure 11 and 12 in appendix. This comparison illustrates whether the excess losses are thin tailed, (i.e. exponential) or fat tailed (i.e. GPD distributed). (Zivot & Wang, 2006) At least for Oman reference distribution of GPD imply better fit as the observations follow the reference line (figure 12 in appendix). For other markets, the superiority is harder to interpret. Overall, the fit looks good. For SP500 reference line of GPD shows better adjustment to returns as exponential one (Figure 22 in appendix).

Figures from 16 to 19 in appendix assess the GPD nature of the excesses. Figures 16 and 17 in appendix show QQ plot and ACF of residuals, which should be i.i.d. GPD distributed**.** Plot of excess distribution and plot of tail of the underlying distribution are shown in Figures 18 and 19 in appendix.

QQ plot of residuals shows that the plotted points do not deviate much from the diagonal, indicating that the exceedances are GPD distributed. In addition, there is not much evidence of autocorrelation in the residuals (Figure 17 in appendix). However, ACF of Kuwait shows some autocorrelation, implying that returns are serially dependent. Excess distribution graph corresponds to the shape of GPD, indicating that GPD provides a good approximation for the excess returns (Figure 18 in appendix). Here again market index of Kuwait show minor deviations. Tail probability estimate in Figure 19 in appendix shows minor deviations from a straight line. Tail plot of the GPD is fitted together with the empirical tail given by the actual data points in logarithmic scale. Oman’s tail plot indicates Pareto behavior, as the other markets show thin tailed features i.e. Weibull and Gumbel. The developed market shows that excess follows GPD distribution, as the QQ-plot and ACF of residuals indicate. Excess distribution shows firm fit and tail plot shows decreasing slope indicating Pareto behavior (Figure 24 in appendix).

Graphical tools are not easy to interpret and there is always question about the adequacy of sample. Up to this point, distributional and time series assumptions have been considered in detail. Henceforth tail risk is estimated and interpreted.

4.7. Risk estimates

The first impression of results indicates much higher risk estimates of ES than VaR. Historical methods results lower tail risk estimates than other tail risk methods, and the difference comes as the quantile increases to 99 %. The highest VaR estimates provide MVaR (Table 3). Maximum drawdown shows similar features among all markets and standard drawdown is just above 50%. Kataoka Safety First suggests the best performing index is that of Kuwait, where KSF of 95 (99) percent illustrates that 1,52% (2,20%) negative loss won't exceed as the 95% (99%) confidence level. SP500 index in general indicates lower risk estimates compared to Gulf markets.

Table 3 shows one day horizon tail risk estimate for the tail probabilities of 0.05, 0.01 and 0.001, which corresponds to 5%, 1% and 0.1 % VaR and ES, respectively. It seems worth to mention that the thresholds in POT approach vary along different indexes. The result of non-parametric VaR and ES estimates divides markets to two distinguishable different groups. Estimates of Oman and Kuwait can be generalized as comparable to SP500- index, whereas UEA and Saudi show much higher risk relation.

In POT approach, confidence levels of 95 and 99 percent divide markets quite similarly as already mentioned in the cases of standard VaR and ES approaches. The higher confidence level implies relatively higher risk estimates for Gulf markets than in more developed markets of SP500- index, also the relative exposure of market indexes is evident, e.g. some markets share same risk level at lower confidence level while at higher level the exposure differs. Indeed, estimates of Oman and Saudi are vulnerable to tail risk more differently at 99 percent level than beyond that level. More intense tail risk estimation at confidence level of 99.9 percent suggests that the aforementioned markets share equal risk exposure. The higher confidence level denotes analogous risk level for Kuwait and SP500 index, market index of UAE remains between two groups.

MVaR and TR_{CF} continue the same expression of tail risk than earlier results where markets divides to different groups according to risk level. Confidence level of 95 percent manifests relation between Kuwait- and SP500- index, while UAE and Saudi persist more Gulf market specific. The higher risk level shows different orientation between Gulf– markets and developed markets. MVaR result of Oman was not reliable, which causes absence of TR_{CF} tail risk estimates.

KSF separates the Saudi index as more extreme market from the rest of indexes. Here KSF gives uniform results even though perspective of tail risk might be harder to distinguish. Estimation of MDD mixes the former allusion of tail risk on behalf of SP500 as it participate in to the category of higher risk markets than based on the majority of risk comparisons.

Overview of results favor tail risk estimates based on the Cornish-Fisher expansion over standard VaR and ES. In addition, higher confidence level brings out important characteristics of markets. KSF can be employed with the non-parametric historical VaR-and ES-and POT-based tail risk methods, which cannot distinguish tail risk between different markets in a way that characteristics and nature of specific market comes out.

	Oman	UAE	Kuwait	Saudi	SP500
VaR_95	1,28 %	1,55 %	1,29 %	2,17 %	1,70 %
VaR_99	2,76 %	3,47 %	3,30 %	4,96 %	2,77 %
ES_95	2,35 %	2,74 %	2,46 %	4,12 %	2,42 %
ES_99	4,46 %	5,03 %	4,45 %	7,19 %	3,71 %
POT_VaR_95	1.48%	1.32%	1.18%	2.22%	2.01%
POT_VaR_99	2.76%	3.54%	3.18%	5.25%	2.76%
POT_VaR_99.9	6.78%	7.16%	6.18%	9.73%	4.94%
POT_ES_95	2.42%	2.72%	2.43%	4.11%	2.55%
POT_ES_99	4.53%	5.10%	4.48%	7.19%	3.71%
POT_ES_99.9	11.15%	8.97%	7.54%	11.74%	7.08%
MVaR_95	-0,26 %	1,53 %	1,45 %	2,11 %	1,65 %
MVaR_99	17,88 %	5,72 %	4,83 %	9,46 %	3,44 %
TR_CF_95	na.	3,16 %	2,99 %	4,54 %	2,55 %
TR_CF_99	na.	7,77 %	6,03 %	10,52 %	4,62 %
MDD	71.28%	55.94%	24.21%	67.21%	52.43%
KSF_95	-1.72%	-1.85%	-1.52%	-2.39%	-1.70%
KSF_99	-2.45%	-2.65%	-2.20%	-3.42%	-2.42%

Table 3 Risk estimates from period 1994-2001 to end of 2006

4.8. Backtesting

Backtesting is conducted to see how the risk measures behave under future market conditions. Time period is chosen from 2007 to 2014. The result of Oman remains incomplete due to unreliable estimates of previous section (Table 4).

	Oman	UAE	Kuwait	Saudi	SP500
VaR_95	**1**	**1**	**1**	0	**1**
VaR_99	**1**	**1**	0	0	**1**
ES_95	0	0	0	0	0
ES_99	0	0	0	0	**1**
POT_VaR_95	0	**1**	**1**	0	**1**
POT_VaR_99	**1**	**1**	0	0	**1**
POT_VaR_99.9	**1**	0	0	0	**1**
POT_ES_95	0	0	0	0	0
POT_ES_99	0	0	0	0	**1**
POT_ES_99.9	0	0	0	0	**1**
MVaR_95		**1**	0	0	**1**
MVaR_99		0	0	0	**1**
TR_CF_95		0	0	0	0
TR_CF_99		0	0	0	0
MDD					
KSF_95	0	0	0	0	**1**
KSF_99	**1**	**1**	**1**	**1**	**1**

Table 4 Backtest from 2007 to 2014. Number 1 refers violation and number 0 no violation.

Backtesting shows that non-parametric VaR induce violation likewise Peaks Over Threshold method of Extreme Value Theory. Expected Shortfall turns out almost well-behaving method, as Tail Risk based VaR along with MVaR that also represents adequate measure of risk. Unexpectedly, SP500 causes more violations than other markets, even though SP500 act as proxy for more developed markets. Saudi market index was characterized the most extreme with Oman index in the basic statistics analysis. By contrast it shows almost no violations in backtesting and acts as the best performing market among all.

Performance of results remains mainly Gulf market specific, meaning SP500 does not share same reliance of risk due to large number of violations, at least at this time period and especially for higher confidence levels. Indeed, SP500 index share comparatively moderate risk estimation which experienced more violations consequently giving it distinctive disposition.

5. Conclusions

The preliminary idea was to employ EVT as a sole model to estimate tail risk in the Gulf markets, but the study ended up to comprise a set of tail risk measures. The idea formulated to provide comparable picture of tail risk among Gulf markets and more developed financial markets. Returns on Gulf markets show firmly extreme features, basically meaning highly peaked shape, as compared to developed markets. More in depth analysis of the returns at extreme level showed features of fat tails and clustering of volatility. The fit of the returns shows that Generalized Pareto distribution provides best fit for Oman and SP500, whereas other Gulf market returns may follow some thinner tailed distribution.

Risk estimates indicated superiority of skewness and kurtosis adjusted VaR methods based on normal distribution compared to historically simulated VaR and EVT-based VaR measures. Surprisingly the developed market index shows more fragile exposure to unexpected events than GULF markets. Reasons can be guessed, centralization might be the one which ties everything together causing rapid spread of effects.

The comprehensive evaluation of investment risks should center around magnitude of extremes rather than quantity. In future studies, this point should be carried out by analyzing the most extreme observations in more detail. Moreover, more competent risk measures could be considered, such as dynamic EVT-GARCH model, where the EVT is applied to residuals of GARCH model.

References

Akgiray V. & Booth G .G. (1988). The Stable-Law Model of Stock Return. *Journal of Business & Economic Statistics, vol. 6 (1).*

Alexander C. (2003). Market Models: A Guide to Financial Data Analysis. John Wiley & Sons.

Ang A., Chen J. & Xing Y. (2006). Downside Risk. *The Review of Financial Studies, vol. 19 (4).*

Arztner P., Delbaen F., Eber J. & Heath D. (1999). Coherent Measure of Risk. Mathematical Finance vol. 9 (3).

Bali T. G. (2003). An Extreme Value Approach to Estimating Volatility and Value at Risk. *The Journal of Business, vol. 76 (1).*

Bali T. G. (2007). A Generalized Extreme Value Approach to Financial Risk Measurement. *Journal of Money, Credit and Banking. 39 (7), p.1613-1649.*

Bali T. G., Demirtas K. O. & Levy H. (2009). Is There an Intertemporal Relationship between Downside Risk and Expected Returns? *Journal of Financial and Quantitative Analysis, vol. 44 (4).*

Bali T. G. & Gokcan S. (2004). Alternative Approaches to Estimating VAR for Hedge Fund Portfolios. *Intelligent Hedge Fund Investing: Alternative Approaches to Estimating VaR for Hedge Fund Portfolios.* Risk books.

Bali T. G. & Neftci S. N. (2003). Disturbing Extremal Behavior of Spot Rate Dynamics. *Journal of Empirical Finance, Vol.10 (4).*

Barone-Adesi G. & Kostas G. (2001). Non parametric VaR Techniques. Myths and Realities. *Economic Notes, 30 (2), p.167-181.*

Beirlant J., Dierckx G. & Guillou A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli vol. 11 (6).*

Billingsley P. (1995) Probability and Measure, 3rd edition. *John Wiley & Sons.*

Blattberg R. C. & Gonedes N. J. (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *The Journal of Business, vol. 47 (2).*

Byström H. N. E. (2004). Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis, 13, p.133-152.*

Campbell R. A. & Kräussl R. (2007). Revisiting the home bias puzzle: Downside equity risk. *Journal of International Money and Finance, vol.26.*

Carmona R. (2014). Statistical Analysis of Financial Data in R. Springer.

Chan K. F. & Gray P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*. *22, p.283-300.*

Chen Q., Giles D. & Feng H. (2012). The extreme-vale dependence between the Chinese and other international stock markets. *Applied Financial Economics, 22.*

Cheng P. (2005). Asymmetric Risk Measures and Real Estate Returns. *The Journal of Real Estate Finance and Economics vol. 30 (1).*

Choi P. & Min I. (2011). A Comparison of Conditional and Unconditional Approaches in Value-at-Risk Estimation. The *Japanese Economic Review, vol. 62 (1).*

Christoffersen P. (2003). Elements of Financial Risk Management. *Academic Press.*

Cirillo P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications, 392 (23), p.5947-5962.*

Clauset A., Shalizi C. R., & Newman M. E. J. (2009). Power-Law Distributions in Empirical Data. *Society for Industrial and Applied Mathematics Review vol. 51 (4).*

Coleman T. S. (2012). Quantitative Risk Management: A Practical Guide to Financial Risk. John Wiley & Sons

Coles S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

Corelli A. (2012). The tale of the tail: extreme-value patterns of financial returns. The *Journal of Risk Finance, 13 (5), p.507-516*.

Cotter J. & Dowd K. (2011). Extreme global equity market risk. *Journal of derivatives & Hedge Funds, vol. 17 (4).*

Da Costa Lewis N. (2012). Market Risk Modelling: Applied Statistical Methods for Practitioners. Risk books.

Daníelsson J. (2011). Financial Risk Forecasting. John Wiley & Sons.

Danielsson J. & De Vries C. (2000). Value-at-Risk and Extreme Returns. *Annales d'économie et de statistique, vol.60.*

Ding Y. & Zhang B. (2009). Optimal portfolio of safety-first models. *Journal of Statistical Planning and Inference, 139.*

Dowd K. (2002). Measuring market risk. John Wiley & Sons.

Duffie D. & Pan J. (1997) An Overview of Value at Risk. *Journal of Derivatives,* Vol. 4.

Eberlein E. & Keller U. (1995). Hyberbolic Distributions in Finance. *Bernoulli, vol.1 (3).*

Eftekhari B., Pedersen C. S. & Satchell S. E. (2000) On the volatility of measures of financial risk: an investigation using returns from European markets. *The European Journal of Finance, vol. 6.*

Elton E. J., Gruber M. J., Brown S. J. & Goetzmann W. N. (2009). Modern Portfolio Theory and Investment Analysis, 8th Edition. *John Wiley and Sons.*

Fama E. F. (1963). Mandelbrot and the Stable Paretian Hypothesis. *The Journal of Business, vol.36.*

Fan J. & Gu J. (2003). Semi parametric estimation of Value at Risk. *Econometrics Journal, Vol. 6.*

Favre L. & Galeano J. (2002). Mean-Modified Value-at-Risk optimization with Hedge Funds. *The Journal of Alternative Investments, vol. 5 (2).*

Finkenstädt B. & Rootzén. (2004). Extreme Values in Finance, Telecommunications, and the Environment. Chapman & Hall CRC.

Franke J., Härdle W. K. & Hafner C. M. (2011). Statistics of Financial Markets: An Introduction, Springer.

Gencay R. & Selcuk F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting, 20, p.287-303.*

Gencay R., Selcuk F. & Ulugulyagci A. (2003). High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance Mathematics and Economics, 33, p. 337-356*

Gettinby G. D., Sinclair C. D., Power D. M. & Brown R. A. (2006). An analysis of the distribution of extremes in indices of share returns in the US, UK and Japan from 1963 to 2000. *International Journal of Finance & Economics, vol. 11 (2).*

Gregoriou G. N. (2009). Encyclopedia of Alternative Investments. *CRC Press.*

Hols M. C. & de Vries C. (1991). The Limiting Distribution of Extremal Exchange Rate Returns. *Journal of Applied Econometrics, vol. 6(3).*

Homescu C. (2014). Tail risk protection in asset management. Social Science Research Network.

Huisman R., Koedijk K. & Pownall R. (1998). VaR-x: Fat tails in financial risk management. *Journal of risk, Vol. 1.*

Huynh H. T., Lai V. S. & Soumaré I. (2008). Stochastic Simulation Applications in Finance with MATLAB Programs. *John Wiley and Sons*

Hwang S. & Pedersen C. S. (2004). Asymmetric risk measures when modeling emerging markets equities: evidence for regional and timing effects. *Emerging Markets Review, vol 5.*

Hyung N. &de Vries C. G. (2005). Portfolio Diversification Effects of Downside Risk. Journal of Financial Econometrics, vol. 3 (1).

Isogai T. (2014). Benchmarking of Unconditional VaR and ES Calculation Method: A comparative Simulation Analysis with Truncated Stable Distribution. *Bank of Japan Working Paper Series, No 14 E 1.*

Jeyasreedharan N., Alles L. & Yatawara N. (2009). The asymptotics of extreme returns in the Australian stock market. *22nd Australasian Finance and Banking Conference 2009*.

Kim Y. S., Rachev S. T., Bianchi M. L. & Fabozzi F. J. (2008). Financial market models with Lévy processes and time-varying volatility. *Journal of Banking & Finance, vol. 32.*

Leal R. P. C. & Vaz de Melo Mendes B. (2005). Maximum Drawdown: Models and Applications. *The Journal of Alternative Investments, vol. 7 (4).*

LeBaron B. & Samanta R. (2005). Extreme Value Theory and Fat Tails in Equity Markets. *Social Science Research Network.*

Longin F. M. (1996). The Asymptotic Distribution of Extreme Stock Market Returns. *The Journal of Business, vol. 69 (3).*

Lópex de Prado M. & Peijan A. (2004). Measuring Loss Potential of Hedge Fund Strategies. The Journal of Alternative Investments, Vol. 7 (1).

Makhwiting M. R., Sigauke C. & Lesaoana M. (2014). Modelling Tail Behavior of Returns Using the Generalized Extreme Value Distribution. *Economics, Management, and Financial Markets 9(1), p. 41-52.*

Mandelbrot B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, vol. 36.*

Mandira S. Thomas S. & Shah A. (2003). Selection of Value-at-Risk models. *Journal of Forecasting, Vol. 22 (4).*

Manganelli S. & Engle R. (2001). Value at Risk Models in Finance. *European Central Bank Working paper no.75*

Markowitz H. (1952). Portfolio Selection. *The Journal of Finance, Vol. 7 (1).*

McNeil A. J. & Frey R. (2000). Estimation of tail related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance, vol. 7 (3-4).*

McNeil A. J., Frey R. & Embrechts P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.

Novak S. Y. (2012). Extreme Value Methods with Applications to Finance. *Monographs on Statistics and Applied Probability 122.* A Chapman & Hall Book.

Onour I.(2009). Extreme Risk and Fat-tails Distribution Model: Empirical Analysis. Munich Personal RePEc Archive.

Pedersen C. S. & Satchell S. E. (2002). On the foundation of performance measures under asymmetric returns. *Quantitative Finance , vol. 2.*

Pelagatti M. (2016). Time Series Modelling with Unobserved Components. *CRC Press.*

Post T. & van Vliet P. (2006). Downside risk and asset pricing. *Journal of Banking & Finance, vol. 30.*

Pätäri E. (2011). Does the Risk-Adjustment Method Matter at All in Hedge Fund Rankings? *International Research Journal of Finance and Economics, Issue 75.*

Rachev S. T., Menn C. &Fabozzi F. J. (2005). Fat-Tailed and Skewed Asset Return Distributions. John Wiley & Sons.

Rachev S. T., Mittnik S., Fabozzi F. J., Focardi S. M. & Jasi´c T. (2007). Financial Econometrics: From Basics to Advanced Modeling Techniques. John Wiley & Sons.

Rachev S. T., Stoyanov S. V. & Fabozzi F. J. (2008). Advanced Stochastic Models, Risk Assessment and Portfolio Optimization: The Ideal Risk, Uncertainty and Performance Measures. John Wiley & Sons.

Rachev S. T., Stoyanov S. V. & Fabozzi F. J. (2011). A Probability Metrics Approach to Financial Risk Measures. John Wiley & Sons.

Rockafellar R. T. & Uryasev S. (2002). Conditional value –at-risk for general loss distributions. *Journal of Banking & Finance, vol. 26 (7).*

Roy A. D. (1952). Safety First and the Holding of Assets. *Econometrica , vol.20 (3).*

Sortino F. A. & Price L. N. (1994). Performance Measurement in a Downside Risk Framework. *The Journal of investing, vol. 3 (3).*

Stoaynov S. V. (2010). A Moment Expansion of Downside Risk Measures. *EDHEC.*

Stoyanov S. V., Rachev S. T., Racheva-Iotova B. & Fabozzi F. J. (2011). Fat-tailed models for risk estimation. *Journal of Portfolio Management, 37/2.*

Tolikas K. (2011) The rare event risk in African emerging stock markets. *Managerial Finance 37(3). p.275-294.*

Zikovic S. & Aktan B. (2009). Global financial crisis and VaR performance in emerging markets: A Case of EU candidate states – Turkey and Croatia. *Proceedings of Rijeka Faculty of Economics, vol.27 (1).*

Zivot E. & Wang J. (2006). Modelling Financial Time Series with S-Plus 2nd Edition. Springer.

Zoglat A., Adlouni S. El., Ezzahid E., Amar A., Okou C. G. & Badaoui F.(2013). Statistical methods to expect extreme values: Application of POT approach to CAC40 return index. *International Journal of Statistics & Economics 10 (1).*

Appendix

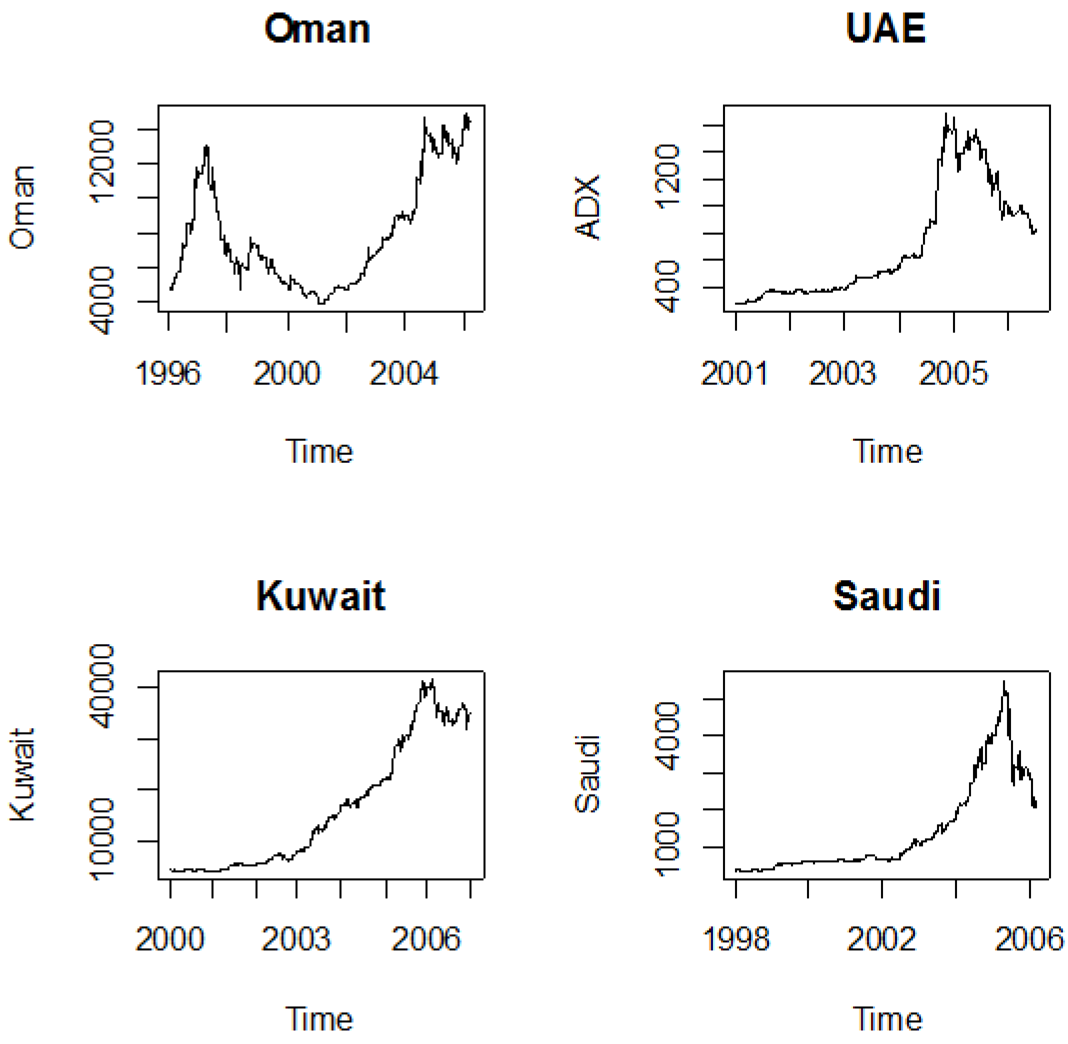

Figure 2. Price indexes

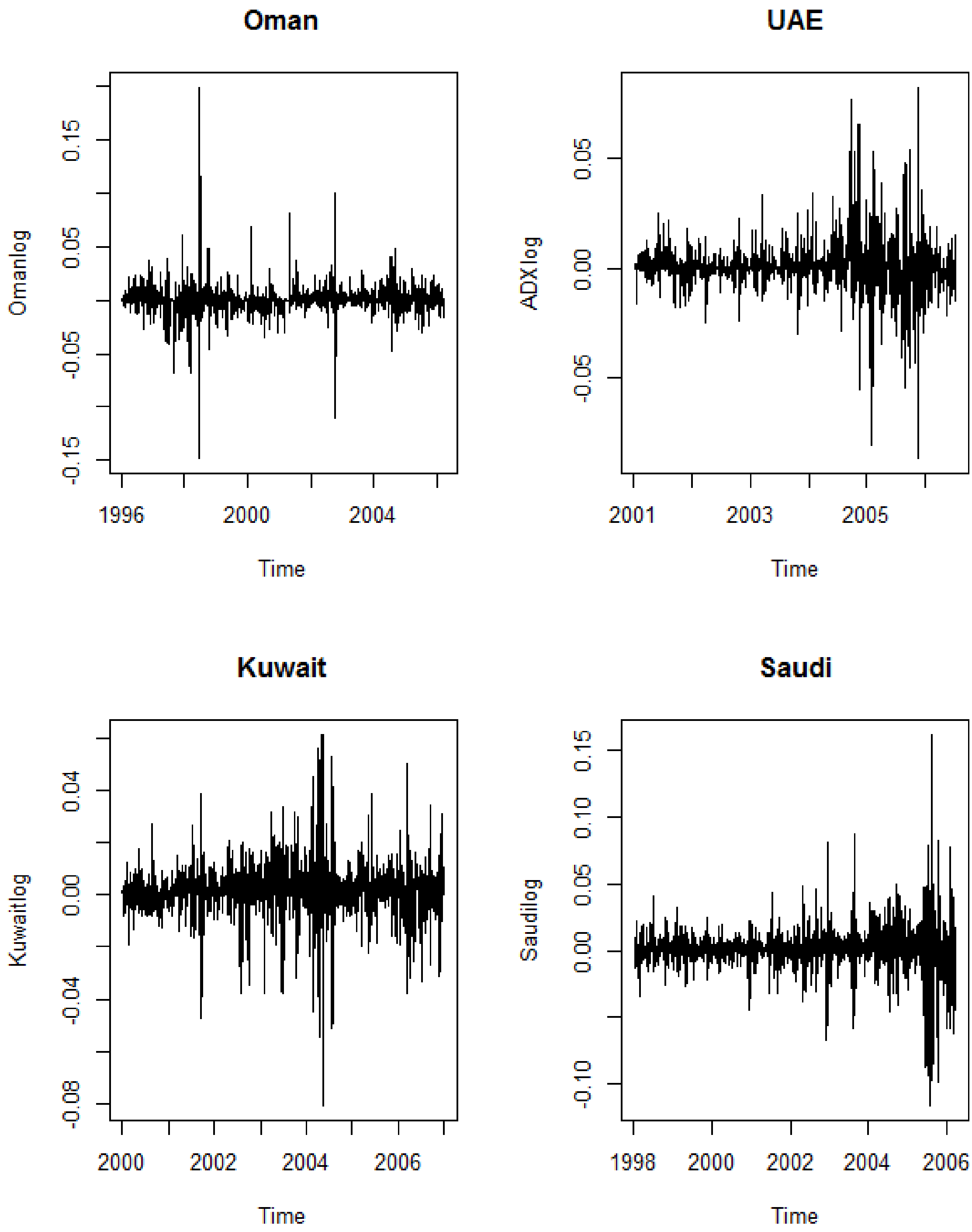

Figure 3. Return indexes

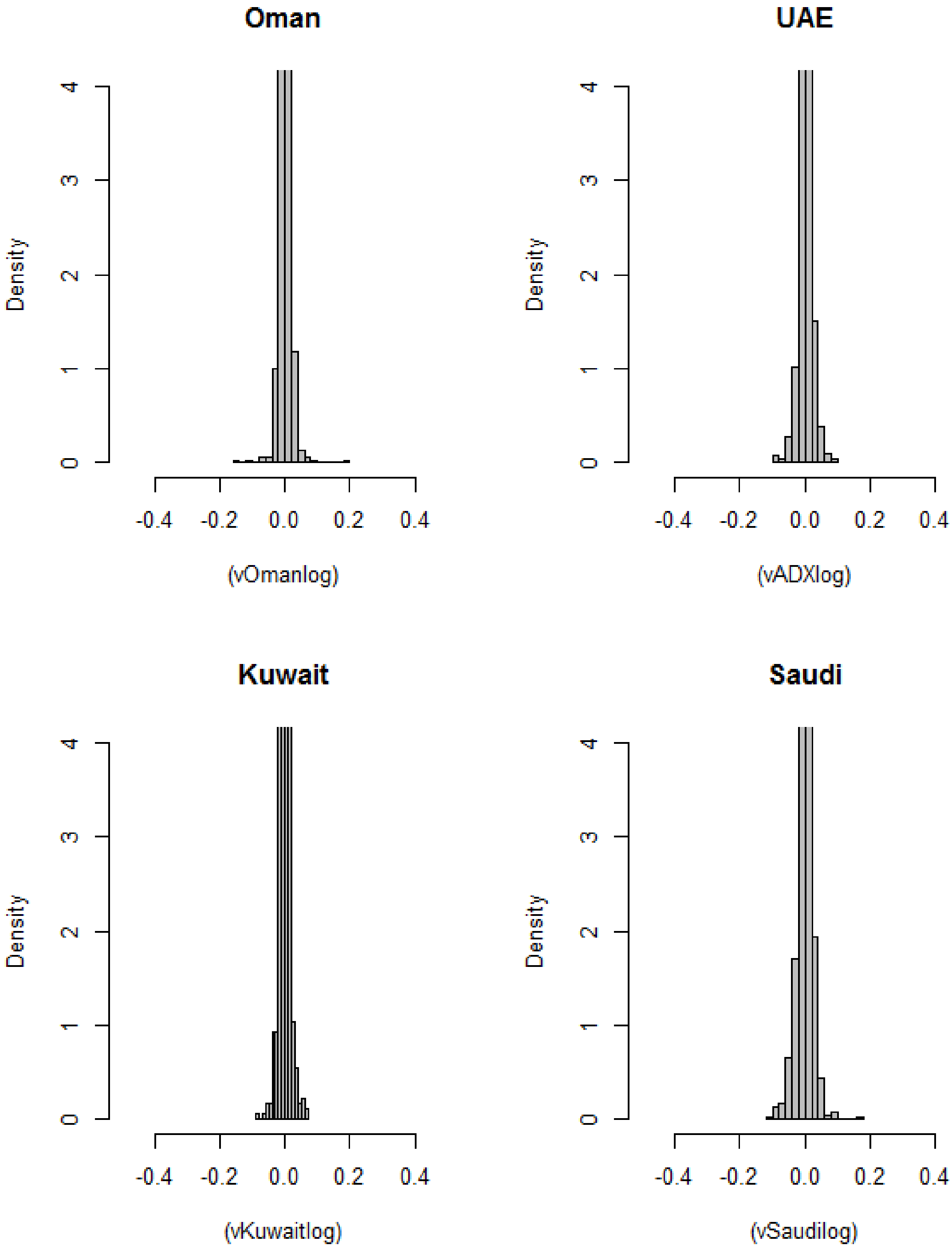

Figure 4. Histogram of log returns

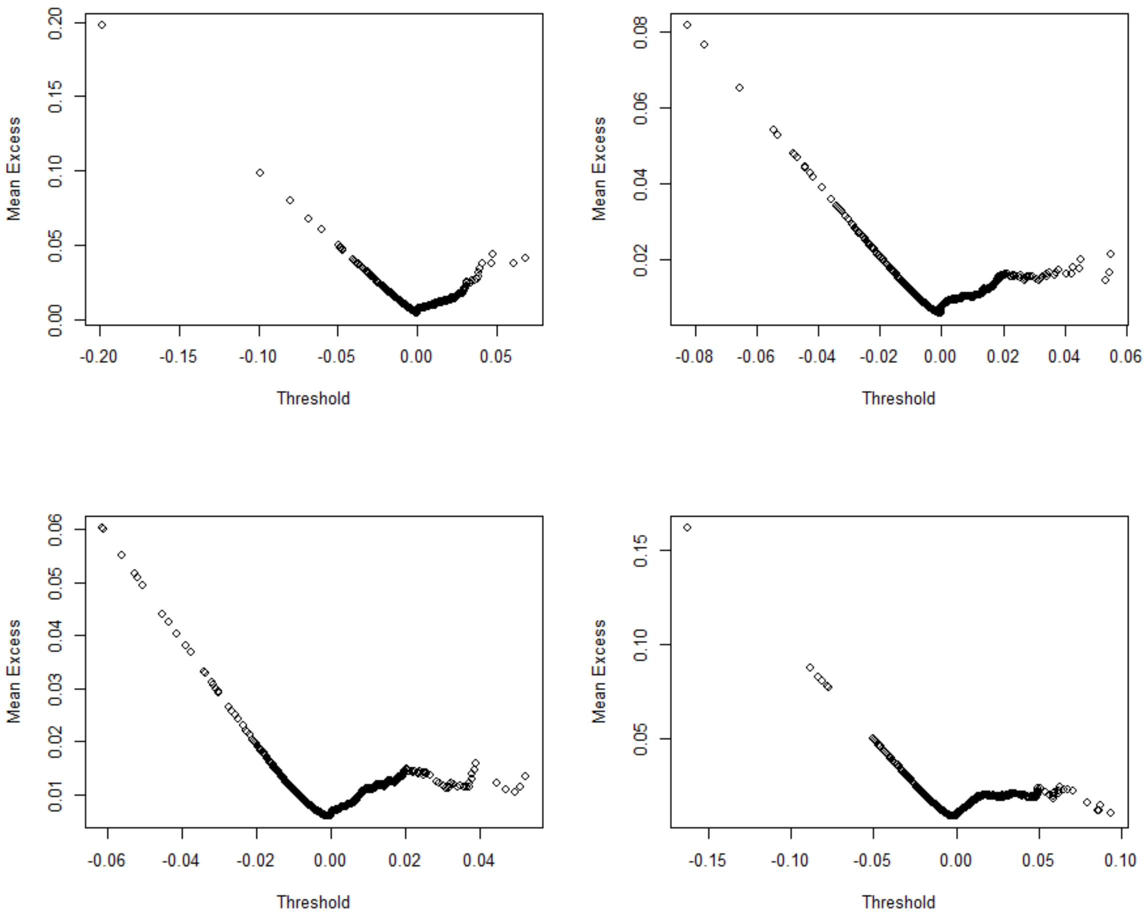

Figure 5. Mean excess plot

(1. Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

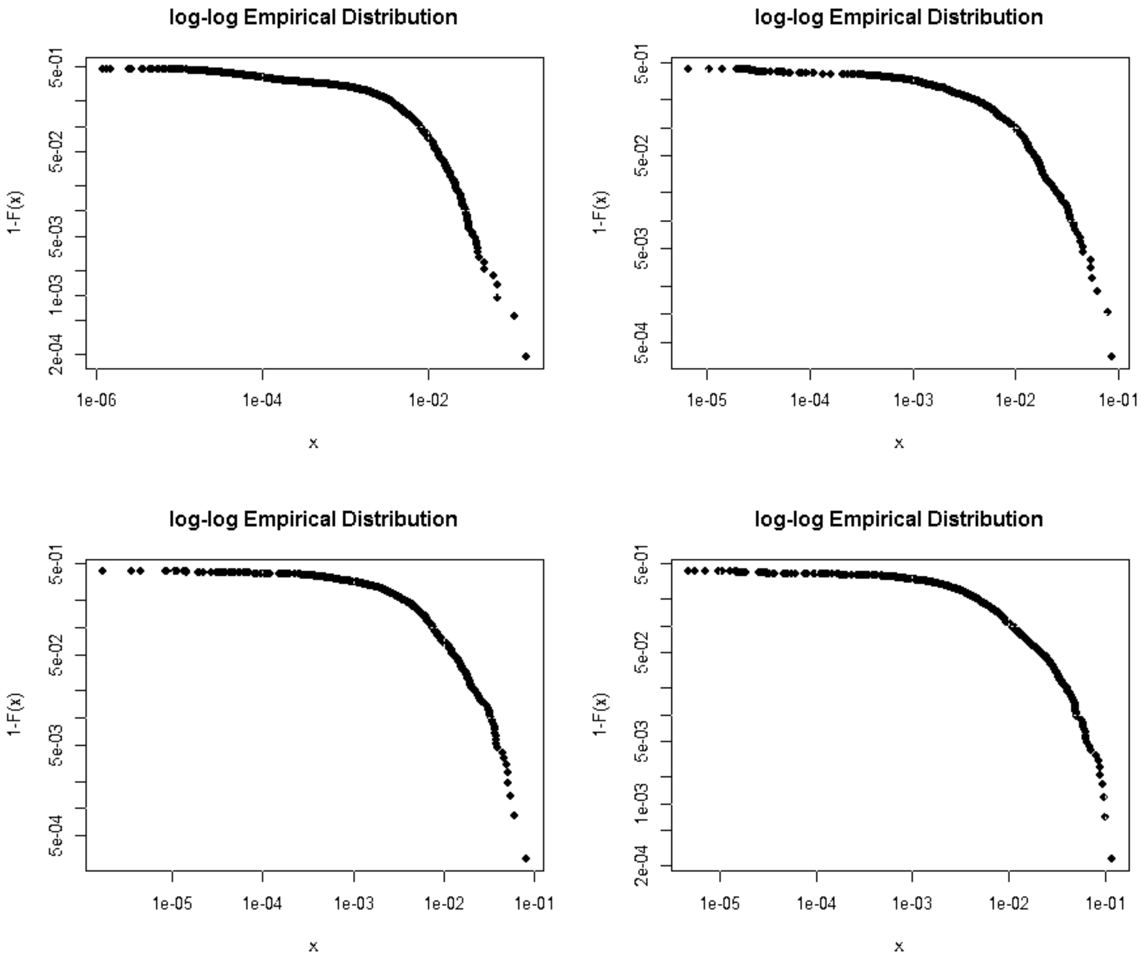

Figure 6. Empirical Distribution plot (Zipf plot)

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

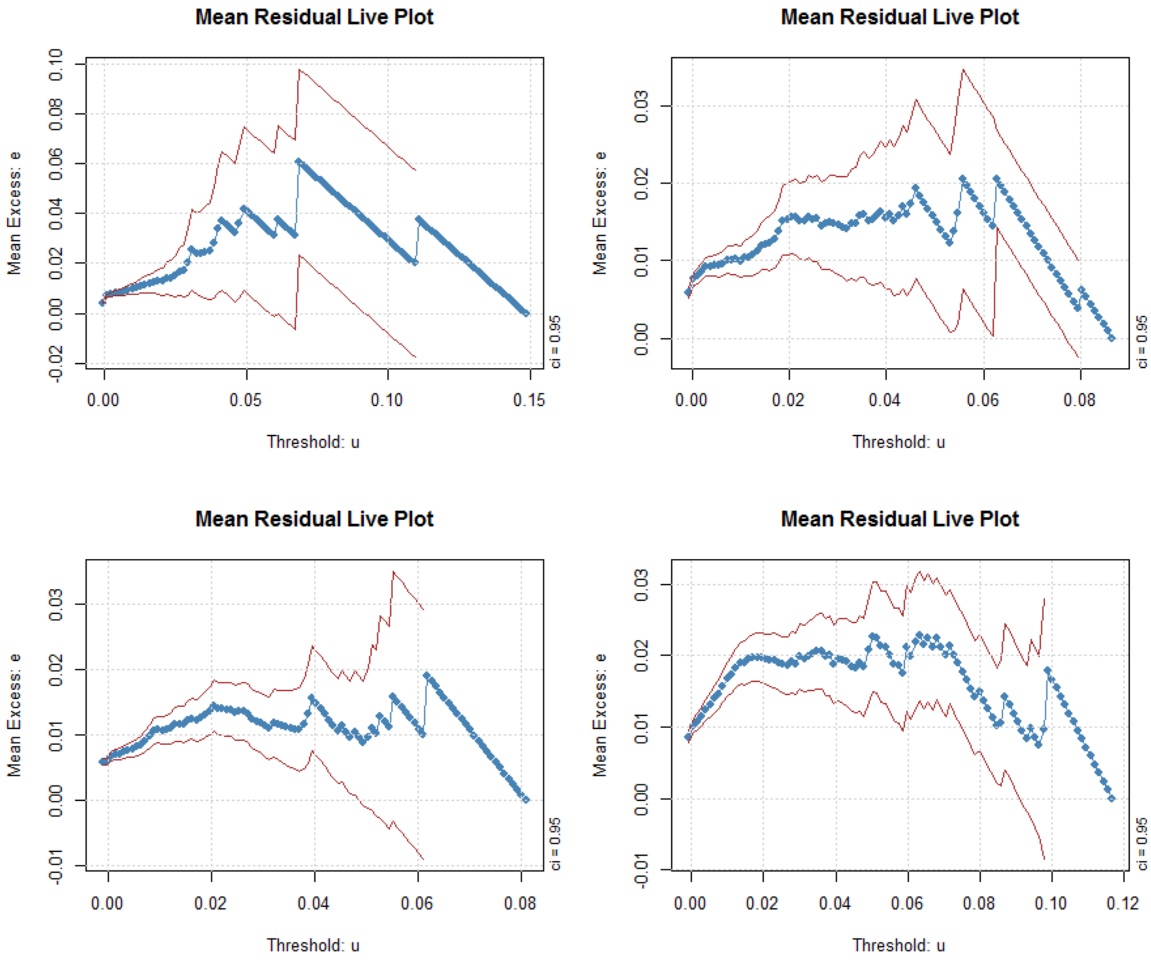

Figure 7. Mean Residual Life Plot

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

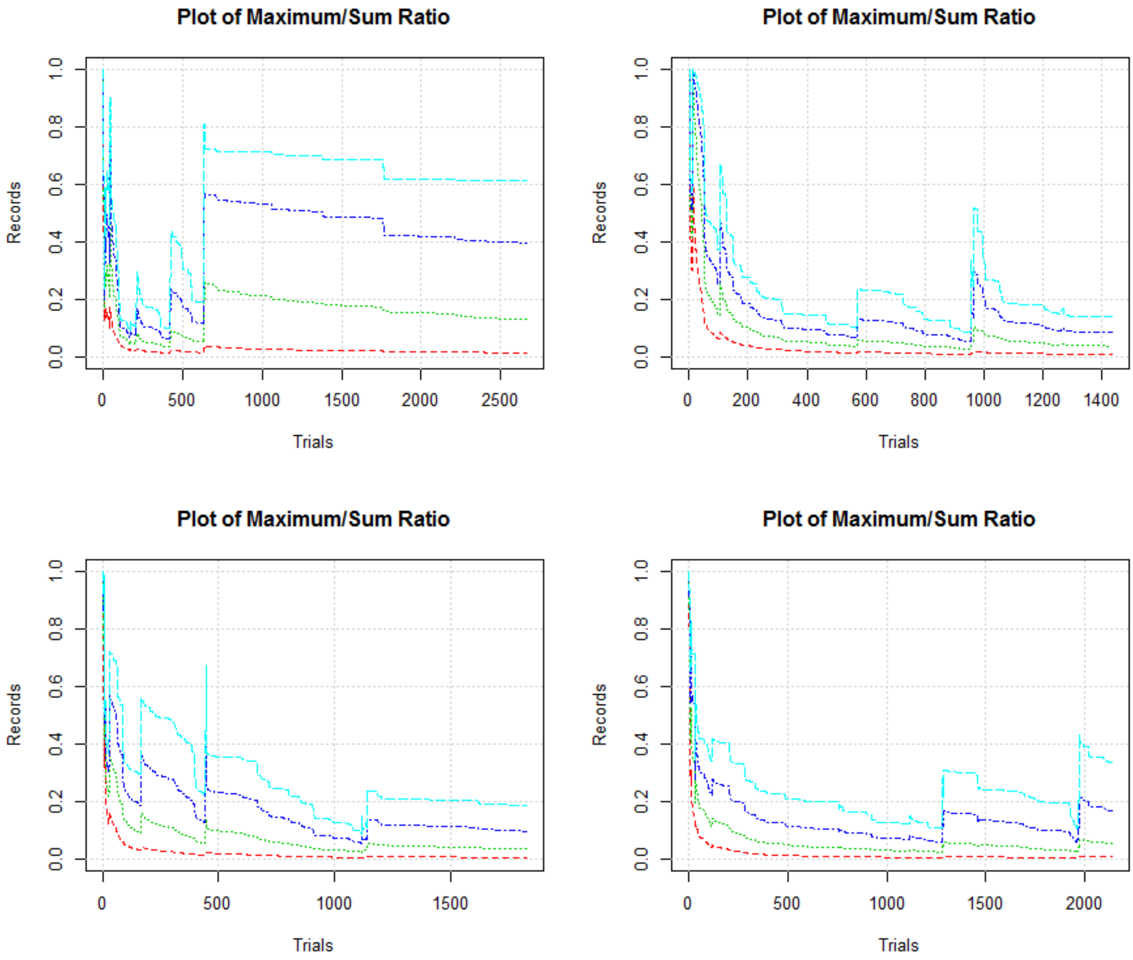

Figure 8. Ms Plot.

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right). First moment colored in red, 2nd col. green, 3nd col. blue and 4th col. light blue)

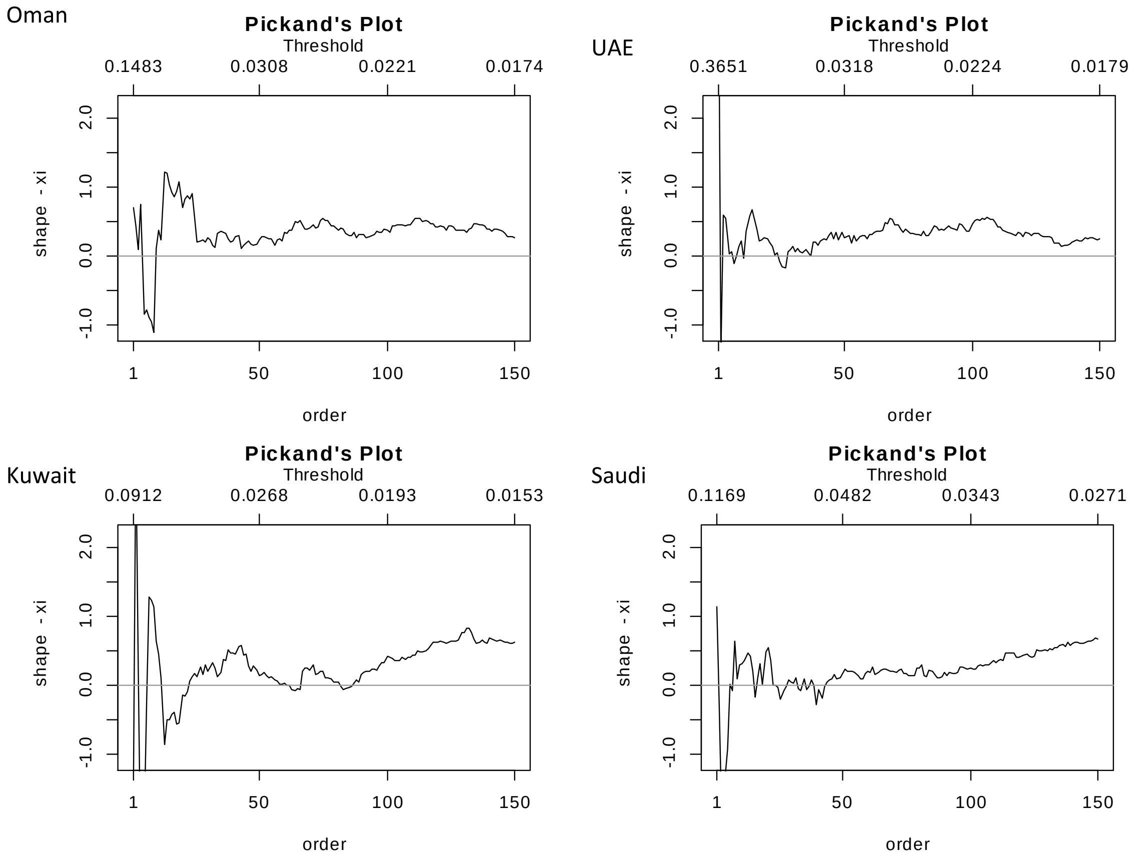

Figure 9. Pickands plot ξ

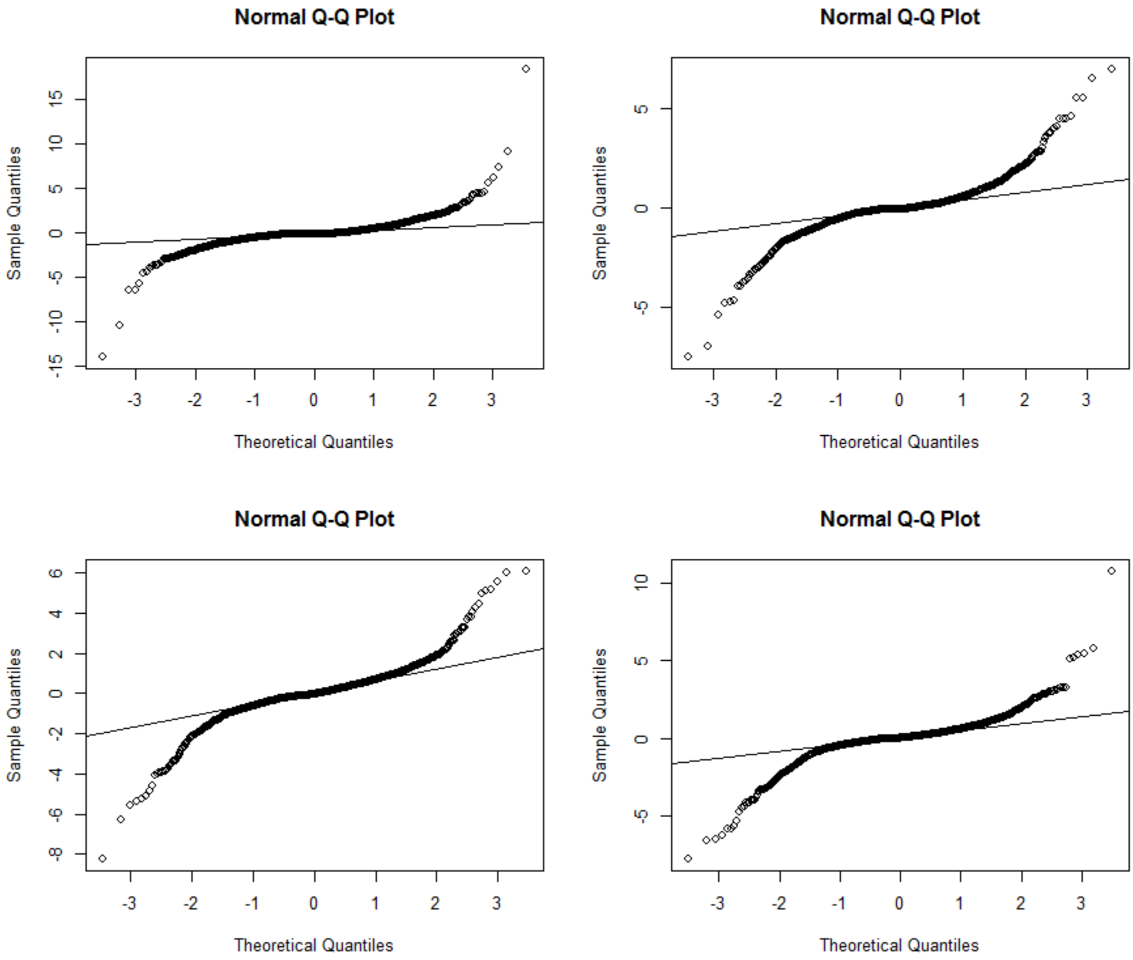

Figure 10. Standardized QQ plot:

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

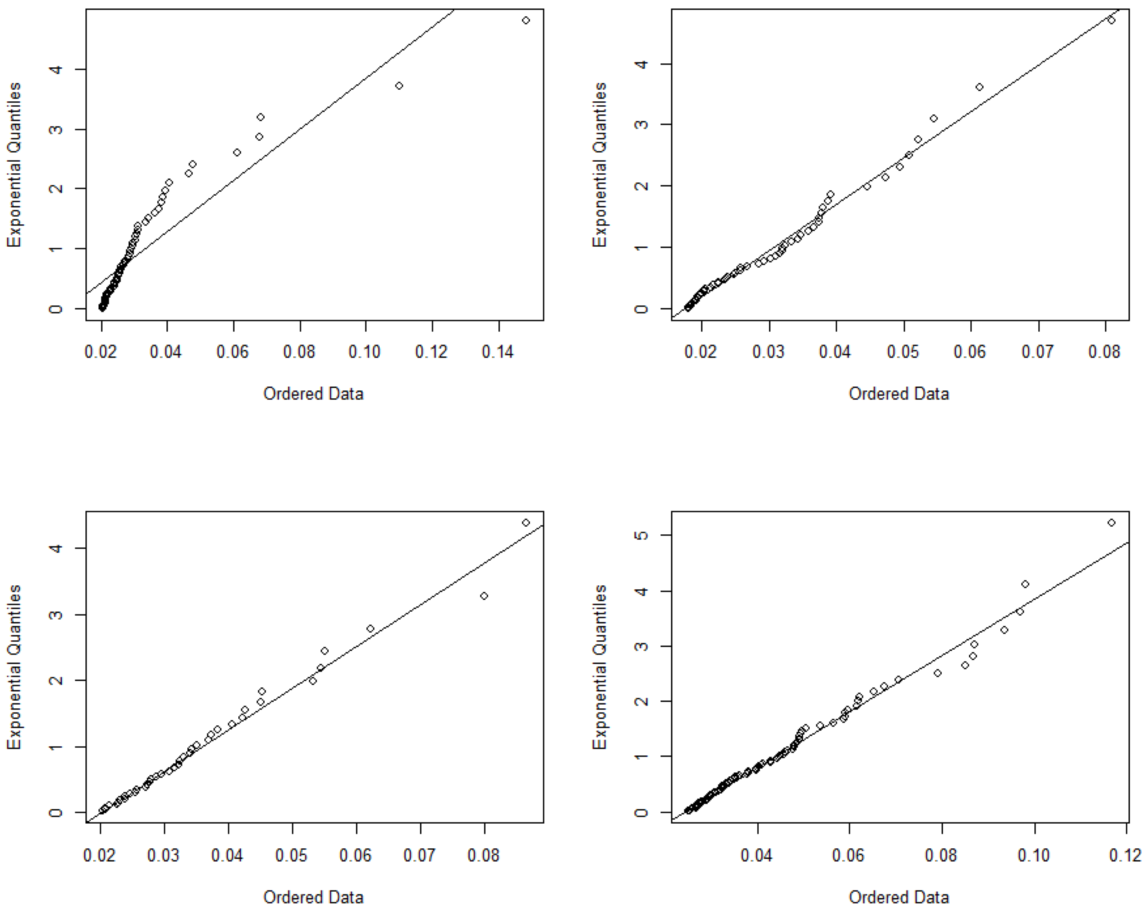

Figure 11. Q-Q plot with exponential reference line

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

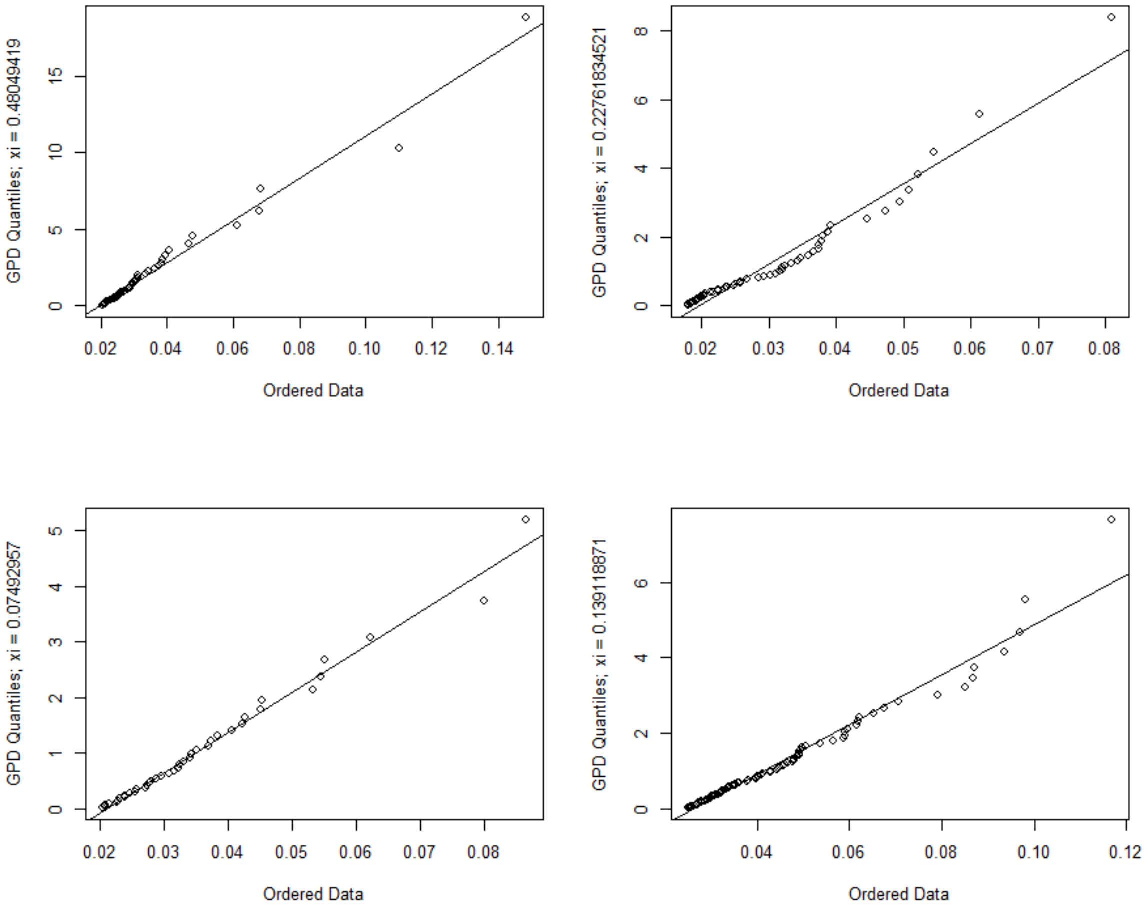

Figure 12. Q-Q plot with GPD reference line modified with shape parameter

(1.Oman (top left), 2. UAE (top right), 3. Kuwait (bottom left), 4.Saudi (bottom right))

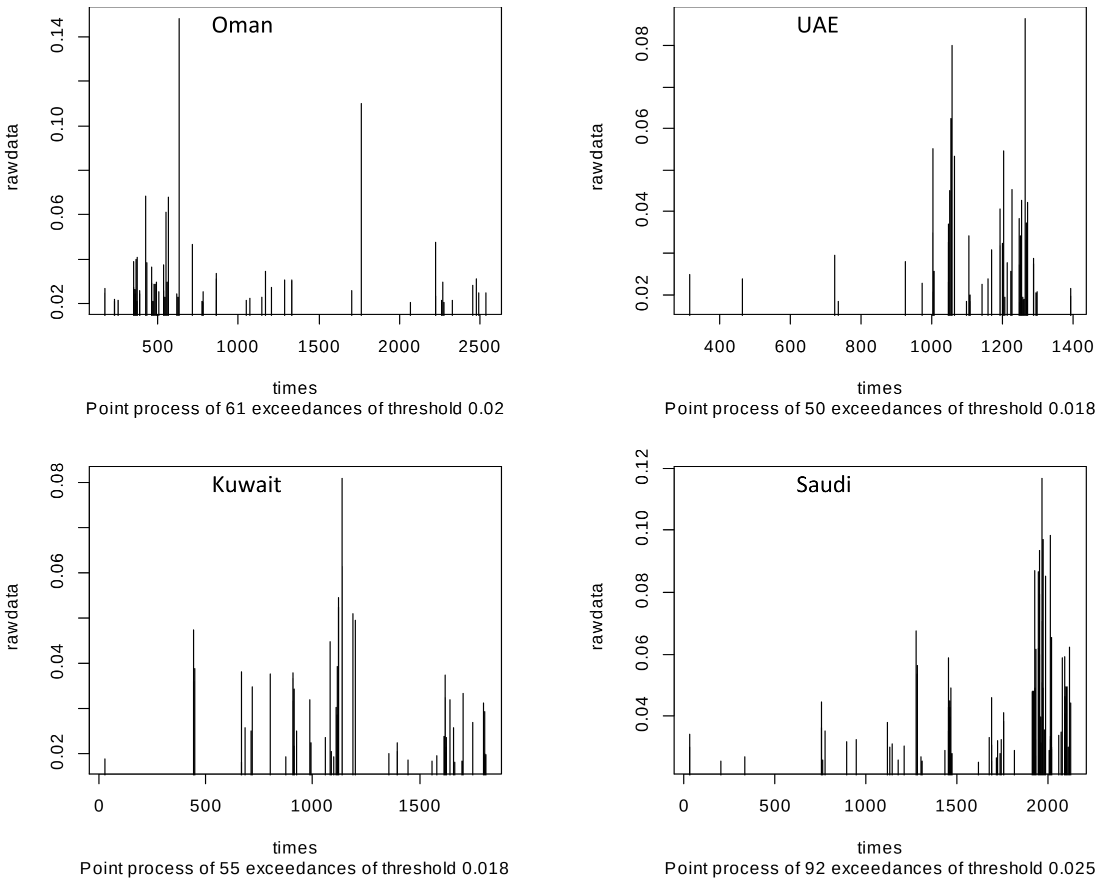

Figure 13. Point process exceedance

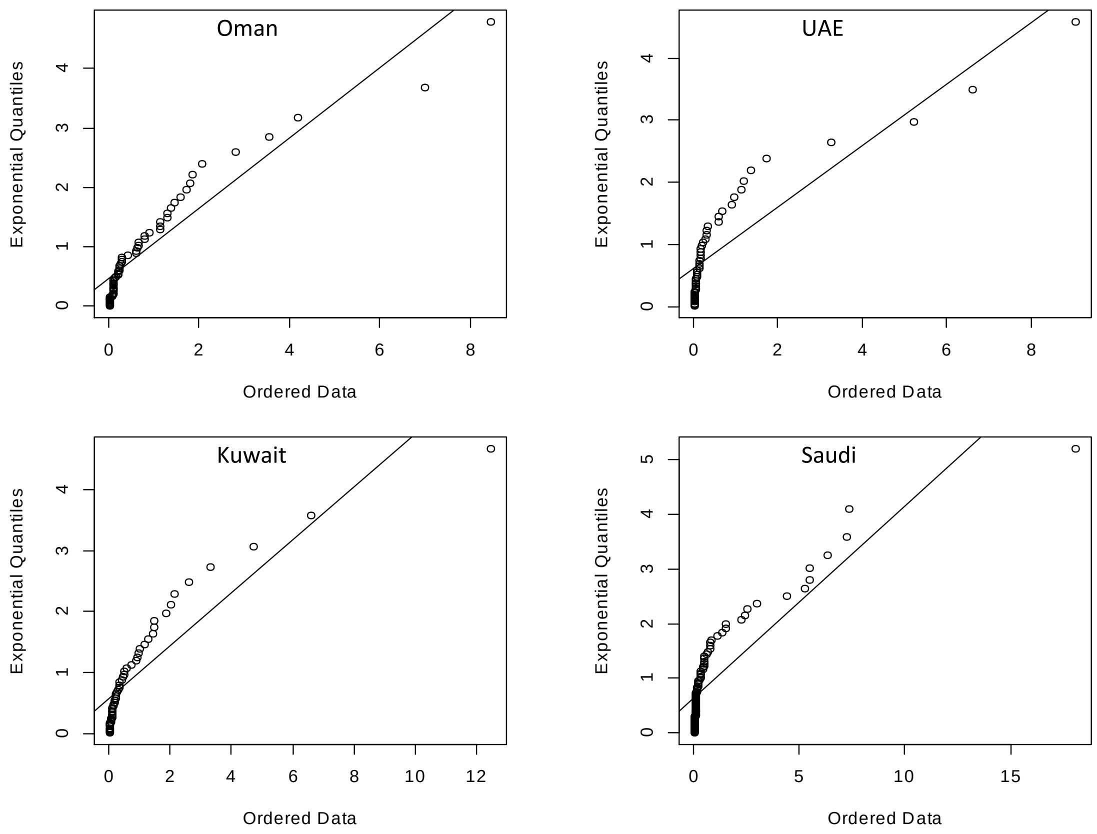

Figure 14. Quantile plot of gaps

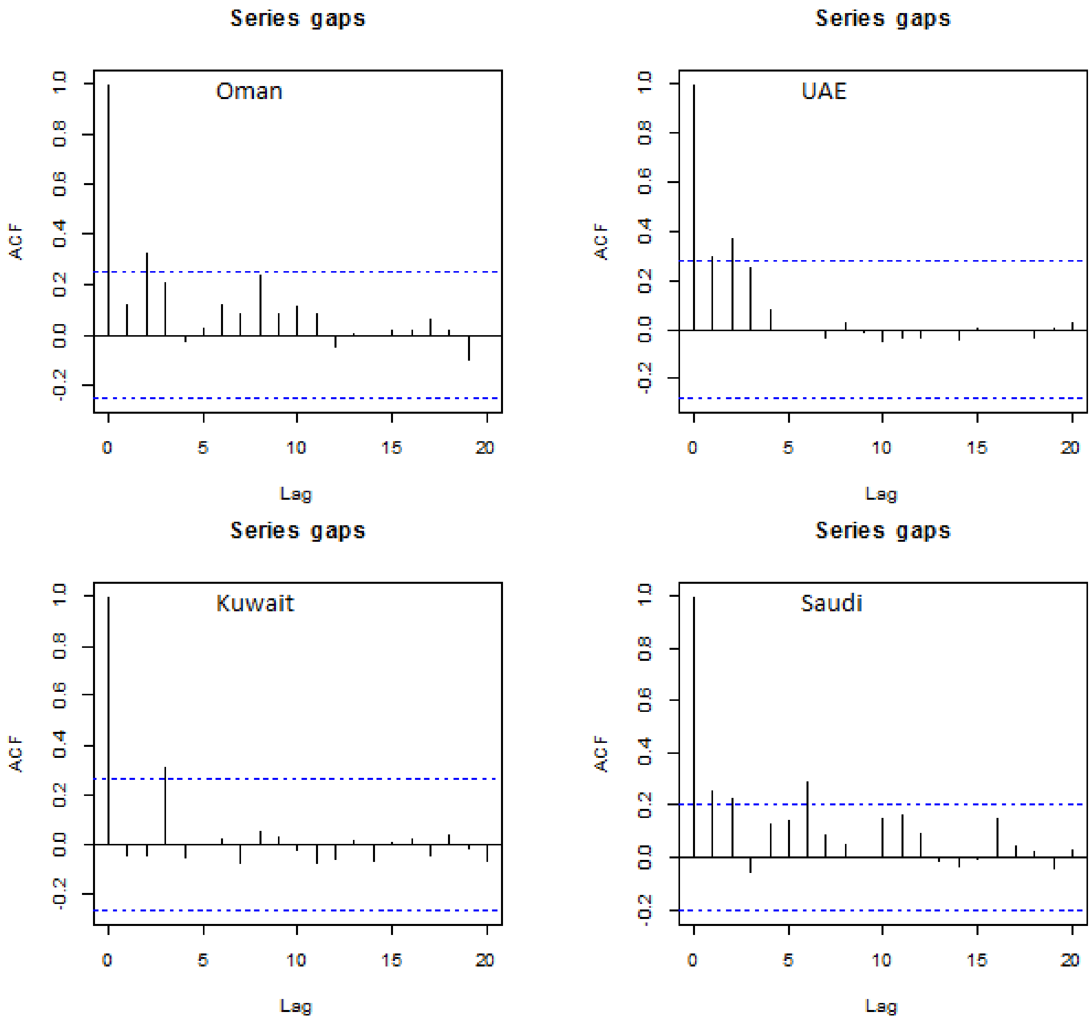

Figure 15. ACF of gaps

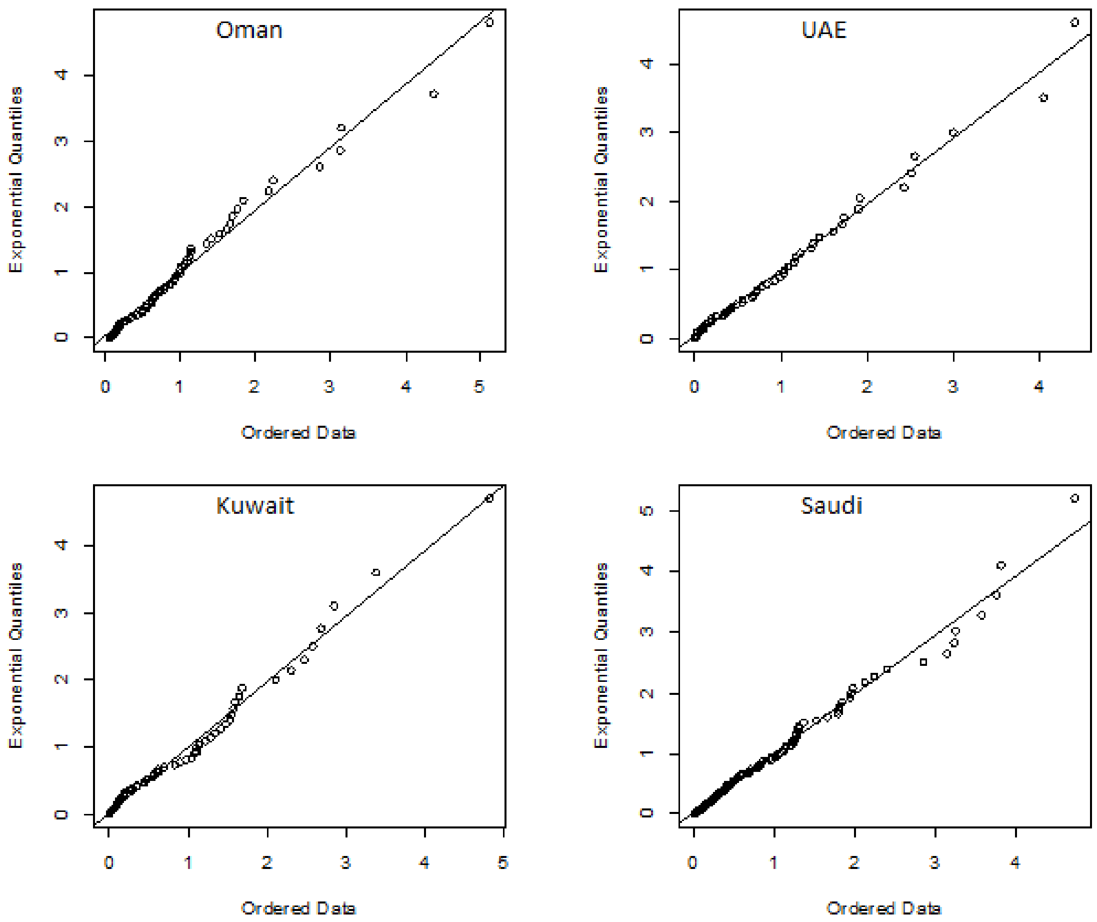

Figure 16. Q-Q plot of residuals

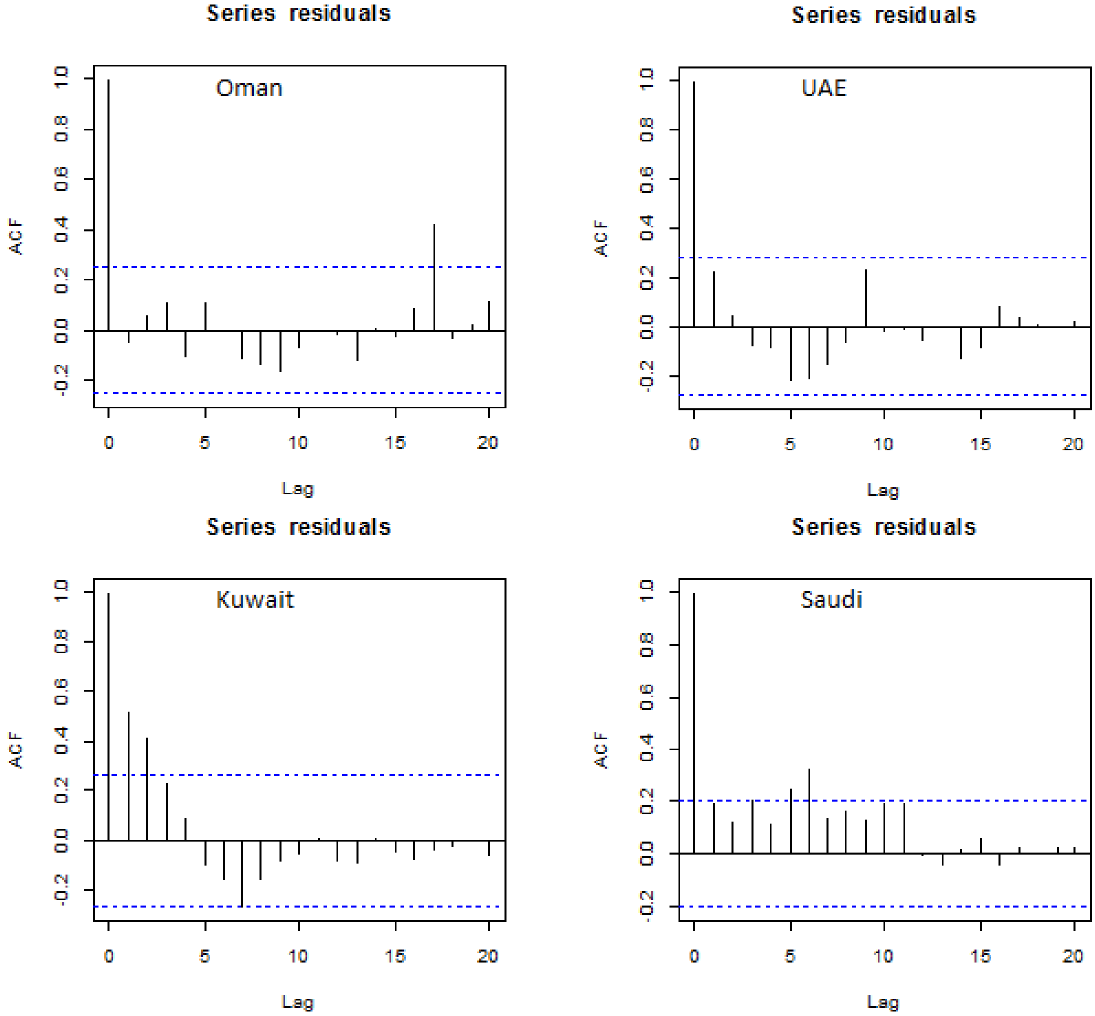

Figure 17. ACF of Residuals

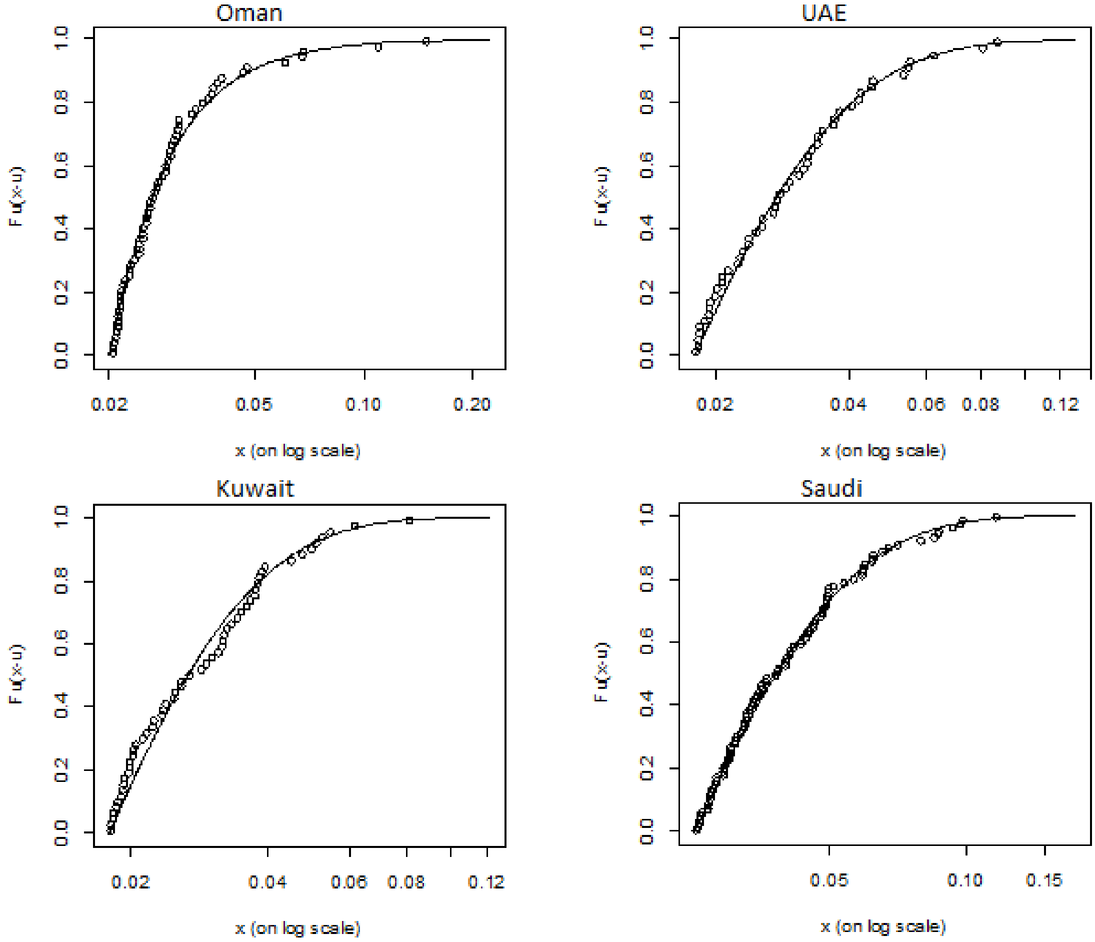

Figure 18. Excess distribution

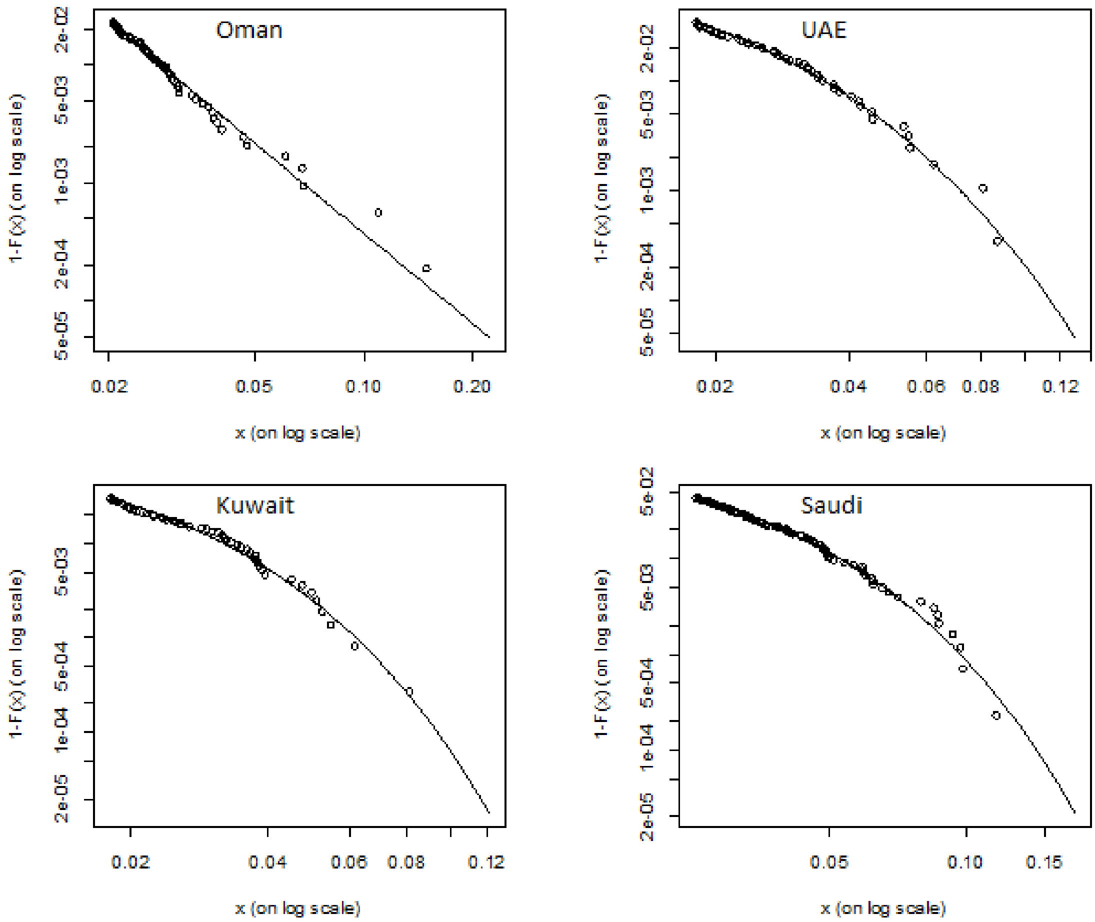

Figure 19. Tail plot

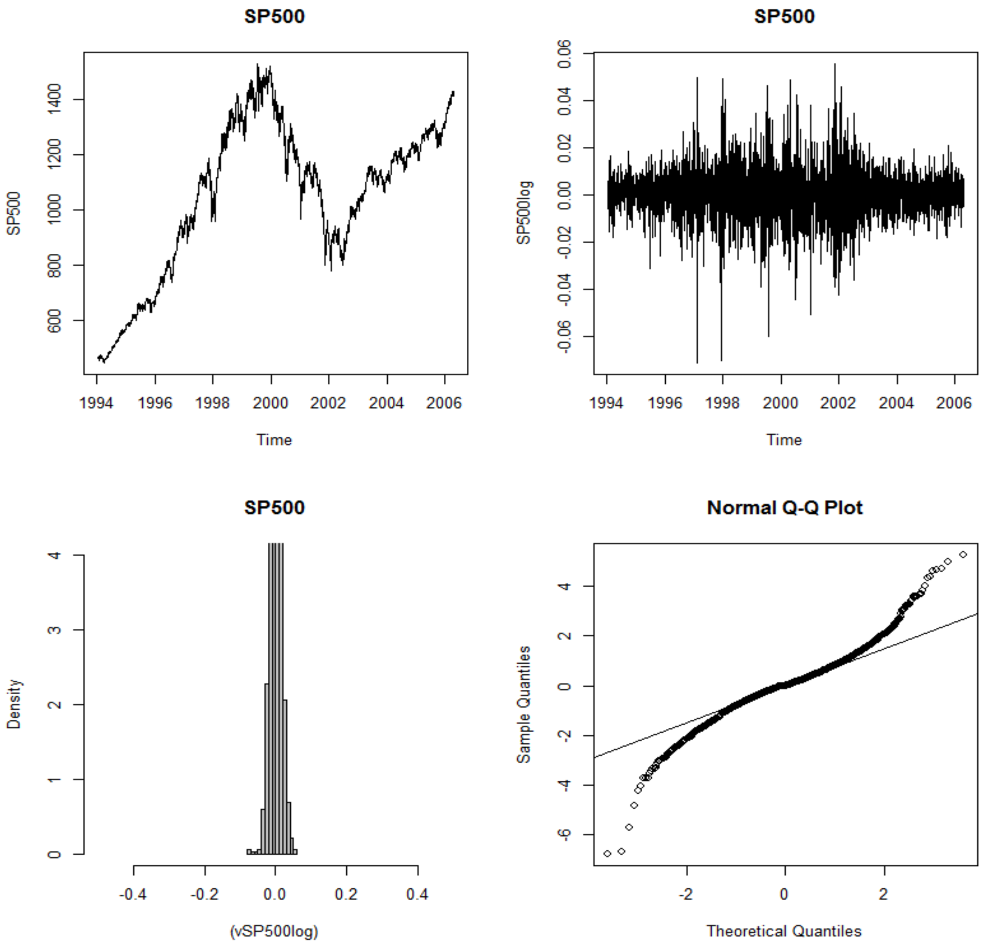

Figure 20. SP500 Price, Return, Histogram and Standardized QQ plot.

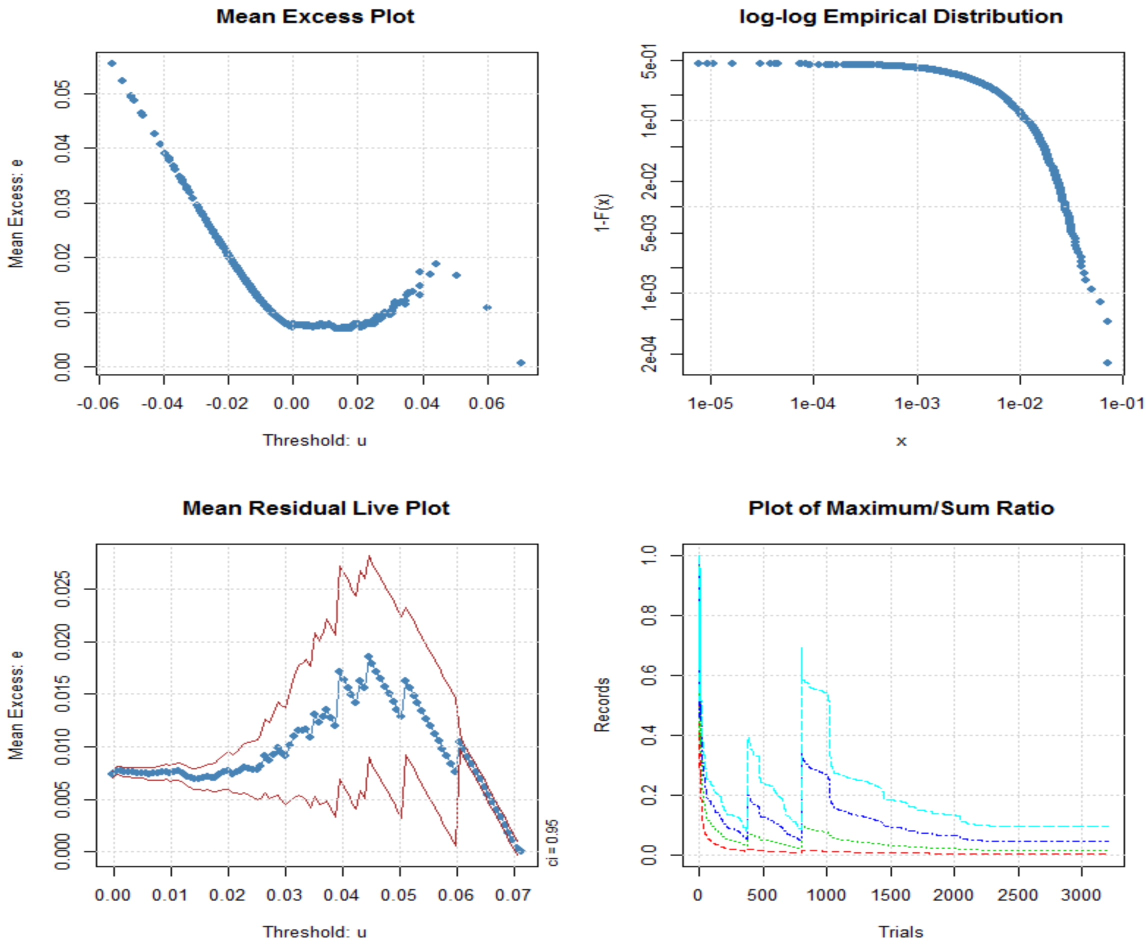

Figure 21. SP500 ME-plot, Zipf-plot, MRL-plot and MS-plot

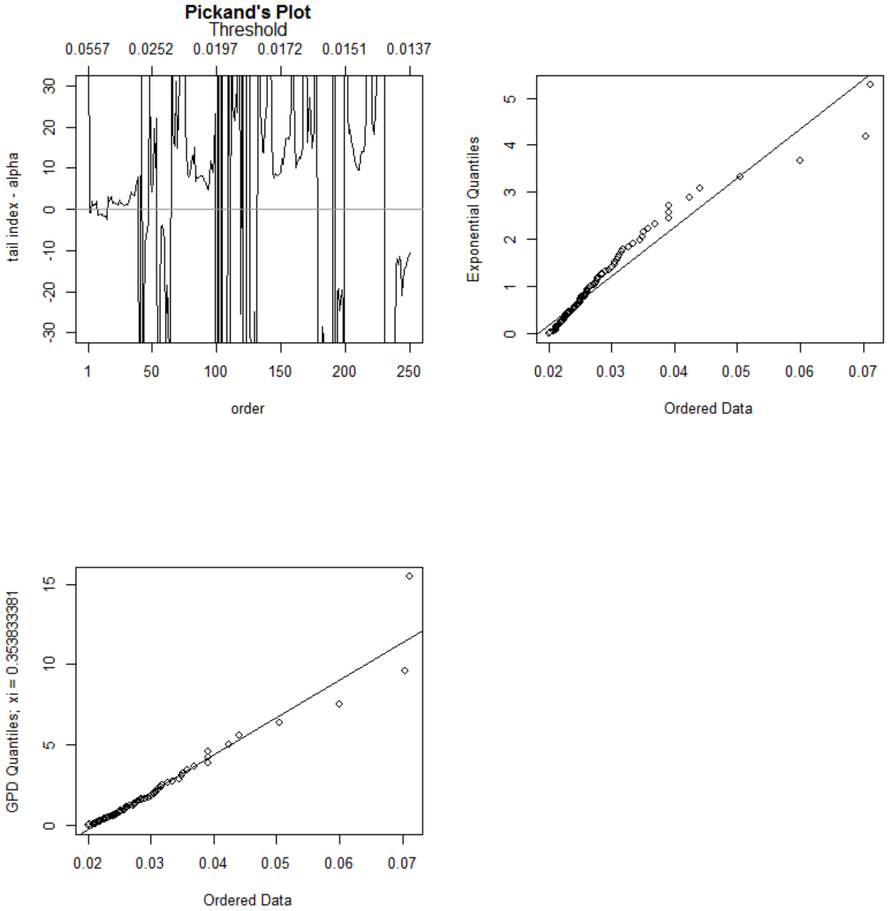

Figure 22. SP500 Pickands plot, QQ-plot with exponential reference line and QQ-plot with GPD reference line

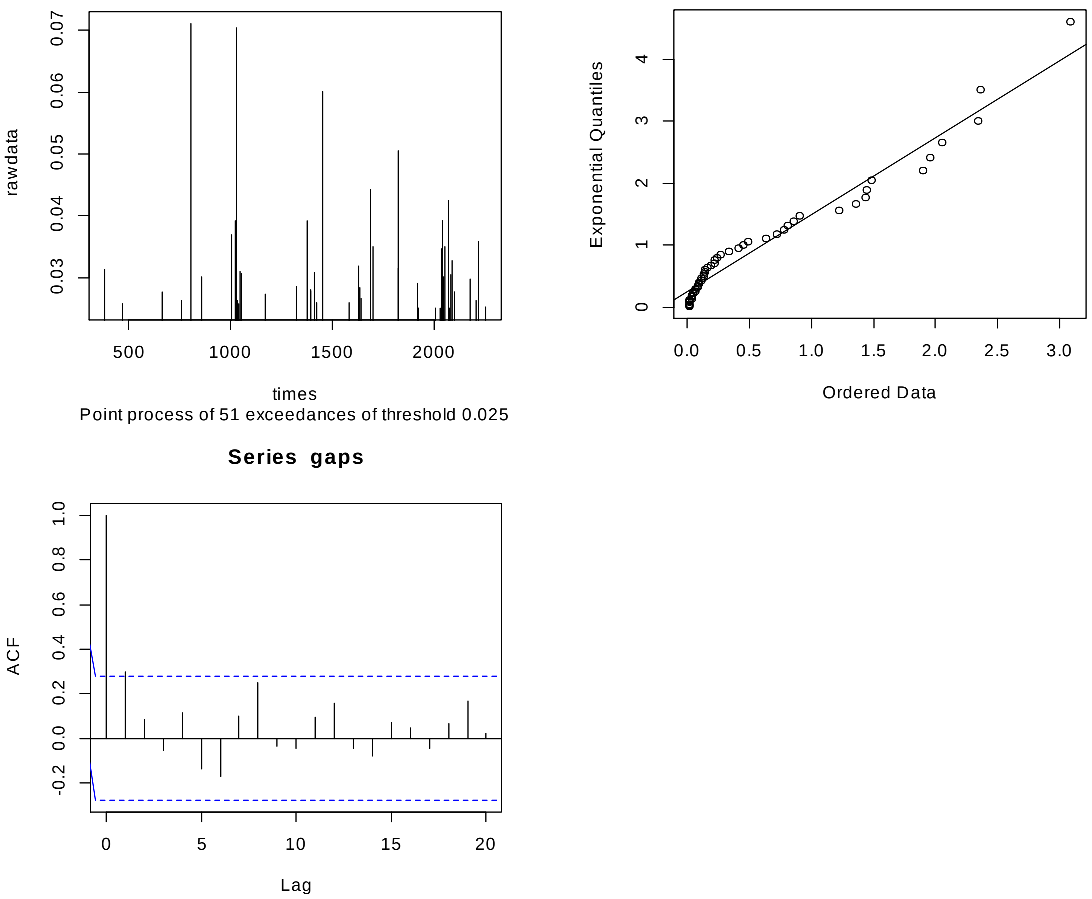

Figure 23. SP500 Point process exceedance, Quantile plot of Gaps and ACF plot of Gaps

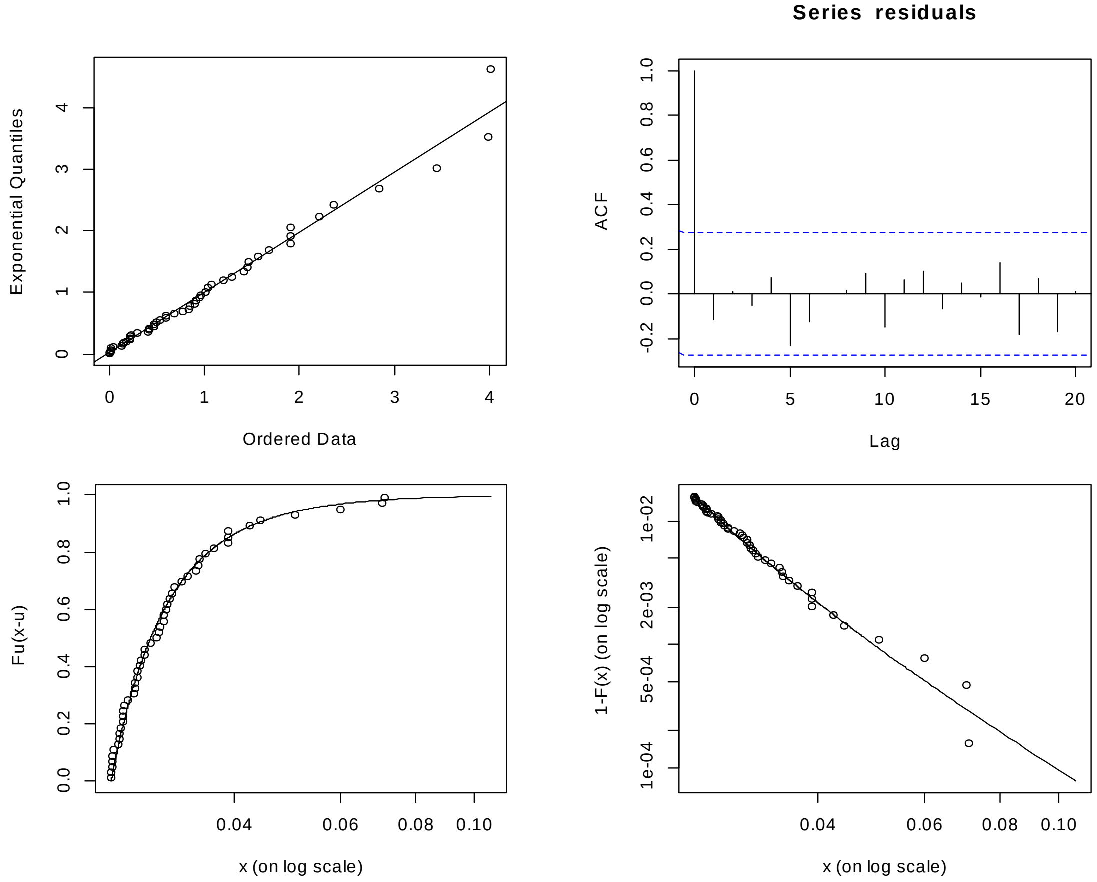

Figure 24. SP500 Quantile plot of residuals, ACF of residuals, Excess distribution and Tail plot

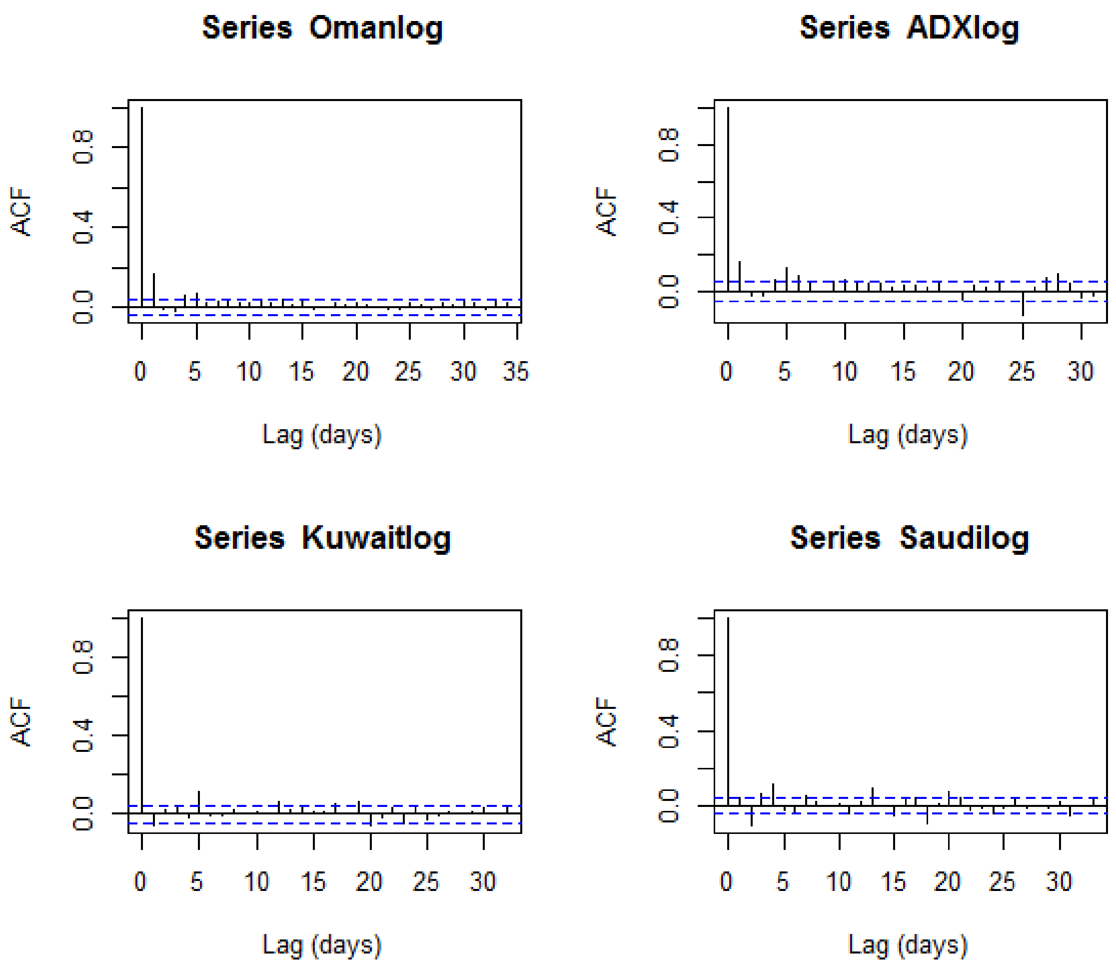

Figure 25. ACF of returns.

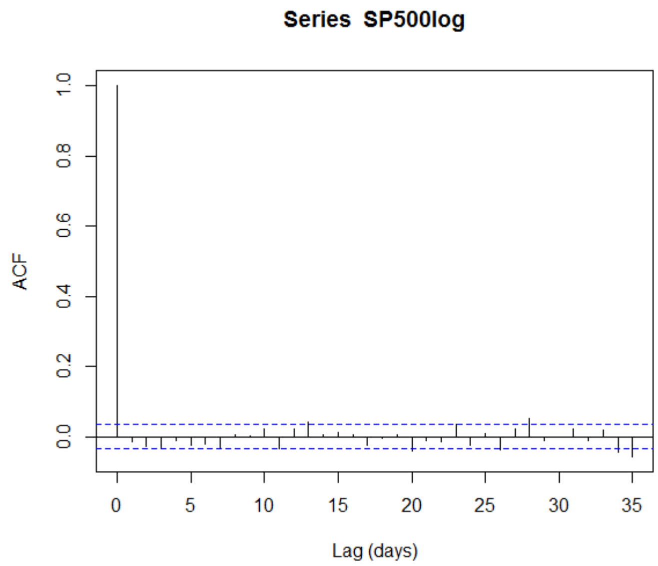

Figure 26. SP500 ACF of returns.

	extremal.index
Oman	0.7121094
UAE	0.2954943
Kuwait	0.5311003
Saudi	0.2238991
SP500	0.1937023

Table 5 Extremal Index