EXTREME VALUE APPROACH IN ANALYZING STOCK RETURNS

IN ISTANBUL STOCK EXCHANGE

GÖZDE ERHAN ÜNAL

BOĞAZİÇİ UNIVERSITY

2006

Extreme Value Approach in Analyzing Stock Returns

in Istanbul Stock Exchange

Dissertation submitted to the

Graduate Institute of Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Administration

by

Gözde Erhan Ünal

Boğaziçi University

2006

ACKNOWLEDGEMENT

First, I thank my advisor Vedat Akgiray, for his continuous support in the Ph.D. program and for making me aware of 'Extreme Value Theory' and all his encouragements and guidance. Besides my advisor, I would like to thank the rest of my thesis committee. I also would like to thank my classmates Ali Coşkun and Kadri Özgüneş, who have been by my side all along.

Let me also say 'thank you' to the following people at Istanbul Bilgi University: A special thanks goes to Harald Schmidbauer, who voluntarily mentored me in my study, for teaching me R-project, and supporting me with his own program codes and valuable comments; Vehbi Sinan Tunalıoğlu, for his help on R-project and Linux; Selime Sezgin, Özlem Kaçar and Özgür Kökalan, who shared their room with me, for their understanding, support and help.

Last, but not least, I thank my family: My husband Erkan for his help in constructing my database and patience during the time I spent in front of the computer. My parents, Vuslat Başaran, and Aytaç Erhan, for giving me life in the first place, for educating me and for unconditional support and encouragement to pursue my interests. My brother Fethi, for 24 hours available support on Linux and hardware problems. 'Eight of us' for listening to my complaints and frustrations, and for always believing in me.

VITA

GÖZDE ERHAN ÜNAL

Personal Profile

Born on 14.02.1976, Istanbul.

Experience

Instructor Bogazici University, Department of International Trade	Oct 2005-Present
Instructor Istanbul Bilgi University	Oct 2005-Present
Teaching Assistant Istanbul Bilgi University, Bus. Adm. and Int. Finance	Feb 2004-Sep 2005
Senior Associate, Corporate Finance İş Investment Securities Inc.	May 2000-Feb 2004
Associate, Corporate Finance Finans Invest Securities Inc.	Feb 2000-May 2000
Research Assistant Bogazici University, Management	Sep 1999-Feb 2000
Teaching Assistant Koc University, Mathematics	Sep 1997-Jun 1999

Education

M.B.A. Bogazici University Degree: 3rd , GPA 3.83	Sep 1997-Jun 1999
BS Mathematics, Koc University Degree: 2nd, GPA 3.43, Full Scholarship	Sep 1993-Jun 1997
Beyoglu Anatolian High School Honor	Sep 1989-Jun 1993

ABSTRACT

The study aims to model the tails of daily returns of securities being traded in ISE via techniques developed by Extreme Value Theory and compute VaR. The performances of classical VaR forecasting methods of Historical Simulation and RiskMetrics™ are compared with the models estimated using Peaks over Threshold (POT) approach, which is put forward by Extreme Value Theory. POT approach incorporates estimating the tail index of Generalized Pareto distributions(GPD). As well as having used nonparametric Hill and Dekkers estimators, also parametric Maximum Likelihood Estimate approach is applied in estimating the tail index of GPD. VaR has been computed with these various approaches mentioned for six stocks being traded in ISE, the ISE National 100 index, and an artificial price weighted index. The models are classified as successful if they satisfy both criteria of unconditional and conditional coverage. Those VaR models that satisfy both criteria of success have also been tested in terms of a Quantile Loss function. The models that gave lowest loss values are preferred.

Among the approaches used in the study, the models that fit Genaralized Pareto distributions to the lower tail are found to outperform the classical Historical Simulation and RiskMetrics approaches.

KISA ÖZET

Bu çalışmada esas olarak İMKB’da işlem gören hisse senetlerinin günlük getirileri için dağılımların alt kuyruğunun modellemesi ve bu dağılıma ilişkili olarak Riske Maruz Değer (RMD) hesaplamalarının yapılması hedeflenmiştir. Daha klasik yöntemler olan Tarihi Simulasyon, RiskMetrics™ yöntemlerinin RMD hesaplamalarındaki performsları Uç Değer Teorisinin sunduğu Eşik Ötesi Gözlemler ile tahmin edilen modellerin performansı ile kıyaslanmıştır. Eşik Ötesi Gözlemler yaklaşımı Genelleştirilmiş Pareto Dağılımı ile kuyruk endeksinin kestirimini içermektedir. Çalışmada, parametrik olmayan Hill ve Dekkers kestiricilerinin yanısıra Maksimum Olabilirlik Kestirimi yaklaşımı ile de kuyruk endeksi tahmin edilmiştir. Yukarıda bahsedilen yaklaşımlar ile RMD, İMKB’de işlem gören altı hisse senedi, IMKB 100 endeksi ve fiyat ağırlıklı olarak hesaplanan bir başka endeks için hesaplanmıştır. Bu modeller arasında koşulsuz ve koşullu kapsama olasılığı kriterlerini sağlayanlar başarılı olarak sınıflandırılmıştır. Her iki başarı kriterini de sağlayan modeller, bir Quantile Kayıp fonksiyonu değerleri hesaplanarak birbirleri arasında test edilmiştir. En düşük kayıp değerlerini veren modeller tercih edilmiştir.

Çalışmada kullanılan yaklaşımlar arasında genelleştirilmiş Pareto dağılımı kullanılarak hesaplanan modeller, Tarihi Simülasyon ve RiskMetrics yaklaşımları ile hesaplanan modellerden daha başarılı bulunmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

ARCH	Autoregressive Conditional Heteroscedasticity
ASE	Athens Stock Exchange
AVE20	price weighted index of 20 companies from ISE
BIS	Bank for International Settlements
BT	Binomial test
CLR	Conditional Likelihood Ratio
CT	contingency table
C-VaR	Conditional Value-at-Risk
EPM	Elemental Percentile Method
ERM	Exchange Rate Mechanism
ES	expected shortfall
EVD	extreme value distributions
EVT	extreme value theory
EWMA	exponentially weighted moving average
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GL	Generalized Logistic
GPD	generalized Pareto distributions
HS	Historical Simulation
i.i.d.	independent and identically distributed
ISE	Istanbul Stock Exchange
ISE100	Istanbul Stock Exchange National 100 index
MLE	Maximum Likelihood Estimate
P&L	Profit and Loss

POT	Peaks over Threshold
PWM	Probability Weighted Moments
QL	quantile loss
RM	RiskMetrics™
SV	stochastic volatility
ULR	Unconditional Likelihood Ratio
VaR	value-at-risk
Φ	Gaussian Distribution

1 Introduction

Extreme movements in the prices of financial assets are of great concern to both investors and the entire economy. A single negative return, or a combination of several smaller returns can create serious liquidity problems. The amount of capital insolvency or portfolio shrinking these liquidity problems cause, can further lead to the bankruptcy of the investor. Furthermore, the whole economy may be affected if enough investors experience such losses.

There is an increasing need for modeling of events that cause larger shocks to the underlying financial system for effective risk management. The crisis occurrence frequency is getting higher and higher as time goes by. The time interval between two consequent crises is consistently shrinking. Below is a list of the crisis that had shaken the Turkish economy in the last fifteen years.

1992-1993	ERM (Exchange Rate Mechanism)
1994	Latin America, Mexico-Tequila
1994	Currency Crisis Turkey
1997	Southeast Asia
1998	Russia
2000	Liquidity
2001	Twin Crisis (Banking and Currency)

Table 1: List of the years of recent crisis that affected Turkey

This list supports the argument that the crisis occurrence is getting more and more frequent. Furthermore, these crisis also revealed how easily the impact of a crisis occurring at some part of the world, can be spread to the rest of the global economy. This trend, called the contagion effect, pushes investors to be more alert to risks not only local but also global.

Value-at-Risk (VaR) is a concept used in measuring the extreme risk mentioned above. The concept is widely used in the finance industry and the firms try to minimize their insolvency risk by sustaining capital reserves that calculated VaR levels indicate. Hence correct VaR calculations play an important role in decision making in the finance community. One of the major difficulties in VaR calculations is to estimate the tails of return distributions where there is a limited number of observations.

In the last few years, Extreme Value Theory (EVT) methods received attention from the finance professionals as the approach promises to estimate the tails of the financial asset returns. The main idea behind these methods is to directly model the tail without the need to make any assumptions regarding the center of the distribution.

This study focuses on using EVT in estimating VaR for several stocks and indices from Istanbul Stock Exchange. Therefore, the VaR concept and EVT will be discussed in more detail in the next introduction sections.

1.1 VaR-Value at Risk Concept

Value-at-Risk is a measure used for quantifying extreme risk. VaR measures how much can be lost with p% probability (95% or 99% for example) over a given period of time (one day or a fortnight for example), i.e. Pr (Loss >VaR) $\leq (1 - p)$. Thus, having VaR calculated, this number summarizes information about the risk of a portfolio.

VaR is shown with the red line in Figure 1. For the given example in the plot, the daily returns are assumed to have a density function shown with the blue line. At 99% percent VaR level, the calculated VaR is 18.6%. This means that for each dollar invested, one can lose \$0.186 or greater the next day with a probability of one in a hundred.

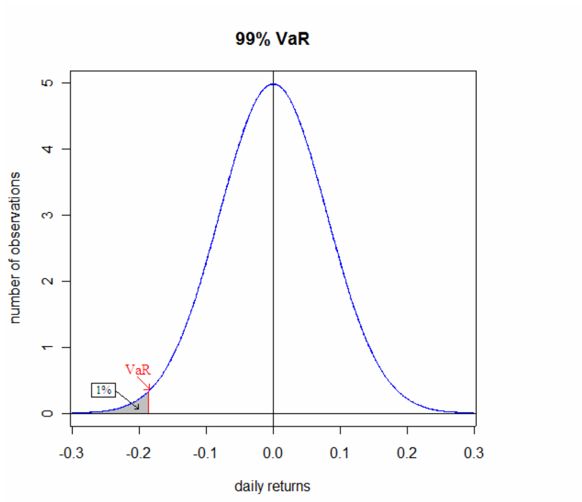

Figure 1: Value at Risk

VaR gained almost immediate acceptance from both regulators and practitioners because of its simplicity and intuitive appeal. As of January 1, 1998, the Basle Committee on Banking Supervision requires banks with significant market exposure to report their daily VaR as a basis to determine their minimum capital adequacy requirements. In the United States, VaR is also accepted by the Securities Exchange Commission as an appropriate measure for corporations to report the risk of their business.

Whilst Basle Committee compels financial institutions to calculate VaR for determining their level of capital adequacy, the Committee does not force institutions to use a prescribed single model. Financial institutions are free to use for example variance-covariance, historical simulation, Monte Carlo methods or a combination of these methods. Each financial institution is encouraged to have its own internal model so that computations in this field evolve better and better.

However, the RiskMetrics™ approach developed by J.P. Morgan became the most popular one. And many financial institutions today follow their approach in estimating the VaR for their portfolios.

Even though, there are more elaborate risk measures such as "Expected Shortfall" (ES) or "Conditional VaR" (C-VaR) being discussed in academic studies, their application in the financial institutions is still limited. The reason may be that C-VaR or ES calculations are not necessitated by regulatory bodies or that computations require an estimate of the whole tail of the firm-wide loss distribution.

For now, VaR is used as a standard risk measure around the world today. So, it became very important to measure VaR accurately for risk management practices. The main difficulty in calculations lies in estimating the tails where there are only very limited observations. The existing models for estimating overall distributions can be divided into the following two categories: 'parametric models' and 'non-parametric models'.

Parametric models try to fit the returns to various known statistical distributions (e.g. normal, symmetric stable, student *t*) or mixture of statistical distributions. VaR estimates are derived once the density function is fitted.

However, the true density function form is often not known and can be time-varying. Conditional models like Autoregressive Conditional Heteroscedasticity (ARCH) and Stochastic Volatility (SV) assume price changes to follow a statistical distribution conditionally. They model the variance of returns as a time dependent process. The advantage of these models is that they consider both fat tails and the dependence problem of the financial data.

Generally, the true distribution function can not be known due to the complexity of the system. The non-parametric approaches are therefore used to model the density function of a given circumstance. Without any distributional assumptions, the data is fit directly according to some statistical criteria. However, the non-parametric models do not work in tails where data is sparse and it is impossible to extrapolate to areas where there is no sample data.

1.2 Extreme Value Theory

Both parametric models and non-parametric models discussed above achieve the best overall fit at the expense of tail fitting. Outlying observations on the tails of distributions occur very rarely but indeed they have important influence on the behavior of the whole model. EVT, an interdisciplinary field in probability theory and statistics, brings in statistical techniques and introduces appropriate mathematical models for learning the probabilities of these outliers.

The main advantage of EVT is that the only focus is on the tail of distribution. The concern is not to estimate the correct shape of the whole distribution function. The central observations are ignored and only extreme observations are taken into consideration. Another advantage of EVT is that since it is a parametric approach it enables extrapolation, i.e. the estimated model can project for values that are beyond those observed so far. Other risk measures of C-VaR and ES can also be estimated using EVT. Furthermore, it can capture event risks, such as crashes and currency devaluations.

The preface of the famous book named *Modelling Extreme Events* start as follows: “In a recent issue, *The New Scientist* ran a cover story under the title: ‘Mission improbable. How to predict the unpredictable?’ In it, the author describes a group of mathematicians who claim that EVT is capable doing just that: predicting the occurrence of rare events, outside the range of available data…..” (Embrechts et al. 1997: VII)

EVT gives us a ground to better explain the outlier or extreme events, which the need for it especially in finance is growing more and more as the number of so-called extreme events has increased considerably.

1.2.1 Purpose and essential conditions

The purpose of the statistical theory of extreme values is to model observed extremes in samples of some specified size, or to estimate the number of extreme data points in a related group.

The essential conditions to study EVT are only that; the phenomenon being measured is a stochastic (random) variable, the initial distribution from which the samples with extreme values have been drawn remains constant from one set of samples to the next, i.e. extreme events are drawn from a common population, and the observed extremes are statistically independent.

Accordingly, EVT does not impose assumptions on the distribution function of the population. For example, the returns need not be assumed to be normally distributed whilst using EVT.

1.2.2 Asymptotic Approach

EVT uses asymptotic theory to base its inferences about extreme events in the tails. The theory tells what the distribution of extreme values should look like in the limit as the sample size increases. For example, in the case that the true density function of some return series is not known and the essential conditions listed above hold, the theorem tells that the distribution of extreme returns converge asymptotically to one of the three types of GEV functions which will be described in Section 2.1.2.

In analyzing stock returns, usually high density (near continuous-time observed) data are used. The distributions are heavy-tailed and return data exhibit clustering of extremes and long-range dependence. Even though there is no universally accepted model that explains all of these phenomena, asymptotic approach of EVT can help estimating the tails using large sample sizes.

1.2.3 Extreme Value Theory and Finance

Extremal events in finance have the advantage that they are mostly quantifiable in units of money as opposed to other extremal occurrences, such as floods and earthquakes, which might cost the loss of lives. However, even in finance there are market crashes that also have a non-quantifiable component: these events can cause a larger shock to the underlying financial system. Contagion risk or the risk that a securities market decline in one country will spread to another, causing serious market losses, or that the failure of one large financial institution will lead to the failure of others.

As it will be mentioned in Literature Review, EVT was first applied by engineers. Following engineers, the finance community has also found out that using EVT may be helpful in risk management. The applications of EVT on financial time series data for quantifying risk measures will be mentioned in Section 2.3.

Daily or weekly reported prices such as stocks, foreign currencies, or commodities such as crude oil, cotton, sugar, etc can all be considered as financial time series data. Firms perform risk management to guard against the risk of loss due to the fall in prices of financial assets held or issued by the company. What is of importance here are the magnitudes of the changes in prices, rather than the average variations. A single, extremely negative return or a sequence of smaller negative returns, can lead a company to bankruptcy. Hence generally companies feel the need to set trading limits. In order to determine these limits, probabilities of extreme negative returns need to be estimated.

Researchers with the aim of studying these trading limits find themselves in the contradictory situation that extreme risks are, by definition, rare; whereas significant statistical results can only be achieved if a sufficient number of these events can be analyzed. Unfortunately, in most mature markets the number of data in real cases is relatively limited. But as ISE is a trading platform in an emerging market, the securities traded in this exchange have high volatilities. Consequently, the dataset (ISE returns) that will be analyzed in this study is rich in extreme returns and losses. EVT places emphasis on the tail behavior as the basis for the analysis. In other words, one is interested in the distribution of the largest order statistic, say M_n, which gives the likelihood of an extreme realization. For this reason, this study focuses on the tail of the distribution. Estimating the lower tail of the distribution is especially important in the correct calculation of VaR figures.

2 Literature Review

In literature, the financial return data is observed to have alternation between periods of tranquility and volatility. Periods of persistent high volatility are followed by periods of persistent low volatility. This feature of financial returns is known as volatility clustering.

Most importantly, there is empirical evidence that distributions of returns can possess heavy (fat) tails so that a careful analysis of returns is required. Fat-tailed distributions exhibit more probability mass in the tails than distributions such as the standard normal distribution. This means that extremely high and low realizations will occur more frequently than under the hypothesis of normality.

Mandelbrot (1963) and Fama (1965) were the first to use heavy tail distributions in finance. They proposed infinite variance models such as the stable distribution to model daily stock returns. However, Akgiray and Booth (1987) investigated the tail behavior and found that the tails of stable distribution are too thick to fit the empirical data.

As this study aims to model the tails of daily returns of securities being traded in ISE via techniques developed by EVT and estimate VaR, the literature review concentrates on the following three topics: EVT, VaR and former studies that used EVT in VaR estimations.

2.1 Extreme Value Theory

2.1.1 Background of EVT

Comparing to the general statistics history, extreme value statistics history is quite brief. Initially, statisticians were concerned with studying the behavior of the masses rather than studying rare events. Kinnison (1985), who offers a thorough history of extreme value theory, acknowledges Fourier in 1824 for the oldest remarks in the statistical literature about extreme values. Fourier had the following remark for the Gaussian distribution that the probability of a deviation being more than three times the square root of two standard deviations from the means is about 0.00002 and hence the observation associated with this deviation could be neglected.

This can be the source of the common but untrue statistical rule of thumb that plus or minus three standard deviations from the mean should be considered the maximum range of valid samples from a Gaussian distribution, irrespective of the sample size taken.

However, as the sample size increases, the largest value encountered in a sample will similarly increase. This is due to the fact that there is more opportunity for values in the tails to occur.

In 1877, Helmert added that the probability of surpassing any specified value depends on the size of the sample. If the distribution that was being sampled is unlimited, no matter how small the probability of the limits given by a rule, then the largest or smallest sample observation is also unlimited.

As a result, the study of extreme value theory is an attempt to describe the relationship between the size and magnitude of the observed extreme values. Now, we can see that the three sigma rule is far conservative for small samples and too liberal for large samples.

Dodd (1923) started the modern day study of statistical extremes with his paper. Fréchet (1927) and Fisher-Tippett (1928) were the followers. Papers of de Finetti (1930), Gumbel (1935) and von Mises (1936) are a few of the most quoted ones written on the subject.

Especially the paper of Fisher-Tippett (1928) is considered as the foundation of asymptotic theory of extreme value distributions. The theorem in the paper specifies the form of the limit distribution for centered and normalized maxima. They argue that in the existence of a normalizing constant and a centering constant, the distribution follows one of the three families of extreme value distributions which will be described in Section 2.1.3.

Since Fisher and Tippett, a large number of books and articles have been written on extreme value theory. Gnedenko (1943) was the first to provide a rigorous proof. De Haan (1970) subsequently applied regular variation as an analytic tool. Then, Weissman (1975) presented a shorter version of de Haan's proof. Weissman (1975) generalized the problem to include non-identically distributed observations. He later in 1978 derived estimators when only the k largest observations of a sample size n are available. De Haan (1981) focused on constructing a confidence interval for estimating the minimum of a function using order statistics.

Beirlant et. al. (1996), Gumbel (1958), and Pfeifer (1989) concentrate more on the statistical methods based on extreme value theory.

Anderson (1970, 1980), Arnold, Balakrishnan and Nagaraja (1992), and Gordon, Schilling and Waterman (1986) studied on extremes for discrete distributions. And Adler (1990), Berman (1992), and Leadbetter et al. (1983) studied extremes for continuous time processes, in particular the Gaussian distribution.

Leadbetter, Linger, and Rootzén (1983) concentrated on extremes of stationary sequences and processes, along with extremes of dependent variables. Galambos (1987) studies the weak and strong limit theory for extremes of independent and identically distributed (i.i.d.) observations.

Resnick (1987) has worked primarily on i.i.d. observations and extreme value theory. The theory of the regularly varying functions is the main analytic tool and the point process theory is his basic probabilistic tool.

Resnick (1987), Reiss (1993), and Galambos (1987) are the first ones to include results based on multivariate extremes in their studies.

Contrary to the classical limit theorems that primarily concern the weak convergence of distribution functions, the main results of Reiss (1989) are formulated in terms of the variational and Hellinger distance. A collection of proceedings from a conference on extreme value theory edited by Hüsler and Reiss (1989) includes recent developments and extensions to multivariate extremes.

The application of extreme value theory began in the middle 1930s with the work of E.J. Gumbel, first in Germany and then in the U.S. when World War II engulfed Europe. Gumbel's first application was the consideration of the longest duration of life, or older age (Gumbel, 1935). Then, he showed that the statistical distribution of floods, long studied by engineers, could be understood by the use of extreme value theory (1941). These procedures have been extensively applied to other areas of science. For example, Tiago de Oliveira (1983) presents the proceedings from the 1983 NATO Advanced Study Institute on Statistical Extremes and Applications to obtain a complete perspective of the field, along with a series of applications.

The book *Statistics of Extremes,* written by Gumbel and published in 1958, pulled interest in the engineering community. Since then, as scientists, applied engineers, mathematicians and statisticians used the theory in their daily practices, many advances in the extreme value theory have taken place.

Extreme value theory had been mostly used in the biological, engineering and environmental studies. For these fields, extreme or unusual conditions are more important than usual conditions. Several examples of fields extreme value theory has decisive role in are listed in Table 2. However, a longer list of references of the use of extreme value theory in engineering is given by Castillo (1988).

Field of study	**EVT has decisive role in**
Carcinogenesis studies	determining maximum dose of chemotherapy
Structural engineering	estimating the probability of occurrence of extreme winds and earthquakes
Ocean engineering	probability distribution of the largest waves
Hydraulics engineering	flood frequency analysis
Pollution studies	determining a critical level of pollution concentration
Meteorology	Study of extreme events such as very high or low temperature, rainfall, sea levels, wind speeds, hurricanes, etc.
Material strength	size effect analysis and possible and reliable extrapolation making
Fatigue strength	estimating the number of cycles to failure under the action of repetitive loads
Electrical strength	estimating lifetimes based on random voltage levels
Corrosion resistance	predicting corrosion failure due to the action of chemical agents

Table 2: Fields that use Extreme Value Theory

Hosking, Wallis, Wood (1985), Hosking, Wallis (1986a, 1986b, 1988), Hosking (1994) have all performed research on extreme value analysis and flood frequency analysis. Barnett and Turkman (1993) present case studies in environmental pollution statistics. Jerkinson (1955) introduced the generalized extreme value distribution while examining meteorological events. Shibata (1994) had an application of extreme value statistics to corrosion.

2.1.2 Univariate Extreme Value Theory

The study aims to estimate the distribution of daily returns of several stocks and indices, which will be described in detail in section 3.1. The portfolio of stock returns, hence their joint distributions are not estimated. This limits the study on a univariate level.

EVT enables two different approaches: the classical approach that uses block maxima data and the more recent approach that uses exceedances over a specified threshold.

In general, for some studies only the annual maximum data may have been recorded. Then working with this annual maxima data does not give the researcher the option of modeling exceedances over a threshold. However, using daily stock returns, the data makes it available that all daily observations are known rather than only the maxima within blocks. Rather than working with data of block maxima, say maxima of daily returns over months, selecting a threshold and working with the data of exceedances over the threshold is a much more efficient way of using data. Yet, one can either assume that daily returns are the maxima over blocks of size of a single day or that daily returns are just point observations.

Let X_i (i=1, 2, n) be a sequence of independent and identically distributed (i.i.d.) random variables. The variables are ordered and reindexed such that order statistics are: $X_1 \leq X_2 \leq_{\ldots} \leq X_{n.}$

If the common distribution function of these sequence of random variables is F then the tail quantile function Q is defined as follows:

$Q\,(n) = \inf\{u; F(u) \geq 1 - 1/n\}$.

The maximum domain of attraction condition that controls EVT is that if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that:

$$\Pr\left\{\frac{X_n - b_n}{a_n} \leq x\right\} \rightarrow G(x) \quad as \quad n \rightarrow \infty$$

for a non-degenerate distribution function $G,$ then G is a member of the generalized extreme value family of distributions (EVD) G_γ.

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right) \qquad \text{Equation 1}$$

Thus, if the tail index γ is estimated, the distribution function of the random variable can also be estimated. The tail index γ is hence the main indicator about the decay of the distribution tail. When $\gamma < 0$, the distribution is classified as Weibull type and it has finite endpoint. When $\gamma = 0$, the distribution is classified as Gumbel type and it decays exponentially. When $\gamma > 0$, the distribution is classified as Fréchet type and it decays polynomially.

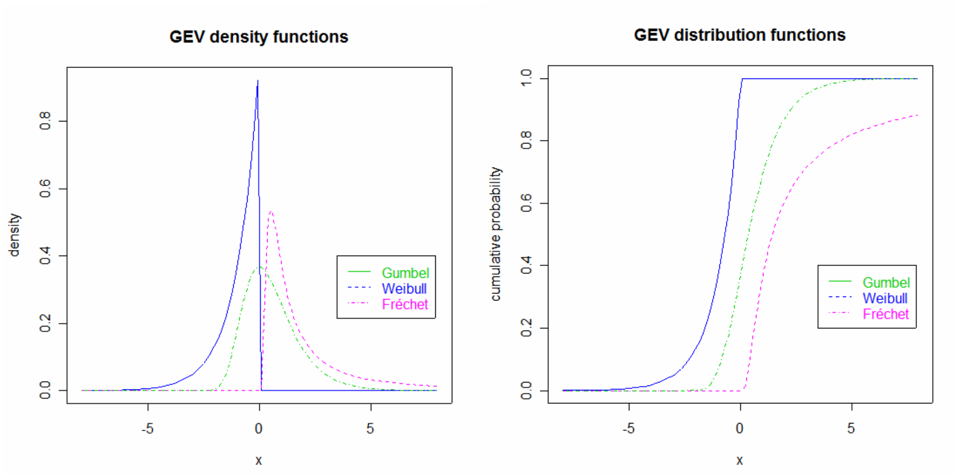

Figure 2: Density and distribution functions of GEV family

Equation 1 is the continuous, unified general model for the family of extreme value distributions. This general formula can be decomposed into three separate formulas by setting $\alpha=1/\gamma$.

Gumbel : $G_0(x) = \left\{ \exp(-e^{-x}) \quad \textit{for all } x \right\}$ Equation 2

Fréchet $\alpha>0$: $G_{1,\alpha}(x) = \begin{cases} \exp(-x^{-\alpha}) & x \geq 0 \\ 0 & x < 0 \end{cases}$ Equation 3

Weibull $\alpha<0$: $G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{-\alpha}) & x \leq 0 \\ 1 & x > 0 \end{cases}$ Equation 4

In this notation each parameter α determines a standard EV distribution. Accordingly, the relevant densities $g=G'$ of the standard EVD functions are:

Gumbel : $g_0(x) = G_o(x)e^{-x}$, for all x;

Fréchet : $g_{1,\alpha}(x) = \alpha G_{1,\alpha}(x)x^{-(1+\alpha)}$, $x \geq 0$;

Weibull : $g_{2,\alpha}(x) = |\alpha| G_{2,\alpha}(x)(-x)^{-(1+\alpha)}$, $x \leq 0$.

As it can be observed in Figure 2, Fréchet densities and the Gumbel density are skewed to the right. Weibull densities, on the other hand, are skewed to the left when α > -3.6, look symmetrical if α is close to -3.6, and are skewed to the right if α < -3.6. Furthermore, for α > -1, Weibull densities have a pole at zero. The effects of shape parameter α on Fréchet and Weibull densities can be seen in Figure 3.

The left figure shows four Fréchet densities at α level 0.5, 1, 2 and 3. All figures are right skewed as expected and the curves get steeper as α decreases. The right figure, on the other hand, depicts again four Weibull densities. The shape parameters α of these densities are -0.5, -1, -3.6 and -5. Again, the curves with lower absolute values of α are steeper. The skewness properties of Weibull densities at different α levels can also be observed in this figure.

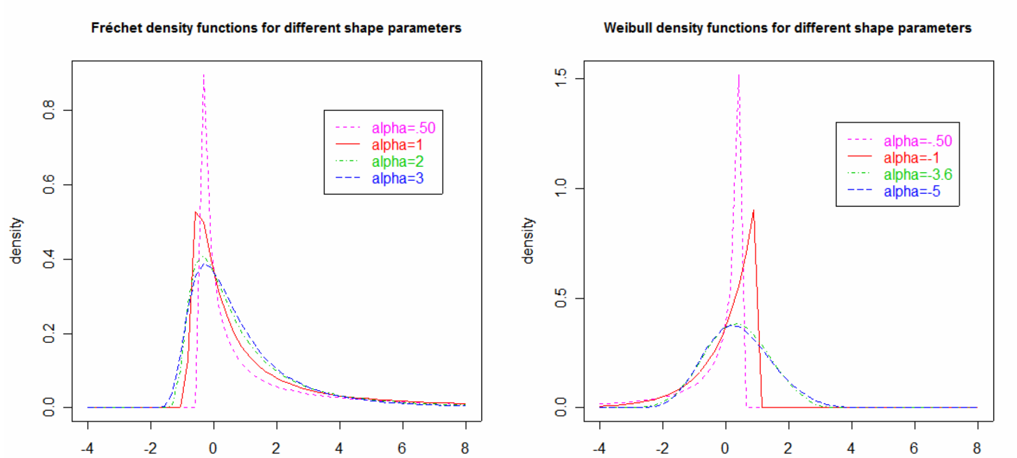

Figure 3: Effect of shape parameters on Fréchet and Weibull densities.

If a random variable X has the distribution function F, then (μ + σ X) has the distribution function $F_{\mu,\sigma}(x)$ =F ((x-μ)/ σ), where μ and σ > 0 are the location and scale parameters. Given the property above and adding the location and scale parameters, μ and σ respectively, full notation of extreme value models are obtained. Then Gumbel, Fréchet and Weibull distribution functions in Equations 2, 3, and 4 in more general context become:

$$G_{0,\mu,\sigma}(x) = \exp(-e^{-(x-\mu)/\sigma})$$

and

$$G_{i,\alpha,\mu,\sigma}(x) = G_{i,\alpha}(\frac{x-\mu}{c}), \qquad i=1, 2.$$

Likewise the density functions for Gumbel, Fréchet, and Weibull are now under this notation;

$$g_{0,\mu,\sigma}(x) = \frac{1}{c} e^{-(x-\mu)/\sigma} \exp(-e^{-(x-\mu)/\sigma})$$

and

$$g_{i,\alpha,\mu,\sigma}(x) = \frac{1}{c} g_{i,\alpha}(\frac{x-\mu}{c}), \qquad i=1, 2.$$

The effects of location and scale parameters on Gumbel function can be observed in Figure 4. The figure on the left shows five different Gumbel densities with five different location parameters, all having σ = 1. Then, the figure on the right shows five different Gumbel densities with five different scale parameters and μ = 0.

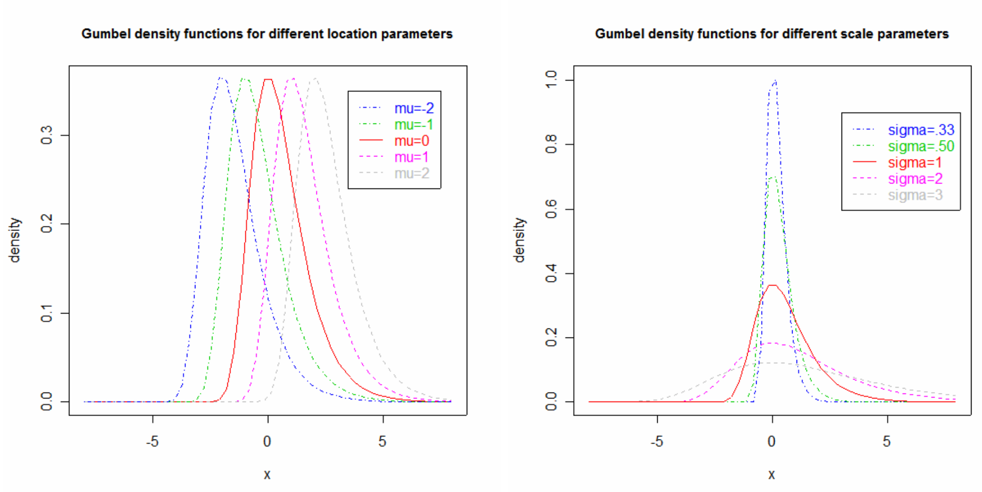

Figure 4: Effects of μ and σ shown on Gumbel density function.

Because of the well-known leptokurtic property of the financial data series, empirical applications in finance are mainly concerned with EVD distributions where $\gamma > 0$, i.e. the Fréchet distribution.

2.1.3 Peaks Over Threshold Model

More recent studies on Extreme Value Theory use data of exceedances rather than working with block maxima or minima. Since this approach uses exceedance data over a given high threshold u, those of the outcomes that exceed u are regarded as extreme events. The EVT model used in this approach is known as the Peaks over Threshold (POT) model or generalized Pareto distribution (GPD) model.

In literature, the theoretical foundation leading to this generalized Pareto model is attributed to Pickands (1975). This model was further developed in the studies of Smith (1987) and Leadbetter (1991). Tajvidi (1996) also studied statistical estimation of the parameters of the model. The main framework of GPD approach these researchers had developed is as follows:

There is an analytic relationship between GPD family of distribution functions (will be denoted by H) and EVD family of distribution functions G defined in the above section. The relationship is;

$$H(x) = 1 + \log G(x), \qquad \text{if } \log G(x) > -1$$

GPD functions are the only continuous distribution functions F such that for a certain choice of constants b_u and a_u, $F^{[u]}(b_u + a_u x)$ is equal to $F(x)$. Here $F^{[u]}(x)$ is again the exceedances distribution function at u relevant to F (the truncation of F left of u). More explicitly, $F^{[u]}(x) = (F(x) - F(u))/(1 - F(u))$. This property is known as the pot-stability of GPD functions.

$F^{[u]}(x)$ is thus the conditional probability that given an exceedance occurs the outcome will be below x. Equivalently, the conditional probability of an extreme event given that the threshold is exceeded is:

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u + y)}{1 - F(u)} = 1 - F^{[u]}(u + y), \qquad y > 0$$

where X_1, X_2, ... is a sequence of i.i.d. random variables having marginal distribution function F.

As the distribution F is unknown, generalized extreme value approximation to the distribution of maxima can be used. The parametric modeling of exceedance distributions $F^{[u]}$ by GPD is based on a limit theorem again. Since the intention is modeling exceedances at high thresholds, thresholds are considered to tend to the right endpoint of the actual distribution *F*.

Balkema-de Haan (1974) states in their study the following limit theorem:
If $F^{[u]}$ $(b_u + a_u\, x)$ has a continuous limiting distribution as *u* goes to the right endpoint $w\,(F)$ of *F*, then $\left| F^{[u]}(x) - H_\gamma(x) \right| \rightarrow 0$ as $u \rightarrow w\,(F)$.

If the convergence above holds, then *F* is said to belong to the pot-domain of attraction of $H\gamma$. This limit distribution of the exceedances over *u* when $u \rightarrow \infty$ is:

$$H_{\gamma,\sigma}(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma} \qquad \text{Equation 5}$$

Equation 7 is the continuous, unified GPD model. Similar to the case in classical family of extreme value distributions, it can again be decomposed into three separate formulas with α $(= 1 / \gamma)$ notation.

$$\text{Exponential} \quad : H_0(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases} \qquad \text{Equation 6}$$

$$\text{Pareto} \quad \alpha > 0 \quad : H_{1,\alpha}(x) = \begin{cases} 0 & x < 1 \\ 1 - x^{-\alpha} & x \geq 1 \end{cases} \qquad \text{Equation 7}$$

$$\text{Beta} \quad \alpha < 0 \quad : H_{2,\alpha}(x) = \begin{cases} 0 & x < \text{-}1 \\ 1 - \left(-x\right)^{-\alpha}) & -1 \leq x \leq 0 \\ 1 & x > 0 \end{cases} \qquad \text{Equation 8}$$

In this notation, for positive and negative *α,* the GPD type is called Pareto and Beta, respectively. The relevant densities $h = H'$ are:

Exponential : $h_0(x) = e^{-x}$ $\quad x \geq 0$

Pareto : $h_{1,\alpha}(x) = \alpha\ x^{-(1+\alpha)}$ $\quad x \geq 1$

Beta : $h_{2,\alpha}(x) = |\alpha|\ (-x)^{-(1+\alpha)}$ $\quad -1 \leq x \leq 0$

The density and distribution functions are plotted for Exponential, Pareto and Beta functions in Figure 5.

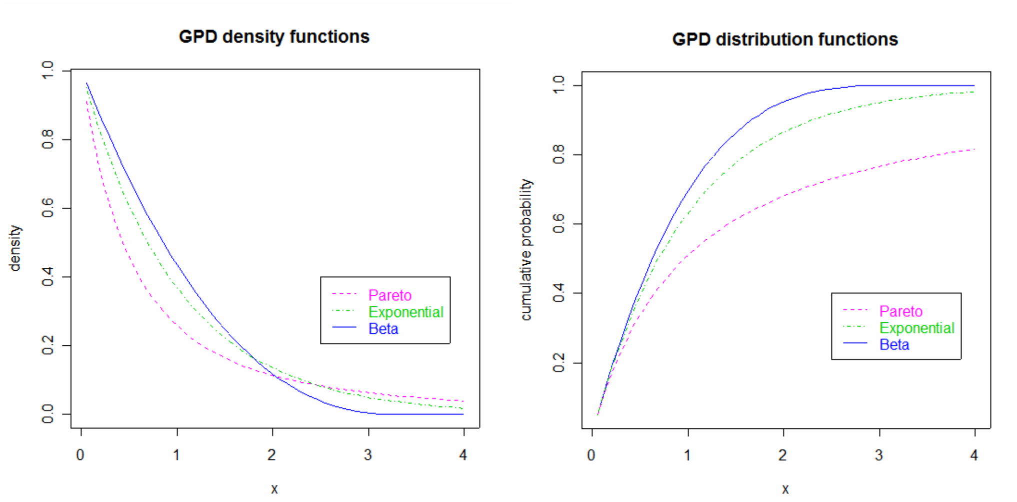

Figure 5: Density and distribution functions of GPD family

Again location and scale parameters μ and $\sigma > 0$ are added to the equations above in order to obtain the full statistical families of GPD functions. Considering the context that the study uses financial data, we are particularly interested in Pareto distribution functions $H_{1,\alpha,0,\sigma}(x) = H_{1,\alpha}(x/\sigma)$ with scale parameter $\sigma > 0$ and fixed location parameter $\mu = 0$.

In Figure 6, the red line on the left diagram is the density function for the Exponential distribution. The other lines are Pareto densities all with location $\mu = 0$

and scale σ = 1, but with different shape parameters γ = 0.5, 1, and 2. In this plot we can see the convergence of Pareto densities to the Exponential density as γ decreases. The right plot of Figure 6, shows Pareto densities all with location μ = 0 and shape index γ = 1, but with four different levels of scales σ ranging from 0.5 to 3.

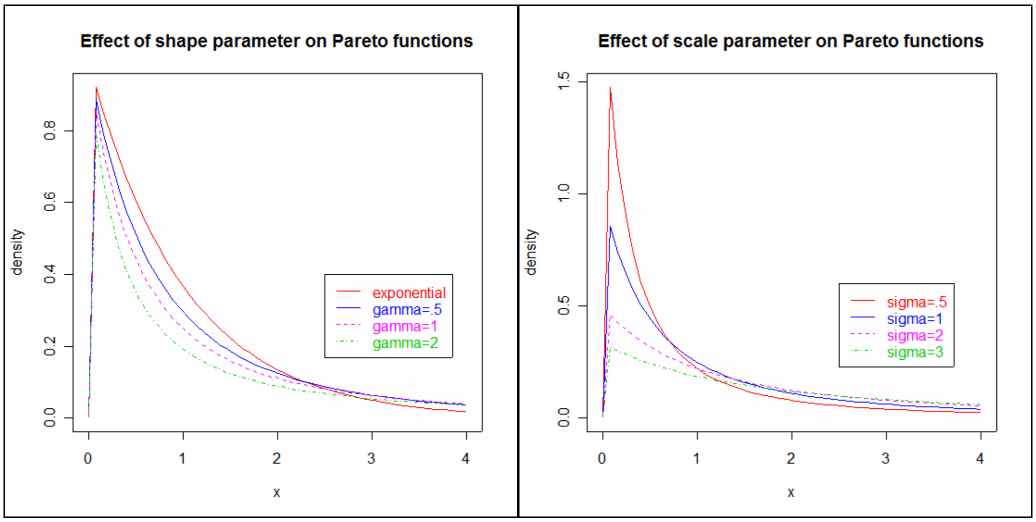

Figure 6: Effects of γ and σ shown on Pareto density function.

In determining the threshold u, the researches face the following handicap. If the threshold is too low, then the asymptotic basis of the model working on maxima is likely to be violated. On the other hand, if the threshold is too high, then the model will be estimated with only a few exceedances and hence the estimators will have high variance. This is known as the bias-variance tradeoff in selecting thresholds.

This threshold choice problem was investigated by Davison and Smith (1990), who have combined their former studies in Davison (1984) and Smith (1984) and recommended the use of mean residual life plot for threshold selection.

Mean residual life plot is the scatterplot of the points:

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u)\right) : u < x_{\max}\right\}, \qquad \text{Equation 9}$$

where $x_1, x_2, ..., x_{n_u}$ are the n_u observations that exceed u and $x_{\max}$ is the maximum of X_i. Above a valid threshold level u, the mean residual life plot is expected to be approximately linear due to the relationship stated in Coles (2001: 79)

$$e(u) = \mathrm{E}(X - u | X > u) = \frac{\sigma_u}{1-\gamma} = \frac{\sigma_{u_0} + \gamma u}{1-\gamma}. \qquad \text{Equation 10}$$

2.1.3.1 Parametric Approaches

After selecting the threshold, the fit of the generalized Pareto distribution over this threshold can be performed by one of the following methods: Maximum likelihood (Smith (1987)), Probability-weighted moments (Hosking and Wallis (1987)), Bayesian analysis methods (Coles and Powell (1996)), Elemental percentile method (Castillo and Hadi (1997)).

Among these four approaches the classical ones which are maximum likelihood and probability weighted moments are the most preferred. However, both have some constraints regarding the tail index γ. In the later studies, Castillo and Hadi (1997) proposed an elemental percentile method (EPM) that does not impose any restrictions on the tail index γ. Also, Coles and Powell (1996) used Bayesian methods to estimate GPD parameters; however these methods have not been excessively explored. Cabras-Castellanos (2005) state that the reason why Bayesian methods have not been explored neither in a non-informative context, nor in an informative one, may be due to the lack of physical meaning of the unknown parameters.

In the following sections, the classical approaches maximum likelihood estimation and probability weighted moments will be discussed in more detail.

2.1.3.1.1 Maximum Likelihood Estimation

Marginal distribution functions of POT models can estimated by maximum likelihood method. The maximum likelihood estimators have been considered by many authors, including Davison (1984), Smith (1984), DuMouchel (1983), Hosking-Wallis (1987) and Grimshaw (1993). Smith (1984) showed that for $\gamma > -0.5$ the likelihood is regular in the sense that the Fisher information matrix exists. A summary of the findings of Smith (1987) is below.

Let X_1, X_2, ... be a sequence of i.i.d. random variables having marginal distribution function F. Assume F is a GPD with parameters γ and σ, then the density function f is:

$$f_{\gamma,\sigma}(x) = \frac{\gamma}{\sigma}\left(1 + \gamma\frac{x}{\sigma}\right)^{-\frac{1}{\gamma}-1}$$

The log-likelihood function of Equation 5 equals

$$\ell((\gamma,\sigma);X) = -n\ln\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{n}\ln\left(1+\frac{\gamma}{\sigma}X_i\right).$$

Solving this likelihood function, the MLE estimates $\hat{\gamma}_n$ and $\hat{c}_n$ are attained. The method is applicable for cases when $\gamma > -0.5$.

Then application of MLE on the n_u excesses of a threshold u, where y_i are the exceedance values ($y_i = X_i - u$), the log-likelihood equation becomes for $\gamma \neq 0$

$$\ell(\sigma, \gamma) = -n_u \ln \sigma - (1 + 1/\gamma) \sum_{i=1}^{n_u} \ln(1 + \gamma y_i / \sigma), \qquad \text{Equation 11}$$

provided $(1 + \gamma y_i / c) > 0$ for $i = 1, \ldots, n_u$; otherwise $\ell(c, \gamma) = -\infty$.

In the case $\gamma = 0$, the log-likelihood is

$$\ell(\sigma) = -n_u \ln \sigma - (\sigma^{-1}) \sum_{i=1}^{n_u} y_i . \qquad \text{Equation 12}$$

Since the log-likelihood above cannot be maximized analytically, numerical techniques are applied for the optimization process.

MLE parameters are consistent and asymptotically efficient as

$$\sqrt{n}\left(\hat{\gamma}_n - \gamma, \frac{\hat{\sigma}_n}{\sigma} - 1 \right) \xrightarrow{d} N(0, M^{-1}), \quad n \to \infty,$$

where,

$$M^{-1} = (1+\gamma)\begin{pmatrix} 1+\gamma & 1 \\ 1 & 2 \end{pmatrix}$$

and $N(\mu, \sum)$ stands for the bivariate normal distribution with mean vector μ and covariance matrix $\sum$.

2.1.3.1.2 Probability-Weighted Moments

The next method with which generalized Pareto distribution can be fitted to data on excesses of high thresholds is the method of probability weighted moments (PWM). Hosking-Wallis (1987) implemented the following probability weighted moment approach for the GPD.

The r^{th} probability weighted moment is

$$\omega_r = \int_0^1 Q(p)p^r dp = E\{Q(p)p^r\}$$

where p has a uniform distribution over [0, 1]. For the order statistics $x_1 \leq x_2 \leq \ldots \leq x_n$ from a random sample of size n, then

$$\omega_r = \frac{1}{n}\sum_{i=1}^{n} x_i p_i^r .$$

For GPD having $\gamma < 1$,

$$\omega_r = E\{X(\overline{G}_{\gamma,\sigma}(X))^r\} = \frac{\hat{c}}{(r+1)(r+1-\hat{\gamma})}, \qquad r = 0, 1, \ldots$$

where X has GPD $G_{\gamma,\sigma}$. Then immediately the parameter estimates obtained are;

$$\hat{\gamma} = 2 - \frac{a_0}{\omega_0 - 2\omega_1}$$

and

$$\hat{\sigma} = \frac{2a_0a_1}{\omega_0 - 2\omega_1}.$$

2.1.3.1.3 Comparison of MLE and PWM

The two classical methods MLE and PWM used for estimation of GPD parameters described above have been compared for GPD distributed data in simulation studies by Hosking-Wallis (1987) and in the Rootzén-Tajvidi (1997) study.

Hosking-Wallis (1987) found that for particularly small sample sizes and for GPD data with shape parameter γ in the range between 0 and 0.4, the PWM method has advantages over the MLE method as PWM estimates showed less mean squared error. However, as the sample size increases the difference becomes less pronounced. Then again, Rootzén-Tajvidi (1997) showed that for heavy tailed data with $\gamma \geq 0.5$ the PWM method gives seriously biased parameter estimates whereas ML estimates are consistent.

2.1.3.2 Nonparametric Approaches

Several different estimators have also been proposed to estimate the tail index. Hill (1975) and Dekkers et al.(1989) are the most well-known among these. These nonparametric estimators use upper order statistics and they will be further described in the following sections.

2.1.3.2.1 Hill Estimator

Hill (1975) proposed the following estimator for γ.

$$H_k = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1} - \log X_{n-k} . \qquad \text{Equation 13}$$

Here, $k = k_n$ is a sequence of positive integers ($1 \leq k < n$) which, in theoretical asymptotic considerations satisfy the conditions $k \to\infty$ and $k/n\to 0$ as $n\to\infty$.

However, plotting Hill estimates $H_{k.n}$ against k, high volatility of the Hill estimate is observed. This makes it very difficult to use the estimator without having a guideline for the choice of k.

Refined estimators of this type and more general classes of estimators in this spirit were discussed in Drees (1995), Drees (1996), and Drees (1998).

2.1.3.2.2 Dekkers Estimator

Dekkers, Einmahl and de Haan (1989) proposed the following estimator, which brings about an adaptation of the Hill estimator to estimate γ.

$$\hat{\gamma}_k^D = H_k + 1 - \frac{1}{2}\left(1 - \frac{H_n^k}{S_k}\right)^{-1} \qquad \text{Equation 14}$$

where

$$S_{k,n} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1} - \log X_{n-k})^2 .$$

2.2 Basle and VaR

After Basle Accord assigned VaR as the risk measurement method, next the question of determining the quality of the models that the practitioners use came up. Basle Committee thus reinforced backtesting procedure and motivated the practitioners to improve the models by imposing low multiplication factors for the models that yield good backtesting results.

Bank for International Settlements (BIS) determines the multiplication factors for capital requirements using the backtest procedure explained in Basle Committee on Banking Supervision (1996).

Accordingly, in the current BIS implementation the practitioners need to produce T=250 VaR forecasts at 99% level. VaR_t denotes the forecast for day t using the information up to time t - 1. Then the following backtesting procedure form the exceedance variable I_t.

$$\begin{Bmatrix} I_t = 1 & if & -r_t > VaR_{t|t-1} \\ I_t = 0 & if & -r_t \leq VaR_{t|t-1} \end{Bmatrix}$$

Equation 15

Then, by the definition above Pr ($I_t = 1$) = Pr ($r_t < -VaR_t$). And by the definition of VaR it follows that $-VaR_t = F_t^{-1}(1-p)$, with F the cumulative distribution function of r_t. Combining the equations, we have Pr (I_t = 1) = (1 − p). Consequently, the distribution of I_t follows a Bernoulli distribution. Then the cumulative distribution of the binomial distribution can be used based on the number of exceedances, i.e. $\sum_{t=1}^{T} I_t$.

Zone	Number of exceedances	Plus factor	Cumulative probability (%)
Green zone	0	0	8.11
	1	0	28.58
	2	0	54.32
	3	0	75.81
	4	0	89.22
yellow zone	5	0.40	95.88
	6	0.50	98.63
	7	0.65	99.60
	8	0.75	99.89
	9	0.85	99.97
red zone	≥10	1.00	99.99

Table 3: BIS multiplication factors

For a sample of 250 forecasts, the table above shows the plus factors added to the constant 3 to determine the multiplication factor, i.e. multiplication factor = 3 + plus factor. For other sample sizes the yellow zone start where the cumulative probability exceeds 95% and the red zone starts where it exceeds 99.99%.

Kerkhof-Melenberg (2004) mentions the following three shortcomings of this procedure. First, the model ignores the clustering phenomenon in the exceedances. This is also discussed by Berkowitz-O'Brien (2002) who have studied daily performances of VaR estimates of six major U.S. companies. The authors state in their study that "…, at times, losses can substantially exceed the VaR, and such events tend to be clustered. This suggests that the banks' models, besides a tendency towards conservatism, have difficulty forecasting changes in the volatility of P&L" (1094). Second, the procedure does not take into account the estimation risk that $-VaR_t = \hat{F}_t^{-1}(1-p)$ is not necessarily equal to $F_t^{-1}(1-p)$ considering the limited size of the data. Finally, there is too much relevant information lost regarding the distribution function by considering only whether the VaR is exceeded or not. Berkowitz (2001) discusses this issue further in detail.

2.3 Empirical VaR calculations using EVT

Although the use of EVT in finance is a recent development, research work that highlights its importance is growing. Longin (1996) was one of the first to apply EVT in finance. In his study he showed that the heavy tails of the daily returns for the S&P500 index over the period 1885 to 1990 could be characterized by the Fréchet distribution. He emphasized the potential of EVT in risk management applications such as VaR estimation, margin setting in futures markets and in regulating capital requirements for financial institutions. In the follow up paper, Longin (2001) introduced the idea that EVT could be used in portfolio management. EVT enables computing the probability of crashes and their associated waiting time period so that these risks could be hedged using derivatives. Longin-Solnik (2001) analyzed extreme returns on a multivariate scale for a group of international stock markets.

Using data on Asian equity markets, Campbell-Koedijk (1999) applied EVT to estimate conditional VaR and compared findings to those based on the RiskMetrics methodology. In this study, the authors concluded that the model provided improved forecasts of the Value-at-Risk and that conditional VaR estimates are better able to capture the nature of downside risk, which is particularly crucial in times of financial crises. Jansen et al (2000) illustrated how EVT can be used for portfolio selection by constructing the optimal portfolio consisting of US bonds and stocks with safety-first approach using VaR-limited downside risk. McNeil and Frey (2000) also proposed an interesting combination of GARCH modeling and EVT to estimate tail-related risk measures for heteroscedastic time series.

Also papers of European interest include Cotter (2001), who used EVT to calculate optimal margins for 12 European stock indices futures and Lux (2001), who used intra-daily data of the DAX index to investigate the asymptotic extreme behavior of returns in the German stock market. More recently, Këllezi-Gilli (2003) applied EVT on thirty-one years of daily returns on Swiss market index to estimate statistical models for several tail-related risk measures. Gettingby et. al. (2004) employed an EVT approach to investigate the distribution of extreme share returns in the UK from 1975 to 2000. They found that the Generalized Logistic (GL) distribution provided an adequate description of both the minima and maxima data. Following this study Tolikas-Brown (2005) used a similar methodology to investigate the asymptotic distribution of the lower tail for daily returns in the Athens Stock Exchange (ASE) over the period 1986 to 2001.

Also, Da Silva-De Melo Mendez (2003) used EVT to analyze ten Asian stock markets and their results suggested that compared to traditional methods estimating VaR via EVT approach is more conservative in determining capital requirements than traditional methods. Ganief-Biekpe (2003) applied POT technique to the South African Rand/Dollar one year futures contract.

3 Methodology

3.1 The Data

The data contain adjusted daily closing prices of 6 companies traded in the ISE, the index value ISE100 and an artificial price-weighted index. The price weighted index is calculated by taking the average of the adjusted prices of 20 companies that are among constituent stocks of the 2005 third quarter ISE National 30 index. The 20 stocks selected are of the most formerly firms that are publicly traded. Below is a list of these 20 stocks' tickers used in computation of the price weighted index.

AKBNK	EREGL	HURGZ	PETKM	TOASO
ARCLK	FINBN	ISCTR	PTOFS	TUPRS
DISBA	FROTO	KCHOL	SISE	VESTL
DOHOL	GARAN	MIGRS	THYAO	YKBNK

Table 4: The list of the stocks used in the computation of price weighted index.

Other than the ISE National 100 and price weighted index of the 20 companies, six individual stocks are selected for equity risk calculations. These six stocks can be grouped into two such that, the first group consists of companies that are well traded. The tickers of this first group are: “YKBNK”, “EREGL”, and “ISCTR”. The other group of poor traded companies is randomly selected among the stocks that are not the constituent stocks of ISE National 100 index. The tickers of this next group are: “BFREN”, “KARTN”, and “GOODY”.

Thus, a total of 8 data series; ISE National 100 index ("ISE100"), price weighted index of the 20 companies ("AVE20"), "ISCTR", "YKBNK", "EREGL", "BFREN", "KARTN", "GOODY" are used in computations.

The dataset is available from June 22nd, 1993 onwards. Only "BFREN" stock data begins at January 4th, 1994. Hence, 3066 daily returns are calculated for the period between June 22nd, 1993 and October 17th, 2005 regarding the five individual stocks; "ISCTR", "YKBNK", "EREGL", "KARTN", "GOODY", together with "AVE20" and "ISE100" indices. Only for "BFREN" 2928 daily returns are computed for the period between January 4th, 1994 and October 17th, 2005.

3.2 Descriptive Statistics

The stocks traded in ISE are highly volatile and within the period chosen in this study, a daily return as high as 17.8% and a daily loss as high as 20% has been observed. Within the 12 years period, the two highest daily returns on the ISE100 index occurred on consecutive days in December 2000. Only two months later, in February 2001 the first and third highest losses were observed. This shows that the well known clustering property of the financial return data is also present in ISE returns. The same property can also be observed in the individual stocks' daily returns.

Table 5 and Table 6 below list the first five highest and lowest daily returns of the dataset used in the study.

Highest Returns					
	1st	2nd	3rd	4th	5th
XU100	2000-12-05 17.8%	2000-12-06 17.1%	1998-09-18 15.6%	2000-01-04 14.1%	2001-04-27 12.7%
AV20	2000-12-06 17.9%	2000-12-05 16.9%	1998-09-18 15.8%	2000-01-04 15.6%	1997-01-27 14.4%
EREGL	2000-12-05 20.3%	1997-01-27 18.4%	1998-09-18 18.0%	1996-11-20 17.4%	2000-12-06 16.8%
ISCTR	1995-04-21 21.5%	1995-04-24 21.4%	2000-01-04 20.8%	2000-12-05 18.7%	1998-09-18 17.2%
YKBNK	1997-01-24 19.6%	2000-12-06 18.2%	1999-12-10 18.2%	2001-04-27 17.7%	1997-12-29 17.5%
BFREN	2004-05-12 193.1%	2000-12-05 22.9%	1995-02-15 21.2%	2004-10-05 20.0%	2004-09-21 20.0%
GOODY	1999-12-15 21.0%	2000-12-05 18.4%	2000-04-25 18.2%	2000-01-05 18.0%	2001-03-30 18.0%
KARTN	2000-01-17 20.8%	2004-10-26 19.6%	2003-09-30 17.6%	2004-11-02 15.7%	1994-06-01 15.1%

Table 5: Date and scale of highest daily returns

Highest Losses					
	1st	2nd	3rd	4th	5th
XU100	2001-02-21 -20.0%	1998-11-11 -16.1%	2001-02-19 -15.8%	1998-08-27 -14.1%	2003-03-03 -13.3%
AV20	2001-02-21 -18.9%	1998-11-11 -16.1%	2001-02-19 -15.8%	1997-10-27 -13.8%	2003-03-03 -12.5%
EREGL	2001-02-21 -22.3%	1998-01-12 -20.5%	2001-02-19 -18.6%	2000-12-07 -16.8%	1997-10-27 -15.2%
ISCTR	2001-02-21 -20.8%	1998-03-13 -17.4%	1998-11-11 -16.6%	2003-03-03 -16.4%	2000-01-06 -16.0%
YKBNK	2002-06-24 -24.1%	2002-06-26 -23.5%	2002-06-25 -21.7%	2001-02-21 -20.8%	2000-12-07 -18.2%
BFREN	1995-04-24 -20.8%	2004-05-14 -20.5%	2004-05-17 -20.4%	1998-11-11 -19.9%	1994-04-26 -19.6%
GOODY	1998-11-11 -17.2%	1994-02-10 -15.4%	2000-12-07 -14.7%	2001-02-19 -14.6%	1997-01-28 -14.1%
KARTN	2003-10-02 -16.3%	1997-01-28 -14.7%	1995-04-24 -14.1%	1997-10-28 -13.6%	2000-11-28 -13.6%

Table 6: Date and scale of lowest daily returns

In Table 7, it can be seen that all of the eight data series' daily logarithmic returns have positive means. RiskMetrics assumes that daily returns have a mean of zero. However, testing the null hypothesis that the mean of daily returns is zero is rejected at 5% significance level.

For ISE100 the skewness is slightly negative indicating that the returns are not exactly symmetric but in fact skewed to the left. For "YKBNK" and "AV20" the daily returns are symmetric and for the rest of the return series, the skewness is positive. Positive skewness implies that the returns are skewed to the right. The companies in the poor traded group, i.e. "BFREN", "KARTN", and "GOODY" have the highest positive skewness. "BFREN" in particular has the highest skewness.

The heavy tailed property of financial returns can also be observed in the data series used in the study. The excess kurtosis for all eight data series is greater than 0 (implying kurtosis is greater than three). This suggests that the tails of the return distribution is heavier than the normal distribution. Again as in the case of skewness, the poorly traded stocks have higher kurtosis values compared to well traded stocks.

	XU100	*AV20*	*EREGL*	*ISCTR*	*YKBNK*	*BFREN*	*GOODY*	*KARTN*
Mean	0.18%	0.20%	0.21%	0.25%	0.22%	0.30%	0.17%	0.21%
Stan. Error	0.05%	0.06%	0.08%	0.08%	0.09%	0.11%	0.06%	0.06%
Stan. Dev.	3.04%	3.33%	4.21%	4.35%	4.76%	5.92%	3.56%	3.54%
Kurtosis	3.51	2.49	2.05	2.23	2.05	383.92	3.49	3.32
Skewness	-0.08	0.01	0.11	0.34	0.00	12.02	0.39	0.47
Minimum	-19.98%	-18.93%	-22.31%	-20.76%	-24.12%	-20.76%	-17.19%	-16.25%
Maximum	17.77%	17.87%	20.25%	21.54%	19.63%	193.11%	21.03%	20.76%
Count	3066	3066	3066	3066	3066	2928	3066	3066
Conf. Lev (95%)	0.11%	0.12%	0.15%	0.15%	0.17%	0.21%	0.13%	0.13%

Table 7: Descriptive statistics of daily log returns

3.3 Normality Tests

The descriptive statistics above point out that the daily returns may not be normally distributed. Checking for normality, the Kolmogorov-Smirnov statistic with Lilliefor's significance correction, Shapiro–Wilk statistics, and Jarque-Bera statistic also known as the Bowman-Shelton statistic all strictly reject normality at very low p-values.

	Kolmogorov-Smirnov			Shapiro-Wilk			Jarque-Bera	
	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	Sig.
XU100	0.054	3066	7.1E-23	0.963	3066	3.0E-27	1577	0
AV20	0.049	3066	1.1E-18	0.973	3066	1.9E-23	795	2.3E-173
EREGL	0.088	3066	1.3E-64	0.973	3066	9.2E-24	545	4.7E-119
ISCTR	0.091	3066	8.1E-68	0.969	3066	3.4E-25	693	3.8E-151
YKBNK	0.084	3066	2.1E-58	0.972	3066	3.1E-24	536	4.5E-117
GOODY	0.124	3066	1.4E-129	0.947	3066	8.9E-32	1633	0
KARTN	0.142	3066	3.7E-171	0.939	3066	1.1E-33	1519	0
BFREN	0.142	2928	1.1E-163	0.658	2928	1.4E-60	18052414	0

Table 8: Normality tests of daily log returns

3.4 VaR Estimation Models Used in the Study

Knowing that the daily logarithmic returns of ISE are not normally distributed and heavy tailed, the VaR for the two indices as well as the six individual stocks described in Section 3.1, are estimated with different approaches: Historical simulation (HS), RiskMetrics approach (RM), nonparametric GPD approach, and parametric GPD approach. The more modern classes of EVT models known as Peaks over Threshold models are applied in estimating the distribution of the tail.

The models are applied on the data using a rolling window style analysis. A window width is specified for each model: HS, RM, nonparametric GPD or parametric GPD.

Then a range of data in the specified window width is drawn from the beginning of the whole data range which consists of daily log returns. This first window is used to forecast VaR for day ‘windowsize+1’. This window is then moved forward by one day, and the data in this new window is used to predict the distribution function for day ‘windowsize+2’. This process is repeated until the last observation of the dataset is reached. Thus having *n* observations in a dataset, a sequence of (*n* - windowsize) VaR forecasts are calculated.

The R-Project, an open-source statistical software is used for the following analysis.

3.4.1 Historical Simulation

Historical simulation is a nonparametric approach that estimates the distribution of the return series as the empirical distribution of the past returns. It is simply finding the quantile of a given past data series and considering this quantile as the VaR for the next day. This approach assumes that the distribution of returns for the next days is the same as the distribution for the past historical data.

For the study, the quantiles have been computed using a windowsize of past 250 daily returns, which corresponds approximately to the number of trading days in a year. Two different types of probability functions have been used in the quantile estimations:

Type I $p(k) = k/n$

Type II $p(k) = (k\text{-}1)/(n\text{-}1)$

Using these probability functions 95% and 99% VaR estimates for the next day are computed by taking the 1% and 5% quantiles of past 250 daily returns respectively. The 250 day window data is rolled over each day such that VaR levels of the next days can be estimated.

The main drawback of this nonparametric approach is that it is impossible to extrapolate for loss levels that have never occurred before. Furthermore, at very low quantile levels depending on the windowsize being used, the quantile estimate will have very high variance; hence the estimates will tend to be inefficient. On the other hand, as financial returns have alternating periods of high and low volatility, widening the windowsize makes it impossible to reflect the effect of changing volatility on the estimates.

3.4.2 RiskMetrics™ approach

RiskMetrics is a methodology developed by J.P. Morgan, which estimates market risk measured by value-at-risk. J.P. Morgan (1996) published the techniques they apply in their methodology in RiskMetrics technical document in very detail. The backbone of the methodology is that the volatility clustering phenomenon mentioned in Section 1 is taken in hand via choosing an autoregressive moving average process to model the price process. The particular autoregressive moving average that is incorporated in RiskMetrics avoids the problem of uniformly weighted moving averages. The model gives more weight to the more recent returns observed using the so called exponentially weighted moving average (EWMA) method.

The volatility estimate that the model chooses to use is:

$$\sigma^2_{t+1|t} = \frac{\sum_{\tau=0}^{\infty} \lambda^{\tau} r^2_{t-\tau}}{\sum_{\tau=0}^{\infty} \lambda^{\tau}} = (1-\lambda)\sum_{\tau=0}^{\infty} \lambda^{\tau} r^2_{t-\tau} .$$

In the equation above the λ parameter used by RiskMetrics is 0.94. This λ parameter is determined by an optimization procedure that produces the best backtesting results. In the methodology described above, the number of past returns that effectively affect the volatility estimate is 75 days. As for the past 75th day return, the exponential weight is as low as 0.94^{75}, which is less than 0.01. Hence, the model represents a finite memory of the market.

The notation $\sigma^2_{t+1|t}$ emphasizes that the volatility estimated on a given day t is actually used as a predictor for the volatility of the next day t+1. For equity risk calculations, the model estimates the return series to be normally distributed conditional on the information set at time t, which consists of the past return series available at time t. Thus, the probability distribution function for the next day is estimated as a normal distribution with a mean of zero. The standard deviation for that day's normal distribution is calculated by the EWMA model described above.

The daily VaR at confidence level p (e.g. 95% or 99%) is then calculated by multiplying $\sigma_{t+1|t}$ with the p quantile of the standard normal distribution.

Additionally, RiskMetrics uses the technique described above to measure VaR of individual assets as well as portfolios of assets. In the case of portfolios containing no options, i.e. linear portfolios, the volatility is estimated by multiplying portfolio weights by the covariance matrix of EWMA of asset returns. The portfolio volatility is thus,

$$\sigma_p^2 = \sum_{i,j} w_i w_j \sigma_{i,j}$$

where w_i is the weight of asset i in the portfolio. Equivalently, however, one can calculate the return of the portfolio first and then apply the EWMA technique directly to the calculated portfolio returns.

The very strong assumption of conditional normality of RiskMetrics has received many criticisms. The well-known leptokurtic property of financial returns, imply that VaR estimates of the RiskMetrics methodology will tend to underestimate the true VaR.

However, as the model is easy to apply and it captures the changing volatility conditions via EWMA model, the model is very widely used and popular.

3.4.3 Nonparametric Peaks over Threshold approach

The Hill estimator and a variation of Hill estimator used by Dekkers et al (1989) both described in section 2.1.3.2 is applied to the data. A windowsize of 250 days is used in determining the marginal distribution. For example, for determining the distribution function for day 251, the returns between day 1 and day 250 are considered. At 90% and 94% threshold levels, the number of order statistics used, k, is 25 and 15 respectively. Next, the tail index parameter γ is estimated by either one of the two famous estimators, Hill or Dekkers.

3.4.3.1 Hill Estimator

If Hill estimator is used, then γ is estimated by Equation 13. As mentioned in 2.1.3.2.1, we expect the Hill plot of k against the estimated tail index parameter $\hat{\gamma}^{(H)}$ to be volatile. This property of Hill estimates on one of the samples we have used can be seen in Figure 7. The figure shows how $\hat{\gamma}^{(H)}$ changes as number of exceedances used in estimate increases.

The figure below is a sample Hill plot of "ISE100", which uses the data window of 250 daily returns between "28th September 2000" and "2nd October 2001", to predict the distribution function for the day "3rd October 2001". This distribution is selected randomly among all the distributions that are estimated for the forecast window.

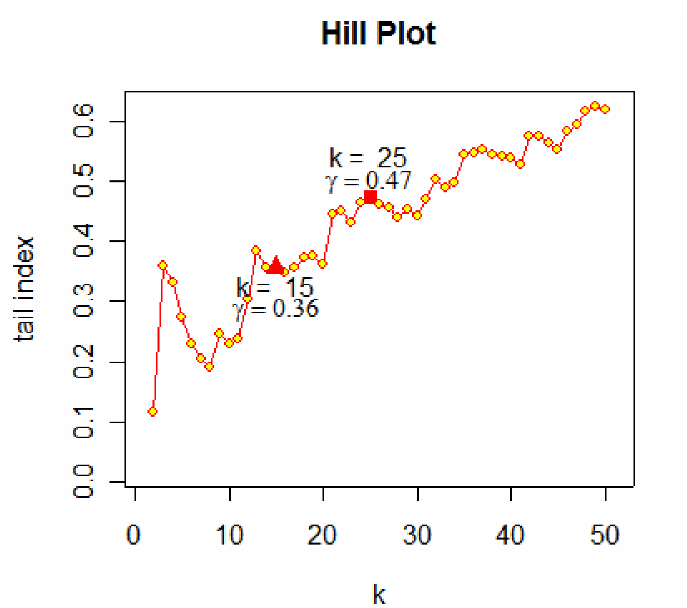

Figure 7: A sample Hill plot

However, since a rolling data window approach is applied, selecting the best choice of k regarding $\hat{\gamma}^{(H)}$ being estimated for each day in the forecast period (3066 – 250 = 2816 days) is not feasible. Hence, 90% and 94% threshold percentiles for the tail are set as a rule of thumb for selection of k.

X_{k+1} of the ordered statistics is taken as the threshold. Then, for the two different cases of tail estimation at 10% and 6%, the thresholds are 26th and 16th largest losses observed in the data window, respectively.

Having $\hat{\gamma}^{(H)}$ as the Hill tail index estimate and X_{k+1} as the threshold, the estimated distribution function of the tail becomes:

$$\hat{G}(x) = \frac{k}{n}\left(\frac{x}{X_{k+1}}\right)^{-1/\hat{\gamma}^{(H)}} .$$

Then the quantile x_p defined by $G(x_p)=p$ is estimated by

$$\hat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-\hat{\gamma}^{(H)}} X_{k+1} .$$

Consequently, the estimated p-level VaR is computed by

$$VaR_p = \left(\frac{n}{k}(1-p)\right)^{-\hat{\gamma}^{(H)}} X_{k+1}$$

where n is the windowsize of 250 days and k is either 25 or 15 depending on the 90% or 94% threshold level selected.

3.4.3.2 Dekkers Estimator

If the estimator recommended by Dekkers et al. (1989) is used, then γ is estimated by Equation 14.

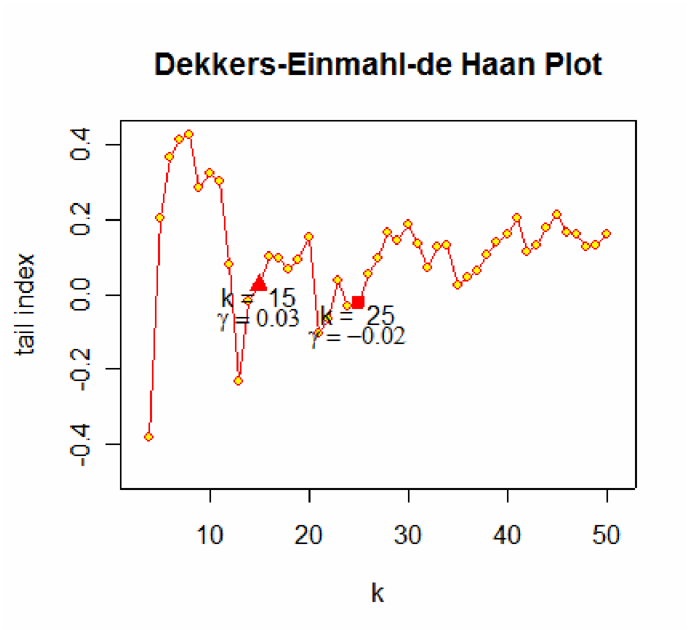

Figure 8: A sample Dekkers plot

The figure above is a plot of $\hat{\gamma}^{(D)}$ tail index estimates against k number of exceedances chosen. The data window used in the plot is the same as that of Figure 7's: daily returns of "ISE100" between "28th September 2000" and "2nd October 2001".

As the rolling data window argument is still valid in this approach, again the same rule of thumb used in Hill estimate is set: k is chosen such that there are 10% and 6% observations in the tail.

Having $\hat{\gamma}^{(D)}$ as the Dekkers tail index estimate and X_i as the ordered statistics then the quantile x_p can be estimated by the following equation (Embrechts et al. 1997: 350)

$$\hat{x}_p = \left(\frac{\left(\frac{k}{n(1-p)} \right)^{\gamma} - 1}{1 - 2^{-\hat{\gamma}^{(D)}}} \right) (X_k - X_{2k}) + X_k .$$

Consequently, the estimated p-level VaR is also computed by this equation where n is the windowsize of 250 days and k depends on the 90% or 94% threshold level selected.

3.4.4 Parametric Peaks over Threshold approach

As mentioned in Section 2.1.3.1.1, the MLE is known to have the two important properties of consistency and asymptotic efficiency when $\gamma > -0.5$. And since the data series this study is applied to is heavy tailed and hence we expect γ to be positive, usage of MLE in estimating parameters is found appropriate.

Comparisons of PWM and MLE discussed in Section 2.1.3.1.2 also guide us to using MLE approach since the database consists of heavy tailed high frequency daily return data.

A windowsize of 250 days have been used in Historical Simulation and nonparametric GPD methods. The selection of this 250 days windowsize was deliberate to be consistent with BIS regulations, encouraging the use of recent data covering the past year. In the RiskMetrics approach, the most recent 75 days data is used effectively in the calculations due to the EWMA method the model applies.

However the optimization procedure used in maximum-likelihood estimation in POT approach can only be applied with a certain amount of data. With too few data optimization can not be achieved. Hence larger windowsizes of 750, 1000, and 1250 days are taken in POT applications. Taking larger windowsizes also enable the database to cover large shocks observed in past three to five years (approximately 750 days to 1250 days).

Selecting the threshold in POT models is as crucial as selecting k in upper order statistics. Two different approaches in determining the threshold has been studied. The first approach is the classical approach of calculating the quantile of the window data being used.

The idea in second approach is close to the methodology that McNeil-Frey (1999) used. McNeil-Frey (1999) combined pseudo-maximum-likelihood fitting of GARCH models to estimate the current volatility and EVT for estimating the tail of the innovation distribution of the GARCH model. Rather than using a GARCH approach however, in this study only the quantile of the normal distribution fit to the data is calculated to be taken as the threshold.

Thresholds are taken either at 90% or 94% quantiles. In the case that quantile approach is applied in determining the threshold, then 10% or 6% quantile of returns in the window data are used.

Then again, if normal distribution approach is applied in determining the threshold then first the mean $\hat{\mu}$ and standard deviation $\hat{c}$ of the window data are calculated. Using 90% and 94% quantiles of the standardized normal distribution (Z_q=1.282552 and Z_q=1.554774 respectively) the threshold u is determined by $u = \hat{\mu} + \hat{c}\ Z_q$.

After the threshold u is assigned, the losses greater than the threshold value are extracted. Suppose that the values $y_1,...,y_{n_u}$ are the n_u excesses of a threshold u. Next, the log-likelihood functions in Equation 11 and Equation 12 are maximized using Nelder-Mead (1965) method. Initial values for the parameters σ and γ to be optimized over are taken 1 and 0.3 respectively.

Once the parameters σ and γ are estimated, having the threshold u and the distribution function of the tail, VaR can also be estimated by,

$$VaR_p = \hat{x}_p = u + \frac{\sigma}{\gamma}\left(\left(\frac{n}{n_u}(1-p)\right)^{-\hat{\gamma}} - 1\right).$$

In summary, using the steps mentioned above, the study is repeated using windowsizes of 750, 1000 and 1250 days. Thresholds are either the empirical quantiles of the window data or quantiles of normal distribution fit to the window data.

As stated in 2.1.3, threshold selection is crucial in correct estimation of the tail index. One guideline is choosing the threshold such that the mean excess function is linear after the selected u. However, this guideline is impossible to be applied for each estimated distribution as the mean excess plot has to be visually inspected and the threshold subjectively determined thereafter. Hence, consistent with the rule of thumb used in nonparametric GPD estimations, 90% and 94% thresholds levels are again used in parametric GPD model estimations.

To give an example of the mean residual life plot, a forecast day for one of the data series needs to be selected. To be consistent, the tail fit of distribution functions concerning the same forecast day ("3rd October 2001") and the same data series ("ISE100" index) that is used in Hill and Dekkers plots (in Figure 7 and Figure 8) is depicted. The sample mean residual plots in Figure 9 use 1000 days windowsize.

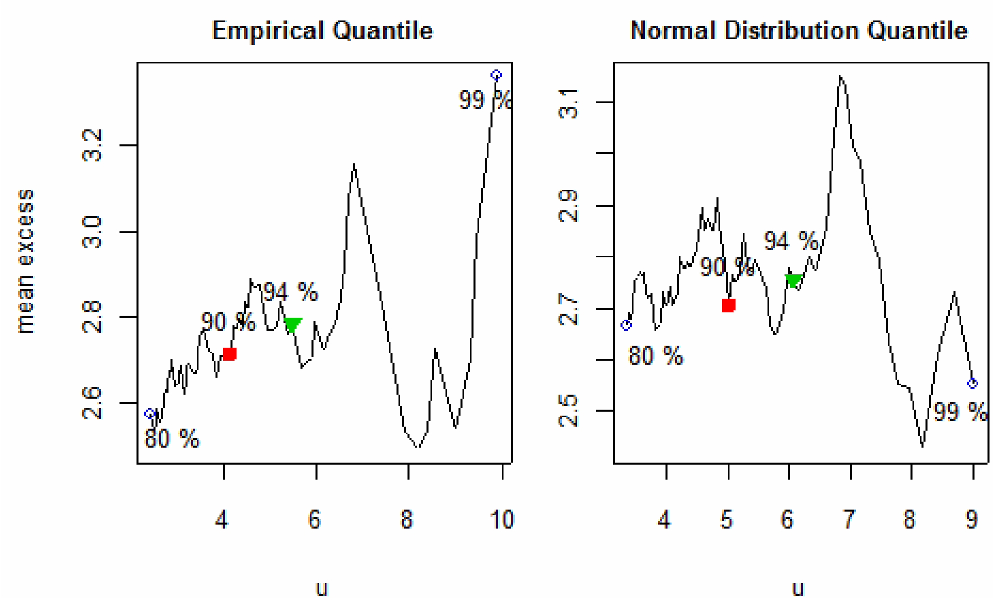

Figure 9: Sample mean residual life plots

The unstable behavior especially towards the higher values of *u* is typical and makes the precise interpretation of the plot difficult (Embrechts et.al. 1997: 298). For this particular day the mean residual life plot is implemented for "ISE100" data series, taking 90% or 94% threshold levels doesn't give the impression that the thresholds are problematic. For other dates and other series the residual life plots are similar to these plots in the sense that for higher values of threshold levels at 95% or above, the mean excess function becomes very volatile.

Once a GPD model is estimated, the estimated parameters $\hat{\gamma}$ and $\hat{c}$ can be used in Equation 10, to compute the expected exceedance $e(u)$ given that a loss level higher than u is observed. Thus Conditional-VaR or ES measure would be computed.

3.5 Model Forecasts

HS estimates the VaR as the empirical quantile of the past returns. The other approaches, on the other hand aims to estimate a distribution function for the next day's return and the 95% or 99% quantiles of the estimated distribution function is used to determine the corresponding VaR.

Extreme Value Theory claims that the models proposed are advantageous in estimating the tail of the distribution function as the focus is on the extreme observations and central values are ignored. It is worthwhile to depict the fit of the estimated distributions to empirical data used. However, as the study estimates thousands of models considering the number of days in the forecast period, it is impractical to show the fit for each estimated model.

To show the fit of estimated distributions again the same forecast day ("3rd October 2001") and the same data series ("ISE100" index) is used. Six different distribution functions and their fit to the lower tail of the empirical data are plotted. The vertical lines show 99% VaR estimations.

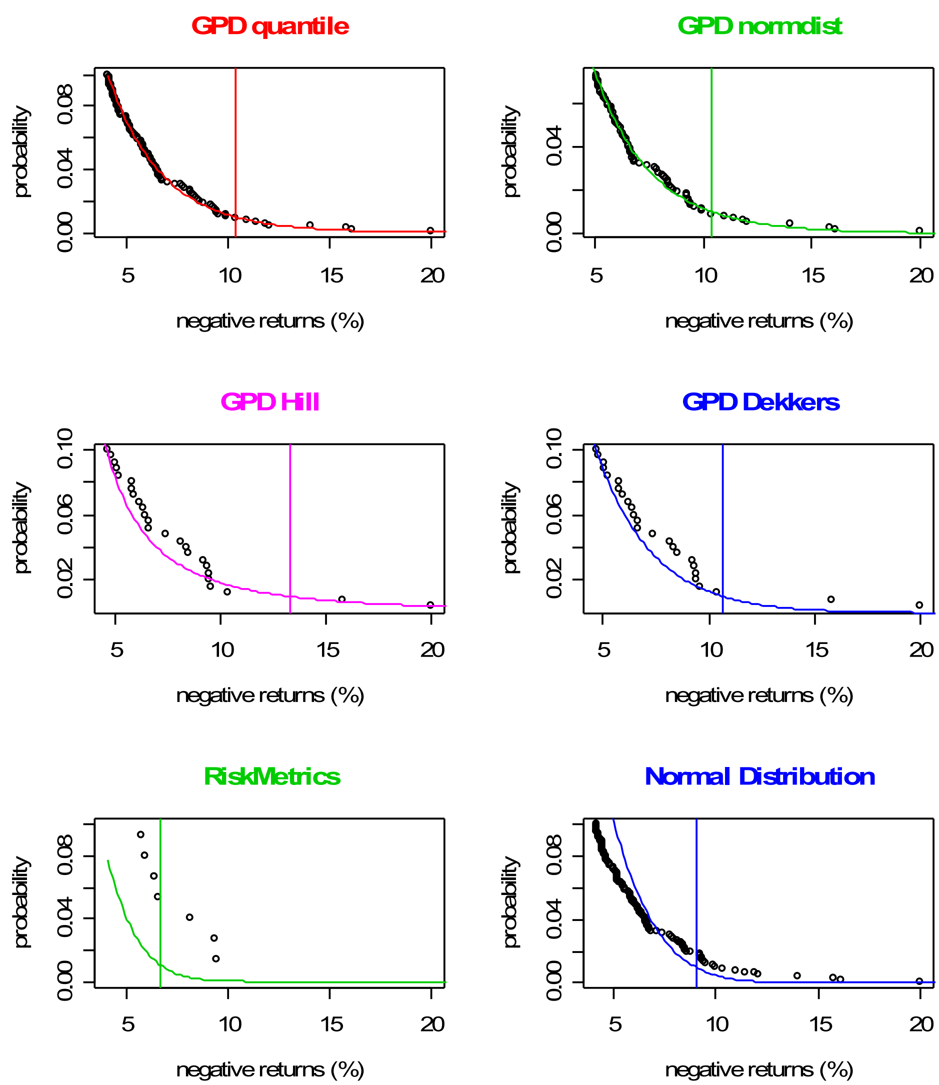

Figure 10: Tail fit of distribution functions estimated

In Figure 10, the first two upper plots show the fit of estimated generalized Pareto density functions, using normal distribution and quantile methods for 90% threshold selection. The windowsize used for the estimation is 1000 days and lowest 100 daily returns of the past 1000 days are plotted as the empirical data in the lower tail.

The two plots in the middle show GPD distributions using Hill and Dekkers estimates. As the windowsize used in the estimations is 250 days, 25 highest daily losses are used to plot the empirical data.

The left lower plot show the tail of the distribution function estimated by RM approach. The only parameter estimated in this approach is the volatility which is estimated by EWMA method. As only past 75 daily returns are needed in the estimation process, only the lowest 7 returns are plotted in the tail.

The right lower plot is simply the tail of the normal distribution function with a mean and standard deviation equal to those of past 1000 daily returns.

Visually one can clearly inspect that the best fit is achieved by the GPD models. As expected, the heavy tails cannot be estimated effectively using normal distributions. Since the probabilities of extreme observations are estimated with lower probabilities, VaR at high quantiles are seriously underestimated.

3.6 Backtesting

Parallel to the current BIS implementation, the computed VaR forecasts for day t using the information up to time $t-1$ are backtested against the daily returns of day t. The exceedance variable I_t is computed for each model using Equation 15.

The performance tests for the models described in the following section use these backtesting results. If a model is adequate the VaR is expected to have an exceedance rate of (1-VaR level). For instance, for a forecast horizon of 1000 days and 99% VaR level, the VaR forecasts are expected to be exceeded in 10 incidences. Furthermore, an adequate VaR model is expected to handle the clustering volatility phenomenon. Then an acceptable model would adapt to the new conditions of changing volatility quickly enough that successive exceedances are not dependent.

3.7 Selection Criteria for Models

In order to choose among the VaR estimation models described in Section 3.4., five criteria have been evaluated. The two of these criteria have been put forth by Christoffersen (1998). These are unconditional and conditional likelihood ratios. The null hypothesis being tested in unconditional likelihood ratio test is similar to the null hypothesis that of binomial proportion test. Also, conditional likelihood ratio test is asymptotically equivalent to Pearson’s chi-squared independence test. Because the application and motivation is much simpler for the Binomial proportion test and Pearson’s chi-squared independence test, these two tests have also been applied on the models. So, in the study two groups of tests will be applied to test if the models satisfy the two unconditional and conditional coverage criteria. The first group consists of the unconditional and conditional likelihood ratio tests. And in the next group of tests, the binomial proportion and a test of independence using contingency tables appear.

For models that satisfy both criteria in either group tests, a loss function has been computed to compare successful models pairwise.

3.7.1 Likelihood Ratio Statistics

Unconditional likelihood ratio test is used to determine whether the frequency of violations are in the range of their expected values for given VaR levels, i.e. whether the coverage is correct. Christoffersen (1998) states that testing this unconditional coverage of the forecasts for the interval being analyzed is not sufficient and that conditional coverage should also be tested. Conditional coverage likelihood ratio tests the null hypothesis that successive observations are statistically independent.

3.7.1.1 Unconditional Coverage

Likelihood ratio statistic for the test of "unconditional coverage"

$$LR_{uc} = -2\log\left[\frac{p^{n_1}(1-p)^{n_0}}{\hat{\pi}^{n_1}(1-\hat{\pi})^{n_0}}\right] \sim \chi^2_{(1)}$$

where

p = VaR level

n_1 = Number of violations (number of 1's in backtesting)

n_0 = Number of compliances (number of 0's in backtesting)

$\hat{\pi} = \dfrac{n_1}{n_0 + n_1}$, the MLE of p.

3.7.1.2 Conditional Coverage

Likelihood ratio statistic for the test of "correct conditional coverage"

$$LR_{cc} = -2\log\left[\frac{p^{n_1}(1-p)^{n_0}}{\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{01})^{n_{00}}\,\hat{\pi}_{11}^{n_{11}}(1-\hat{\pi}_{11})^{n_{10}}}\right] \sim \chi^2_{(2)}$$

where

n_{ij} = Number of i values followed by a j value in the I_t series ($i, j = 0, 1$)

$\pi_{ij} = \Pr\{I_t = i \mid I_{t-1} = j\}$ $(i, j = 0, 1)$

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}$$

$$\hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

3.7.2 Testing for Proportions and Association

In order to test unconditional coverage, binomial proportion test is applied in this group of selection criteria. To test independence (conditional coverage), contingency table framework is applied.

3.7.2.1 Binomial Proportion Test

The null hypothesis that proportion of compliances i.e. the occurrence of actual returns not exceeding the VaR forecast, is equal to the VaR level is tested.

The test statistic = $\dfrac{(n/T) - (1 - VaRlevel)}{\sqrt{VaRlevel(1 - VaRlevel)/T}}$

where

n = number of violations, i.e. the actual day return exceeding VaR forecasted for that day.

T = number of days that VaR is forecasted = length of data series – windowsize

3.7.2.2 Contingency Table Framework

In this case, the hypothesis that no association exists between the successive observations or that these successive observations are independent is tested.

To test the null hypothesis of independence between two successive day forecasts, first the contingency table and the expected values under the independence assumption are calculated. Contingency table used in the study is a matrix with two rows and two columns.

Contingency Table= $\begin{vmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{vmatrix}$

where

n_{00} = the number of occurrences when the actual returns of both successive days do not exceed the VaR level forecasted (compliance on both consecutive days)

n_{01} = the number of occurrences when the actual returns of the first day do not exceed the VaR forecasted but the actual return of the next day exceed the forecasted VaR (compliance followed by violation)

n_{10} = the number of occurrences when the actual returns of the first day exceed the VaR forecasted but the actual return of the next day do not exceed the forecasted VaR (violation followed by compliance)

n_{11} = the number of occurrences when the actual returns of both successive days exceed the VaR level forecasted (violation on both consecutive days)

The expected values are calculated using,

$$E_{ij} = \frac{R_i C_j}{n}$$

where

R_i =row totals

C_j =column totals

$n = n_{00} + n_{01} + n_{10} + n_{11}$

A test of independence of successive day forecasts at a significance level of 5% is based on the decision rule: Reject H_0 if $\sum_{i=0,1}\sum_{j=0,1}\frac{\left(n_{ij}-E_{ij}\right)^2}{E_{ij}} > \chi^2_{1,0.05}$.

3.7.3 Results of Performance Tests

The results of performance tests of these HS, RM, nonparametric GPD and parametric GPD models can all be followed in the Appendix in Tables 15 through 21. In these tables "ws" column stands for windowsize. The p-values of performance tests of the second group of selection criteria; Binomial test and Chi-Square Contingency Table test are given in the columns "BT" and "CT", respectively. Similarly, the p-values of performance tests of the first group of selection criteria; Unconditional Likelihood Ratio test and Conditional Likelihood Ratio test are given in the columns "ULR" and "CLR", respectively.

3.7.3.1 Performance Tests of HS Models

The backtesting results of VaR forecasts using HS is given in Table 15. A model passes the performance test if the null hypothesis of that test cannot be rejected, i.e. if p-value is larger than the chosen significance level of 0.05. In the table it can be observed that Binomial and Unconditional Likelihood Ratio tests, testing the same null hypothesis that the exceedances of the forecasted VaR are in the range of the expected values for given VaR levels, lead to similar conclusions. The columns BT and ULR show that the null hypothesis cannot be rejected 30 out of 32 cases, i.e. HS model is successful in general in this performance criterion. For the 32 HS models computed, accept or reject decisions regarding the null hypothesis never contradicted.

Comparing the unconditional coverage performances of type I and type II percentile methods, whereas type I is better in 99% VaR forecasting, type II is better in 95% VaR forecasting.

The second performance criterion of independence of successive observations is tested under similar null hypothesis by χ^2 Contingency Table and Conditional Likelihood Ratio tests. The regarding tests' p-values given in columns CT and CLR show that among 36 HS models computed only 6 of them pass the second criterion. CT and CLR agree on the successful performance of 4 models.

In sum, models using HS approach are observed to satisfy the first success criterion and fail to satisfy the second criterion in general. And, among the 32 HS models computed in this study there are only 8 models that satisfy either set of performance tests: the first set of performance tests being BT and CT and the second set of performance tests being ULR and CLR.

Equity	**VaR**	**Quantile Method**	**Test Set 1**		**Test Set 2**	
			Criterion I	**Criterion II**	**Criterion I**	**Criterion II**
ISCTR	95	Type I	Pass	Pass	Pass	Fail
BFREN	99	Type I	Pass	Fail	Pass	Pass
GOODY	99	Type I	Pass	Pass	Pass	Pass
ISCTR	99	Type I	Pass	Pass	Pass	Pass
ISE100	99	Type I	Pass	Fail	Pass	Pass
KARTN	99	Type I	Pass	Pass	Pass	Pass
ISCTR	95	Type II	Pass	Pass	Pass	Fail
ISCTR	99	Type II	Pass	Pass	Pass	Pass

Table 9: Successful HS models

The table above shows the performances of the 8 successful HS models on all four tests. It can be observed that 4 of these models satisfy both test sets. Two of them satisfy only the first test set but not the second test set. And again two of the models satisfy the second test set but not the first test.

3.7.3.2 Performance Tests of RM Models

Table 16 lists the p-values of selection tests on VaR forecasts of RM approach. In the table, results of all four tests both for 95% and 99% VaR forecasts are given. The Binomial test and Unconditional Likelihood Ratio test results again direct us to exactly the same conclusions. Both tests reject the same 8 models out of 16 models computed. Among the 8 successful models satisfying unconditional coverage requirement, 5 are for 99% VaR and the remaining 3 for 95% VaR. Regarding the second success criterion, there are 6 models that pass χ^2 Contingency Table test and 4 models that pass the Conditional Coverage test.

Comparing RM models with HS, HS models are observed to be much better in terms of unconditional coverage whereas RM models satisfy the conditional coverage requirement better than HS.

Equity	**VaR**	**Test Set 1**		**Test Set 2**	
		Criterion I	**Criterion II**	**Criterion I**	**Criterion II**
BFREN	99	Pass	Pass	Pass	Pass
ISCTR	99	Pass	Pass	Pass	Pass
KARTN	99	Pass	Pass	Pass	Pass
YKBNK	99	Pass	Fail	Pass	Pass

Table 10: Successful RM models

In Table 10, among 16 models -computed for 95% and 99% VaR levels of 8 data series- those that satisfy either set of performance criteria are listed. Thus, a total of 4 models are considered to be successful.

3.7.3.3 Performance Tests of Nonparametric GPD Models

In Tables 17 and 18, p-values of the tests on VaR forecasts of the GPD models using Hill (1975) and Dekkers et. al. (1989) estimates are given respectively. In these tables it can again be observed that BT and ULR selects the same models successful. Among 32 models computed with Hill estimate 27 of them are successful in unconditional coverage. Similarly, among the other 32 models that are computed with Dekkers estimate 24 of them are successful.

The second success criterion of conditional coverage again picks a fewer number of models as successful. Whereas the contingency table test selects 6 models as successful among the models estimated by the Hill estimate, it selects 8 models among the models estimated by Dekkers. The other test of conditional coverage, i.e. CLR test categorizes 7 and 6 models as successful models amongst those models that use Hill and Dekkers estimates respectively.

Even though Dekkers estimate is successful in satisfying the unconditional coverage requirements in all models at 95% VaR level, the performance is very poor at 99% level. All GPD models using Hill or Dekkers estimates are poor in predicting 95% VaR and are better in 99% VaR level in terms of conditional coverage requirement.

	VaR	Equity	Th	Test Set 1		Test Set 2	
				BT	**CT**	**ULR**	**LR**
Hill	95	ISCTR	90	Pass	Pass	Pass	Fail
		ISCTR	94	Pass	Pass	Pass	Fail
	99	EREGL	94	Pass	Fail	Pass	Pass
		GOODY	90	Pass	Pass	Pass	Pass
		ISCTR	90	Pass	Pass	Pass	Pass
		ISCTR	94	Pass	Pass	Pass	Pass
		ISE100	90	Pass	Fail	Pass	Pass
		KARTN	90	Pass	Pass	Pass	Pass
		KARTN	94	Pass	Fail	Pass	Pass
Dekkers	95	ISCTR	90	Pass	Pass	Pass	Pass
		ISCTR	94	Pass	Pass	Pass	Fail
		KARTN	94	Pass	Fail	Pass	Pass
	99	EREGL	90	Pass	Pass	Pass	Pass
		GOODY	94	Pass	Pass	Pass	Pass
		ISCTR	90	Pass	Pass	Pass	Pass
		KARTN	90	Pass	Pass	Pass	Pass

Table 11: Successful models using nonparametric estimators

Among 64 models computed, 16 models pass either set of performance tests. Five of these successful models forecast 95% VaR whereas 11 of them forecast 99% VaR. Among the 5 successful models forecasting 95% VaR, 3 of them use 94% threshold level. However, at 99% VaR level, 7 out of 11 successful models use 90% threshold level. Nine of these 16 successful models use Hill estimate and the remaining 7 use Dekkers estimate.

However these results do not give a clear guidance of which estimate –either Hill or Dekkers- to use or which threshold level to select in forecasting 95% and 99% VaR.

3.7.3.4 Performance Tests of parametric GPD Models

The performance tests of parametric GPD models are given in tables 19 through 22. Table 19 and Table 20 lists the results of the tests for parametric GPD models that use normal distribution in determining the threshold. Table 19 lists results for the 90% threshold level and Table 20 lists results for 94% level. Table 21 and Table 22, on the other hand, list the p-values of performance tests of parametric GPD models that determine the threshold using the quantile approach. The tables list the results of models with thresholds calculated at 90% and 94% quantiles respectively.

In Table 21 and Table 22, there are 48 (8 data series x 2 VaR levels x 3 windowsizes) model test results. However, in the tables showing GPD test results of the models that use normal distribution percentiles as thresholds, there are fewer models. Among the data series studied, “BFREN” has highest volatility and kurtosis (Please see Table 7). As the volatility is too high for “BFREN”, the threshold set as the 90% and 94% quantiles of normal distribution, leave too few exceedances to work with. Hence, with few data in hand the optimization process can not be achieved and maximum likelihood estimate cannot be attained. A similar situation comes up again in the case of “AVE20” index when 750 days windowsize and 94% threshold level is used for normal distribution percentile. In this particular case, again no maximum likelihood estimate could be attained. Thus, Table 19 and Table 20 list the results of 42 and 40 models that could be computed respectively.

Yet again, whilst first criterion of unconditional coverage is easily satisfied by 139 out of 178 models, the second criterion is satisfied by 52 models by CT and 57 models by CLR test.

threshold model	th %	equity	Ws	Test Set 1 BT	Test Set 1 CT	Test Set 2 ULR	Test Set 2 LR
Normal Distribution	94	EREGL	1000	Pass	Fail	Pass	Pass
		ISCTR	750	Pass	Pass	Pass	Pass
			1000	Pass	Pass	Pass	Pass
			1250	Pass	Pass	Pass	Pass

Table 12: Successful parametric GPD models forecasting 95% VaR

Among 89 models that forecast 95% VaR, only 4 of them pass either set of performance tests. These successful models that forecast 95% VaR all use 94 percentile of the normal distribution as the threshold return level. VaR of "ISCTR" can be forecasted by any of the 750, 1000 or 1250 windowsize levels. 95% VaR for "EREGL" can be successfully forecasted if the windowsize is selected as 1000 days.

	Normdist		**quantile**		**# of models**
	90	**94**	**90**	**94**	
Ws=750	AVE20 EREGL ISCTR ISE100 (7)	ISCTR (6)	AVE20 BFREN EREGL ISCTR ISE100 (8)	AVE20 BFREN EREGL ISCTR ISE100 (8)	**15** (29)
Ws=1000	AVE20 EREGL GOODY ISCTR ISE100 KARTN (7)	AVE20 EREGL GOODY ISCTR ISE100 KARTN (7)	AVE20 ISCTR (8)	AVE20 BFREN ISCTR ISE100 KARTN YKBNK (8)	**20** (30)
Ws=1250	EREGL GOODY ISCTR ISE100 KARTN (7)	AVE20 EREGL GOODY ISCTR ISE100 KARTN (7)	AVE20 EREGL ISCTR ISE100 (8)	EREGL GOODY ISCTR ISE100 KARTN (8)	**20** (30)
# of models	**15** (21)	**13** (20)	**11** (24)	**16** (24)	**55** (89)

Table 13: Successful parametric GPD models forecasting 99% VaR

At 99% VaR level there are 55 successful models out of 89 computed models. Table 13 shows the distribution of these successful models. The numbers in parentheses show the total number of models computed for each section.

In those sections where high threshold levels at 1000 and 1250 days windowsizes are used, the models performed well almost for every data series computed. All data series but “YKBNK”, that is seven out of eight data series could be forecasted successfully by the models that use 1000 day windowsize and normal distribution in determining threshold both at 90% and 94% quantile and also those that use 1250 days windowsize and 94% quantile of normal distribution for threshold selection.

3.7.4 Loss Function

For these successful models that pass either group of tests on unconditional and conditional coverage, a loss function measuring their performance is calculated. This loss function was proposed first in the study of Angeledis-Benos (2004). The authors named the loss function as Quantile Loss (QL) function. The QL function has the following form:

$$\psi_{t+1}^{QL} = \begin{cases} \left(y_{t+1} - VaR_{t+1|t}\right)^2, & if\ \ y_{t+1} < VaR_{t+1|t} \\ \left(Quantile\{y, 100p\}_1^T - VaR_{t+1|t}\right)^2, & if\ \ y_{t+1} \geq VaR_{t+1|t} \end{cases}$$

This loss function penalizes a model such that if a violation occurs i.e. $y_{t+1} \geq VaR_{t+1|t}$ then the penalty is the square of the loss that exceeds VaR forecasted. In the other cases the penalty is the squared distance between the *p*-quantile of the realized future returns and the estimated VaR i.e. $\left(Quantile\{y,100p\}_1^T - VaR_{t+1|t}\right)^2$.

In order to overcome the problem that one can never know the "true" VaR whilst working with real financial data Angelidis-Benos (2004) proxied the "true" VaR by the empirical distribution of the realized returns. The authors claim that the proxy they use at least meets the unconditional coverage requirement as the total number of violations will be equal to the expected one. Thus having set a proxy for the "true" VaR the authors penalized the models if the estimated forecasts of the model diverged from the "true" VaR.

3.8 Successful Models and their Loss Functions.

3.8.1 Picking the Best Model

Having calculated the daily loss values for VaR estimations of 89 successful models, next these models are being compared pairwise. Two selected models are compared using the following methodology described in Sarma et al.'s 2004 paper:

> Testing for the superiority of a model *vis-a-vis* another in terms of the loss function
>
> Consider two VaR models model *i* and model *j*. The superiority of model *i* over model *j* with respect to a certain loss function can be tested by performing a one-sided sign test. The null hypothesis is:
>
> H_0: $\{\theta=0\}$
>
> against the one-sided alternative hypothesis:
>
> H_1: $\{\theta<0\}$
>
> where θ is the median of the distribution of z_t defined as $z_t = l_{it} - l_{jt}$ where l_{it} and l_{jt} are the values of a particular loss function generated by model *i* and model *j* respectively for the day *t*. Here z_t is known as

the loss differential between model i and model j at time t. Negative values of z_t indicate a superiority of model i over j.

Testing procedure

Define an indicator variable

$$\psi_t = \begin{cases} 1 & if \quad z_t \geq 0 \\ 0 & if \quad z_t < 0 \end{cases}$$

The sign statistic S is the number of non-negative z's:

$$S_{ij} = \sum_{t=1}^{T} \psi_t$$

If z_t is i.i.d. then the exact distribution of S_{ij} is binomial with parameters (T, 0.50) under the null hypothesis. For large samples the standardized version of the sign statistic S_{ij} is asymptotically standard normal:

$$S_{ij}^a = \frac{S_{ij} - 0.5T}{\sqrt{0.25T}} \sim N(0,1) \text{ asymptotically}$$

H_0 is rejected at the 5% level of significance if $S_{ij}^a < -1.66$. Rejection of H_0 would imply that model i is significantly better than model j in terms of the particular loss function under consideration; otherwise model i is not significantly better than model j. (344)

Having the methodology to compare the performances of models regarding the QL function, for each data series combinations of two models among successful models are selected. These two models' performances are compared against each other. Next, models are sorted in terms of their performances in these comparison tests. The best models that outperform all the remaining models are described in the following section.

3.8.2 Best Estimated Models

Table 14 lists the results of comparison tests of 87 successful model forecasts based on their performances regarding the QL function. The study shows that the best models are always GPD models.

On the 8 data series this empirical study is applied only for 3 of them 95% VaR forecast methods is acceptable given the unconditional and conditional coverage criteria set by the study. On the other hand, at 99% VaR level for all data series there are models satisfying these success criteria. These statements can also be followed in the tables in the Appendix on pages 74 through 81.

Equity	VaR	number of successful model	Best Model			
			Approach	Method	Threshold percent	window-size
EREGL	95	1	GPD	Normdist	94	1000
ISCTR	95	9	GPD	Normdist	94	1000
			GPD	Normdist	94	1250
KARTN	95	1	GPD	Dekkers	94	250
KARTN	99	11	GPD	Normdist	90	1250
GOODY	99	8	GPD	Normdist	90	1250
EREGL	99	11	GPD	Normdist	90	1000
AVE20	99	9	GPD	Normdist	90	1000
YKBNK	99	2	GPD	Quantile	94	1000
ISE100	99	12	GPD	Quantile	90	1250
ISCTR	99	18	GPD	Quantile	90	1000
BFREN	99	5	GPD	Quantile	90	750

Table 14: Best models

For "EREGL" and "KARTN" at 95% there is a single model satisfying both unconditional and conditional coverage requirement. Hence the QL functions of these models are not needed to be compared with a second model.

For “ISCTR” however there are nine models that satisfy the performance tests at 95% VaR level. For these 9 successful models, 36 different combinations of model pairs can be made. After testing all these 36 pairs of models against each other, we conclude that two GPD models listed in Table 14 outperform the rest of the seven models. When these two most successful models are tested against each other, the test does not give a significant result that one outperforms the other. Hence we conclude that these two models are the best models in forecasting 95% VaR for “ISCTR”. Both GPD models estimate the parameters using exceedances over the 94% quantile of the normal distribution of different data windows of past 1000 days and 1250 days.

At 99% VaR forecasting, for all of the 8 data series, a single model significantly outperforms the remaining models. GPD models that use 90% quantile of the normal distribution of data windows of past 1250 days are best in forecasting 99% VaR of “GOODY” and “KARTN”. Also the models that again use 90% quantiles of normal distribution as thresholds but 1000 day windowsizes are best in forecasting 99% VaR of “EREGL” and “AV20”.

In general among models that determine thresholds as quantiles of normal distribution, at the same windowsizes 90% level thresholds are better than 94% thresholds.

Among the two successful models forecasting 99% VaR of “YKBNK”, the GPD model using the threshold of 94% quantile of the empirical distribution of data windows of past 1000 days is better.

GPD models using 90% empirical quantiles as thresholds are best in forecasting 99% VaR for “ISE100”, “ISCTR,” and “BFREN” at 1250, 1000, and 750 day windowsizes respectively.

Among the parametric GPD models that best forecast 10 different VaR levels (95% VaR for “EREGL” and “ISCTR”, and 99% VaR for all 8 data series) six of them use past 1000 daily returns.

Considering Table 14 unfortunately no clear-cut recipe can be given for windowsizes and threshold selection of GPD models. However, it can clearly be said that parametric GPD models are powerful tools in estimating the lower tail of ISE stock returns.

In Figures 11 through 18 in the Appendix, all the models computed for all 8 data series are plotted such that their performances can be compared. For each data series there are four plots available, depicting unconditional and conditional coverage performances of 95% and 99% VaR forecasts against the median values of the computed loss functions.

In the scatterplots, “r” is used as the point character for a model estimated by RiskMetrics approach. Likewise “s” is the point character used for HS models. The characters “h” and “d” denote GPD models using Hill and Dekkers estimators respectively. Similarly, the letters “q” and “n” are used for GPD models taking thresholds of the empirical or Gaussian quantiles respectively.

The unconditional coverage performance of a model is better if the percent of exceedances are close to (1 –VaR level) shown with the horizontal line. The conditional coverage figures were calculated using χ^2-statistic as described in 3.7.2.2. Since the null hypothesis being tested is aimed to be accepted, lower values of χ^2-statistics are preferred. Thus, models that are closer to the vertical lines and simultaneously at the most left position outperform others.

In general, plots reveal the fact that for the chosen loss function GPD models are superior. Even though conditional and unconditional coverage performance of each parametric GPD model tried at different windowsizes (750, 1000 and 1250 days) and different thresholds cannot be guaranteed, especially at 99% level it is almost always possible to find a GPD model outperforming other approaches.

4 Conclusion

As the case in Basel, if just unconditional coverage is considered as the success criterion of a VaR forecast model, then one should rather simply employ Historical Simulation. Particularly Type II quantiles should be used for forecasting 95% VaR and Type I quantiles for forecasting 99% VaR. However, if conditional coverage is desired as well as unconditional coverage, then other approaches such as RM and GPD should be considered.

Extreme Value Theory, emphasizes that the models perform best in explaining the extreme tail events. Otherwise, the asymptotic approach of the theory is violated.

Nonparametric and parametric GPD models are observed to perform better in predicting 99% VaR, in line with the extreme value theory.

In fact, all approaches (HS, RM, nonparametric GPD and parametric GPD) perform better at 99% VaR forecasting rather than 95% VaR forecasting given the conditions that they satisfy not only unconditional coverage but also conditional coverage.

Particularly, parametric GPD models are found to estimate VaR more efficiently than other approaches both for the poorly and well traded group of stocks in ISE together with the indices the empirical study is applied to.

Considering the better performance of GPD models, the study shows that RM approach is inadequate for VaR estimation in emerging market conditions like ISE. Comparing the nonparametric and parametric models, the main advantage of nonparametric approach is that it enables to work with smaller windowsizes, whereas MLE approach of GPD cannot be optimized if too few extreme data is used.

Considering the tail distribution plots and results of loss function tests, both visually and computationally GPD is verified to estimate VaR better than the other approaches. Using MLE also enables computation of the standard errors of estimated parameters so that confidence intervals for VaR forecasts can also be computed.

GPD models enable estimating tail distribution functions, which are not only valuable in forecasting correct VaR but also in that the models make it possible to compute other important risk measures like Conditional VaR or Expected Shortfall.

It is also possible to extrapolate the probability of an extreme incident beyond the ones observed so far which is very important in stress tests. It is also possible to estimate how long it will take for a particular extreme event to happen.

5 Appendix

	EQUITY	VaR	Ws	BT	CT	ULR	CLR
Percentile Method Type I	AVE20	95	250	25%	0%	23%	0%
	BFREN	95	250	27%	0%	21%	0%
	EREGL	95	250	33%	0%	30%	0%
	GOODY	95	250	53%	0%	50%	1%
	ISCTR	95	250	48%	17%	35%	1%
	ISE100	95	250	53%	0%	50%	0%
	KARTN	95	250	53%	0%	50%	1%
	YKBNK	95	250	92%	0%	89%	0%
	AVE20	99	250	56%	0%	49%	0%
	BFREN	99	250	100%	0%	98%	8%
	EREGL	99	250	100%	0%	93%	0%
	GOODY	99	250	100%	15%	92%	53%
	ISCTR	99	250	70%	10%	62%	42%
	ISE100	99	250	70%	0%	62%	6%
	KARTN	99	250	44%	6%	37%	28%
	YKBNK	99	250	85%	0%	77%	1%
Percentile Method Type II	AVE20	95	250	100%	0%	97%	0%
	BFREN	95	250	84%	0%	81%	0%
	EREGL	95	250	99%	0%	96%	0%
	GOODY	95	250	80%	0%	77%	0%
	ISCTR	95	250	94%	11%	90%	2%
	ISE100	95	250	92%	0%	89%	0%
	KARTN	95	250	87%	0%	83%	2%
	YKBNK	95	250	49%	0%	47%	0%
	AVE20	99	250	57%	0%	51%	0%
	BFREN	99	250	15%	0%	14%	1%
	EREGL	99	250	1%	0%	2%	0%
	GOODY	99	250	2%	0%	3%	1%
	ISCTR	99	250	18%	40%	17%	29%
	ISE100	99	250	13%	0%	12%	0%
	KARTN	99	250	45%	0%	40%	1%
	YKBNK	99	250	57%	0%	51%	1%

Table 15: Performance tests of HS models.

EQUITY	VaR	BT	CT	ULR	CLR
AVE20	95	8%	1%	7%	0%
BFREN	95	0%	1%	0%	0%
EREGL	95	0%	0%	0%	0%
GOODY	95	4%	0%	3%	0%
ISCTR	95	2%	31%	2%	4%
İSE100	95	41%	0%	38%	0%
KARTN	95	4%	16%	3%	4%
YKBNK	95	9%	0%	8%	0%
AVE20	99	0%	0%	0%	0%
BFREN	99	7%	50%	7%	15%
EREGL	99	6%	0%	6%	1%
GOODY	99	0%	1%	0%	0%
ISCTR	99	9%	46%	8%	13%
İSE100	99	0%	13%	0%	1%
KARTN	99	9%	46%	8%	13%
YKBNK	99	82%	0%	75%	12%

Table 16: Performance tests of RM approach.

EQUITY	VaR	th	ws	BT	CT	ULR	CLR
AVE20	95	90	250	4%	0%	4.6%	0%
AVE20	95	94	250	49%	0%	46%	0%
BFREN	95	90	250	43%	0%	41%	0%
BFREN	95	94	250	91%	0%	88%	0%
EREGL	95	90	250	5%	0%	5%	0%
EREGL	95	94	250	94%	0%	90%	0%
GOODY	95	90	250	0%	0%	0%	0%
GOODY	95	94	250	34%	0%	32%	0%
ISCTR	95	90	250	34%	27%	32%	2%
ISCTR	95	94	250	94%	23%	90%	3%
ISE100	95	90	250	7%	0%	7%	0%
ISE100	95	94	250	80%	0%	77%	0%
KARTN	95	90	250	1%	1%	2%	0%
KARTN	95	94	250	30%	0%	28%	1%
YKBNK	95	90	250	1%	0%	1%	0%
YKBNK	95	94	250	44%	0%	42%	0%
AVE20	99	90	250	56%	0%	49%	0%
AVE20	99	94	250	8%	0%	8%	0%
BFREN	99	90	250	15%	0%	14%	1%
BFREN	99	94	250	1%	0%	1%	0%
EREGL	99	90	250	56%	0%	49%	4%
EREGL	99	94	250	13%	2%	12%	7%
GOODY	99	90	250	70%	10%	62%	42%
GOODY	99	94	250	34%	0%	31%	1%
ISCTR	99	90	250	85%	12%	77%	48%
ISCTR	99	94	250	57%	27%	51%	53%
ISE100	99	90	250	85%	0%	77%	7%
ISE100	99	94	250	34%	0%	31%	0%
KARTN	99	90	250	100%	18%	93%	56%
KARTN	99	94	250	34%	1%	31%	11%
YKBNK	99	90	250	45%	0%	40%	0%
YKBNK	99	94	250	9%	0%	8%	0%

Table 17: Performance tests of the GPD models using Hill estimator.

EQUITY	VaR	th	ws	BT	CT	ULR	CLR
AVE20	95	90	250	100%	0%	97%	0%
AVE20	95	94	250	79%	0%	76%	0%
BFREN	95	90	250	46%	0%	43%	0%
BFREN	95	94	250	20%	0%	18%	0%
EREGL	95	90	250	92%	0%	89%	0%
EREGL	95	94	250	33%	0%	30%	0%
GOODY	95	90	250	6%	0%	6%	0%
GOODY	95	94	250	53%	0%	50%	1%
ISCTR	95	90	250	38%	62%	35%	55%
ISCTR	95	94	250	38%	17%	35%	1%
ISE100	95	90	250	61%	0%	58%	0%
ISE100	95	94	250	66%	0%	62%	0%
KARTN	95	90	250	66%	1%	62%	4%
KARTN	95	94	250	100%	4%	97%	17%
YKBNK	95	90	250	92%	0%	89%	0%
YKBNK	95	94	250	72%	0%	69%	0%
AVE20	99	90	250	13%	0%	12%	0%
AVE20	99	94	250	2%	0%	3%	0%
BFREN	99	90	250	0%	0%	1%	0%
BFREN	99	94	250	0%	0%	1%	0%
EREGL	99	90	250	34%	33%	31%	42%
EREGL	99	94	250	34%	1%	31%	0%
GOODY	99	90	250	0%	0%	0%	0%
GOODY	99	94	250	45%	30%	40%	48%
ISCTR	99	90	250	6%	51%	6%	14%
ISCTR	99	94	250	1%	65%	1%	3%
ISE100	99	90	250	2%	0%	3%	0%
ISE100	99	94	250	2%	0%	3%	0%
KARTN	99	90	250	25%	51%	23%	32%
KARTN	99	94	250	0%	27%	0%	0%
YKBNK	99	90	250	13%	0%	12%	1%
YKBNK	99	94	250	6%	0%	6%	0%

Table 18: Performance tests of the GPD models using Dekkers estimator.

EQUITY	VaR	ws	BT	CT	ULR	CLR
AVE20	95	750	0%	0%	0%	0%
AVE20	95	1000	0%	0%	0%	0%
AVE20	95	1250	0%	2%	0%	0%
EREGL	95	750	17%	0%	15%	0%
EREGL	95	1000	13%	0%	11%	0%
EREGL	95	1250	0%	79%	0%	0%
GOODY	95	750	57%	0%	54%	0%
GOODY	95	1000	79%	0%	75%	0%
GOODY	95	1250	12%	0%	10%	1%
ISCTR	95	750	3%	6%	2%	2%
ISCTR	95	1000	3%	2%	2%	1%
ISCTR	95	1250	1%	6%	0%	0%
ISE100	95	750	62%	0%	59%	0%
ISE100	95	1000	32%	0%	29%	0%
ISE100	95	1250	4%	1%	3%	0%
KARTN	95	750	85%	0%	82%	0%
KARTN	95	1000	28%	0%	25%	0%
KARTN	95	1250	12%	0%	10%	0%
YKBNK	95	750	99%	0%	95%	0%
YKBNK	95	1000	71%	0%	67%	0%
YKBNK	95	1250	15%	0%	13%	0%
AVE20	99	750	20%	1%	15%	9%
AVE20	99	1000	73%	4%	65%	32%
AVE20	99	1250	3%	85%	1%	4%
EREGL	99	750	53%	0%	47%	8%
EREGL	99	1000	91%	67%	82%	80%
EREGL	99	1250	9%	81%	5%	14%
GOODY	99	750	0%	0%	0%	0%
GOODY	99	1000	15%	1%	14%	5%
GOODY	99	1250	81%	70%	71%	80%
ISCTR	99	750	67%	5%	59%	33%
ISCTR	99	1000	43%	72%	35%	56%
ISCTR	99	1250	63%	72%	54%	72%
ISE100	99	750	21%	1%	19%	7%
ISE100	99	1000	74%	0%	66%	7%
ISE100	99	1250	47%	74%	38%	60%
KARTN	99	750	9%	0%	9%	1%
KARTN	99	1000	31%	21%	28%	34%
KARTN	99	1250	63%	10%	56%	40%
YKBNK	99	750	67%	0%	61%	1%
YKBNK	99	1000	91%	0%	82%	4%
YKBNK	99	1250	47%	0%	38%	1%

Table 19: Performance tests GPD models (threshold is 90% quantile of Φ)

EQUITY	VaR	ws	BT	CT	ULR	CLR
AVE20	95	1000	4%	0%	3%	0%
AVE20	95	1250	0%	7%	0%	0%
EREGL	95	750	17%	0%	15%	0%
EREGL	95	1000	23%	3%	21%	6%
EREGL	95	1250	0%	42%	0%	0%
GOODY	95	750	69%	0%	65%	1%
GOODY	95	1000	97%	0%	93%	0%
GOODY	95	1250	15%	0%	13%	0%
ISCTR	95	750	50%	5%	48%	16%
ISCTR	95	1000	89%	8%	85%	26%
ISCTR	95	1250	37%	6%	34%	14%
ISE100	95	750	24%	0%	21%	0%
ISE100	95	1000	44%	0%	40%	0%
ISE100	95	1250	3%	0%	2%	0%
KARTN	95	750	5%	0%	5%	0%
KARTN	95	1000	87%	0%	83%	0%
KARTN	95	1250	37%	0%	34%	0%
YKBNK	95	750	12%	0%	12%	0%
YKBNK	95	1000	29%	0%	27%	0%
YKBNK	95	1250	32%	0%	28%	1%
AVE20	99	1000	14%	77%	9%	22%
AVE20	99	1250	5%	83%	2%	7%
EREGL	99	750	100%	0%	92%	1%
EREGL	99	1000	100%	7%	100%	43%
EREGL	99	1250	9%	81%	5%	14%
GOODY	99	750	9%	0%	9%	1%
GOODY	99	1000	15%	1%	14%	5%
GOODY	99	1250	81%	70%	71%	80%
ISCTR	99	750	39%	2%	32%	20%
ISCTR	99	1000	43%	72%	35%	56%
ISCTR	99	1250	23%	77%	16%	34%
ISE100	99	750	21%	0%	19%	0%
ISE100	99	1000	91%	0%	83%	6%
ISE100	99	1250	33%	75%	26%	47%
KARTN	99	750	29%	0%	26%	1%
KARTN	99	1000	74%	12%	66%	45%
KARTN	99	1250	100%	6%	91%	40%
YKBNK	99	750	84%	0%	76%	1%
YKBNK	99	1000	91%	0%	82%	4%
YKBNK	99	1250	47%	0%	38%	1%

Table 20: Performance tests GPD models (threshold is 94% quantile of Φ)

EQUITY	VaR	ws	BT	CT	ULR	CLR
AVE20	95	750	0%	0%	0%	0%
AVE20	95	1000	0%	0%	0%	0%
AVE20	95	1250	0%	6%	0%	0%
BFREN	95	750	68%	0%	64%	0%
BFREN	95	1000	41%	0%	38%	0%
BFREN	95	1250	8%	0%	6%	0%
EREGL	95	750	28%	0%	25%	0%
EREGL	95	1000	13%	0%	11%	0%
EREGL	95	1250	0%	79%	0%	0%
GOODY	95	750	15%	0%	10%	0%
GOODY	95	1000	9%	0%	9%	0%
GOODY	95	1250	50%	0%	46%	4%
ISCTR	95	750	3%	6%	2%	2%
ISCTR	95	1000	1%	1%	1%	0%
ISCTR	95	1250	1%	5%	0%	0%
ISE100	95	750	69%	0%	65%	0%
ISE100	95	1000	57%	0%	53%	0%
ISE100	95	1250	2%	0%	1%	0%
KARTN	95	750	64%	0%	60%	0%
KARTN	95	1000	50%	0%	46%	0%
KARTN	95	1250	27%	0%	24%	0%
YKBNK	95	750	24%	0%	21%	0%
YKBNK	95	1000	50%	0%	46%	0%
YKBNK	95	1250	12%	0%	10%	0%
AVE20	99	750	100%	9%	91%	45%
AVE20	99	1000	22%	76%	15%	0%
AVE20	99	1250	23%	77%	16%	34%
BFREN	99	750	94%	0%	85%	6%
BFREN	99	1000	4%	0%	4%	0%
BFREN	99	1250	100%	0%	97%	3%
EREGL	99	750	29%	0%	26%	8%
EREGL	99	1000	74%	0%	66%	0%
EREGL	99	1250	15%	79%	9%	23%
GOODY	99	750	0%	0%	0%	0%
GOODY	99	1000	0%	0%	0%	0%
GOODY	99	1250	15%	0%	14%	4%
ISCTR	99	750	100%	11%	92%	48%
ISCTR	99	1000	91%	5%	82%	38%
ISCTR	99	1250	63%	72%	54%	72%
ISE100	99	750	29%	0%	26%	8%
ISE100	99	1000	22%	0%	20%	0%
ISE100	99	1250	47%	74%	38%	60%
KARTN	99	750	0%	0%	0%	0%
KARTN	99	1000	0%	57%	0%	1%
KARTN	99	1250	2%	0%	2%	0%
YKBNK	99	750	67%	0%	61%	1%
YKBNK	99	1000	43%	0%	39%	1%
YKBNK	99	1250	81%	0%	71%	0%

Table 21: Performance tests GPD models (threshold is 90% empirical quantile)

EQUITY	VaR	ws	BT	CT	ULR	CLR
AVE20	95	750	0%	0%	0%	0%
AVE20	95	1000	0%	0%	0%	0%
AVE20	95	1250	0%	3%	0%	0%
BFREN	95	750	75%	0%	71%	0%
BFREN	95	1000	26%	0%	22%	0%
BFREN	95	1250	8%	0%	6%	0%
EREGL	95	750	14%	0%	12%	0%
EREGL	95	1000	13%	0%	11%	0%
EREGL	95	1250	0%	71%	0%	0%
GOODY	95	750	78%	0%	74%	0%
GOODY	95	1000	89%	0%	85%	1%
GOODY	95	1250	18%	1%	16%	2%
ISCTR	95	750	3%	19%	2%	4%
ISCTR	95	1000	4%	3%	3%	2%
ISCTR	95	1250	1%	6%	0%	0%
ISE100	95	750	49%	0%	46%	0%
ISE100	95	1000	38%	0%	34%	0%
ISE100	95	1250	2%	0%	1%	0%
KARTN	95	750	69%	0%	65%	0%
KARTN	95	1000	32%	0%	29%	0%
KARTN	95	1250	10%	0%	8%	0%
YKBNK	95	750	33%	0%	30%	0%
YKBNK	95	1000	57%	0%	53%	0%
YKBNK	95	1250	15%	0%	13%	0%
AVE20	99	750	20%	1%	15%	9%
AVE20	99	1000	9%	79%	5%	14%
AVE20	99	1250	3%	85%	1%	4%
BFREN	99	750	77%	0%	69%	7%
BFREN	99	1000	51%	0%	45%	6%
BFREN	99	1250	93%	0%	83%	4%
EREGL	99	750	100%	9%	91%	45%
EREGL	99	1000	74%	0%	66%	0%
EREGL	99	1250	9%	81%	5%	14%
GOODY	99	750	0%	0%	0%	0%
GOODY	99	1000	0%	0%	0%	0%
GOODY	99	1250	63%	63%	56%	66%
ISCTR	99	750	83%	7%	74%	40%
ISCTR	99	1000	57%	70%	49%	67%
ISCTR	99	1250	33%	75%	26%	47%
ISE100	99	750	40%	0%	36%	8%
ISE100	99	1000	58%	0%	51%	7%
ISE100	99	1250	33%	75%	26%	47%
KARTN	99	750	21%	0%	19%	1%
KARTN	99	1000	22%	25%	20%	28%
KARTN	99	1250	63%	10%	56%	40%
YKBNK	99	750	67%	0%	61%	1%
YKBNK	99	1000	100%	0%	100%	5%
YKBNK	99	1250	81%	0%	71%	0%

Table 22: Performance tests GPD models (threshold is 94% empirical quantile)

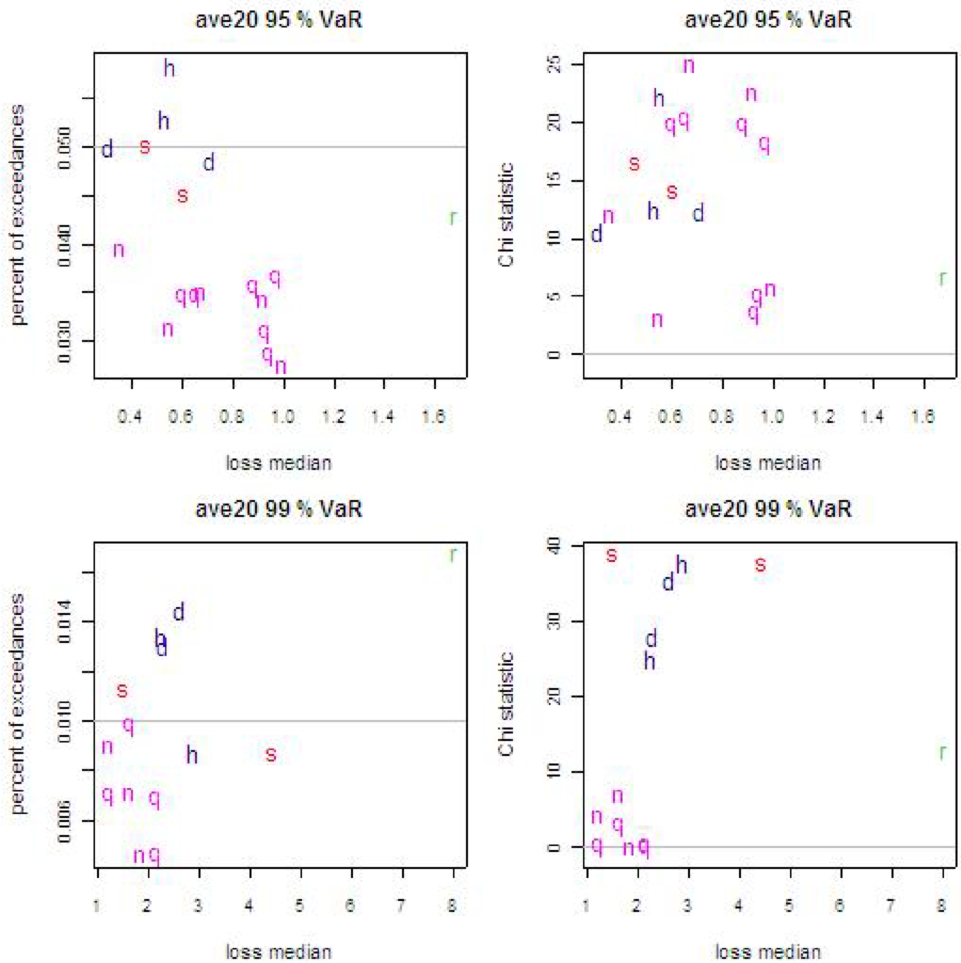

Figure 11: Performances of VaR models at 99% and 95% for ‘AVE20’ data.

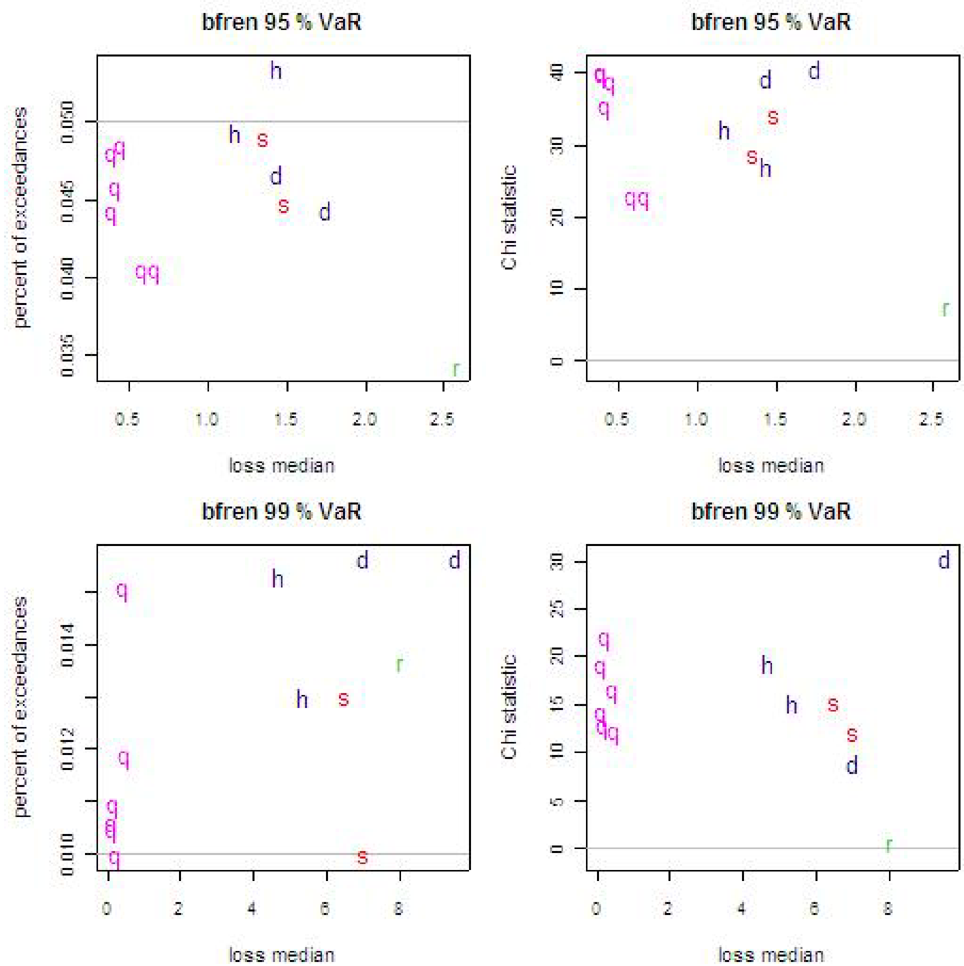

Figure 12: Performances of VaR models at 99% and 95% for ‘BFREN’ data.

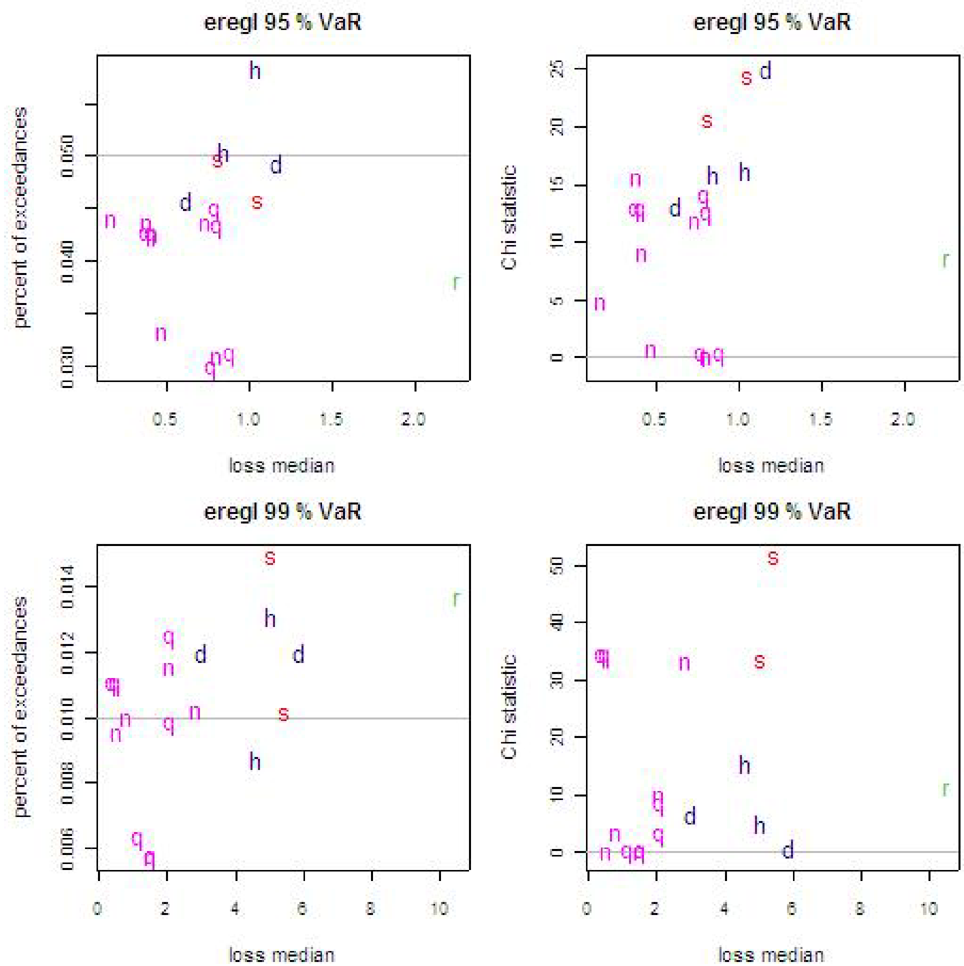

Figure 13: Performances of VaR models at 99% and 95% for ‘EREGL’ data.

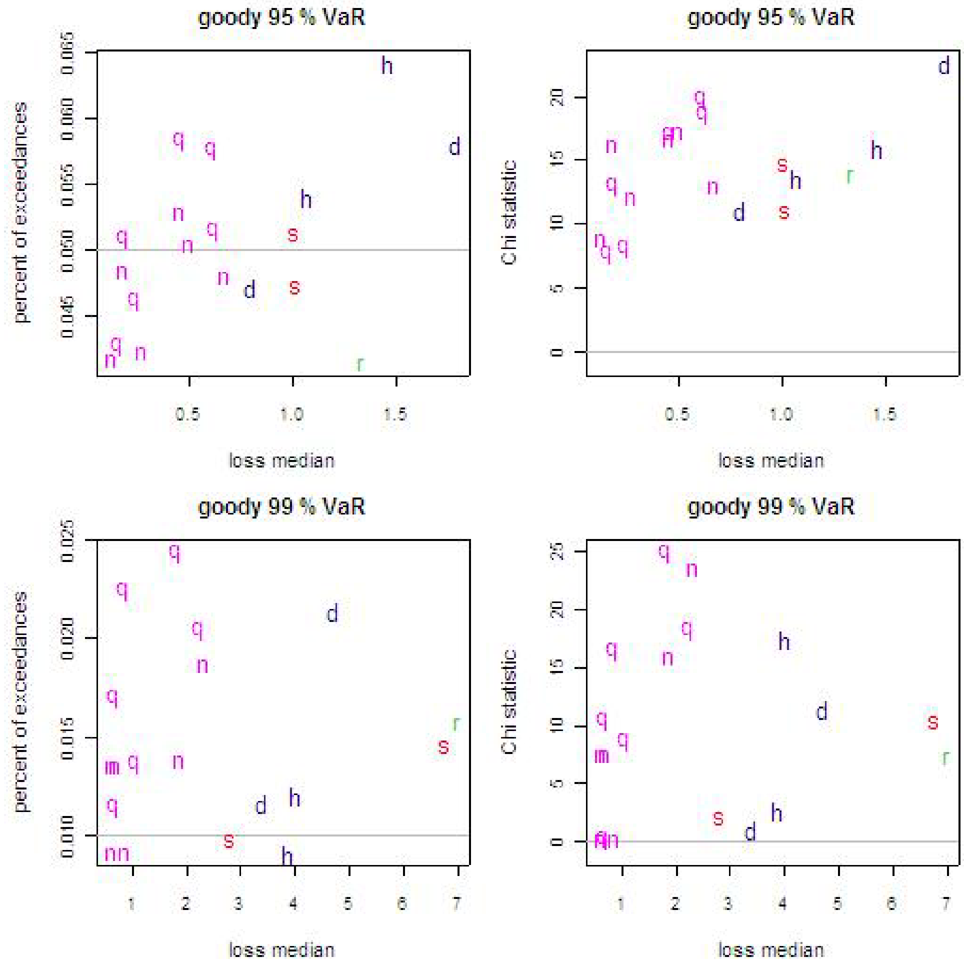

Figure 14: Performances of VaR models at 99% and 95% for ‘GOODY’ data.

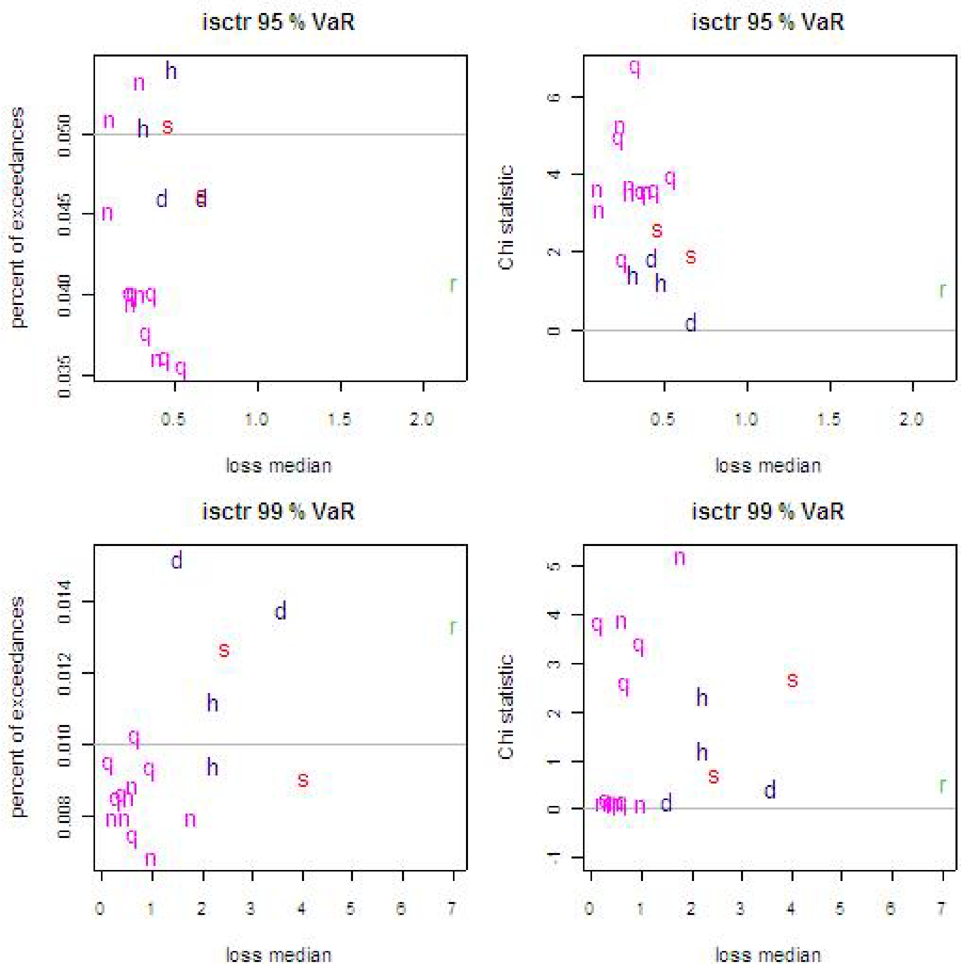

Figure 15: Performances of VaR models at 99% and 95% for ‘ISCTR’ data.

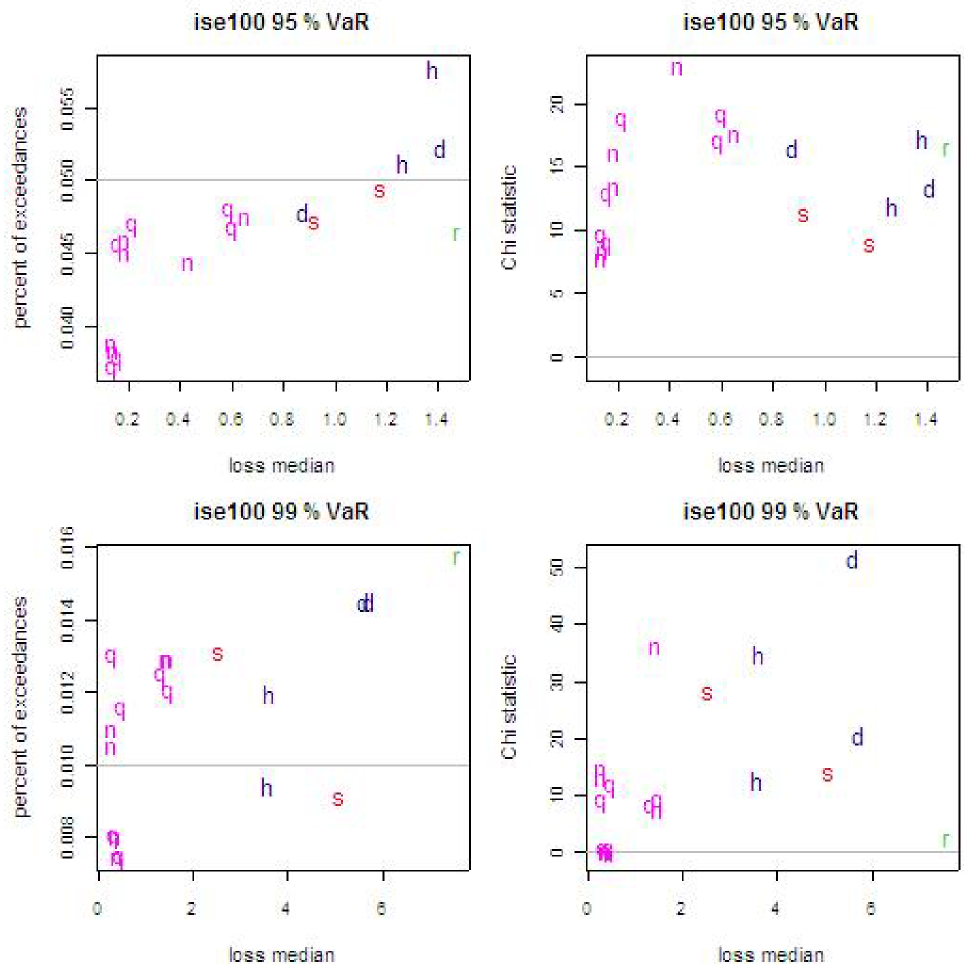

Figure 16: Performances of VaR models at 99% and 95% for ‘ISE100’ data.

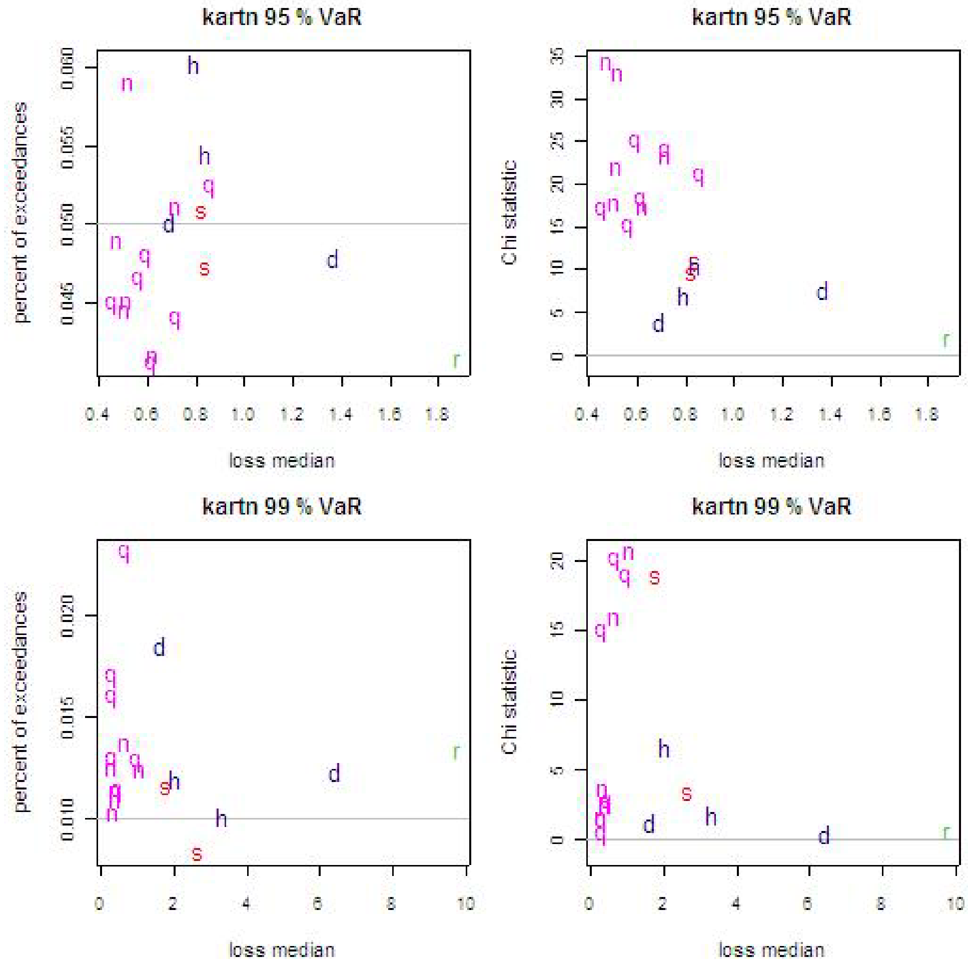

Figure 17: Performances of VaR models at 99% and 95% for 'KARTN' data.

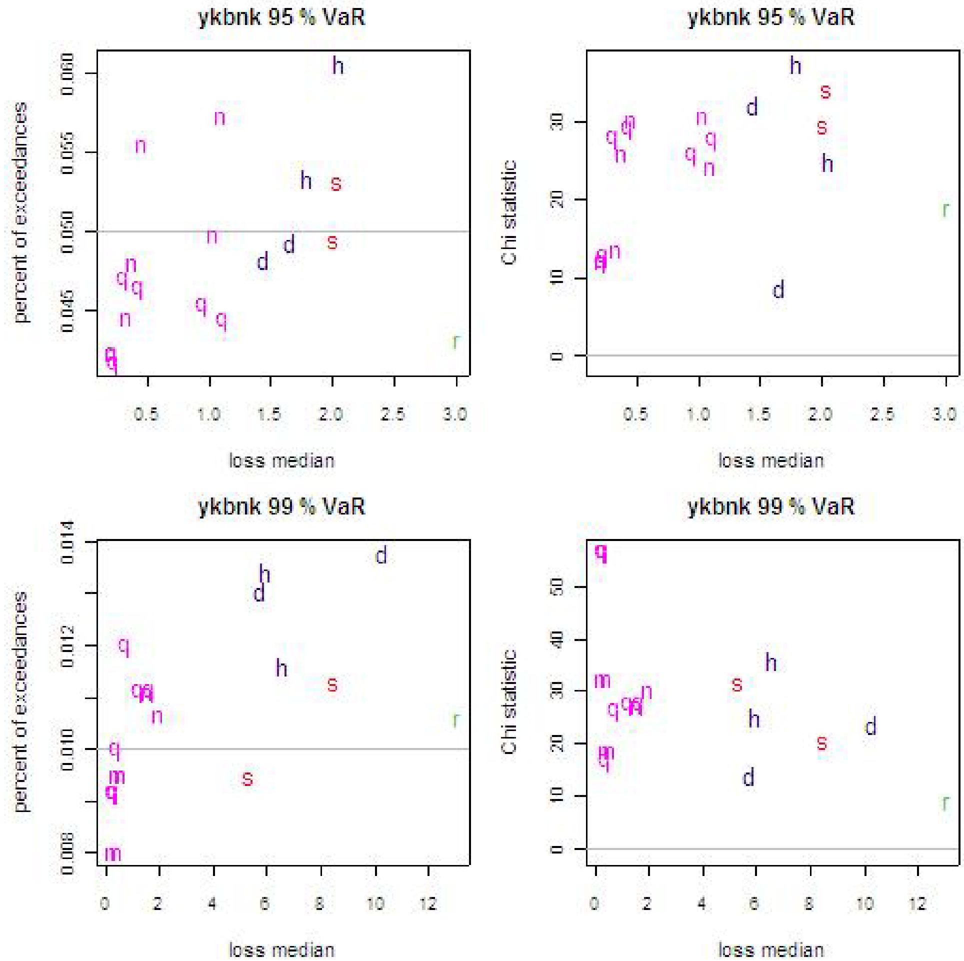

Figure 18: Performances of VaR models at 99% and 95% for ‘YKBNK’ data.

6 Bibliography

Adler R.J. (1990) An Introduction to Continuity Extrema and Related Topics for General Gaussian Processes. Institute of Mathematical Statistics Lecture Notes--Monograph Series **12**.

Anderson C.W. 1970. Extreme value theory for a class of discrete distributions with applications to some stochastic processes *Journal of Applied Probability* **7** 99-113.

Anderson C.W. (1980) Local limit theorems for the maxima of discrete random variables. *Math. Proc. Camb. Philos. Soc.* **88** 161-165.

Arnold B.C. Balakrishnan N. and Nagaraja H.N. (1992) *A First Course In Order Statistics* Wiley: New York.

Bachelier L. (1900) Théorie de la Spéculation. *Ann. Sci. école norm. Sup.* **III**-17 21-86.

Balkema A.A. and de Haan L. (1972) On R. von Mises' Condition for the Domain of Attraction of $\exp(-e{-x})^1$. *Annals of Mathematical Statistics.* **43** 1352-1354.

Balkema A.A. and de Haan L. (1974) Residual life time at great age. *Ann. Probab.* **2** 792-804.

Barnett V. and Turkman K.F. (1993) *Statistics for the Environment*. Wiley New York.

Basle Committee on Banking Supervision (1995a) *An internal model-based approach to market risk capital requirements*. Technical report. Basle Committee on Banking Supervision Basle Switzerland.

Basle Committee on Banking Supervision (1995b) *Planned supplement to the capital accord to incorporate market risks*. Basle Committee on Banking Supervision Basle Switzerland.

Basle Committee on Banking Supervision (1996) *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements.* Bank for International Settlements Basle Switzerland.

Beirlant J. Teugels JL and Vynckier P. (1996) *Practical Analysis of Extreme Values* Leuven University Press Leuven

Berkowitz J. (2001) Testing density forecasts with applications to risk management *Journal of Business and Economic Statistics* **19** 465-474.

Berkowitz J. and O'Brien J. (2002) How accurate are value-at-risk models at commercial banks? *Journal of Finance* **57** 1093-1111.

Berman S.M. (1992) *Sojouns and Extremes of Stochastic Processes.* Wadsworth Belmont California.

Blattberg R.C. and Gonedes N.J. (1974) A Comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business.* **47** 244-280.

Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307-327.

Campbell R. and Koedijk K. G. (1999) Capturing downside risk in financial markets: the case of the Asian Crisis. *Journal of International Money and Finance* **18** 853-870.

Castillo E. (1988) *Extreme Value Theory in Engineering*. Academic Press Boston.

Castillo E. and Hadi A. S. (1997) "Fitting the Generalized Pareto Distribution to Data" *Journal of the American Statistical Association* 92 1609-1620.

Christoffersen P. (1998) *Evaluating Interval Forecasts*. International Economic Review **39** 841-862.

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

Coles S. and Powell E. (1996) Bayesian methods in extreme value modelling: a review and new developments *International Statistical Review* 64 119–136.

Cotter J. (2001) Margin Exceedences for European Stock Index Futures Using Extreme Value Theory *Journal of Banking and Finance* **25** 1475-1502.

Danielsson J. and De Vries C.G. (2000) Value at risk and extreme returns. *Annales d' Economie et de Satistique* **60** (October-December) 239-270.

Davison A.C. (1984) Modelling excesses over high thresholds with an application. In *Statistical Extremes and Applications* J. Tiago de Oliveira (ed.) Reidel Dordrecht 461-482.

Da Silva A.L.C and De Melo Mendez B.V. (2000) Value at risk and Extreme Returns in Asian Stock Markets. *International Journal of Business* **8** 17-40.

De Finetti B (1930) Funzione caratteristica di un fenomeno aleatorio. *Memorie della R. Accademia dei Lincei.* **6** 86–133. (Reimprinted in de Finetti (1981) 265–315.)

De Finetti B (1981) *Scritti [1926–1930]*. Padova: Cedam.

De Haan L. (1970) On Regular Variation and Its Applications to the Weak Convergence of Sample Extremes Math. Centre Tracts **32** Amsterdam.

De Haan L. (1981) Estimation of the minimum of a function using order statistics. *Journal of the Americal Statistical Association.* **76** 467-469.

De Haan L. and Resnick S.I. (1980) A simple asymptotic estimate for the index of a stable distribution. *Journal of the Royal Statistical Society.* **42** 83-87.

De Haan L. Resnick S.I. Rootzén H and De Vries C.G. (1990) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH-processes. *Stoch. Proc Appl.* **32** 213-224.

Dekkers A. Einmahl J. and de Haan L. (1989) A moment estimator for the index of an extreme-value distribution *Annals of Statistics* **17** 1833–1855.

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

Drees H. (1995) Refined pickands estimators of the extreme value index *Annals of Statistics* **23** 2059–2080.

Drees H. (1996) Refined pickands estimators with bias correction *Comm. Statist. Theory Methods* **25** 837–851.

Drees H. (1998) On smooth statistical tail functional *Scandinavian Journal of Statistics* **25** 187–210.

De Vries C.G. (1994) Stylized facts of nominal exchange returns. *The Handbook of International Economics R. Van der Ploeg (ed.)* Blackwell Oxford pp. 335-389.

DuMouchel W. (1983) Estimating the Stable Index σ in Order to Measure Tail Thickness.

Annals of Statistics **11** 1019-1036.

Embrechts P. Klüppelberg C. And Mikosch T. (1997) *Modelling Extreme Events for Insurance and Finance*. Springer Verlag.

Engle R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987-1007.

Fama E.F. (1963) Mandelbrot and the stable Paretian hypothesis. *The Journal of Business.* **36** 420-429.

Fan Z. (1999) An empirical examination of extreme value theory methods in VaR estimation Department of Statistics. Leonard N. Stern School of Business NYU 2 (PhD thesis).

Feller E. (1971) An introduction to probability theory and its applications Volume II. John Wiley & Sons New York.

Fisher R.A. and Tippett L.H.C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24** 180-190.

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

Galambos J. (1987) *Asymptotic Theory of Extreme Order Statistics* 2nd edition Krieger Malabar Florida.

Ganief S. and Biekpe N. (2003) Measuring Market Risk Using Extreme Value Theory. An Empirical Study Using South African Rand/Dollar One-Year Future Contract. *African Finance Journal* **5** 68-86.

Gençay R. Selçuk F. (2001) Overnight Borrowing Interest Rates and Extreme Value Theory. *Bilkent University Department of Economics Discussion Paper* 01-03.

Gnedenko BV (1943) Sur la distribution limité du terme maximum d'une série aléatoire' *Annals of Mathematics* **44** 423-453.

Gordon L. Schilling M.F. and Waterman M.S. (1986) An extreme value theory for long head runs. *Probability Theory Related Fields* **72** 279-287.

Grimshaw S.D. (1993) Computing Maximum Likelihood Estimates for the Generalized Pareto distribution. *Technometrics* **35** 185-191.

Gumbel E. J. (1935) Les Valeurs Extremes des Distributions Statistiques *Annales de l'lnstitut Henri Poincarh* **4** 115.

Gumbel E.J. (1941) The return period of flood flows. *Annals of Mathematical Statistics* **12** 163-190.

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

Gumbel E.J. Lieblein J. (1954) Some Applications of Extreme Value Methods. *The Americal Statistician* **8** 14-17.

Hamao Y. Masulis R.W and Ng V. (1990) Correlations in price changes and volatility across international stock markets. *The Review of Financial Studies* **3** 281-307.

Hill B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3** 1163-1174.

Hilliard J.E. (1979) The relation between equity indices on world exchanges. *The Journal of*

Finance **34** 103-114.

Hols M. Vries C.G. de (1991) The limiting distribution of extremal exchange rate returns. *Journal of Applied Econometrics.* **6** 287-302.

Hosking J.R.M. (1984) Testing whether the shape parameter is zero in the generalized extreme-value distribution. *Biometrika* **71** 367-374.

Hosking J.R.M. (1994) The four-parameter kappa distribution. *IBM Journal of Research Development* **38** May 251-258.

Hosking J.R.M. Wallis J.R. (1986a) Paleoflood Hydrology and Flood Frequency Analysis.bu *Water Resources Research* **22** April 543-550.

Hosking J.R.M. Wallis J.R. (1986b) The Value of Historical Data in Flood Frequency Analysis. *Water Resources Research* **22** April 1606-1612.

Hosking J.R.M. Wallis J.R. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* **29** 339-349.

Hosking J.R.M. Wallis J.R. (1988) The Effect of Intersite Dependence on Regional Flood Frequency Analysis. *Water Resources Research* **24** April 588-600.

Hosking J.R.M. Wallis J.R. Wood E.F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251-261.

Husler J. and Reiss R.D. (1989)Maxima of normal random vectors: between independence and complete dependence. *Statistical Probability Letters* **7** 283–286.

Jansen D.W. Koedijk K.G. De Vries C.G. (2000) Portfolio Selection with Limited Downside

Risk. *Journal of Empirical Finance* 7 247-269.

Jenkinson A.F. (1955) The frequency distribution of annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteorol. Soc.* **81** 158-171.

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

J.P. Morgan. (1996) Riskmetrics technical manual. J.P.Morgan Bank New York.

Këllezi E. and M. Gilli (2003) An Application of Extreme Value Theory for Measuring Risk. Working Paper University of Geneva.

Kerkhof J. and Melenberg B. (2004) Backtesting for risk-based regulatory capital. *Journal of Banking and Finance* **8** 1845-1865.

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

Leadbetter M.R. (1991) On a basis for ‘peaks over threshold’ modeling. *Statistical Probability Letters* **12** 357-362.

Leadbetter M.R. Lindgren G. and Rootzén H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer Berlin.

Longin FM (1996) The Asymptotic Distribution of Extreme Stock Market Returns *Journal of Business* **69** 383-408.

Longin FM (2001) Portfolio Insurance and Market Crashes. *Journal of Asset Management* **2** 136-161.

Longin FM and Solnik B (2001) Extreme Correlation of International Equity Markets. *Journal of Finance* **56** 649-676.

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

Lux T. (2001) The Limiting Extremal Behaviour of Speculative Returns: An Analysis of Intra-daily Data from the Frankfurt Stock Exchange. *Applied Financial Economics* **11** 299-315.

Mandelbrot B. (1963) The variation of certain speculative prices. *Journal of Business* **36** 394-419.

McCulloch J.H. (1978) Continuous Time Processes with Stable Increments. *The Journal of Business* **51** 601-619.

McNeil A.J. and Frey R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series. *Journal of Empirical Finance* **7** 271-300.

Mises R.von (1936) La distribution de la plus grande de *n* valeurs. *Rev. Math Union Interbalcanique* **1** 141-160. Reproduced in Selected Papers of Richard von Mises *Amer. Math. Soc.* **2** (1964) 271-294.

Nelder J. A. and Mead R. (1965) A simplex algorithm for function minimization. *Computer Journal* **7** 308-313.

Newbold P. et al. (1995) *Statistics for Business and Economics* Pearson Education New Jersey.

Pfeifer D. (1989) Einführung in die Extremwertstatistik. Teubner Stuttgard.

Pickands J. (1975) Statistical inference using extreme value order statistics. *Annals of Statistics* **3** 119-131.

R Development Core Team (2004) R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria. ISBN 3-900051-07-0 URL http://www.R-project.org.

Reiss R.D. (1989) Approximate Distributions of Order Statistics. Springer New York.

Reiss R.D. (1993) *A Course on Point Processes.* Springer New York.

Reiss R.D. and Thomas M. (1997) *Statistical Analysis of Extreme Values* Birkhauser Berlin.

Resnick S.I. (1992) Extreme Values Regular Variation and Point Processes. Springer New York.

Rootzén H. and Tajvidi N. (1997) Extreme value statistics and wind storm losses: A case study *Scandinavian Actuarial Journal* **1** 70-94.

Sarma M. et al (1993) Selection of Value-at-Risk models. *Journal of Forecasting;* 22 337-358.

Shibata T. (1994) Application of extreme value statistics to corrosion. *Extreme Value Theory and Applications*. Proceedings of Gaitghersburg Conference Vol II 327-336.

Smith R.L. (1984) Threshold methods for sample extremes. In *Statistical Extremes and Applications* edited by J. Tiago de Oliveira Reidel Dordrecht 621-638.

Smith R.L. (1987) Estimating Tails of Probability Distributions. *Annals of Statistics.* **15**

1175-1207.

Smith R.L. Weissman I. (1994) Estimating the extremal index. *Journal of Royal Statistical Society.* **56** 515-528

Tajvidi N. (1996) Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions. PhD Thesis Department of Mathematics University of Gothenburg.

Talgat Daniyarov (2005) VaR: Value at Risk estimation. R package version 0.2.

Taylor S.J. (1986) *Modeling Financial Time Series*. Wiley Chichester.

Tiago de Oliveira J. (1997) *Statistical Analysis of Extremes.* Pendor.

Tolikas K. and Brown R. A. (2005) The Distribution of the Extreme Daily Share Returns in the Athens Stock Exchange. *European Journal of Finance* Forthcoming http://ssrn.com/abstract=597921

Weissman I. (1975) Multivariate extremal processes generated by independent non-identically distributed random variables. *Journal of Applied Probability* **12** 477-487.

Weissman I. (1978) Estimation of parameters and large quantiles based on the *k* largest observations. *Journal of American Statistics Association* **73** 812-815.